||Jai Sri Gurudev||

B G S INSTITUTE OF TECHNOLOGY

BG Nagara-571448, Mandya

Department of Computer Science and Engineering

COURSE FILE

Course Coordinator	:	M J PRASANNA KUMAR
Designation	:	Assistant Professor
Course Name	:	Machine Learning
Course Code	:	17CS73
Academic Year	:	2020-2021(ODD)
For the period	:	01/09/2020 to 16/01/2021

M.J Prasanna Kumar

Signature of Course Coordinator

B. K. Raghu

Signature of HOD

HOD
Dept. of Computer Science & Engg
B.G.S. Institute of Technology
B.G. Nagar - 571 448
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

B G S Institute of Technology

VISION

BGSIT is committed to the cause of creating tomorrow's engineers by providing quality education inculcating ethical values.

MISSION

- Imparting quality technical education by nurturing a conducive learning environment.
- Offering professional training to meet industry requirements.
- Providing education with a moral - cultural base and spiritual touch

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION

To produce engineers by possessing good technical knowledge and ethics through quality education and research.

MISSION

M1: Achieve excellence by providing good infrastructure and competent faculty.

M2: Strengthening the technical, soft skills, leadership qualities and ethical values to meet the industry requirements.

M3: Facilitate experimental learning through research projects

B. K. Raghu

H O D
Dept. of Computer Science & Engg.
B.G.S. Institute of Technology,
B.G. Nagar 571 448.
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

PROGRAM OUTCOMES (POs)

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change

B. K. Raghu
HOD
Dept. of Computer Science & Engg.
B.G.S. Institute of Technology,
B.G. Nagar 571 448
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PEO 1: Graduates will be pursuing successful career and higher education.

PEO 2: Graduates will be able to apply the knowledge of programming skills to solve the real-world problems.

PEO 3: Graduates will display professional ethics to work in a team and lead the team by effectively communicating the ideas.

PEO 4: Graduates will practice lifelong learning.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO 1: Ability to apply Mathematical Methodologies, Management Principles and Ethics, Electronics and Embedded Systems and Programming Technologies to solve real time problems.

PSO 2: Ability to apply software design and development practices to develop software in emerging areas such as Internet of Things, Data Management, Social Networking and Security, Cloud and High-Performance Computing.

B. K. Raghu

HOD
Dept. of Computer Science & Engg.
B. M. S. Institute of Technology,
P.O. Yelahanka, Bangalore-64, Bangalore Dist
Karnataka (INDIA)

Revised Academic Calendar of VTU, Belagavi for ODD Semester of 2020-21 (Tentative)

	I Sem B. E. / B. Tech. / B. Arch./B.Plan	I sem M.Tech./MBA /MCA/M.Arch.	III, V B. E. /B. Tech./B.Plan/ B.Arch & VII sem BPlan /BArch & IX Sem B. Arch.	VII Sem B. E. /B. Tech	III & V Sem MCA	III Sem MBA	III Sem M. Tech.	III Sem M. Arch.
Commencement of ODD Semester	14.12.2020	Will be announced later	01.09.2020	01.09.2020	01.09.2020	01.09.2020	01.09.2020	01.09.2020
Last Working day of ODD Semester	25.03.2021		16.01.2021	16.01.2021	16.01.2021	16.01.2021	16.01.2021	16.01.2021
Practical Examinations	29.03.2021 Onwards#		21.01.2021 Onwards#	21.01.2021 Onwards#	08.02.2021 Onwards#	--	21.01.2021 Onwards#	--
Theory Examinations	12.04.2021 To 30.04.2021		08.02.2021 To 27.03.2021	08.02.2021 To 27.03.2021	21.01.2021 To 06.02.2021	21.01.2021 To 19.02.2021	28.01.2021 To 13.02.2021	21.01.2021 To 06.02.2021
Internship			---	29.03.2021 To 10.04.2021	---	---	---	---
Internship Viva-Voce			---	---	---	---	15.02.2021 To 22.02.2021	---
Professional training / Organization study			---	---	---	22.02.2021 To 03.04.2021	---	---
Commencement of EVEN Semester	03.05.2021		29.03.2021	12.04.2021	15.02.2021	05.04.2021	23.02.2021	08.02.2021

NOTE

- VII Semester B. E. / B. Tech. students shall have to undergo **Internship** as per circular of University **VTU/Aca/2019-20/85, dated 12.05.2020.**
- I Semester B. E/ B. Tech / B. Arch Students shall compulsorily undergo **Induction Program** for 01 Weeks.
- The classroom sessions for all the semesters would be in **ONLINE mode/blended mode** until further orders.
- The institute needs to function for **six days** a week with additional hours (**Saturday is a full working day).**
- If any of the above dates are declared to be a holiday then the corresponding event will come into effect on the next working day.
- (#) Notification regarding the Calendar of Events relating to the conduct of **University Examinations** will be issued by the Registrar (Evaluation) from time to time.
- Academic Calendar may be modified based on guidelines/directions issued in the future by MHRD/UGC/AICTE/State Government.
- Revised Academic Calendar is also applicable for **Autonomous Colleges.**
- The MBA students are permitted to carry out **project work** in blended mode (ONLINE/OFFLINE). More emphasis on OFFLINE mode wherever feasible.

04-12-20

REGISTRAR

DEPARTMENT OF COMPUTER SCEINCE & ENGINEERING

ACADEMIC YEAR : 2020-21	ONLINE	ODD SEMESTER
CLASS : VII CSE 'A' & 'B'	LAB TIME TABLE	SUBJECT NAME : Machine Learning Laboratory
		SUBJECT CODE : 17CSL76

DAY	09:30 am-10:30 am	11:00 am - 12:00 pm	12:30 pm – 01:30 pm	01:30 pm - 02:30 pm	02:30 PM - 03:30 PM	03:30 pm – 05:30 pm		
MONDAY		ML		LUNCH BREAK				
TUESDAY					ML			
WEDNESDAY	ML							
THURSDAY					ML			
FRIDAY	ML							
SATURDAY					ML LAB - ALL BATCH			

CODE	SUBJECT	STAFF
17CSL76	Machine Learning Laboratory	Prasanna Kumar M J
17CS73	Machine Learning	M J Prasanna Kumar

B. K. Rag
H.O.D
H O D
Dept. of Computer Science & Engg
B.G.S. Institute of Technology
B.G. Nagar - 571 448
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

Time Table Co-ordinator

Principal
PRINCIPAL
B.G.S. INSTITUTE OF TECHNOLOGY
B.G. NAGAR - 571 448

||Jai Sri Gurudev||

BGS INSTITUTE OF TECHNOLOGY, B G NAGAR
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
COURSE OUTCOMES AND CO-PO MAPPING

Course coordinator :M J PRASANNA KUMARA **Sem: VII**
COURSE CODE: 17CS73 **COURSE NAME: Machine Learning**

CO1	Choose the learning techniques to investigate various training examples
CO2	Describe the characteristics of learning techniques and solve problems associated with
CO3	Apply effectively neural networks for appropriate applications
CO4	Apply Bayesian techniques and derive effectively learning rules
CO5	Evaluate hypothesis and investigate instant based learning and reinforced learning

PSO1	Ability to apply Mathematical Methodologies, Management Principles and Ethics, Electronics and Embedded Systems and Programming Technologies to solve real time problems.
PSO2	Ability to apply software design and development practices to develop software in emerging areas such as Internet of Things, Data Management, Social Networking and Security, Cloud and High-Performance Computing.

COs	POs												PSOs	
	PO 1	PO 2	PO3	PO 4	PO5	PO 6	P0 7	PO 8	PO 9	P01 0	PO1 1.	PO1 2	PSO 1	PSO 2
CO1	3	3	3	1	-	-	-	-	1	-	-	2	2	2
CO2	3	3	3	1	-	-	-	-	1	-	-	1	2	2
CO3	3	3	3	1	-	1	-	-	1	-	-	1	2	2
CO4	3	3	3	1	-	-	-	-	1	-	-	2	2	2
CO5	3	3	3	1	-	-	-	-	1	-	-	1	2	2
AVG	**3**	**3**	**3**	**1**		**1**	**-**	**-**	**1**	**-**	**-**	**1.4**	**2**	**2**

MAPPING	LEVEL	JUSTIFICATION
CO1-PO1	3	Knowledge of various machine learning approaches involves solving complex engineering problems
CO1-PO2	3	Principles of mathematics and engineering sciences are used in various aspects of machine learning approaches
CO1-PO3	3	Using the knowledge of supervised learning concepts, we can

B. K. Raghu
H.O.D
Dept. of Computer Science & Engg.
B.G.S. Institute of Technology,
B.G. Nagar - 571 448
Nagamangala Tq, Mandya Dist
Karnataka (INDIA)

		design and develop solutions for complex engineering problems
CO1-PO4	1	Supervised learning concepts can be used to conduct experiments to provide valid conclusions
CO1-PO9	1	Expertise developed, which will enable the student to become a productive member of a design team
CO1-PO12	2	The student will become aware of the need for lifelong learning and the continued upgrading of technical knowledge
CO1-PSO1	2	Knowledge of various machine learning approaches involves solving complex engineering problems
CO1-PSO2	2	machine learning approaches involves Data collection
CO2-PO1	3	Knowledge of theoretical foundations of decision trees involves solving complex engineering problems
CO2-PO2	3	Principles of mathematics and engineering sciences are used in theoretical foundations of decision trees to identify best split and Bayesian classifier to label data points.
CO2-PO3	3	Knowledge of theoretical foundations of decision trees to identify best split can be used to design and develop solutions for complex engineering problems
CO2-PO4	1	Theoretical foundations of decision trees to identify best split and Bayesian classifier to label data points. knowledge can be used to design and conduct experiments to provide valid conclusions
CO2-PO9	1	Expertise developed, which will enable the student to become a productive member of a design team
CO2-PO12	1	The student will become aware of the need for lifelong learning and the continued upgrading of technical knowledge
CO1-PSO1	2	Decision trees involves solving mathematical engineering problems
CO1-PSO2	2	Decision trees approaches involves Data collection
CO3-PO1	3	Knowledge of classifier models, neural network applications helps in solving complex engineering problems
CO3-PO2	3	Principles of mathematics and engineering sciences are used in various aspects of neural network and classifier models

L. K Hiremath
Dept. of Computer Science & Engg.
B. G. S. Institute of Technology,
B. G. Nagar 571 448
Mandya Dist

CO3-PO3	3	Knowledge of neural network and classifier models can be used to design and develop solutions for complex engineering problems
CO3-PO4	1	Various neural network and classifier models knowledge can be used to design and conduct experiments to provide valid conclusions
CO3-PO6	1	Knowledge of theoretical foundations of neural network to identify best split to label data points. will help understand issues and societal problems related to cybercrimes and computer hacking
CO3-PO9	1	Expertise developed, which will enable the student to become a productive member of a design team
CO3-PO12	1	The student will become aware of the need for lifelong learning and the continued upgrading of technical knowledge
CO1-PSO1	2	Knowledge of neural network applications helps in solving c engineering problems
CO1-PSO2	2	Neural network application problems approaches involves Data collection
CO 4-PO1	3	Knowledge of theoretical foundations of Bayesian network involves solving complex engineering problems
CO 4-PO2	3	Principles of mathematics and engineering sciences are used in Bayesian classifier to label data points.
CO 4-PO3	3	Knowledge of theoretical foundations of Bayesian classifier is used to design and develop solutions for complex engineering problems
CO 4-PO4	1	Theoretical foundations of Bayesian classifier to label data points. knowledge can be used to design and conduct experiments to provide valid conclusions
CO4-PO9	1	Expertise developed, which will enable the student to become a productive member of a design team

B. K. Rao
HOD
Dept. of Computer Science & Engg.
B.G.S. Institute of Technology,
B.G. Nagar - 571 448.
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

CO4-PO12	2	The student will become aware of the need for lifelong learning and the continued upgrading of technical knowledge
CO1-PSO1	2	Knowledge of theoretical foundations of Bayesian network involves solving engineering problems
CO1-PSO2	2	Bayesian network approaches involves Data collection and management
CO5-PO1	3	Study of HMM involves solving complex engineering problems
CO5-PO2	3	Study of HMM involves principles of mathematics and engineering
CO5-PO3	3	Sequence emission probability evaluation knowledge can be used to design and develop solutions for complex engineering problems
CO5-PO4	1	State sequence identification and sequence emission probability evaluation skills can be used to design and conduct experiments to provide valid conclusions
CO5-PO9	1	Expertise developed, which will enable the student to become a productive member of a design team
CO5-PO12	1	The student will become aware of the need for lifelong learning and the continued upgrading of technical knowledge
CO1-PSO1	2	Study of HMM involves solving Mathematical engineering problems
CO1-PSO2	2	Evaluate hypothesis involves Data collection and management

Signature of Course Coordinator **Signature of Module Coordinator** **Signature of HoD**

Dept. of Computer Science & Engg.
B.G.S. Institute of Technology,
B.G. Nagar - 571 448.
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

Seventh Semester B.E. Degree Examination, Dec.2018/Jan.2019
Machine Learning

Time: 3 hrs. Max. Marks: 80

Note: ***Answer FIVE full questions, choosing ONE full question from each module.***

Module-1

1 a. Specify the learning task for 'A checkers learning problem'. (03 Marks)

b. Discuss the following with respect to the above,
   (i) Choosing the training experience.
   (ii) Choosing the target function and
   (iii) Choosing a function approximation algorithm. (09 Marks)

c. Comment on the issues in machine learning. (04 Marks)

OR

2 a. Write candidate elimination algorithm. Apply the algorithm to obtain the final version space for the training example. (10 Marks)

Sl. No.	Sky	Air temp	Humidity	Wind	Water	Forecast	Enjoy sport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

b. Discuss about an unbiased Learner. (06 Marks)

Module-2

3 a. What is a decision tree & discuss the use of decision tree for classification purpose with an example. (08 Marks)

b. Write and explain decision tree for the following transactions: (08 Marks)

Tid	Refund	Martial status	Taxable Income	Cheat
1	Yes	Single	125 K	No
2	No	Married	100 K	No
3	No	Single	70 K	No
4	Yes	Married	120 K	No
5	No	Divorced	95 K	Yes
6	No	Married	60 K	No
7	Yes	Divorced	220 K	No
8	No	Single	85 K	Yes
9	No	Married	75 K	No
10	No	Single	90 K	Yes

OR

4 a. For the transactions shown in the table compute the following :
   (i) Entropy of the collection of transaction records of the table with respect to classification.
   (ii) What are the information gain of a_1 and a_2 relative to the transactions of the table? (08 Marks)

Instance	1	2	3	4	5	6	7	8	9
a_1	T	T	T	F	F	F	F	T	F
a_2	T	T	F	F	T	T	F	F	T
Target class	+	+	–	+	–	–	–	+	–

b. Discuss the decision learning algorithm. (04 Marks)

c. List the issues of decision tree learning. (04 Marks)

Module-3

5 a. Draw the perceptron network with the notation. Derive an equation of gradient descent rule to minimize the error. (08 Marks)

b. Explain the importance of the terms : (i) Hidden layer (ii) Generalization (iii) Overfitting (iv) Stopping criterion (08 Marks)

OR

6 a. Discuss the application of Neural network which is used for learning to steer an autonomous vehicle (06 Marks)

b. Write an algorithm for back propagation algorithm which uses stochastic gradient descent method. Comment on the effect of adding momentum to the network. (10 Marks)

Module-4

7 a. What is Bayes theorem and maximum posterior hypothesis? (04 Marks)

b. Derive an equation for MAP hypothesis using Bayes theorem. (04 Marks)

c. Consider a football game between two rival teams. Team 0 and Team 1. Suppose Team 0 wins 95% of the time and Team 1 wins the remaining matches. Among the games won by team 0, only 30% of them come from playing on teams 1's football field. On the otherhand, 75% of the victories for team 1 are obtained while playing at home. If team 1 is to host the next match between the two teams, which team will most likely emerge as the winner? (08 Marks)

OR

8 a. Describe Brute-force MAP learning algorithm (04 Marks)

b. Discuss the Naive Bayes classifier (04 Marks)

c. The following table gives data set about stolen vehicles. Using Naive bayes classifier classify the new data (Red, SUV, Domestic) (08 Marks)

Table

Color	Type	Origin	Stolen
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

Module-5

9 a. Write short notes on the following

(i) Estimating Hypothesis accuracy

(ii) Binomial distribution (08 Marks)

b. Discuss the method of comparing two algorithms. Justify with paired to tests method. (08 Marks)

OR

10 a. Discuss the K-nearest neighbor language. (04 Marks)

b. Discuss locally weighted Regression. (04 Marks)

c. Discuss the learning tasks and Q learning in the context of reinforcement learning. (08 Marks)

Seventh Semester B.E. Degree Examination, Dec.2019/Jan.2020
Machine Learning

Time: 3 hrs. Max. Marks: 80

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

1 a. What do you mean by well-posed learning problem? Explain with example. (04 Marks)
b. Explain the various stages involved in designing a learning system in brief. (08 Marks)
c. Write Find_S algorithm and discuss the issues with the algorithm. (04 Marks)

OR

2 a. List the issues in machine learning. (04 Marks)
b. Consider the given below training example which finds malignant tumors from MRI scans.

Example	Shape	Size	Color	Surface	Thickness	Target concept
1	Circular	Large	Light	Smooth	Thick	Malignant
2	Circular	Large	Light	Irregular	Thick	Malignant
3	Oval	Large	Dark	Smooth	Thin	Benign
4	Oval	Large	Light	Irregular	Thick	Malignant
5	Circular	Small	Light	Smooth	Thick	Benign

Show the specific and general boundaries of the version space after applying candidate elimination algorithm. (Note: Malignant is +ve, Benign is –ve). (08 Marks)
c. Explain the concept of inductive bias in brief. (04 Marks)

Module-2

3 a. Discuss the two approaches to prevent over fitting the data. (08 Marks)
b. Consider the following set of training examples:

Instance	Classification	a_1	a_2
1	1	1	1
2	1	1	1
3	0	1	0
4	1	0	0
5	0	0	1
6	0	0	1

(i) What is the entropy of this collection of training examples with respect to the target function classification?
(ii) What is the information gain of a_2 relative to these training examples? (08 Marks)

OR

4 a. Define decision tree. Construct the decision tree to represent the following Boolean functions:
i) $A \wedge \neg B$ ii) $A \vee [B \wedge C]$ iii) A XOR B (06 Marks)
b. Write the ID3 algorithm. (06 Marks)
c. What do you mean by gain and entropy? How it is used to build the decision tree. (04 Marks)

Module-3

5 a. Define perceptron. Explain the concept of single perceptron with neat diagram. (06 Marks)

b. Explain the back propagation algorithm. Why is it not likely to be trapped in local minima? (10 Marks)

OR

6 a. List the appropriate problems for neural network learning. (04 Marks)

b. Discuss the perceptron training rule and delta rule that solves the learning problem of perceptron. (08 Marks)

c. Write a remark on representation of feed forward networks. (04 Marks)

Module-4

7 a. Explain Naïve Bayes classifier. (08 Marks)

b. Explain brute force MAP learning algorithm. (08 Marks)

OR

8 a. Discuss Minimum Description Length principle in brief. (08 Marks)

b. Explain Bayesian belief networks and conditional independence with example. (08 Marks)

Module-5

9 a. Define: (i) Simple Error (ii) True Error (04 Marks)

b. Explain K-nearest neighbor learning algorithm. (08 Marks)

c. What is reinforcement learning? (04 Marks)

OR

10 a. Define expected value, variance, standard deviation and estimate bias of a random variable. (04 Marks)

b. Explain locally weighted linear regression. (08 Marks)

c. Write a note on Q-learning. (04 Marks)

Seventh Semester B.E. Degree Examination, June/July 2019
Machine Learning

Time: 3 hrs. Max. Marks: 80

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

1 a. Define machine learning. Describe the steps in designing learning system. (08 Marks)
b. Write Find-S algorithm and explain with example. (04 Marks)
c. Explain List-Then-Eliminate algorithm. (04 Marks)

OR

2 a. List out any 5 applications of machine learning. (05 Marks)
b. What do you mean by hypothesis space, instance space and version space? (03 Marks)
c. Find the maximally general hypothesis and maximally specific hypothesis for the training examples given in the table using candidate elimination algorithm. (08 Marks)

Day	Sky	Air Temp	Humidity	Wind	Water	Forecast	Enjoy Sport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Module-2

3 Construct decision tree for the following data using ID3 algorithm.

Day	A1	A2	A3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	No

(16 Marks)

OR

4 a. Explain the concept of decision tree learning. Discuss the necessary measure required to select the attributes for building a decision tree using ID3 algorithm. (08 Marks)
b. Discuss the issues of avoiding over fitting the data, handling continuous data and missing values in decision trees. (08 Marks)

Module-3

5 a. Explain artificial neural network based on perception concept with diagram. (06 Marks)
b. What is gradient descent and delta rule? Why stochastic approximation to gradient descent needed? (04 Marks)
c. Describe the multilayer neural network. Explain why back propagation algorithm required. (06 Marks)

OR

6 a. Derive the back propagation rule considering the output layer and training rule for output unit weights. (08 Marks)

b. What is squashing function 3 why is it needed? (04 Marks)

c. List out and explain in briefly representation power of feed forward networks. (04 Marks)

Module-4

7 a. Explain maximum a posteriori (MAP) hypothesis using Bayes theorem. (06 Marks)

b. Estimate conditional probabilities of each attributes {colour, legs, height, smelly} for the species classes: {M, H} using the data given in the table. Using these probabilities estimate the probability values for the new instance – (Colour = Green, Legs = 2, Height = Tall and Smelly = No) (10 Marks)

No	Colour	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

OR

8 a. Explain Naive Bayes classifier and Bayseian belief networks. (10 Marks)

b. Prove that how maximum likelihood (Bayesian learning) can be used in any learning algorithms that are used to minimize the squared error between actual output hypothesis and predicted output hypothesis. (06 Marks)

Module-5

9 a. Explain locally weighted linear regression. (08 Marks)

b. What do you mean by reinforcement learning? How reinforcement learning problem differs from other function approximation tasks. (05 Marks)

c. Write down Q-learning algorithm. (03 Marks)

OR

10 a. What is instance based learning? Explain K-Nearest neighbour algorithm. (08 Marks)

b. Explain sample error, true error, confidence intervals and Q-learning function. (08 Marks)

Machine Learning - 15CS73 Question Bank

Module 1- Introduction to ML and Concept Learning

Introduction to Machine Learning (Chapter 1)

1. Define Machine Learning. Discuss with examples why machine learning is important.
2. Discuss with examples some useful applications of machine learning.
3. Explain how some areas/disciplines that influenced the machine learning.
4. What do you mean by a well–posed learning problem? Explain the important features that are required to well–define a learning problem.
5. Define learning program for a given problem. Describe the following problems with respect to Tasks, Performance and Experience: a. Checkers Learning Problems
   b. Handwritten Recognition Problem
   c. Robot Driving Learning Problem
6. Describe in detail all the steps involved in designing a learning system.
7. Discuss the perspective and issues in machine learning.

Concept Learning (Chapter 2)

8. Define Concept and Concept Learning. With example explain how the Concept Learning task determines the Hypothesis for given target concept.
9. Discuss Concept learning as search with respect to General to specific ordering of hypothesis.
10. Describe Find S Algorithm. What are the properties and complaints of Find S.
11. Illustrate Find S Algorithm over *EnjoySport* concept. Training instances given below.

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

12. Define Consistent *Hypothesis* and *Version Space*. With example explain Version Space and Representation of version Space.
13. Describe List the Eliminate Algorithm.
14. Explain the candidate elimination algorithm.

15. Trace Candidate-Elimination algorithm on the following data.

a)

Origin	*Manufacturer*	*Color*	*Decade*	*Type*	*Example Type*
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive

b)

Origin	*Manufacturer*	*Color*	*Decade*	*Type*	*Example Type*
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive
Japan	Toyota	Green	1980	Economy	Positive
Japan	Honda	Red	1990	Economy	Negative

c)

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

16. Explain the inductive biased hypothesis space, unbiased learner and the futility of Bias Free Learning. Describe the three types of learner.

17. What is the role of a function approximation algorithm? How does learner system estimate training values and adjusts weights while learning?

18. Describe in brief: Version spaces and Candidate –Elimination Algorithm.

19. Define Inductive Learning Hypothesis.

20. Describe Inductive Systems and Equivalent Deductive Systems

21. Rank the following three types of leaners according to their biases: a. Rote Learner

b. Candidate Elimination Learner

c. Find S Learner.

Module 2- Decision Tree Learning

Decision Tree Learning (Chapter 3)

1. Explain the following with examples: a. Decision Tree,
   b. Decision Tree Learning
   c. Decision Tree Representation.

2. What are appropriate problems for Decision tree learning? OR
   What are the characteristics of the problems suited for decision tree learning?
3. Explain the concepts of entropy and information gain.
4. Describe the ID3 algorithm for decision tree learning with example.

OR

What is the procedure of building Decision tree using ID3 with Gain and Entropy. Illustrate with example.

OR

What do you mean by Gain and Entropy? How is it used to build the Decision tree in algorithm? Illustrate using an example.

5. Give Decision trees to represent the Boolean Functions: a. A && - B
   b. A V [B && C]
   c. A XOR B
   d. [A&&B] V [C&&D]

6. Consider the following set of training examples.

a. What is the entropy of this collection of training example with respect to the target function classification?

b. What is the information gain of A2 relative to these training examples?

Instance	Classification	A1	A2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

7. Discuss Hypothesis Space Search in Decision tree Learning.

8. Discuss Inductive Bias in Decision Tree Learning. Differentiate between two types of biases. Why prefer Short Hypotheses?

9. What are issues in decision tree learning? Explain briefly How are they overcome? a. Discuss the following issues in detail:
   a. Avoiding overfitting in Decision Trees
   b. Incorporating Continuous valued attributes
   c. Handling Training Examples with Missing attribute values.
   d. Handling Attributes with Different costs.

10. Other solved examples in the class.

Module 3- Artificial Neural Networks
Artificial Neural Networks (Chapter 4)

1. What is Artificial Neural Network?
2. What are the types of problems in which Artificial Neural Network can be applied.
3. Write a note on Representational Power of Perceptron
4. What is linearly in separable problem? Design a two-layer network of perceptron to implement a) X OR Y b) X AND Y
5. Explain the concept of a Perceptron with a neat diagram.
6. Discuss the Perceptron training rule.
7. Define Delta Rule.
8. Under what conditions the perceptron rule fails and it becomes necessary to apply the delta rule
9. What do you mean by Gradient Descent?
10. Derive the Gradient Descent Rule.
11. What are the conditions in which Gradient Descent is applied.
12. What are the difficulties in applying Gradient Descent.
13. Explain the importance of Stochastic Gradient Descent
14. Differentiate between Gradient Descent and Stochastic Gradient Descent
15. Differentiate between Gradient Descent and Perceptron training rule.
16. Derive the Backpropagation rule considering the training rule for Output Unit weights and Training Rule for Hidden Unit weights
17. Write the algorithm for Back propagation.
18. Explain how to learn Multilayer Networks using Backpropagation Algorithm.
19. What is Squashing Function?
20. Briefly explain the following with respect to Backpropagation a) Convergence and Local Minima of MLP

    b) Representational Power of Feedforward Networks

    c) Generalization, Overfitting, and Stopping Criterion

Module 4- Bayesian Learning

Bayesian Learning (Chapter 6)

1. Define (i) Prior Probability (ii) Conditional Probability (iii) Posterior Probability

2. Define Bayesian theorem? What is the relevance and features of Bayesian theorem? Explain the practical difficulties of Bayesian theorem.

3. Consider a medical diagnosis problem in which there are two alternative hypotheses: 1. That the patient has a particular form of cancer (+) and 2. That the patient does not (-). A patient takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, .008 of the entire population have this cancer. Determine whether the patient has Cancer or not using MAP hypothesis.

4. Explain Brute force Bayes Concept Learning

5. Define MAP hypothesis. Derive the relation for h_{MAP} using Bayesian theorem.

6. What are Consistent Learners?

7. Discuss Maximum Likelihood and Least Square Error Hypothesis.

8. Describe Maximum Likelihood Hypothesis for predicting probabilities.

9. Describe the concept of MDL. Obtain the equation for h_{MDL}

10. What is conditional Independence?

11. Explain Naïve Bayes Classifier with an Example.

12. Explain the Gradient Search to Maximize Likelihood in a neural Net.

13. What are Bayesian Belief nets? Where are they used?

14. Explain Bayesian belief network and conditional independence with example.

15. Explain the concept of EM Algorithm. Discuss what are Gaussian Mixtures.

MODULE-5 : EVALUATION HYPOTHESIS, INSTANCE BASED LEARNING, REINFORCEMENT LEARNING

1. Explain the basic definitions of sampling theory.
2. Explain the binomial distribution in detail.
3. Explain the nominal distribution in detail.
4. What is instance based learning?
5. **Define Central Limit Theorem**

B. K. Rag

BGS Institute of Technology

Department: Computer Science and Engineering/Information Science and Engineering

Test: I

Semester: VII **Section**: A/B USN:

Subject Name & Code: Machine Learning Techniques (17CS73)

Instructions

Duration: 60 minutes Max. Marks: 30

i)Select one question from each part.

ii)All questions carry equal marks.

Date: 06:11:2020

Time: 10:00AM TO 11:00AM.

Question Number	Questions	Marks	CO	Levels
	PART - A			
1	a) What is the procedure to building a decision tree using ID3 algorithm, explain with Gain and Entropy. Illustrate with Example?	7.5	2	2
	b) Define Consistent Hypothesis and version space. Write List-then-Eliminate Algorithm	7.5	1	2
	OR			
2	a) Apply the Candidate Elimination algorithm for the given example.	7.5	1	3

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

	b) Explain appropriate problems for Decision tree Learning	7.5	1	2
	PART – B			
3	a) What do you mean by a well posed learning problem? Explain its features with example.	7.5	1	2
	b) Discuss Hypothesis Space Search in Decision tree Learning.	7.5	2	2
	OR			
4	a) Determine the root of the Decision tree using Id3 Algorithm.	7.5	2	3

Sl. No	Student	FirstLast year	Male	workshard	sleep	Target
1	Richard	Yes	Yes	No	Yes	Yes
2	Alan	Yes	Yes	Yes	No	Yes
3	Alison	No	No	Yes	No	Yes
4	Jeff	No	Yes	No	Yes	No
5	Gail	Yes	No	Yes	Yes	Yes
6	Simon	No	Yes	Yes	Yes	No

	b) Explain the Issues of Machine Learning? Explain applications of ML	7.5	1	2

Signature of Staff

B. K. Raghu HOD

Signature of HOD

CBCS Scheme (VTU)

DEPARTMENT: Computer Science & Engg. ~~ELECTRONICS AND COMMUNICATION ENGINEERING~~

Scheme & Solution – TEST – I Date: 06.11.2020

Semester: VII Subject Title: Machine Learning Subject Code: 15CS73

Question Number	Solution	Marks Allocated
1. a)	ID3 (Examples, Target_attribute, Attributes) • Create a Root node for the tree • If all eg are +, Return the single node tree Root, with label + • If all eg. are −, Return the Single-node tree Root, with label − • If Attributes is empty, Return the single-node tree Root, with label = most common value of TA in Eg. • otherwise Begin. • A ← the attribute from attributes the best * classifies Ex • The decision attribute for Root ← A • For each possible value, vi of A • Add a new tree branch below Root, corresponding to the test A=vi • Let Examples vi, be the subset of Ex that have value vi for A • If Ex vi, is empty • Then below this new branch add a leaf node with label = most common value of Target-attribute in Ex • Else below this new branch add the subtree ID3 (Examples vi, Target, A) • End • Return Root. $Entropy(S) = -p_+ \log p_+ - p_- \log_2 p_-$ $Gain(S,A) = Entropy(S) - \sum_{v \in values(A)} \frac{\lvert S_v \rvert}{\lvert S \rvert} Entropy(S_v)$	5½ 1 1
b	1. Version space ← a list containing every hypothesis in H 2. For each training example (x, c(x)) Remove from version space any hypothesis h for which h(x) ≠ c(x) 3. Output the list of hypothesis in version space + Explanation	4½ 3
2. (a)	S_0 ⟨∅, ∅, ∅, ∅, ∅, ∅⟩ S_1 ⟨Sunny, warm, Normal, strong, warm, same⟩ $S_2 S_3$ ⟨Sunny, warm, ?, strong, warm, same⟩ S_4 ⟨Sunny, warm, ?, strong, ?, ?⟩ G_4 ⟨Sunny, ?????⟩ ⟨?, warm, ????⟩ G_3 ⟨Sunny, ?????⟩ ⟨?, warm, ????⟩ ⟨????same⟩ G_1 G_2 ⟨? ? ? ? ? ?⟩	7½

Question Number	Solution	Marks Allocated
2 b)	i) Instances are represented by attribute-value pairs. ii) The target fun has discrete o/p values. iii) Disjuctive descriptions may be required. iv) The training data may contain errors. v) The training data may contain missing attr. values.	$1\frac{1}{2} \times 5 = 7.5$
3. (a)	A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at Tasks in T as measured by P improves with experience E.	2
	Three features are • The class of Tasks • The measure of performance to be improved • Source of Experience. + Explanation	2 $3\frac{1}{2}$
b)	* Searching a space of hypotheses for one that fits the training ex. * The hypothesis space searched by ID3 * Its a simple to complex, hill climbing search.	
4. (a)	1. $E(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$ $= 0.918$ 2. $IG = E(S) - \sum_V \frac{\lvert S_V \rvert}{V} E(S_V)$ $V(\text{Instance}) = 6$ First last year: Y → [3, 0] = 0; NO → [1, 2] $= \frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3} = 0.918$ male: Yes → [2, 2] = 1; NO → [2, 0] = 0 $= 0.918 - [\frac{4}{6} \times 1 + \frac{2}{6} \times 0]$ $= 0.251$	

Question Number	Solution	Marks Allocated

workhard

yes: [3, 1]

NO: [1, 1] = 1

$= \frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}$

$= 0.8112$

$IG = 0.918 - \left[\frac{4}{6} \times 0.8112 + \frac{2}{6} \times 1\right]$

$= 0.043$

sleep

Yes: [2, 2] = 1

NO: [2, 0] = 0

$IG = 0.918 - \left[\frac{4}{6} \times 1\right]$

$= 0.251$

step3 = IG (FirstLastyear) = 0.459

FirstLastyear

yes: [3, 0]

NO: [1, 2]

Step 4 :

	st	male	workhard	sleep	[illegible]
3.	Alisn	NO	yes	NO	Yes
4	Jeff	Yes	NO	YES	NO
6	Simon	Yes	Yes	Yes	NO

V = 3

E(S) = 0.918

male

yes: [0, 2] = 0

NO: [1, 0] = 0

$IG = 0.918 - \left[\frac{2}{3} \times 0 + \frac{1}{3} \times 0\right]$

$= 0.918$

workhard

yes: [1, 1] = 1

NO: [0, 1] = 0

$= 0.918 - \left[\frac{2}{3} \times 1\right]$

$= 0.251$

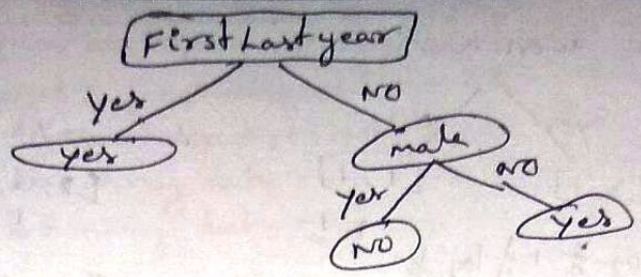
sleep

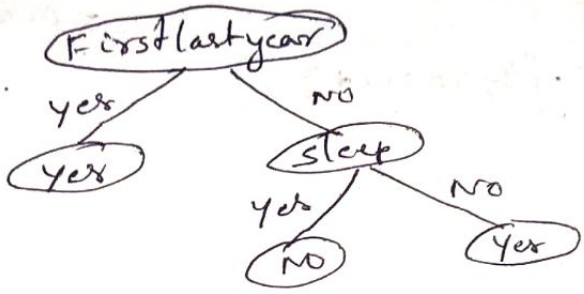
Yes: [0 2] = 0

NO: [1, 0] = 0

$= 0.918$

Question Number	Solution	Marks Allocated





7½

4.

b) Issues in ML.

1. what alg exist for learning general target fns from specific training examples? which alg performs best for which types of problems & representations?
2. How much training data is sufficient?
3. when & how can prior knowledge held by the learner guide the process of generalizing from examples.
4. what is the best strategy for choosing a useful next training ex
5. what is the best way to reduce the learning task to one (or) more fns approximation probs + explanation.

7½

B. K. Rag

HOD
Dept. of Computer Science & Engg
B. G. S. Institute of Technology
B. G. Nagar 571 448
Adichunchanagiri Tq. Mandya Dist
Karnataka (INDIA)

BGSIT BG Nagara	Doc. Title: Internal Test Question Paper		Doc. No.: 06#Form#02b
	Page 1 of 1	Date: 15.10.2018	Rev. No. 00

Internal Test Question Paper Format – CBCS Scheme (VTU)

Name of the Faculty/s: **Prasanna Kumar M J**

Date:08/12/2020 Signature: M.J Prasanna

Reviewer's Signature: B.K. Raghu

NOTE: Only the following information is to be given to the students.

BGS Institute of Technology

Department: Computer Science and Engineering/Information Science and Engineering

Test: II

USN:

Semester: VII **Section**: A/B

Subject Name & Code: Machine Learning Techniques(15CS73)

Instructions

Duration: 60 minutes Max. Marks: 30

i) Select one question from each part.

ii) All questions carry equal marks.

Question Number	Questions	Marks	CO	Levels
	PART - A			
1	a) State the Problems faced in Neural networks?	7.5	CO3	L2
	b) How a single perceptron can be used to represent the Boolean functions such as a AND , OR	7.5	CO3	L3
	OR			
2	a) Derive the Back Propagation Rule with Steps case 1 and case 2 in detail.	7.5	CO3	L3
	b) Explain the concept of Perceptron with neat diagram.	7.5	CO3	L2
	PART – B			
3	a) Discuss the Hypothesis space search and inductive bias w.r.t to back propagation algorithm.	7.5	CO3	L2
	b) Write the difference between stochastic and Gradient Descent technique?.	7.5	CO3	L2
	OR			
4	a) Explain the representation for power of perceptron w.r.t Back propagation algorithm	7.5	CO3	L2
	b) Discuss the following i) Convergence and local minima ii) Over fitting and Under fitting.	7.5	CO3	L2

Note: The Choice question should satisfy same COs and levels.

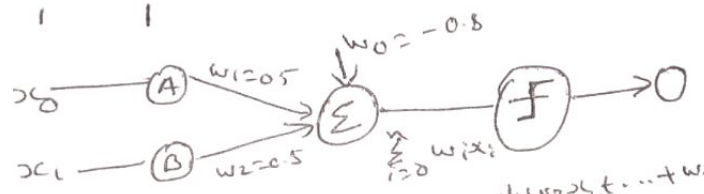
06#Form#02b-0

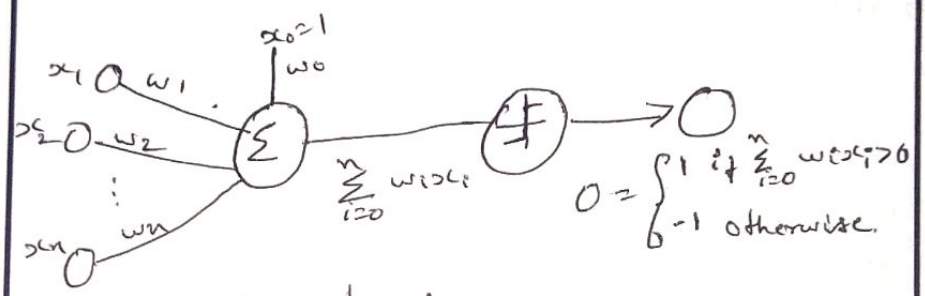
CBCS Scheme (VTU)

DEPARTMENT: COMPUTER SCIENCE & ENGINEERING

Scheme & Solution – TEST – II Date: 08/12/2020

Semester: VII Subject Title: Machine Learning Subject Code: 17CS73

Question Number	Solution	Marks Allocated
1. (a)	Instances are having many attribute value pairs. → The training examples are real valued, descrete valued or a vector of real valued or a vector of real valued and discrete valued attributes → Training function may contain error. + Explanation.	7.5
1. (b)	Boolean func AND A B A∧B 0 0 0 0 1 0 1 0 0 1 1 1 + set $w_0 = 0.8$ $w_1 = 0.5$ $w_2 = 0.5$  $O(x_1 \ldots x_n) = \begin{cases} 1 & \text{if } w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n > 0 \\ -1 & \text{otherwise} \end{cases}$ 1) if $A=0$ & $B=0$ (x_1, x_2) → $-0.8 + (0.5*0) + (0.5*0) = -0.8 < 0$ $= 0$ 2) $A=0$ & $B=1$ = $0.3 < 0$ so $= 0$ 3) $A=1$ & $B=1$ = $-0.3 < 0$ = so 0 4) $A=1$ & $B=1$ = $0.2 > 0$ output $= 1$	½ 3½
	A B A∨B 0 0 0 0 1 1 1 0 1 1 1 1 1) $A=0$ $B=0$ → $-0.3 < 0$ o/p = 0 2) $A=0$ $B=1$ → $0.2 > 0$ o/p = 1 3) $A=1$ $B=0$ → $0.2 > 0$ o/p = 1 4) $A=1$ $B=1$ → $0.7 > 0$ o/p = 1	3½

Question Number	Solution	Marks Allocated
2. (a)	Case 1: Training Rule for output unit weights $\Delta w_{ji} = -\eta \frac{\delta E_d}{\delta w_{ji}} = \eta (t_j - o_j) o_j (1-o_j) x_{ji}$	3½
	Case 2: Training Rule for Hidden unit weights $\delta_j = o_j (1-o_j) \sum_{k \in Downstream(j)} \delta_k w_{kj}$ and $\Delta w_{ji} = \eta \delta_j x_{ji}$	4
2. (b)	 + Explanation.	3½ 4
3. (a)	• Hypothesis space is the n-dimensional euclidean space of the n n/w weights and hyp is continuous • It is different characteristics. + Explanation	7½
(b)	* Standard gradient descent the error is summed over all examples before updating weights, whereas in stochastic gd weights are updated upon examining each training eg * Summing over multiple examples in standard gradient descent requires more computation per weight update step. standard gd is often used with a larger step size per weight update than stochastic gd + explanation	7½

Question Number	Solution	Marks Allocated
4. (a)	Feed forward networks are depends on the width and depth of the networks.	1½
	1. Boolean functions. 2. Continuous functions 3. Arbitrary functions + Explanation	6
4. (b)	* Convergence and local minima • Guaranteed to converge toward some local minimum in E & not necessarily to the global minimum error. * Highly effective ~~fuzz~~ approximator method in practice. * Local minima can be gained by considering the manner in which network weights evolve the no of training iterations increases.	4
	* Over fitting and Under fitting * Underfitting can be avoided by using more data & reducing the features - low variance & High bias. * Overfitting :- It uses more data, inaccurate entry, doesn't categorize. - high variance & low bias.	3½

B. K. Rai

H O D
Dept. of Computer Science & Eng
Institute of Technology

BGSIT BG Nagara	Doc. Title: Internal Test Question Paper		Doc. No.: 06#Form#02b
	Page 1 of 1	Date: 15.10.2018	Rev. No. 00

Internal Test Question Paper Format – CBCS Scheme (VTU)

Name of the Faculty/s: Prasanna Kumar M J

Date:11/01/2021 Signature: M. J Prasanna

Reviewer's Signature: B. K. Ragh

NOTE: Only the following information is to be given to the students.

BGS Institute of Technology

Department: Computer Science and Engineering/Information Science and Engineering

Test: III

USN:

Semester: VII **Section**: A/B

Subject Name & Code: Machine Learning Techniques(17CS73)

Instructions

Duration: 60 minutes Max. Marks: 30

i) Select one question from each part.

Question Number	Questions	Marks	CO	Levels
	PART - A			
1	a) Describe EM Algorithm.	7.5	CO4	L2
	b) Explain the difficulties in evaluating the hypothesis	7.5	CO5	L2
	OR			
2	a) Brute force bayes concept learning.	7.5	CO4	L2
	b) Explain Sampling theorem with example.	7.5	CO5	L2
	PART – B			
3	a) A dataset of n=10 samples and the sample error error(h)=0.4 , Find the probability for x=7. Use the Binomial distribution find the standard deviation and variance.	7.5	CO5	L3
	b) Explain Maximum Likelihood and least square error hypothesis	7.5	CO4	L2
	OR			
4	a) What is locally weighted regression method? Explain briefly.	7.5	CO5	L2
	b) Explain naïve bayes classifier with following example classify (yes, Blue,None) as male or female	7.5	CO4	L3

Name	over 6 ft	Eye color	Hairstyle	Sex
Morgan	Yes	Blue	Long	Female
Bob	NO	Brown	None	male
Vincent	Yes	Brown	Short	male
Amanda	NO	Brown	Short	Female
Reid	NO	Blue	Short	male
Lauren	NO	purple	Long	Female
Elise	Yes	Brown	Long	Female

ii)

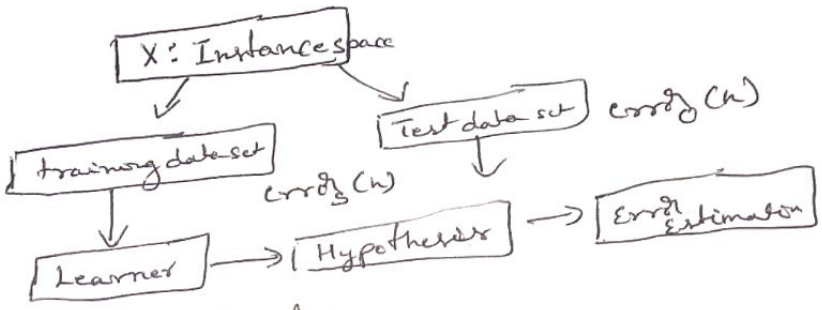
Name	over 6ft	Eyecolor	Hairsty	Sex
Amanda	NO	Brown	Short	?

CBCS Scheme (VTU)

DEPARTMENT: COMPUTER SCIENCE & ENGINEERING

Scheme & Solution – TEST – III Date: 11/01/2021

Semester: VII **Subject Title: Machine Learning** **Subject Code:** 17CS73

Question Number	Solution	Marks Allocated
1. (a)	It can be used even for variables whose value is never directly observed. → Its widly used approach to learning in the presence of unobserved variables. $E[z_{ij}] = \frac{P(x=x_i \mid \mu=\mu_j)}{\sum_{i=1}^{n} P(x=x_i \mid \mu=\mu_n)}$ $\mu_j \leftarrow \frac{\sum_{i=1}^{m} E[z_{ij}]\, x_i}{\sum_{i=1}^{m} E[z_{ij}]}$	3 4½
(b)	 + Explanation	3 4½
2. (a)	Alg: Given an instance space X (Training data) - A hypothesis space H - The target concept $C: X \to \{0,1\}$ - A sequence of instances. $\langle\langle x_1, d_1\rangle, \langle x_2, d_2\rangle \ldots \langle x_m, d_m\rangle$ Step 1: For each hypothesis h in H, find the posterior probability $P(h/D) = \frac{P(D/h) \cdot P(h)}{P(D)}$ Step 2: output the highest posterior probability hmap $\boxed{h_{map} = \underset{h \in H}{\operatorname{argmax}}\, P(h/D)}$ + Expl	7½

Question Number	Solution	Marks Allocated
2. (b)	1. Random variable: 2. Probability distribution 3. mean value or Expected outcome: 4 variance: + Explanation	7½
3. (a)	$P(x=r) = {}^{n}C_{r}\, \text{error}_s(h)^{r} (1-\text{error}_s(h))^{n-r}$ (1) $P(x=7) = {}^{10}C_{7} (0.4)^{7} (1-0.4)^{10-7}$ $= 0.04$ (2) variance $= n \cdot p \cdot (1-p)$ $= 10 \cdot \text{error}_s(h)\, (1-\text{error}_s(h))$ $= 10 * 0.4 * 0.6$ $= 2.4$ (3) standard deviation $= \sqrt{Var}$ $= \sqrt{2.4} = 1.54$	
(b)	case 1: The data are mutually linearly dependent. + Explanation case 2: Data are mutually independent or non linear. + Explanation	3½ 4
4. (a)	The phrase "locally weighted regression" is called local because the func is approximated based only on data near the query point, weighted because the contribution of each training example is weighted by its distance and regression because this is the term used widely in the	

Question Number	Solution	Marks Allocated
	~~Stastecal~~ Statistical learning community for the problem of approximating real-valued funs	3
	+ Explain with derivations.	4½

4. (b)

Step1: Using Naive Bayes classifier.

Male : p(male) . p(Amanda|male) p(NO|male).
p(Brown|male) . p(short|male)

Female: p(Female) . p(Amanda/Female) . p(NO|Female)
p(Brown|Female) . p(short/Female).

Step2: Construct the frequency table.

Nmale = 3
p(Nmale) = 3/7
= 0.42

Nfemale = 4
p(Nfemale) = 4/7
= 0.57.

	Male	Female	p(Ai/male)	p(Ai/Female)
Amanda	0	1	0/3 = 0	1/4 = 0.25
NO	2	2	2/3 = 0.66	2/4 = 0.5
Brown	2	2	2/3 = 0.66	2/4 = 0.5
short	2	1	2/3 = 0.6	1/4 = 0.25

Step3: Apply formula step1

male : 0.428 X 0 X 0.6 X 0.6 X 0.6
= 0

Female = 0.571 X 0.25 X 0.5 X 0.5 X 0.25
= 0.0089

step4:

male = $\frac{0}{0+0.0089}$
= 0%

Female = $\frac{0.0089}{0+0.0089}$
= 1 = 100%

B.K.Raj
HOD
Dept. of Computer Science
B.G.S. Institute of Te
B.G. Nagar - 57
Ramanagara Tq, M
Karnataka (IN

|| Jai Sri Gurudev||

B G S INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCEINCE & ENGINEERING

Academic Year: 2020– 2021 **For the Period:** 01/09/2020 to 16/01/2021

Assignment I Questions

Faculty Name: M J PRASANNA KUMARA
Semester: 7th
Course Name: Machine Learning
Course Code: 17CS73

Sl. No.	Questions	COs	Level
1	Write the decision tree using ID3 algorithm, explain with Gain and Entropy. Illustrate with Example?	CO2	L3
2	What are the Issue in Decision tree learning, How they are overcome?	CO2	L2
3	Explain the Concept of Designing and Learning System?	CO1	L2
4	Explain the Issues of Machine Learning? Explain applications of ML	CO1	L2
5	Explain Briefly hypothesis Space Search in Decision Tree Learning?	CO2	L3
6	Illustrate with example candidate elimination algorithm.	CO1	L2

B. K. Rag
HOD
Dept. of Computer Science & Engg
B.G.S. Institute of Technology
B.G. Nagar - 571 448
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

|| Jai Sri Gurudev||

B G S INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCEINCE & ENGINEERING

Academic Year: 2020– 2021 **For the Period:** 01/09/2020 to 16/01/2021

Assignment II Questions

Faculty Name: M J PRASANNA KUMARA
Semester: 7th
Course Name: Machine Learning
Course Code: 17CS73

Sl. No.	Questions	COs	Level
1	State the Problems faced in Neural networks?	CO3	L3
2	Explain the Brute Force MAP hypothesis learner technique?	CO4	L2
3	Derive the Back Propagation Rule with Steps in detail	CO3	L3
4	Write the difference between stochastic and Gradient Descent technique?	CO3	L3
5	Define perceptions ? explain the representation for power of perceptrons?	CO3	L3
6	Explain likelihood least-squared error hypotheses?	CO4	L2
7	Describe the concept of MDL.	CO4	L2
8	Explain EM algorithm	CO4	L2

B. K. Raghu
HOD
Dept. of Computer Science & Engg
B.G.S. Institute of Technology
B.G. Nagar - 571 448
Bellalangala Tq. Mandya Dist
Karnataka (INDIA)

|| Jai Sri Gurudev||

B G S INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCEINCE & ENGINEERING

Academic Year: 2020– 2021 **For the Period:** 01/09/2020 to 16/01/2021

Assignment III Questions

Faculty Name: M J PRASANNA KUMARA
Semester: 7th
Course Name: Machine Learning
Course Code: 17CS73

Sl. No.	Questions	COs	Level
1	1. Explain naïve bias classifier and Bayesian belief network.	CO4	L3
2	Prove that how the maximum likelihood (Bayesian learning) can be used in any learning algorithm that are used to minimize the squared error between actual output hypotheses and predicated output hypothesis.	CO4	L4
3	Explain locally weighted linear regression.	CO5	L3
4	What do you mean by reinforcement learning? How reinforcement learning problem differs from other function approximation tasks.	CO5	L4
5	Write down Q-learning algorithm.	CO5	L3
6	What is instance based learning? Explain k-nearest neighbour algorithm.	CO5	L3
7	Explain sample error, true error, confidence intervals and Q-learning function.	CO5	L3
8	Explain the difficulties in evaluating the hypothesis.	CO5	L2

B. K. Raghu D
Dept. of Computer Science & E
B. G. S. Institute of Technol
B.G. Nagar - 571 448
Nagamangala Tq, Mandya
Karnataka (INDIA)

BGS Institute of Technology

VTU

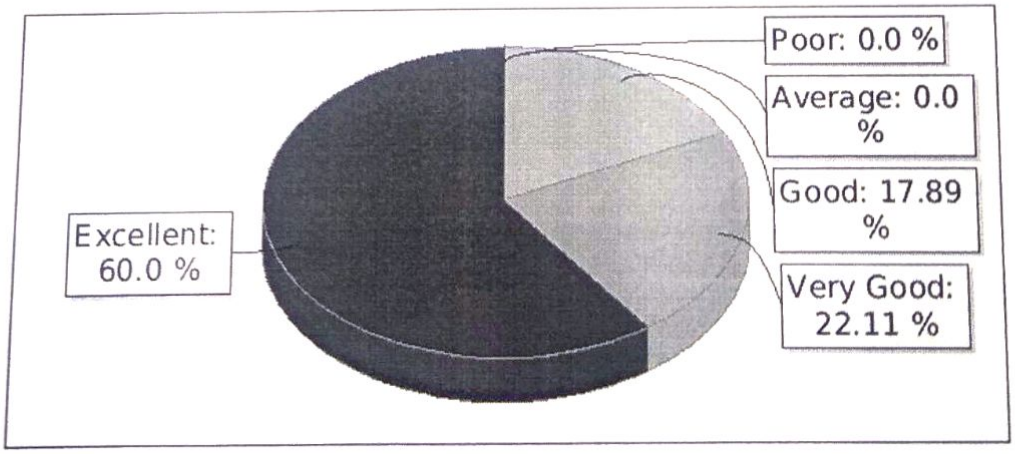
Bengaluru – Hassan National Highway (NH-75), Nagamangala
Taluk, Mandya District, B G Nagar, Karnataka 571448
Ph: 08234 - 288418 , Fax: 288419
Email: principalbgsit@rediffmail.com , Web: www.bgsit.ac.in

Course End Survey 2020-21

Batch : BE , 2017-2021
Staff Name : Mr Prasanna Kumar M J
Subject Code : 17CS73
Subject Name : MACHINE LEARNING

Department :
Semester 7
Date : 18 Jan 2021

No	Questions	Poor	Average	Good	Very Good	Excellent	%	Average Score (5)
		1	2	3	4	5		
	CO							
1	Choose the learning techniques to investigate various training examples	0	0	5	10	23	89.5	4.5
2	Describe the characteristics of learning techniques and solve problems	0	0	7	7	24	88.9	4.4
3	Apply effectively neural networks for appropriate applications	0	0	8	8	22	87.4	4.4
4	Apply Bayesian techniques and derive effectively learning rules	0	0	7	8	23	88.4	4.4
5	Evaluate hypothesis and investigate instant based learning and reinforced learning	0	0	7	9	22	87.9	4.4
	Total Count	0	0	34	42	114	88.4	4.42



B. K. Raghavendra
HOD
Dept. of Computer Science & Engg
B.G.S. Institute of Technology
B.G. Nagar - 571 448
Nagamangala Tq. Mandya Dist
(INDIA)

BGS Institute of Technology

VTU

Bengaluru – Hassan National Highway (NH-75), Nagamangala Taluk, Mandya District, B G Nagar, Karnataka 571448

Ph: 08234 - 288418 , Fax: 288419

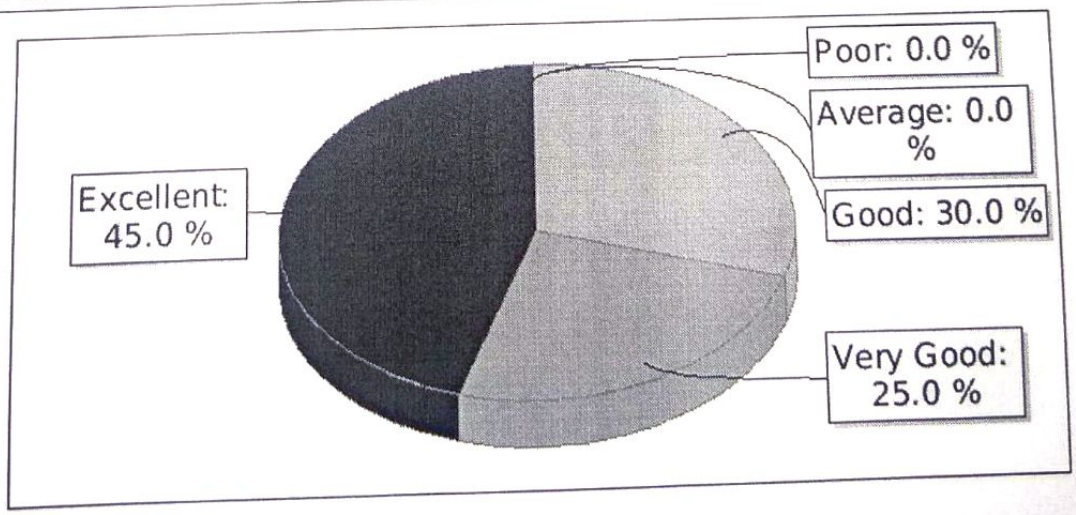
Email: principalbgsit@rediffmail.com , Web: www.bgsit.ac.in

Course End Survey 2020-21

Batch : BE , 2017-2021
Staff Name : Mr Prasanna Kumar M J
Subject Code : 17CS73
Subject Name : MACHINE LEARNING

Department : Computer Science and Engineering,Information Science and Engineering
Semester 7
Date : 18 Jan 2021

No	Questions	Poor	Average	Good	Very Good	Excellent	%	Average Score (5)
		1	2	3	4	5		
	CO							
1	Choose the learning techniques to investigate various training examples	0	0	6	5	9	83	4.2
2	Describe the characteristics of learning techniques and solve problems	0	0	6	5	9	83	4.2
3	Apply effectively neural networks for appropriate applications	0	0	6	5	9	83	4.2
4	Apply Bayesian techniques and derive effectively learning rules	0	0	6	5	9	83	4.2
5	Evaluate hypothesis and investigate instant based learning and reinforced learning	0	0	6	5	9	83	4.2
	Total Count	0	0	30	25	45	83	4.2



B. K. Raghu

Dept. of Computer Science & Eng
B.G. S. Institute of Technology
B.G. Nagar - 571 448
Nagamangala Tq. Mandya D
(INDIA)

BGS Institute of Technology

VTU

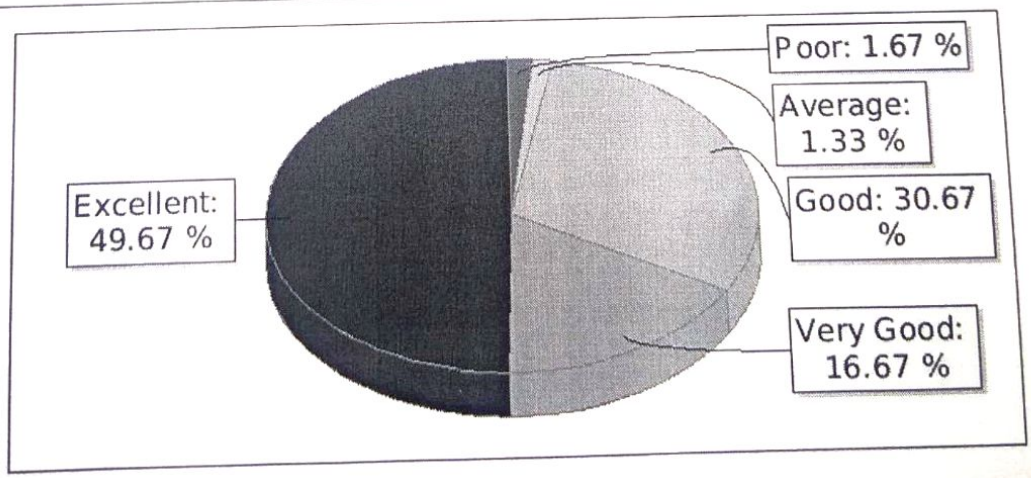
Bengaluru – Hassan National Highway (NH-75), Nagamangala
Taluk, Mandya District, B G Nagar, Karnataka 571448
Ph: 08234 - 288418 , Fax: 288419
Email: principalbgsit@rediffmail.com , Web: www.bgsit.ac.in

Course End Survey 2020-21

Batch : BE , 2017-2021
Staff Name : Mr Prasanna Kumar M J
Subject Code : 17CS73
Subject Name : MACHINE LEARNING

Department : Computer Science and Engineering,Information Science and Engineering
Semester 7
Date : 18 Jan 2021

No	Questions	Poor	Average	Good	Very Good	Excellent	%	Average Score (5)
		1	2	3	4	5		
	CO							
1	Choose the learning techniques to investigate various training examples	1	1	18	8	32	83	4.2
2	Describe the characteristics of learning techniques and solve problems	1	0	19	12	28	82	4.1
3	Apply effectively neural networks for appropriate applications	1	0	19	9	31	83	4.2
4	Apply Bayesian techniques and derive effectively learning rules	1	1	18	11	29	82	4.1
5	Evaluate hypothesis and investigate instant based learning and reinforced learning	1	2	18	10	29	81.3	4.1
	Total Count	5	4	92	50	149	82.3	4.14



B. K. Raghu
HOD
Dept. of Computer Science & Engg
B. G. S. Institute of Technology.
B.G Nagar 571 448
Nagamangala Tq Mandya Dist
Karnataka (NDA)

||Jai Sri Gurudev||

BGS Institute of Technology

BG Nagara-571448, Mandya

Result Analysis

Course Coordinator: Mr. M J Prasanna Kumar

Course Name & Code: Machine Learning 17CS73

Semester & Sec: 7th Sem

Academic Year: 2020-2021

Sl. No.	USN	NAME	IA	SEE	Total	Result
1	4BW17CS001	ABHISHEK URS C J	38	27	65	Pass
2	4BW17CS002	AISHWARYA D	35	33	68	Pass
3	4BW17CS003	AISHWARYA G P	38	29	67	Pass
4	4BW17CS004	AISHWARYA K.P	36	41	77	Pass
5	4BW17CS005	AJAY S	36	24	60	Pass
6	4BW17CS006	AKANKASHA K P	36	41	77	Pass
7	4BW17CS009	ANUPAMA A M	35	39	74	Pass
8	4BW17CS010	ATHFIA FARHEEN N	37	27	64	Pass
9	4BW17CS012	BHAVAN A J	35	24	59	Pass
10	4BW17CS013	BHAVANI N D	35	29	64	Pass
11	4BW17CS014	BHUMIKA M R	37	31	68	Pass
12	4BW17CS015	BINDU H	35	50	85	Pass
13	4BW17CS016	BRUNDA D	35	21	56	Pass
14	4BW17CS017	CHAITHRA R	37	35	72	Pass
15	4BW17CS018	CHAITHRA JAIN H P	35	34	69	Pass
16	4BW17CS020	DEEKSHITHA C	35	42	77	Pass
17	4BW17CS021	DEEPIKA A N	37	46	83	Pass
18	4BW17CS022	DIVYA KHYANI	36	34	70	Pass
19	4BW17CS023	DIVYASHREE K H	37	39	76	Pass
20	4BW17CS024	HARISH GOWDA	37	40	77	Pass
21	4BW17CS025	HARSHITHA Y	36	41	77	Pass
22	4BW17CS026	HEMA D	38	24	62	Pass
23	4BW17CS027	INDU SHREE G J	37	45	82	Pass
24	4BW17CS028	ISHWARAPPA HAVIN	36	35	71	Pass
25	4BW17CS029	JEEVAN R	39	36	75	Pass
26	4BW17CS031	JINASHREE P	36	42	78	Pass
27	4BW17CS032	KARTHIK K P	35	34	69	Pass
28	4BW17CS034	LAKSHMIKANTH GOWDA M	38	35	73	Pass

29	4BW17CS035	MAANYA K V	36	41	77	Pass
30	4BW17CS036	MANJUSHREE C S	37	40	77	Pass
31	4BW17CS037	MANOJ S B	36	13	49	Pass
32	4BW17CS038	MEGHANA K	38	24	62	Pass
33	4BW17CS039	MEGHANA M V	35	33	68	Pass
34	4BW17CS040	NAVYASHREE H D	37	26	63	Pass
35	4BW17CS041	NIKITH G S	36	40	76	Pass
36	4BW17CS042	NOOR AYESHA S	35	30	65	Pass
37	4BW17CS043	POOJA D R	38	39	77	Pass
38	4BW17CS044	POOJA K S	35	21	56	Pass
39	4BW17CS045	POOJASHREE G	38	38	76	Pass
40	4BW17CS046	PRIYADARSHINI P	37	34	71	Pass
41	4BW17CS048	PRIYANKA V L	37	36	73	Pass
42	4BW17CS050	PUNEETH RAJ B S	35	15	50	Pass
43	4BW17CS051	RAHUL B	35	26	61	Pass
44	4BW17CS052	RAKESH C S	35	22	57	Pass
45	4BW17CS053	RAKSHITHA N	36	44	80	Pass
46	4BW17CS054	RAMYA K L	35	41	76	Pass
47	4BW17CS055	RANJITHA B S	37	27	64	Pass
48	4BW17CS056	RITESH KUMAR CHANDA	30	21	51	Pass
49	4BW17CS057	ROHIT KUMAR JHA	30	33	63	Pass
50	4BW17CS058	SAHANA L M	37	29	66	Pass
51	4BW17CS059	SANJANA GOWDA N C	39	47	86	Pass
52	4BW17CS060	SANJAY KUMAR C G	35	27	62	Pass
53	4BW17CS061	SHANKREPPA HANDARGAL	35	24	59	Pass
54	4BW17CS062	SHIFAALI	37	28	65	Pass
55	4BW17CS063	SHRUSTI M	37	27	64	Pass
56	4BW17CS064	SIDDARTH SINGH	37	21	58	Pass
57	4BW17CS065	SINCHANA B R	37	42	79	Pass
58	4BW17CS066	SMITHA B U	37	29	66	Pass
59	4BW17CS084	NAMRATHA	37	39	76	Pass
60	4BW17CS085	NAYANA	35	29	64	Pass
61	4BW17CS086	SOWMYA JAKKULA	37	23	60	Pass
62	4BW18CS403	DHANANJAYA	35	35	70	Pass
63	4BW18CS404	GAGAN B S	38	42	80	Pass
64	4BW18CS406	GIRISH REDDY	30	22	52	Pass
65	4BW18CS410	VIDYASAGAR	30	26	56	Pass
66	4BW16CS055	RAJU M D	37	26	63	Pass
67	4BW17CS067	SMITHA M	35	29	64	Pass
68	4BW17CS068	SNEHA N J	38	33	71	Pass

69	4BW17CS069	SOWNDARYA L T	37	39	76	Pass
70	4BW17CS070	SPOORTHI H	36	41	77	Pass
71	4BW17CS071	SPOORTHI R	37	39	76	Pass
72	4BW17CS072	SPOORTHI C	37	30	67	Pass
73	4BW17CS074	SWATHI D	38	30	68	Pass
74	4BW17CS075	TASMIYA	37	48	85	Pass
75	4BW17CS076	TEJAS RAHUL R	37	23	60	Pass
76	4BW17CS077	THEJAS G C	36	30	66	Pass
77	4BW17CS078	VARALAKSHMI C K	35	23	58	Pass
78	4BW17CS081	YASHASHWINI H M	38	23	61	Pass
79	4BW17CS082	YOGASHREE C R	37	34	71	Pass
80	4BW18CS400	ANUSHA K J	37	27	64	Pass
81	4BW18CS401	BHAVYA J K	38	36	74	Pass
82	4BW18CS402	BINDHUSHREE A C	36	33	69	Pass
83	4BW18CS405	GAYITHRI K	37	31	68	Pass
84	4BW18CS407	GREESHMA M S	37	35	72	Pass
85	4BW18CS408	KALAVATHI R	37	30	67	Pass
86	4BW18CS409	KEERTHI B L	35	34	69	Pass

No. of Students Attended	**86**
No. of Students Absent	**Nil**
No. of Students Passed	**86**
No. of Students Failed	**0**
Pass Percentage	**100%**

Signature of Course Coordinator

B. K. Raghu

Signature of HoD

HOD
Dept. of Computer Science & Engg
B.G.S. Institute of Technology
B.G. Nagar - 571 448
Nagamangala Tq, Mandya Dist
Karnataka (INDIA)

ML Attainment 2017 Scheme 2020-21 Batch

	60%	30%	10%	
	CIE	SEE	CES	OTAL
CO1	2.70	1.06	2.6	2.19
CO2	2.87	1.06	2.6	2.30
CO3	2.81	1.06	2.6	2.26
CO4	3.00	1.06	2.6	2.37
CO5	3.00	1.06	2.6	2.37

CO-PO/PSO Mapping Table

PO/PSO		PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2.19	2	2	2	1					1			2	2	2
CO2	2.30	2	2	2	1					1			1	2	2
CO3	2.26	2	2	2	1		1			1			1	2	2
CO4	2.37	2	2	2	1					1			2	2	2
CO5	2.37	2	2	2	1					1			1	2	2
Sum		10	10	10	5		1			5			7	10	10
Weighted Sum		22.98	22.98	22.98	11.49		2.26			11.49			16.06	22.98	22.98
Attainment		1.53	1.53	1.53	0.77		0.75			0.766			1.071	1.532	1.53

B. K. Raghu
Dept. of Computer Science & Engg.
G.S. Institute of Technology
B.G. Nagar - 571 448
Ramanagata Tq. Mandya Dist
Karnataka (INDIA)

Machine Learning
17CS73

M J Prasanna Kumar

CONTENTS

INTRODUCTION

- **Inducing general functions from specific training examples is a main issue of machine learning.**
- **Concept learning -** a **learning task** in which a **human or machine learner** is **trained to classify objects** by being shown a **set of example objects** along with their **class labels.** The learner **simplifies** what has been observed by **condensin**g it in the form of an **example.**
- **Concept learning** - also known as **category learning**, **concept attainment**, and **concept formation.**

INTRODUCTION ….

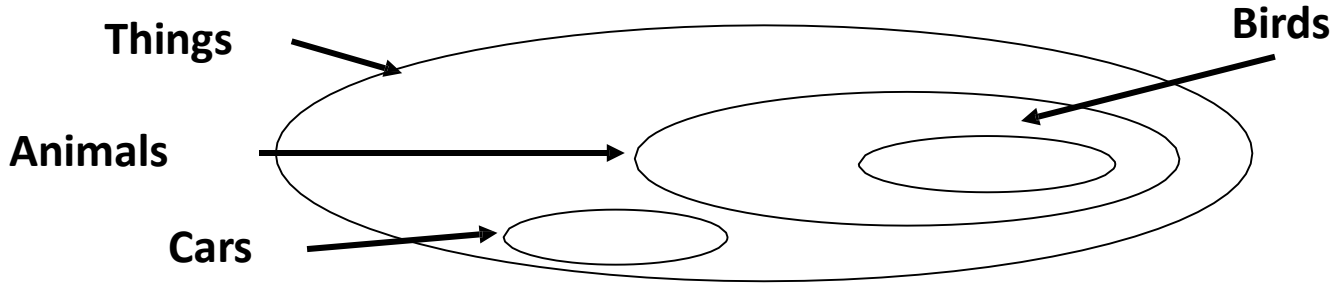
- **Concept Learning: Acquiring the definition of a general category from given sample of positive and negative training examples of the category.**

A Formal Definition for Concept Learning:

- **Inferring a boolean-valued function from training examples of its input and output.**
- **Let us try to learn the definition of a concept from examples.**

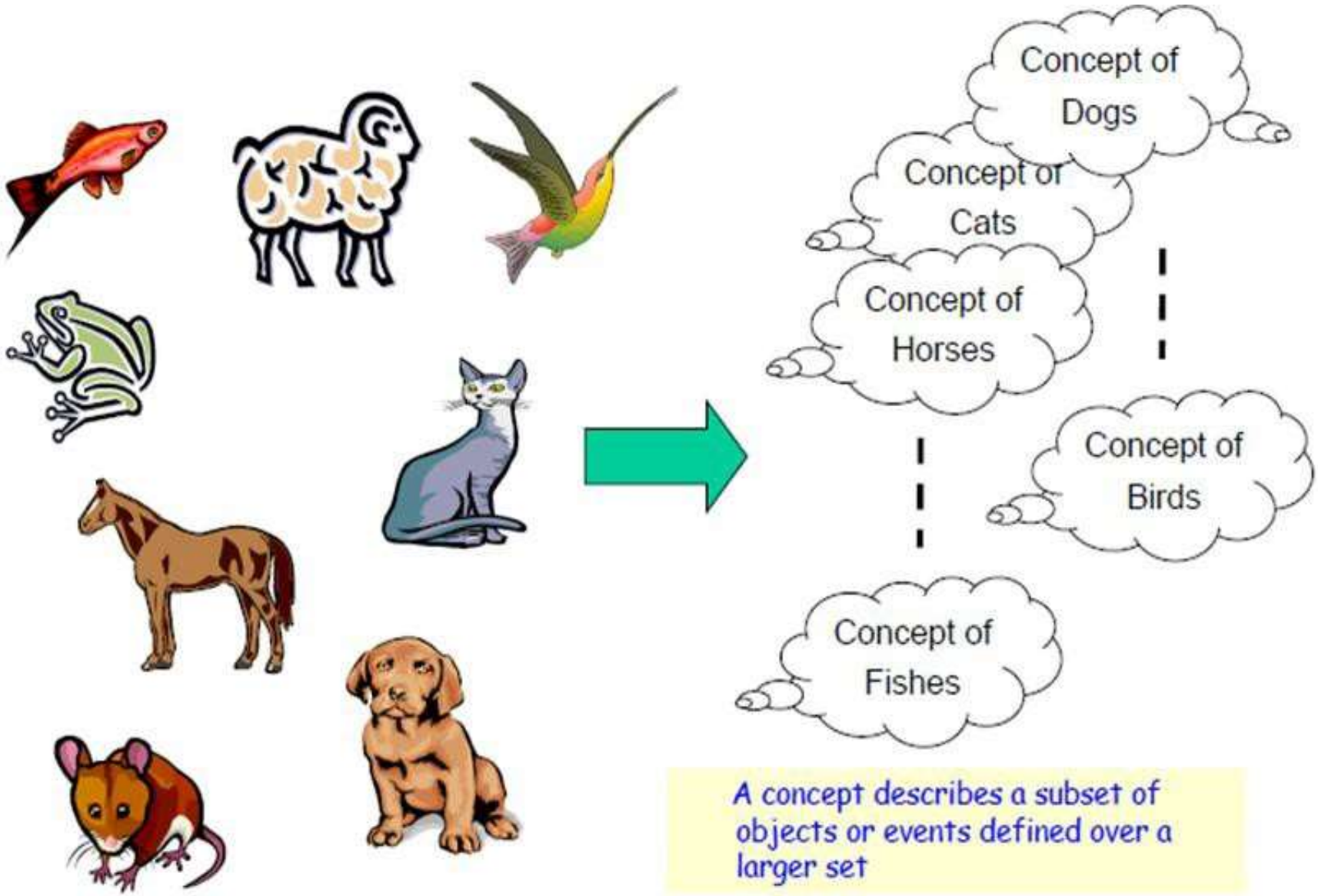
What is a Concept? - Examples

- **An example for concept-learning is the learning of bird-concept from the given examples of birds (positive examples) and non-birds (negative examples).**
- **The concept of a bird is the subset of all objects (i.e., the set of all things or all animals) that belong to the category of bird.**

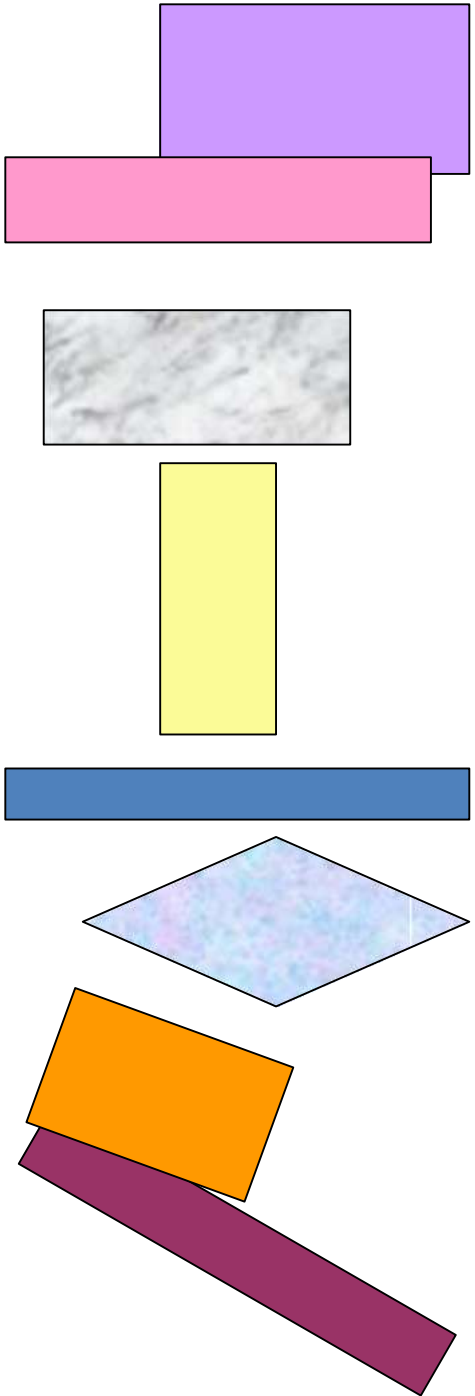


- Each **concept** is a boolean-valued function defined over this larger set. **[Example: a function defined over all animals whose value is true for birds and false for every other animal]**

What is a Concept? – Examples …..

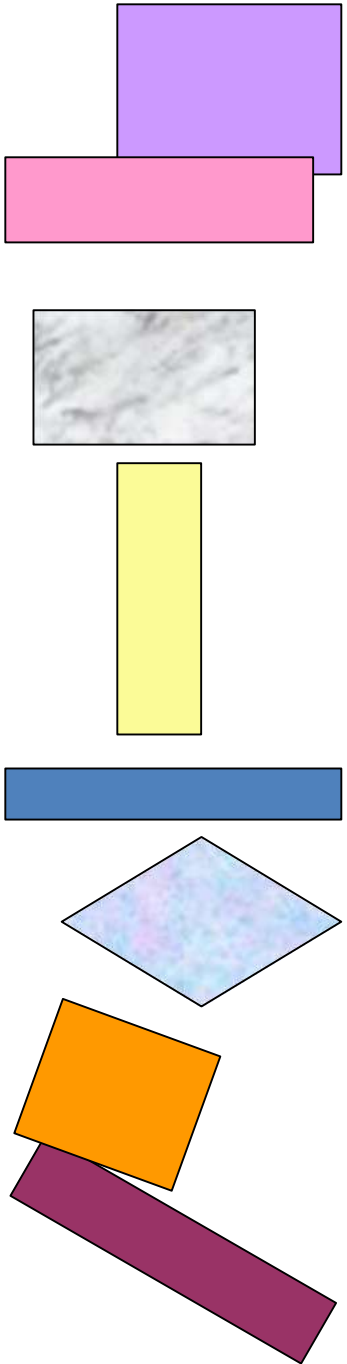


A concept describes a subset of objects or events defined over a larger set

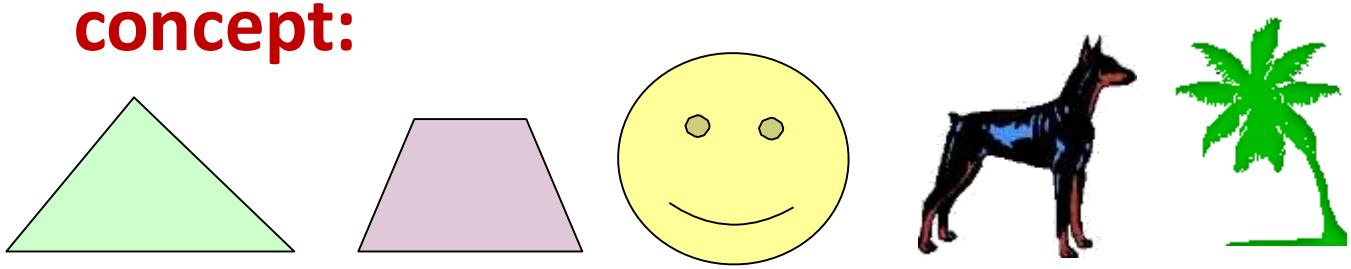


What is a Concept? – Examples …..

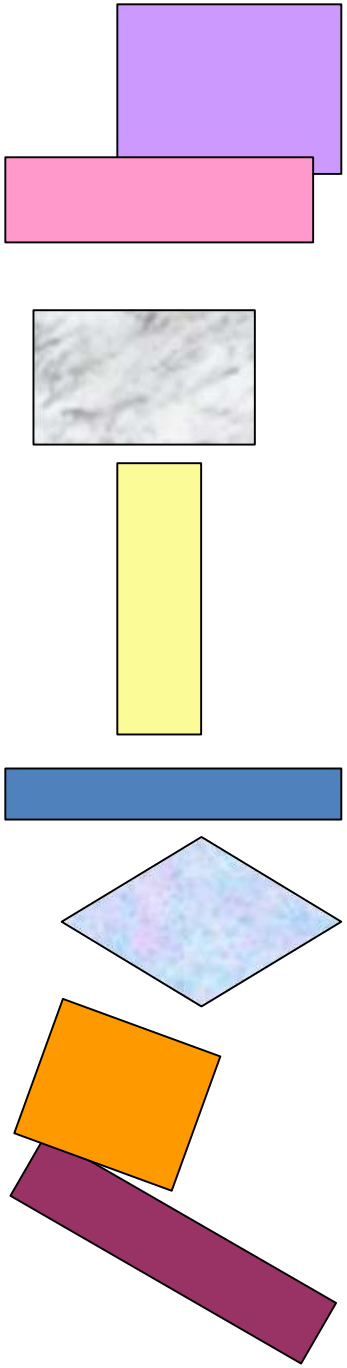
- ✓ Concepts are categories of stimuli that have certain features in common.
- ✓ The shapes on the left are all members of a conceptual category: RECTANGLE.
- ✓ Their common features are (1) 4 lines; (2) opposite lines parallel; (3) lines connected at ends; (4) lines form 4 right angles.
- ✓ But they are different colors and sizes and have different orientations is irrelevant and are not defining features of the concept



- ✓ If a stimulus is a member of a specified conceptual category, it is referred to as a "positive instance".
- ✓ If it is not a member, it is referred to as "negative instance". These are all negative instances of the rectangle concept:

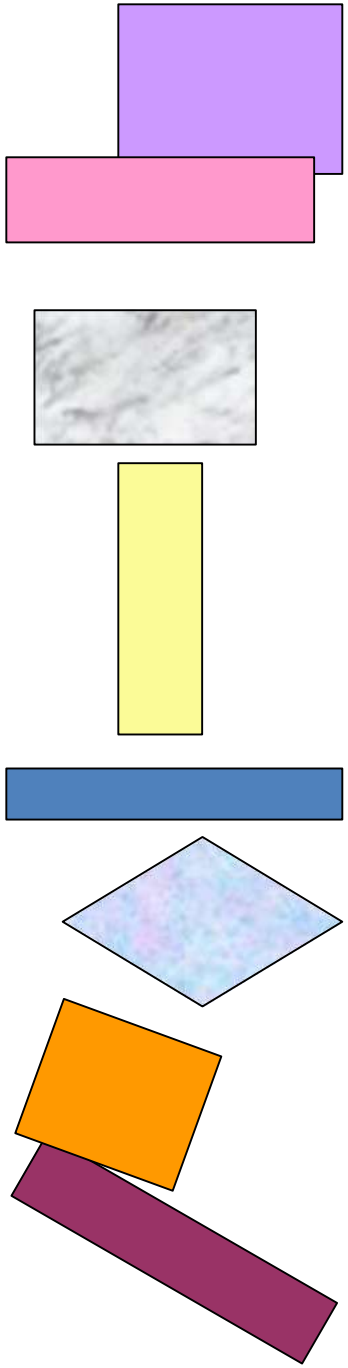


- ✓ As rectangles are defined, a stimulus is a negative instance if it lacks any one of the specified features.



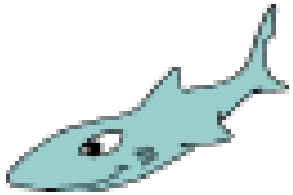
Every concept has two components:

- ✓ **Attributes:** These are features of a stimulus that one must look for to decide if that stimulus is a positive instance of the concept.
- ✓ **A rule:** This a statement that specifies which attributes must be present or absent for a stimulus to qualify as a positive instance of the concept.
- ✓ For rectangles, the attributes would be the four features and the rule would be that all the attributes must be present.



- ✓ The simplest rules refer to the presence or absence of a single attribute. For example, a "vertebrate" animal is defined as an animal with a backbone. Which of these stimuli are positive instances?

+



+

–

+

- ✓ This rule is called affirmation. It says that a stimulus must possess a single specified attribute to qualify as a of a concept.

- ✓ The opposite or “complement” of affirmation is negation. To qualify as a positive instance, a stimulus must lack a single specified attribute.

————— + –

- ✓ An invertebrate animal is one that lacks a backbone. These are the positive and negative instances when the negation rule is applied.

A Concept Learning Task – EnjoySport Training Examples

Example	Sky	Temp	Humidity	Wind	Water	Forecast	EnjoySports
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

ATTRIBUTES

CONCEPT

- A set of example days, and each is described by six attributes.
- The task is to learn to predict the value of EnjoySport for arbitrary day, based on the values of its attribute values -Target concept

A Concept Learning Task – Hypothesis Representation

- **Goal: To infer the "best" concept-description from the set of all possible hypotheses.**
- **Each hypothesis consists of a conjunction of constraints on the instance attributes.**
- **Each hypothesis will be a vector of six constraints, specifying the values of the six attributes**

  (Sky, AirTemp, Humidity, Wind, Water, and Forecast)

A Concept Learning Task – Hypothesis Representation…..

Each attribute will be:

- ✓ **? - indicating any value is acceptable for the attribute (don't care)**
- ✓ **single value – specifying a single required value (ex. Warm) (specific)**
- ✓ **0 - indicating no value is acceptable for the attribute (no value)**
- ✓ **A hypothesis:**

Sky	AirTemp	Humidity	Wind	Water	Forecast
< Sunny,	?,	?,	Strong ,	? ,	Same >

A Concept Learning Task – Hypothesis Representation…..

- ✓ **Most General Hypothesis: Everyday** is a good day for water sports **<?, ?, ?, ?, ?, ?>** (Positive example)
- ✓ **Most Specific Hypothesis: No day** is a good day for water sports **<0, 0, 0, 0, 0, 0>** (No day is Positive example)
- ✓ **EnjoySport concept learning task requires learning the sets of days for which EnjoySport = yes, describing this set by a conjunction of constraints over the instance attributes.**

EnjoySport Concept Learning Task

Given:

- ✓ **Instances X**: Set of all **Possible days,** each described by the attributes
  - **Sky (Sunny, Cloudy, and Rainy)**
  - **Temp (Warm and Cold)**
  - **Humidity (Normal and High)**
  - **Wind (Strong and Weak)**
  - **Water (Warm and Cool)**
  - **Forecast (Same and Change)**

- ✓ **Target Concept (Function) c : EnjoySport : X → {0,1}**
- ✓ **Hypotheses H : Each hypothesis is described by a conjunction of constraints on the attributes.**
- ✓ **Training Examples D : positive and negative examples of the target function** along with their **target concept value c(x).**
- ✓ **Determine : A hypothesis h in H such that h(x) = c(x) for all x in D.**

EnjoySport Concept Learning Task…..

- ✓ **Members of the concept** (instances for which **c(x)=1**) are called ***positive examples***.
- ✓ **Nonmembers of the concept** (instances for which **c(x)=0**) are called ***negative examples***.
- ✓ **H** represents the set of ***all possible hypotheses***. **H** is determined by the human designer's choice of a hypothesis representation.
- ✓ The **goal of concept-learning** is to **find a hypothesis h: X → {0, 1}** such that **h(x)=c(x)** for **all x in D.**
- ✓ **Example of a hypothesis: <?, Cold, High, ?, ?, ?>**

  If the **air temperature is cold and the humidity high** then it is a **good day** for **water sports.**

Concept Learning As Search

- ✓ Concept Learning can be viewed as the **task of searching** through a **large space of hypotheses** implicitly defined by the hypothesis representation.
- ✓ The goal of this search is to find the hypothesis that **best fits** the training examples.
- ✓ The hypothesis space has a **general-to-specific ordering of hypotheses**, and the search can be efficiently organized by taking advantage of a naturally occurring **structure** over the hypothesis space.
- ✓ By selecting a hypothesis representation, the designer of the learning algorithm **implicitly defines the space** of all hypotheses that the program can ever represent and therefore can ever learn.

Enjoy Sport - Hypothesis Space

- ✓ Sky has **3 possible values**, and other 5 attributes have **2 possible values.**
- ✓ There are **96 (= 3.2.2.2.2.2)** distinct **instances in X.**
- ✓ There are **5120 (=5.4.4.4.4.4)** syntactically distinct hypotheses in H.
  - – Two more values for attributes: **? and 0**
- ✓ Every hypothesis containing **one or more 0 symbols** represents the empty set of instances.

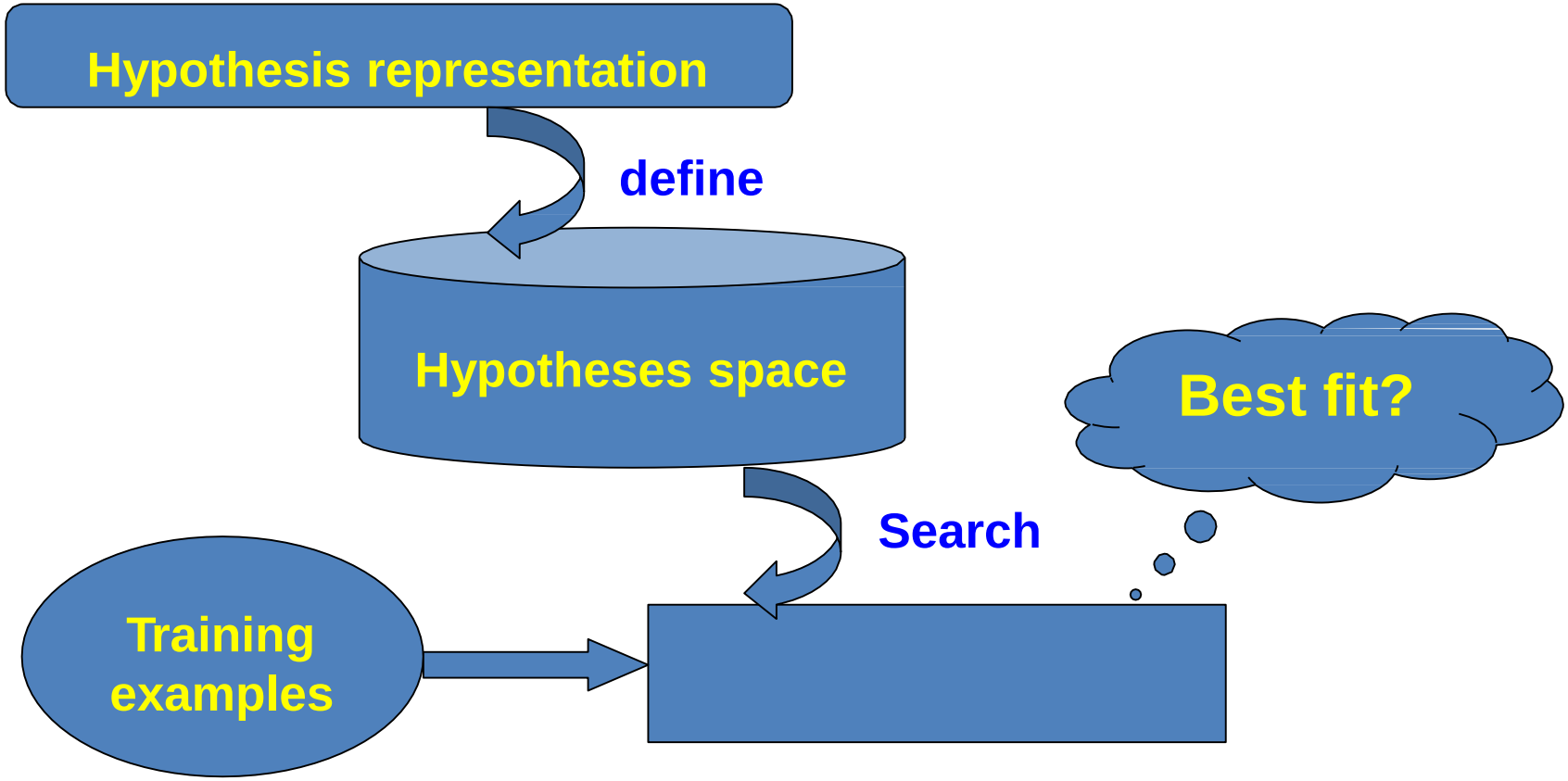
  that is, it classifies every instance as **negative.**

Enjoy Sport - Hypothesis Space…..

- ✓ There are **973 (= 1 + 4.3.3.3.3.3)** semantically distinct hypotheses in H.
  - **Only one more value** for attributes: **?,** and **one hypothesis** representing **empty set** of instances.
- ✓ Although EnjoySport has small, finite hypothesis space, most learning tasks have much larger **(even infinite)** hypothesis spaces.
  - We need efficient search algorithms on the hypothesis spaces.

Concept Learning As Search: General-to-Specific Ordering of Hypotheses

- ✓ The hypothesis space has a **general-to-specific** ordering of hypotheses, and the search can be efficiently organized.



Concept Learning As Search: General-to-Specific Ordering of Hypotheses

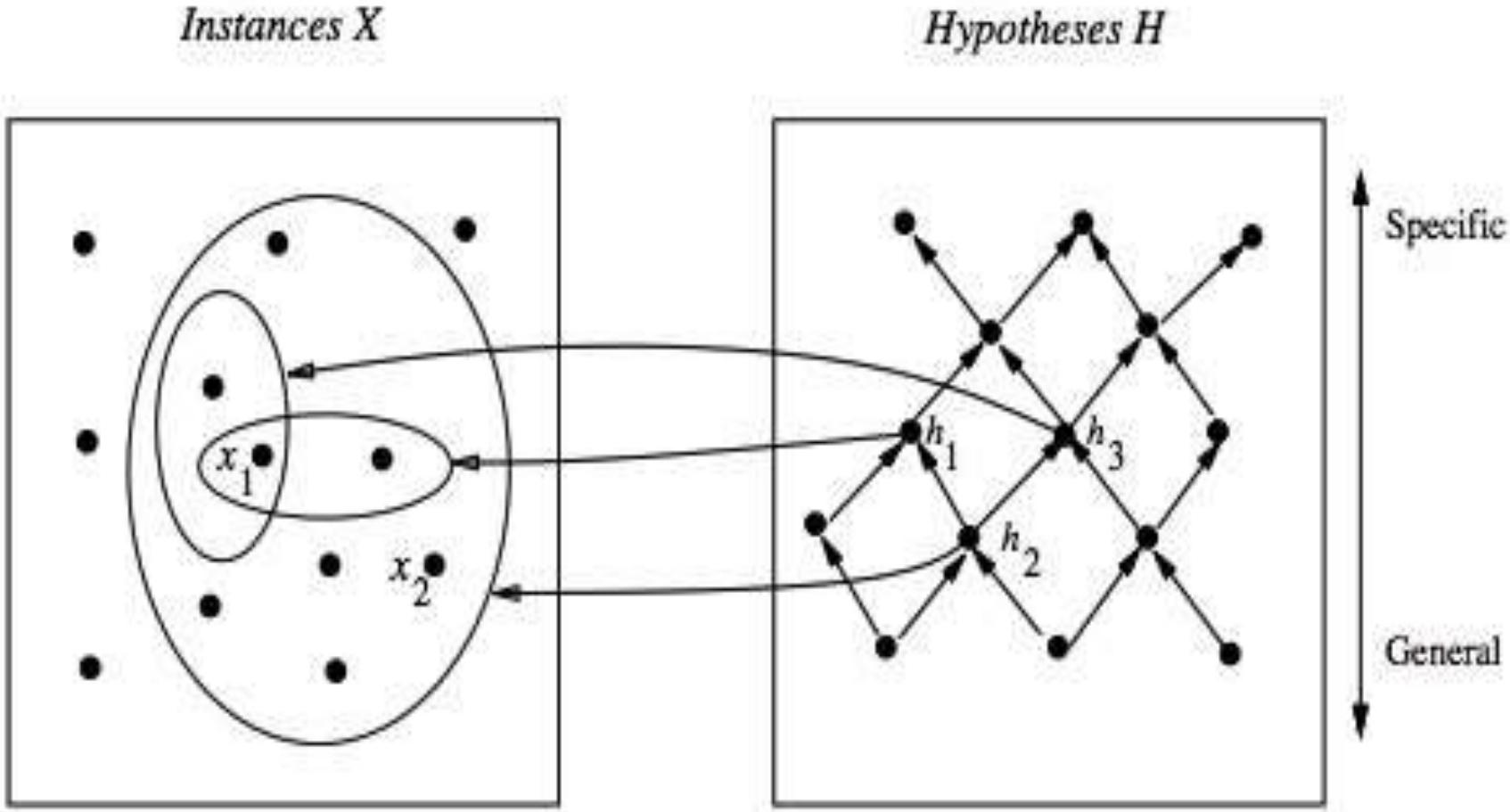
- **Definition:** Let h_j **and** h_k be boolean-valued functions defined over **X.** Then h_j is **more-general-than-or-equal-to** h_k <u>iff</u> for all **x in X,** $[(h_k(x) = 1) \rightarrow (h_j(x)=1)]$
- **Example:**

  h1 = <Sunny, ?, ?, Strong, ?, ?>

  h2 = <Sunny, ?, ?, ?, ?, ?>

  Every instance that are classified as **positive by *h1*** will also be classified as **positive by *h2*** in our example data set. Therefore ***h2* is more general than *h1*.**
- We also use the ideas of **"strictly" -more-general-than,** and **more-specific-than** [Mitchell].

More General than Relation



x_1 = <Sunny, Warm, High, Strong, Cool, Same>

x_2 = <Sunny, Warm, High, Light, Warm, Same>

h_1 = <Sunny, ?, ?, Strong, ?, ?>

h_2 = <Sunny, ?, ?, ?, ?, ?>

h_3 = <Sunny, ?, ?, ?, Cool, ?>

Find –S Algorithm : Finding a Maximally Specific Hypothesis Learning Algorithm

- ✓ **FIND-S Algorithm starts from the most specific hypothesis and generalize it by considering only positive examples.**
- ✓ **FIND-S algorithm ignores negative examples.**
  - As long as the hypotheses space contains a hypothesis that describes the true target concept, and the training data contains no errors, ignoring negative examples does not cause to any problem.
- ✓ **FIND-S algorithm finds the most specific hypothesis within H that is consistent with the positive training examples.**
  - The final hypothesis will also be consistent with negative examples if the correct target concept is in H, and the training examples are correct.

Find –S Algorithm : Finding a Maximally Specific Hypothesis Learning Algorithm…

1. Initialize h to the most specific hypothesis in H
2. For each positive training instance x
   - For each attribute constraint a_i in h

     If the constraint a_i in h is satisfied by x

     Then do nothing

     Else replace a_i in h by the next more general constraint that is satisfied by x
3. Output hypothesis h

Find –S Algorithm

1. Initialize h to meet specific hypothesis

   h= { ‘ø’ , ‘ø’ , ‘ø’ , ‘ø’ , ‘ø’ , ‘ø’ }

2. For each positive example:

   For each attribute in the example:

   if attribute value = hypothesis value

   do nothing

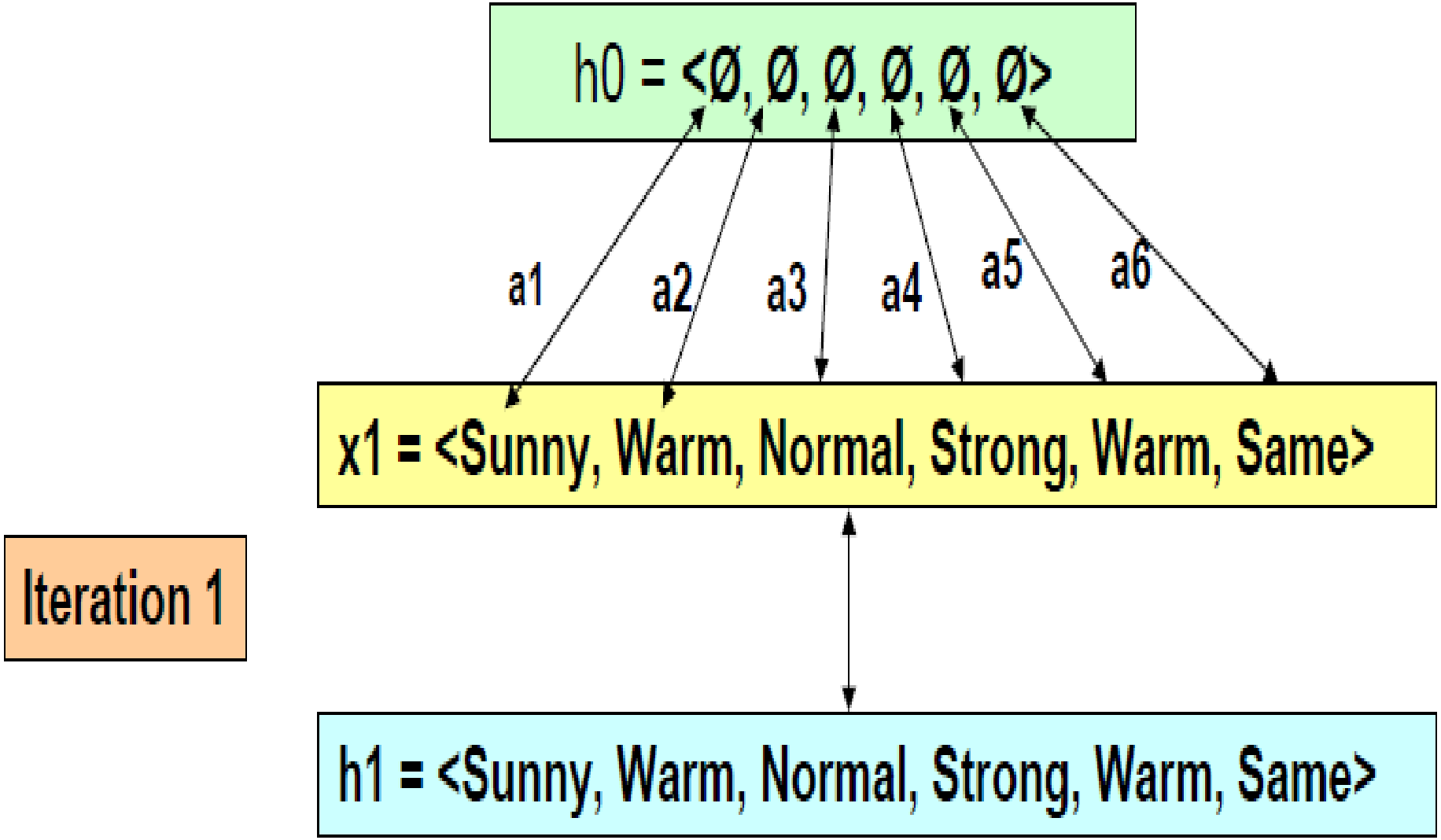
   else

   replace hypothesis value with more general constraint ‘?’

3. Output hypothesis h

Example	Sky	Temp	Humidity	Wind	Water	Forecast	EnjoySports
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Find –S Algorithm: Example



Find –S Algorithm: Example……

h1 = <Sunny, Warm, Normal, Strong, Warm, Same>

Iteration 2

x2 = <Sunny, Warm, High, Strong, Warm, Same>

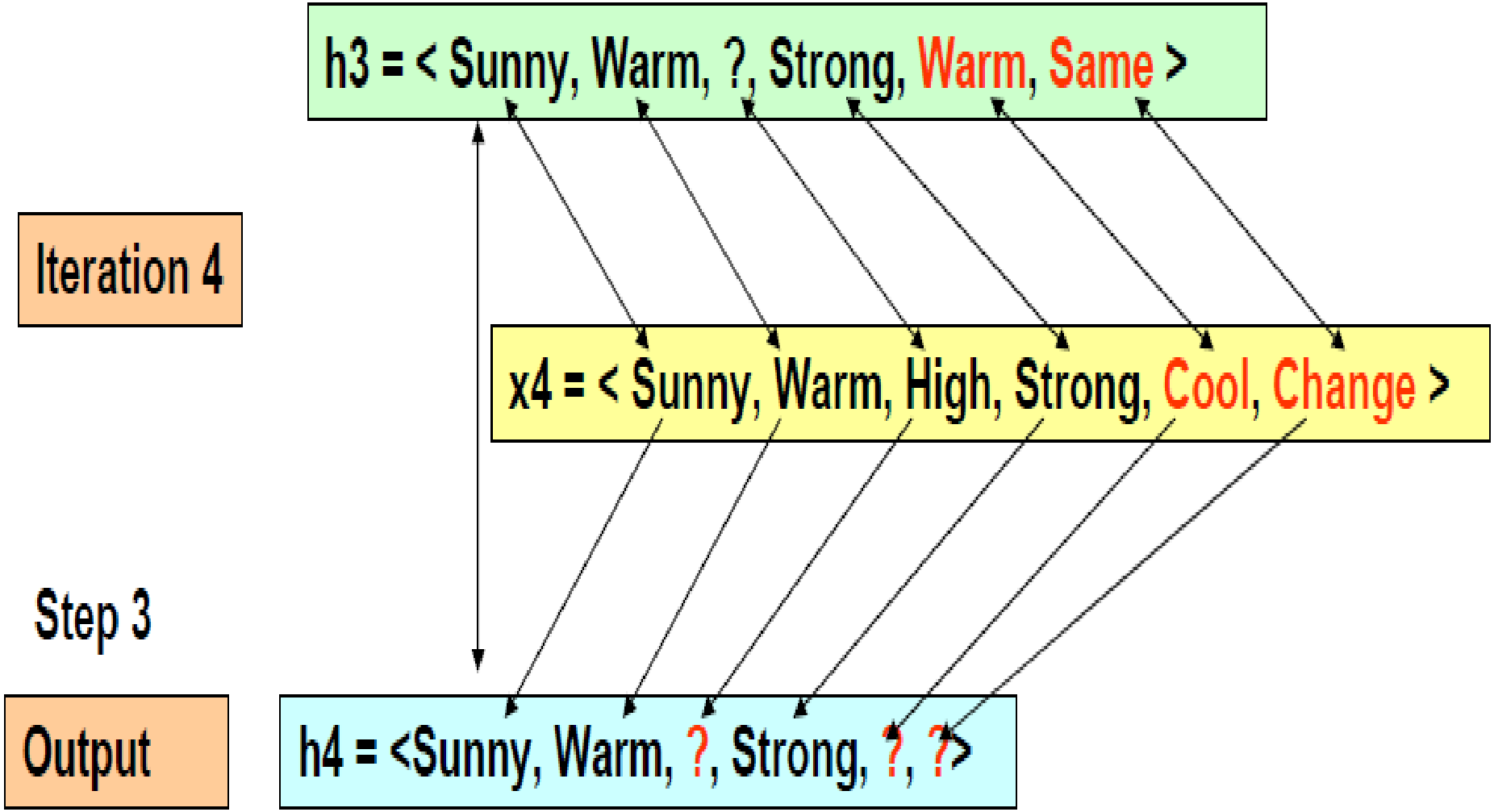
h2 = <Sunny, Warm, ?, Strong, Warm, Same>

Iteration 3

Ignore

h3 = <Sunny, Warm, ?, Strong, Warm, Same>

Find –S Algorithm: Example



Properties and Shortcomings of Find-S

- Find-S is guaranteed to output the most specific hypothesis within H that is consistent with the positive training examples

- Is it a good strategy to prefer the most specific hypothesis?

- What if the training set is inconsistent (noisy)?
- What if there are several maximally specific consistent hypotheses? Find-S cannot backtrack!
- Negative example are not considered

Version Space and CANDIDATE ELIMINATION Algorithm

The key idea in the CANDIDATE-ELIMINATION algorithm is to output a description of the set of all ***hypotheses consistent with the training examples***

Representation

- ***Definition:*** A hypothesis h is **consistent** with **a** set of training examples ***D*** if **and** only if
  h(x) = ***c(x)*** for each example (x, ***c(x))*** in ***D.***

$$Consistent(h, D) \equiv (\forall \langle x, c(x) \rangle \in D)\ h(x) = c(x))$$

Note difference between definitions of *consistent* and *satisfies*

- an example *x* is said to ***satisfy*** hypothesis ***h*** when ***h(x)*** = 1, regardless of whether *x* is a positive or negative example of the target concept.
- an example *x* is said to ***consistent*** with hypothesis ***h*** iff ***h(x)*** = ***c(x)***

Version Spaces …

- **Definition:** A hypothesis *h* is **consistent** with a set of training examples ***D* iff** ***h(x) = c(x)*** for each example **< x, c(x) > in *D*.**

- **Definition:** The **version space**, denoted $VS_{H,D}$ with respect to **hypothesis space *H*** and **training examples *D*,** is the **subset of hypotheses from *H*** **consistent** with the training examples in *D*.

$$VS_{H,D} \equiv \{h \in H \mid Consistent(h, D)\}$$

The Idea

1. An explicit list of hypotheses : List-Then-Eliminate Algorithm

2. A compact representation of hypotheses which exploits the more_general_than partial ordering: Candidate-Elimination Algorithm

List-Then-Eliminate algorithm

Version space as list of hypotheses

1. $\boldsymbol{VersionSpace} \leftarrow$ a list containing every hypothesis in $\boldsymbol{H}$
2. For each training example, $\langle \boldsymbol{x}, \boldsymbol{c}(\boldsymbol{x}) \rangle$

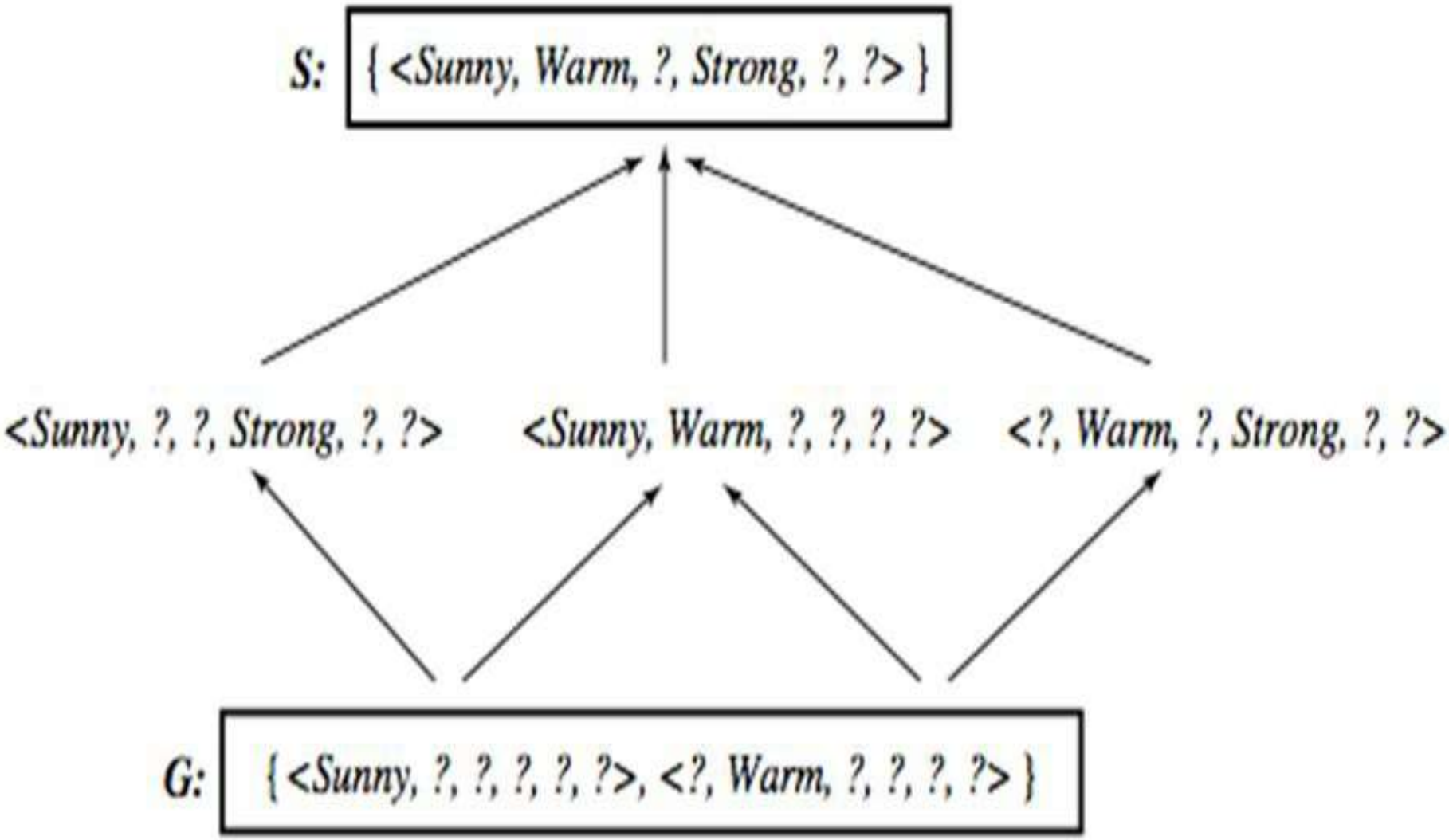
   Remove from $\boldsymbol{VersionSpace}$ any hypothesis h for which $\boldsymbol{h}(\boldsymbol{x}) \neq \boldsymbol{c}(\boldsymbol{x})$
3. Output the **list of hypotheses** in $\boldsymbol{VersionSpace}$

List-Then-Eliminate algorithm…

Problems

- The hypothesis space must be finite
- Enumeration of all the hypothesis, rather inefficient

Representation for Version Space with General and Specific Boundary Sets



Note: **The output of *Find-S* is just ⟨*Sunny, Warm , ?, Strong, ?, ?*⟩**

Candidate Elimination Algorithm

- **Uses Version space**
- **Consider both +ve and –ve results**
- **We have both specific and general hypothesis**

- **For a +ve example**
  - We tend to generalize specific hypothesis

- **For a -v example**
  - We tend to rate general hypothesis more specific

Candidate Elimination Algorithm

Algorithm

1. **Intialize G and S as most general and specific hypothesis**
2. **For each example d:**

    if d is +ve:

    -> Make specific hypothesis more general

    else

    -> Make general hypothesis more specific

Candidate Elimination Algorithm

- **For each training example d, do**
  - ✓ **If d is a positive example**
    - Remove from **G** any hypothesis **inconsistent with d**
    - For each hypothesis **s in S** that **is not consistent with d**
      - Remove **s from S**
      - **Add to S** all minimal generalizations **h of s** such that
        - **h is consistent with d,** and some member of **G** is **more general than h**
      - **Remove from S** any hypothesis that is **more general than** another hypothesis in **S**

Candidate Elimination Algorithm …

- ✓ **If d is a negative example**
  - Remove from **S** any hypothesis **inconsistent with d**
  - For each hypothesis **g in G** that is **not consistent with d**
    - Remove **g from G**
    - **Add to G** all minimal specializations **h of g** such that
      - **h is consistent with d,** and some member of **S is more specific than h**
    - **Remove from G** any hypothesis that is **less general than** another hypothesis in **G**

Initialize G to the set of maximally general hypotheses in H

Initialize S to the set of maximally specific hypotheses in H For

For each training example d, do

- If d is a positive example
  - Remove from G any hypothesis inconsistent with d
  - For each hypothesis s in S that is not consistent with d
    - Remove s from S
    - Add to S all minimal generalizations h of s such that
      - h is consistent with d, and some member of G is more general than h
    - Remove from S any hypothesis that is more general than another hypothesis in S

- If d is a negative example
  - Remove from S any hypothesis inconsistent with d
  - For each hypothesis g in G that is not consistent with d
    - Remove g from G
    - Add to G all minimal specializations h of g such that
      - h is consistent with d, and some member of S is more specific than h
    - Remove from G any hypothesis that is less general than another hypothesis in G

Example: Initially

S_0 : $\langle \varnothing, \varnothing, \varnothing, \varnothing, \varnothing. \varnothing \rangle$

G_0 $\langle ?, ?, ?, ?, ?, ? \rangle$

Example	Sky	Temp	Humidity	Wind	Water	Forecast	EnjoySports
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Example:

after seeing ⟨Sunny, Warm, Normal, Strong, Warm, Same ⟩ +

S_0: ⟨∅, ∅, ∅, ∅, ∅. ∅⟩

S_1: ⟨Sunny, Warm, Normal, Strong, Warm, Same⟩

G_0, G_1 ⟨?, ?, ?, ?, ?, ?⟩

Example	Sky	Temp	Humidity	Wind	Water	Forecast	EnjoySports
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Example:

after seeing ⟨Sunny, Warm, High, Strong, Warm, Same⟩ +

S_1: ⟨Sunny, Warm, Normal, Strong, Warm, Same⟩

S_2: ⟨Sunny, Warm, ?, Strong, Warm, Same⟩

G_1, G_2 ⟨?, ?, ?, ?, ?, ?⟩

Example	Sky	Temp	Humidity	Wind	Water	Forecast	EnjoySports
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Example:

after seeing ⟨Rainy, Cold, High, Strong, Warm, Change ⟩ –

S_2, S_3: ⟨Sunny, Warm, ?, Strong, Warm, Same⟩

G_3: ⟨Sunny, ?, ?, ?, ?, ?⟩ ⟨?, Warm, ?, ?, ?, ?⟩ ⟨?, ?, ?, ?, ?, Same⟩

G_2: ⟨ ?, ?, ?, ?, ?, ?, ?⟩

Example	Sky	Temp	Humidity	Wind	Water	Forecast	EnjoySports
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Example:

after seeing ⟨Sunny, Warm, High, Strong, Cool Change ⟩ +

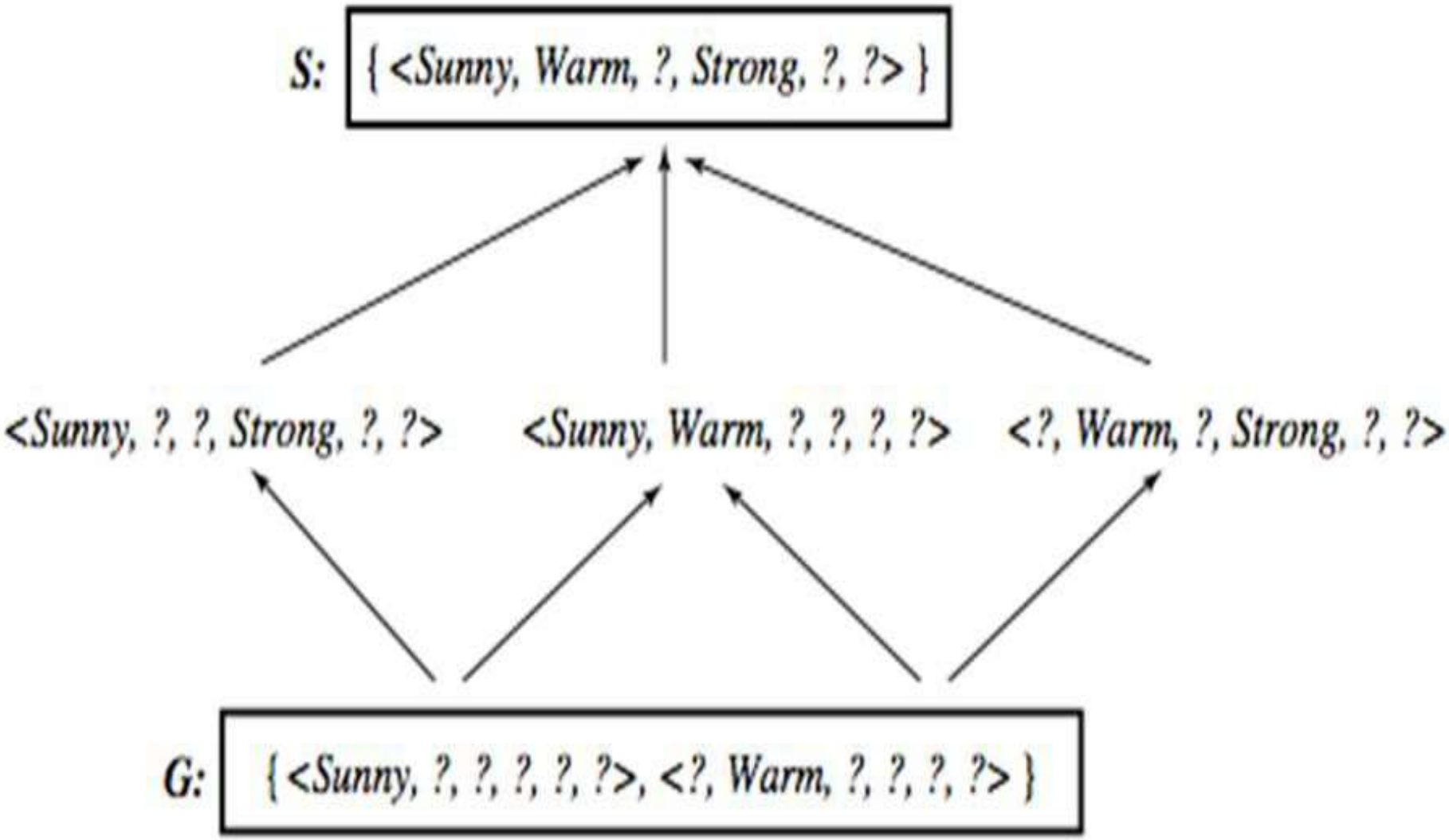
S_3 ⟨Sunny, Warm, ?, Strong, Warm, Same⟩

S_4 **⟨Sunny, Warm, ?, Strong, ?, ?⟩**

G_4: **⟨Sunny, ?, ?, ?, ?, ?⟩ ⟨?, Warm, ?, ?, ?, ?⟩**

G_3: ⟨Sunny, ?, ?, ?, ?, ?⟩ ⟨?, Warm, ?, ?, ?, ?⟩ ⟨?, ?, ?, ?, ?, Same⟩

Representation for Version Space with General and Specific Boundary Sets



Note: **The output of *Find-S* is just ⟨*Sunny, Warm , ?, Strong, ?, ?*⟩**

Example:

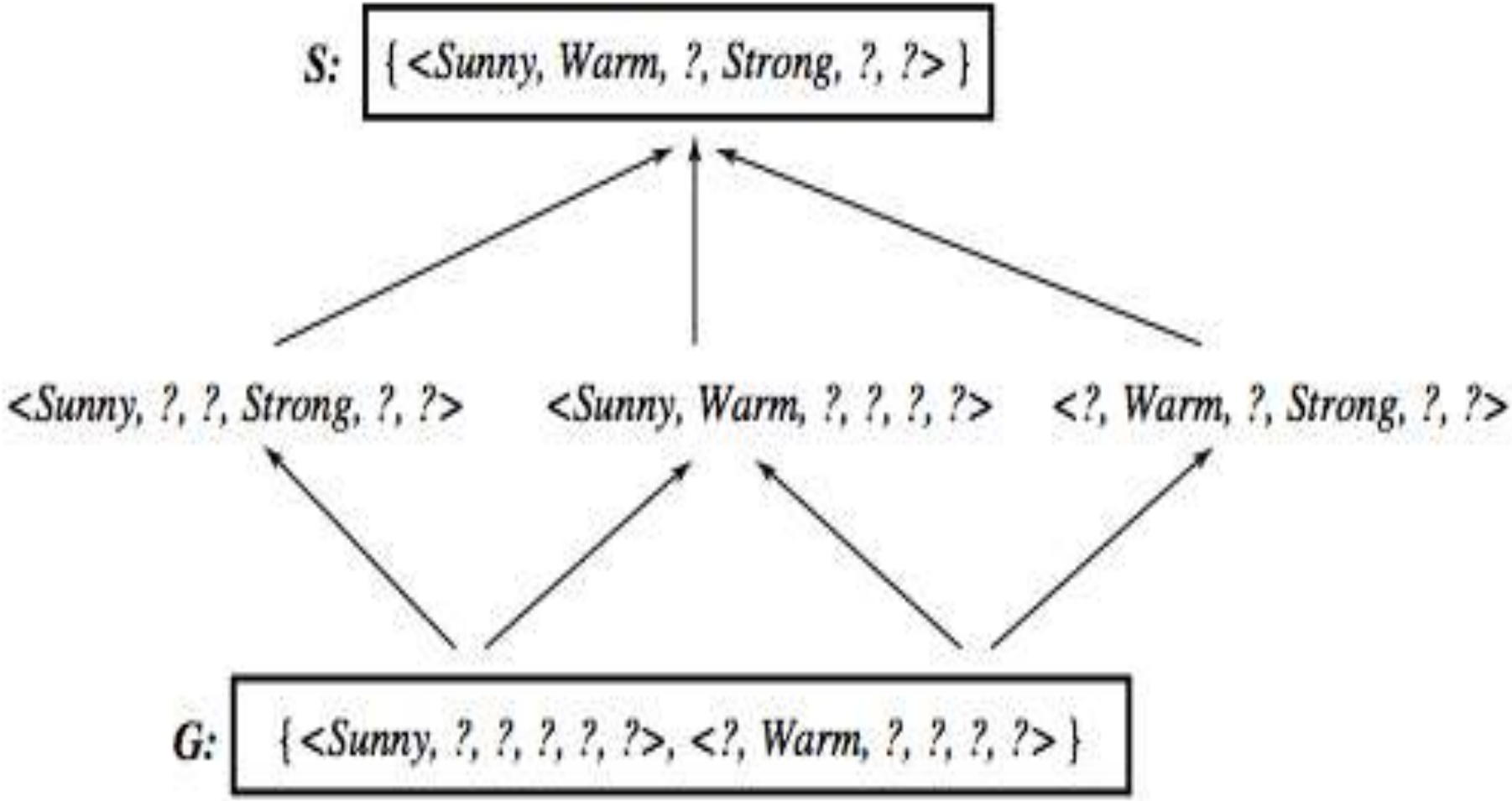
- ✓ Given that there are six attributes that could be specified to specialize G2, why are there only **three new hypotheses** in G3?
- ✓ For example, the hypothesis **h = <?, ?, Normal, ?, ?, ?>** is a minimal specialization of G2 that correctly labels the new example as a **negative example**, but it is not included in G3.
  - The reason for excluding this hypothesis is that it is **inconsistent with S2.**
  - The algorithm determines this simply by noting that **h is not more general than** the current specific boundary, **S2.**

Example:

- ✓ The **S boundary** of the version space forms a **summary** of the previously encountered **positive examples** that can be used to determine whether any given hypothesis is consistent with these examples.

- ✓ The **G boundary** summarizes the information from previously encountered **negative examples**. Any hypothesis more specific than G is assured to be consistent with past negative examples.

Learned Version Space



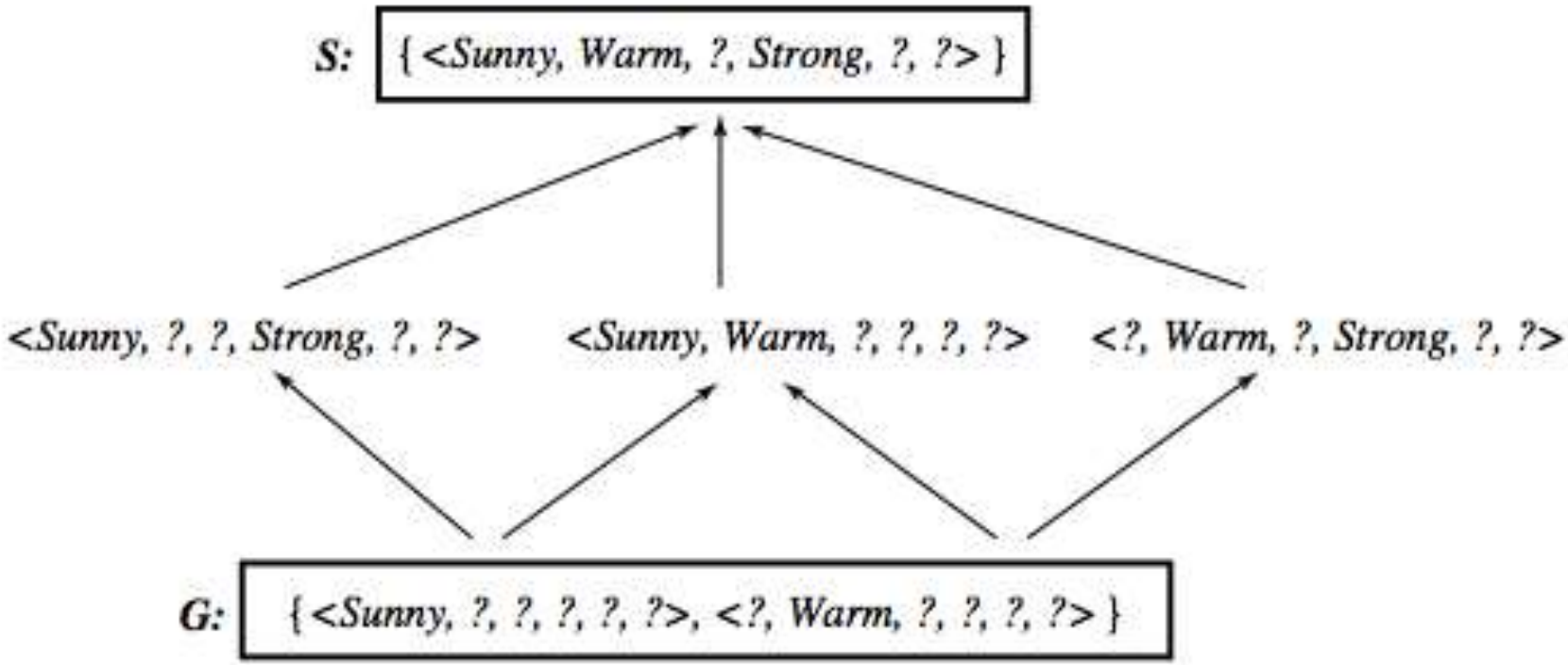
Observations

- ✓ The learned Version Space correctly describes the target concept, provided:
  - There are no errors in the training examples
  - There is some hypothesis that correctly describes the target concept
- ✓ If *S* and *G* converge to a single hypothesis then concept is exactly learned
- ✓ In case of errors in the training, useful hypothesis are discarded, no recovery possible
- ✓ An empty version space means no hypothesis in *H* is consistent with training examples

Ordering on Training Examples

- ✓ The learned version space does not change with **different orderings** of training examples
- ✓ **Efficiency does**
- ✓ **Optimal strategy (if you are allowed to choose)**
  - **Generate instances that satisfy half the hypotheses in the current version space. For example:**
  - ⟨*Sunny, Warm, Normal, Light, Warm, Same*⟩ **satisfies 3/6 hypothesis**
  - **Ideally the *VS* can be reduced by half at each experiment**
  - **Correct target found in** $\lceil log_2 |VS| \rceil$ **experiments**

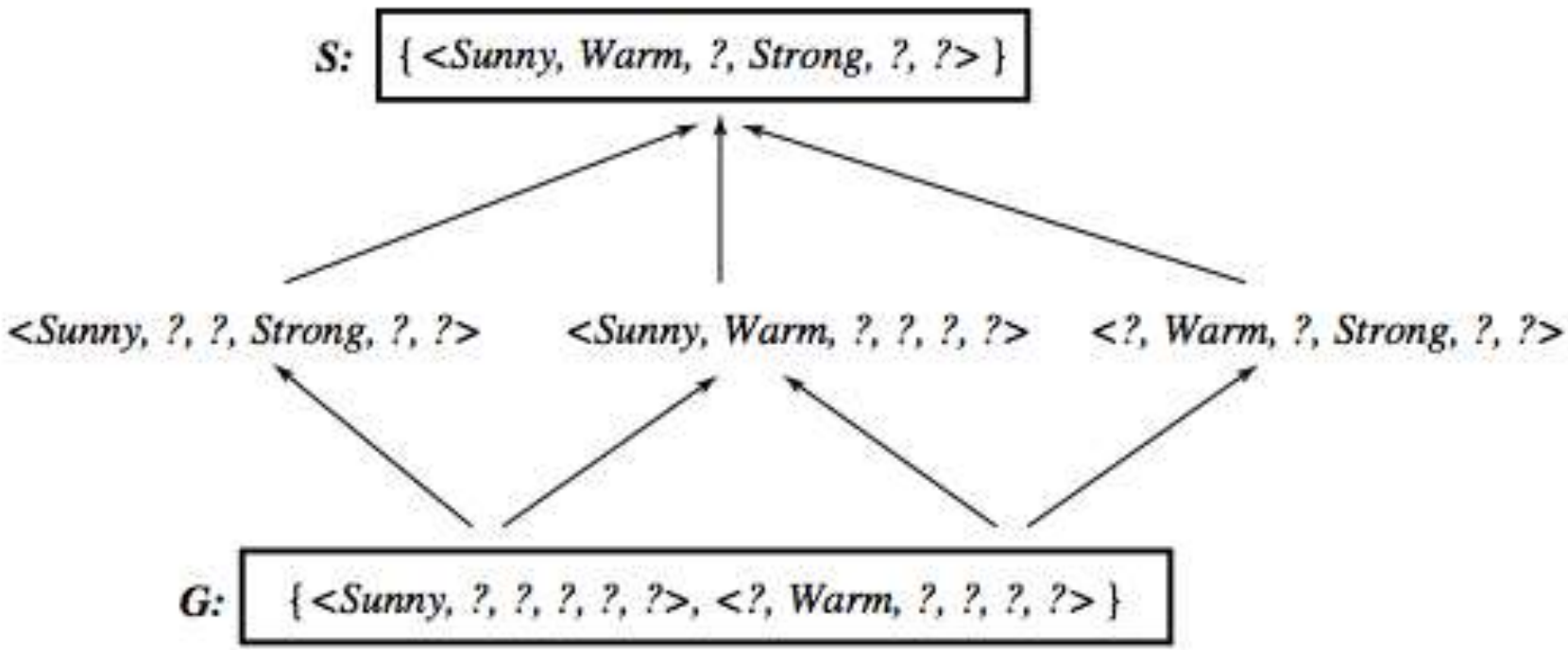
Use of Partially Learned Concepts



$\langle Sunny\ Warm\ Normal\ Strong\ Cool\ Change \rangle$

Classified as positive by all hypothesis, since satisfies any hypothesis in S

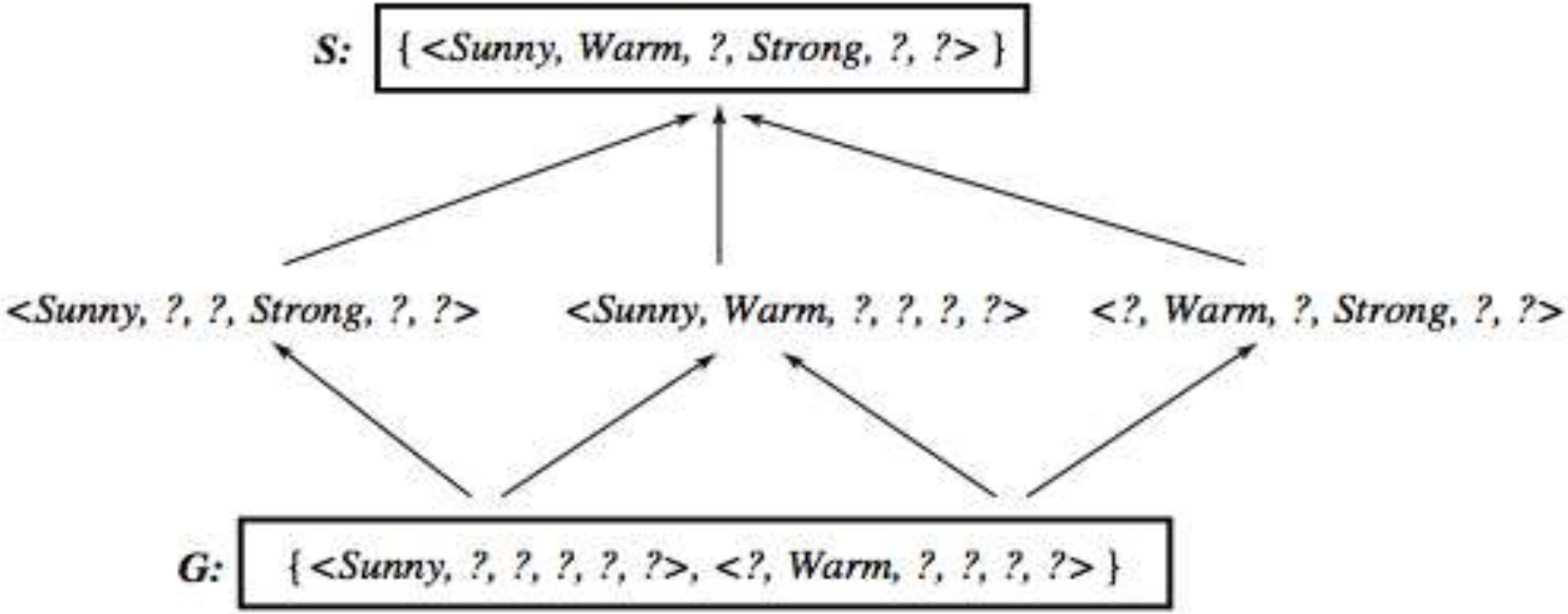
Classifying new examples



$\langle Rainy\ Cool\ Normal\ Light\ Warm\ Same \rangle$

Classified as negative by all hypothesis, since does not satisfy any hypothesis in G

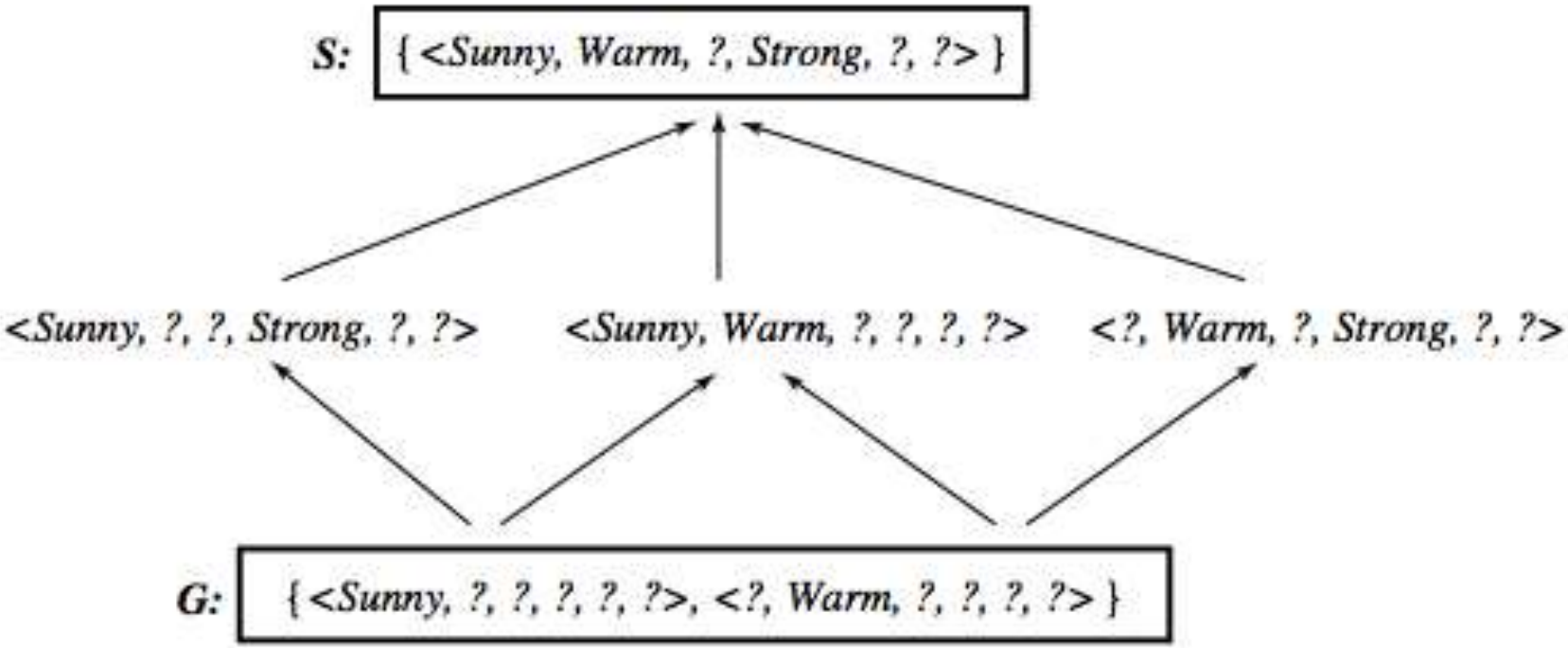
Classifying new examples…..



$\langle Sunny\ Warm\ Normal\ Light\ Warm\ Same \rangle$

Uncertain classification: half hypothesis are consistent, half are not consistent

Classifying new examples…..



⟨Sunny, Cold, Normal, Strong, Warm, Same⟩

4 hypothesis not satisfied, 2 satisfied
Probably a negative instance.

Comparison

Algorithm	Order	Strategy	N/P
FIND-S	Specific-to-general	Top-down	Positive
LIST-THEN-ELIMINATE	General-to-Specific	Bottom-up	Negative
CANDIDATE-ELIMINATION	Bi-directional	Bi-directional	Both

Concept Learning - Summary

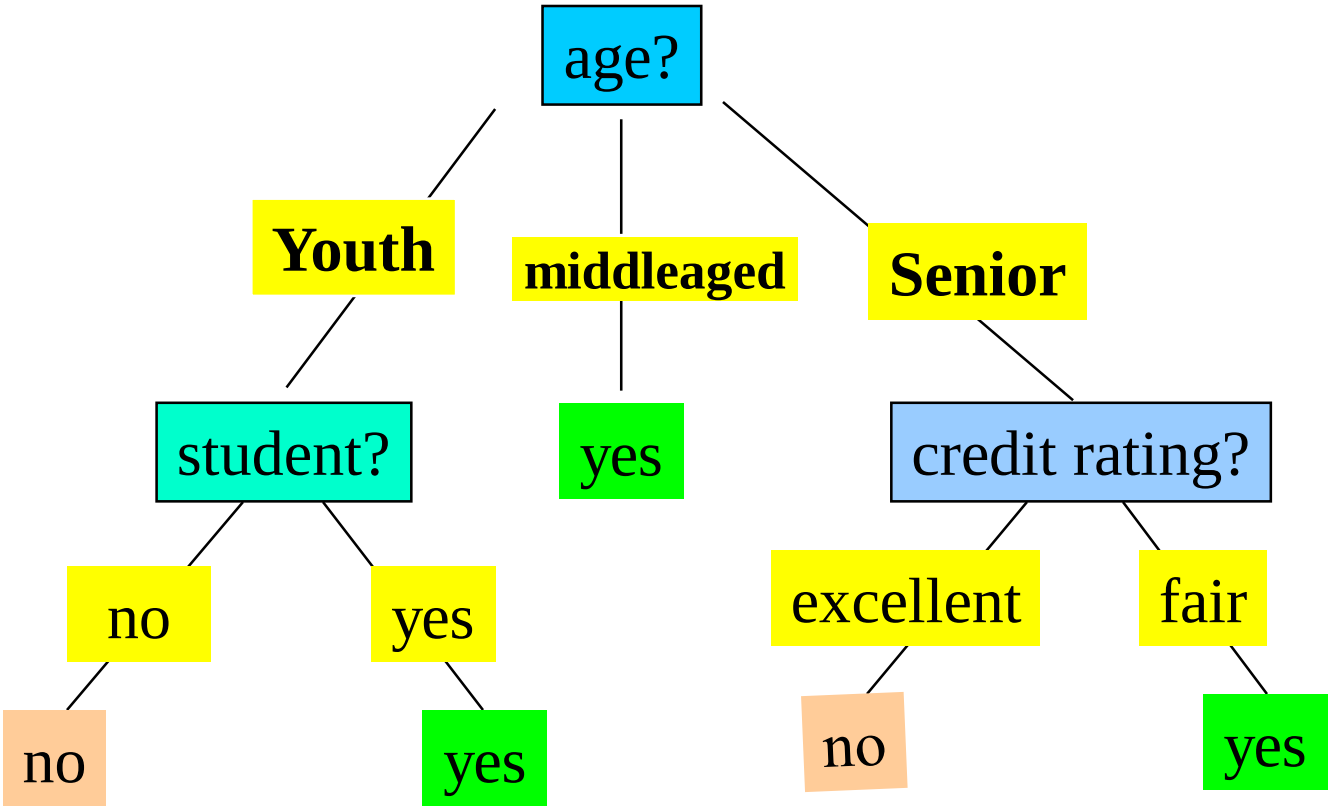
- ✓ **Concept learning can be seen as a problem of searching**
- ✓ The **general-to-specific** partial ordering of hypotheses provides a **useful structure** for **organizing the search** through the hypothesis space.
- ✓ The **FIND-S algorithm** utilizes this **general-to-specific** ordering, performing a **specific-to-genera**l search through the hypothesis space.
- ✓ The **CANDIDATE-ELIMINATION algorithm** utilizes this **general-to-specific** ordering to compute the **version space** by **incrementally** computing the sets of **maximally specific (S) and maximally general (G) hypotheses.**
- ✓ The **CANDIDATE-ELIMINATION algorithm is not robust to noisy data or to situations** in which the unknown target concept is not expressible in the provided hypothesis space.

Decision Trees and ID3 Algorithm

Decision Tree

A flow-chart-like tree structure
Internal node denotes a test on an attribute
Branch represents an outcome of the test
Leaf nodes represent class labels or class distribution
Constructed in a top down recursive divide and conquer approach

Example: A Decision Tree for "*buys_computer*"



Decision Tree Induction: Training Dataset

RID	age	income	student	credit_rating	buys_computer
1	Youth	high	no	fair	no
2	Youth	high	no	excellent	no
3	middle a	high	no	fair	yes
4	Senior	medium	no	fair	yes
5	Senior	low	yes	fair	yes
6	Senior	low	yes	excellent	no
7	middle a	low	yes	excellent	yes
8	Youth	medium	no	fair	no
9	Youth	low	yes	fair	yes
10	Senior	medium	yes	fair	yes
11	Youth	medium	yes	excellent	yes
12	middle a	medium	no	excellent	yes
13	middle a	high	yes	fair	yes
14	Senior	medium	no	excellent	no

Algorithms for Decision tree construction

ID3(Iterative Dichotomiser) by J Ross Quinlay
C4.5 (extension of ID3) by E B Hunt and J Mann
CART (Classification and Regression Trees)

Steps

Decision tree generation consists of two phases

- Tree construction
  - At start, all the training examples are at the root
  - Partition examples recursively based on selected attributes
- Tree pruning
  - Identify and remove branches that reflect noise or outliers

Use of decision tree: Classifying an unknown sample

- Test the attribute values of the sample against the decision tree

Attribute Selection Measures

Information Gain
Gain ratio
Gini index
Selection measures provides a ranking for each attribute describing the given training tuples.
The attribute having the best score for the measure is chosen as the splitting attribute for the given tuples.

Information Gain

ID3 algorithm uses information Gain
Invented by C shannon
Based on entropy (information theory)
Let node N hold the tuples of partition D.
the attribute with the highest information gain is chosen as the spliting attribute for the node N

Entropy

A formula to calculate the homogeneity of a sample.
A completely homogeneous sample has entropy of 0.
An equally divided sample has entropy of 1.
Entropy(s) = - p+log2 (p+) -p-log2 (p-) for a sample of negative and positive elements.
The formula for entropy is:

$$Entropy(S) = \sum_{i=1}^{c} p_i \log_2 p_i$$

Information Gain (IG)

The information gain is based on the decrease in entropy after a dataset is split on an attribute.
Which attribute creates the most homogeneous branches?
First the entropy of the total dataset is calculated.
The dataset is then split on the different attributes.
The entropy for each branch is calculated. Then it is added proportionally, to get total entropy for the split.
The resulting entropy is subtracted from the entropy before the split.
The result is the Information Gain, or decrease in entropy.
The attribute that yields the largest IG is chosen for the decision node.

Information Gain (ID3)

- Expected information (entropy) needed to classify a tuple in D:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

  - Where, p_i be the probability that an arbitrary tuple in D belongs to class C_i, *i.e.*, $p_i = |C_{i,D}|/|D|$
- Information needed (after using A to split D into v partitions) to classify D:

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

- Information gained by branching on attribute A

$$Gain(A) = Info(D) - Info_A(D)$$

Decision Tree Induction: Training Dataset

RID	age	income	student	credit_rating	buys_computer
1	Youth	high	no	fair	no
2	Youth	high	no	excellent	no
3	middle a	high	no	fair	yes
4	Senior	medium	no	fair	yes
5	Senior	low	yes	fair	yes
6	Senior	low	yes	excellent	no
7	middle a	low	yes	excellent	yes
8	Youth	medium	no	fair	no
9	Youth	low	yes	fair	yes
10	Senior	medium	yes	fair	yes
11	Youth	medium	yes	excellent	yes
12	middle a	medium	no	excellent	yes
13	middle a	high	yes	fair	yes
14	Senior	medium	no	excellent	no

Example for finding Information Gain

|D|=14, no. of Classes=2, m=2

- Class C_1: buys_computer = "yes" | C_1 |=9
- Class C_2 : buys_computer = "no" | C_2 |=5

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

$p_i = |C_{i,D}|/|D|$

age	C_1	C_2
youth	2	3
middle a	4	0
senior	3	2

$$Info(D) = Info(9,5) = -\frac{9}{14}\log_2(\frac{9}{14}) - \frac{5}{14}\log_2(\frac{5}{14}) = 0.940$$

RID	age	income	student	credit_rating	buys_computer
1	Youth	high	no	fair	no
2	Youth	high	no	excellent	no
3	middle a	high	no	fair	yes
4	Senior	medium	no	fair	yes
5	Senior	low	yes	fair	yes
6	Senior	low	yes	excellent	no
7	middle a	low	yes	excellent	yes
8	Youth	medium	no	fair	no
9	Youth	low	yes	fair	yes
10	Senior	medium	yes	fair	yes
11	Youth	medium	yes	excellent	yes
12	middle a	medium	no	excellent	yes
13	middle a	high	yes	fair	yes
14	Senior	medium	no	excellent	no

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

$$Info_{age}(D) = \frac{5}{14} Info(D1) + \frac{4}{14} Info(D2) + \frac{5}{14} Info(D3) = 0.694$$

$$Info(D1) = Info(2,3) = -\frac{2}{5}\log_2(\frac{2}{5}) - \frac{3}{5}\log_2(\frac{3}{5})$$

$$Info(D2) = Info(4,0) = -\frac{4}{4}\log_2(\frac{4}{4}) - \frac{0}{4}\log_2(\frac{0}{4})$$

$$Info(D3) = Info(3,2) = -\frac{3}{5}\log_2(\frac{3}{5}) - \frac{2}{5}\log_2(\frac{2}{5})$$

Contd..

$$Gain(age) = Info(D) - Info_{age}(D) = 0.246$$

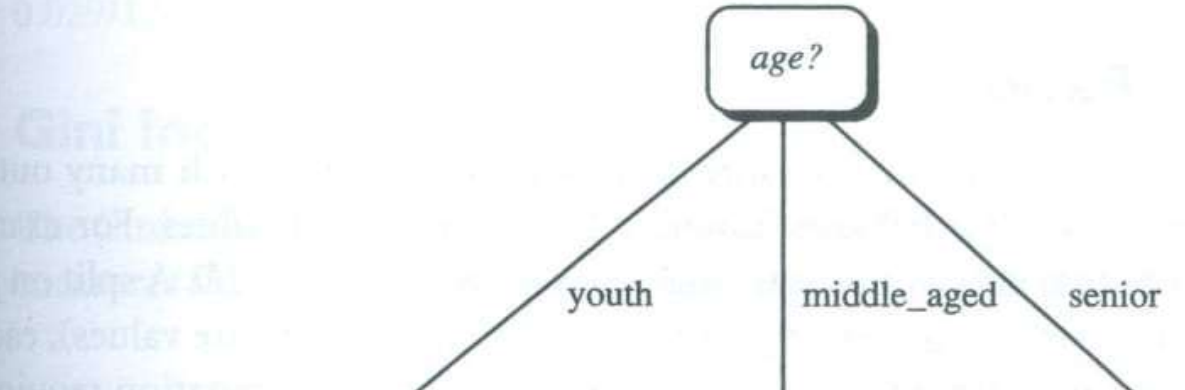
similarly

$$Gain(income) = 0.029$$

$$Gain(student) = 0.151$$

$$Gain(credit\_rating) = 0.048$$

Output after age as splitting attribute

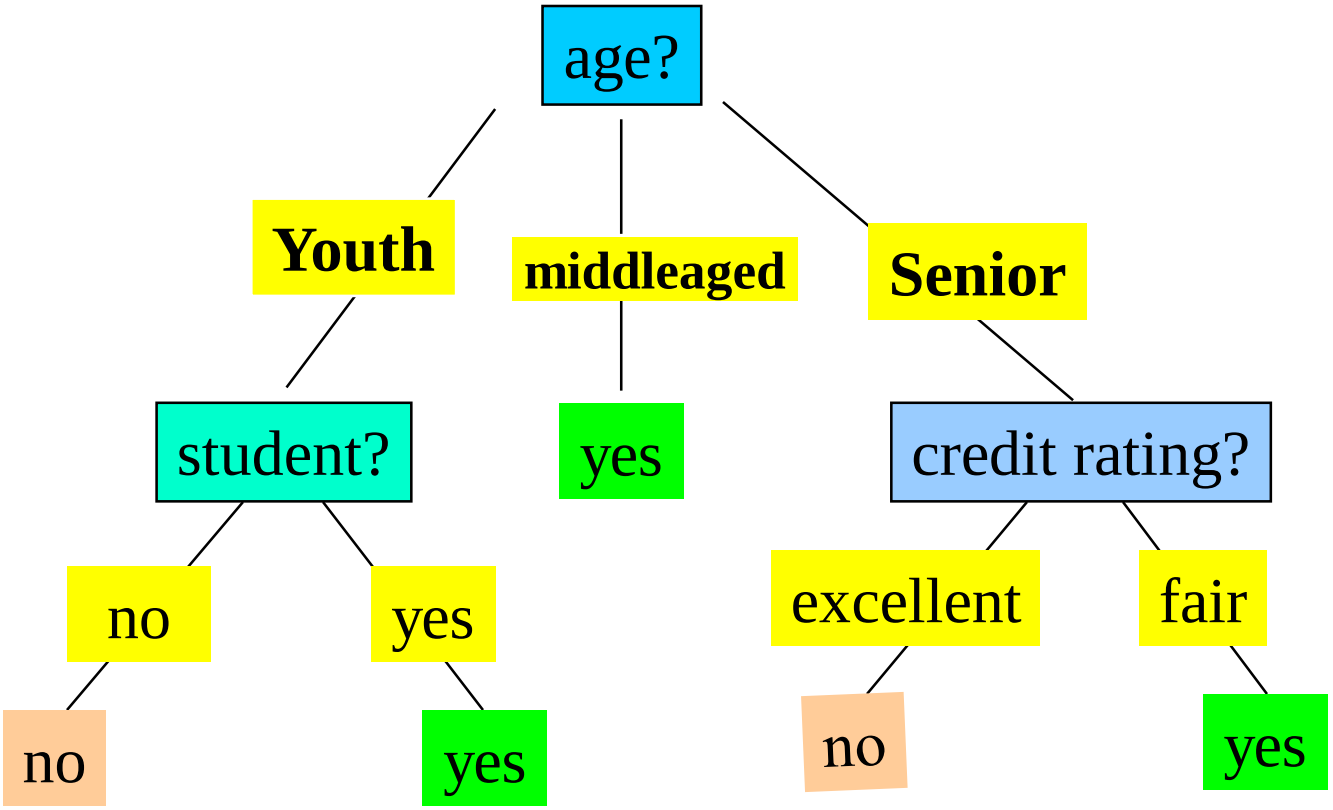


income	*student*	*credit_rating*	*class*
high	no	fair	no
high	no	excellent	no
medium	no	fair	no
low	yes	fair	yes
medium	yes	excellent	yes

income	*student*	*credit_rating*	*class*
medium	no	fair	yes
low	yes	fair	yes
low	yes	excellent	no
medium	yes	fair	yes
medium	no	excellent	no

income	*student*	*credit_rating*	*class*
high	no	fair	yes
low	yes	excellent	yes
medium	no	excellent	yes
high	yes	fair	yes

Output: A Decision Tree for "*buys_computer*"



Algorithm for Decision Tree Induction

Tree is constructed in a top-down recursive divide-and-conquer manner

At start, all the training examples are at the root

Attributes are categorical (if continuous-valued, they are discretized in advance)

Examples are partitioned recursively based on selected attributes

Test attributes are selected on the basis of a heuristic or statistical measure (e.g., information gain)

Time Complexity of ID3 Algorithm is O(n*|D|*log|D|)

Conditions for stopping partitioning

The recursive partitioning stops only when any of the following terminating condition is true

All of the tuples in partition D belong to same class

There are no remaining attributes in which the tuple may be further partitioned. In this case, majority voting is employed. This involves converting node N into leaf and labeled it with the most common class in D.

There are no tuples for a given branch, that is, a partition Dj is empty. In this case, a leaf is created with the majorityclass in D

ID3 Algorithm

Algorithm: Generate_decision_tree. Generate a decision tree from the training tuples of data partition, D.

Input:

- Data partition, D, which is a set of training tuples and their associated class labels;
- *attribute_list*, the set of candidate attributes;
- *Attribute_selection_method*, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a *splitting_attribute* and, possibly, either a *split-point* or *splitting subset*.

Output: A decision tree.

Method:

```
(1)  create a node N;
(2)  if tuples in D are all of the same class, C, then
(3)      return N as a leaf node labeled with the class C;
(4)  if attribute_list is empty then
(5)      return N as a leaf node labeled with the majority class in D; // majority voting
(6)  apply Attribute_selection_method(D, attribute_list) to find the "best" splitting_criterion;
(7)  label node N with splitting_criterion;
(8)  if splitting_attribute is discrete-valued and
         multiway splits allowed then // not restricted to binary trees
(9)      attribute_list ← attribute_list − splitting_attribute; // remove splitting_attribute
(10) for each outcome j of splitting_criterion
     // partition the tuples and grow subtrees for each partition
(11)     let D_j be the set of data tuples in D satisfying outcome j; // a partition
(12)     if D_j is empty then
(13)         attach a leaf labeled with the majority class in D to node N;
(14)     else attach the node returned by Generate_decision_tree(D_j, attribute_list) to node N;
     endfor
(15) return N;
```

Computing Information-Gain for Continuous-Value Attributes

Let attribute A be a continuous-valued attribute

Must determine the *best split point* for A

- Sort the value A in increasing order
- Typically, the midpoint between each pair of adjacent values is considered as a possible *split point*
  - $(a_i+a_{i+1})/2$ is the midpoint between the values of a_i and a_{i+1}
- The point with the *minimum expected information requirement* for A is selected as the split-point for A

Split:

- D1 is the set of tuples in D satisfying A ≤ split-point, and D2 is the set of tuples in D satisfying A > split-point

Gain Ratio for Attribute Selection (C4.5)

Information gain measure is biased towards attributes with a large number of values

C4.5 (a successor of ID3) uses gain ratio to overcome the problem (normalization to information gain)

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right)$$

GainRatio(A) = Gain(A)/SplitInfo(A)

Ex.

$$SplitInfo_A(D) = -\frac{4}{14} \times \log_2\left(\frac{4}{14}\right) - \frac{6}{14} \times \log_2\left(\frac{6}{14}\right) - \frac{4}{14} \times \log_2\left(\frac{4}{14}\right) = 0.926$$

gain_ratio(income) = 0.029/0.926 = 0.031

The attribute with the maximum gain ratio is selected as the splitting attribute

Gini index (CART, IBM IntelligentMiner)

If a data set *D* contains examples from *n* classes, gini index, *gini*(*D*) is defined as

$$gini(D) = 1 - \sum_{j=1}^{n} p_j^2$$

where p_j is the relative frequency of class *j* in *D*

If a data set *D* is split on A into two subsets D_1 and D_2, the *gini* index *gini*(*D*) is defined as

$$gini_A(D) = \frac{|D_1|}{|D|} gini(D_1) + \frac{|D_2|}{|D|} gini(D_2)$$

Reduction in Impurity:

$$\Delta gini(A) = gini(D) - gini_A(D)$$

The attribute provides the smallest $gini_{split}(D)$ (or the largest reduction in impurity) is chosen to split the node (*need to enumerate all the possible splitting points for each attribute*)

Gini index (CART, IBM IntelligentMiner)

Ex. D has 9 tuples in buys_computer = "yes" and 5 in "no"

$$gini(D) = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 = 0.459$$

Suppose the attribute income partitions D into 10 in D_1: {low, medium} and 4 in D_2

$$gini_{income \in \{low,medium\}}(D) = \left(\frac{10}{14}\right) Gini(D_1) + \left(\frac{4}{14}\right) Gini(D_1)$$

$$= \frac{10}{14}(1 - (\frac{6}{10})^2 - (\frac{4}{10})^2) + \frac{4}{14}(1 - (\frac{1}{4})^2 - (\frac{3}{4})^2)$$

$$= 0.450$$

$$= Gini_{income \in \{high\}}(D)$$

but $gini_{\{medium,high\}}$ is 0.30 and thus the

All attributes are assumed continuous-valued

May need other tools, e.g., clustering, to get the possible split values

Can be modified for categorical attributes

Extracting Classification Rules from Trees

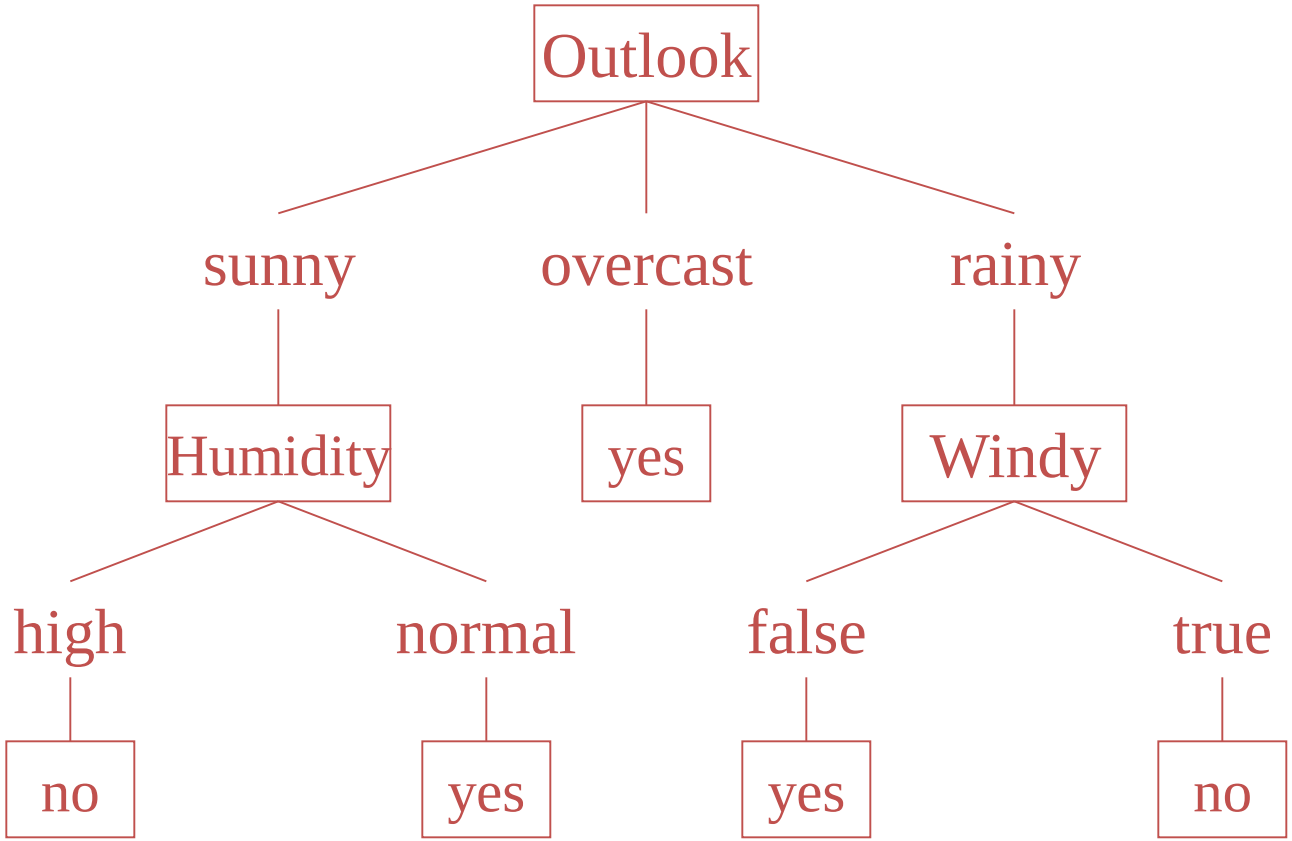
Represent the knowledge in the form of IF-THEN rules
One rule is created for each path from the root to a leaf
Each attribute-value pair along a path forms a conjunction
The leaf node holds the class prediction
Rules are easier for humans to understand
Example

IF *age* = "<=30" AND *student* = "*no*" THEN *buys_computer* = "*no*"
IF *age* = "<=30" AND *student* = "*yes*" THEN *buys_computer* = "*yes*"
IF *age* = "31…40" THEN *buys_computer* = "*yes*"
IF *age* = ">40" AND *credit_rating* = "*excellent*" THEN *buys_computer* = "*yes*"
IF *age* = ">40" AND *credit_rating* = "*fair*" THEN *buys_computer* = "*no*"

Example 2: database: playtennis

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	Normal	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	High	Strong	Yes
D8	Sunny	Mild	Normal	Weak	No
D9	Sunny	Hot	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Cool	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Decision Tree For Playing Tennis



Variable Quality Measures

Outlook

Sunny

Outlook	Temp	Hum	Wind	Play
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Overcast

Outlook	Temp	Hum	Wind	Play
Overcast	Hot	High	Weak	Yes
Overcast	Cool	Normal	Strong	Yes

Rain

Outlook	Temp	Hum	Wind	Play
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Continuous Variables

Temp.	Play
80	No
85	No
83	Yes
75	Yes
68	Yes
65	No
64	Yes
72	No
75	Yes
70	Yes
69	Yes
72	Yes
81	Yes
71	No

Sort →

Temp.	Play
64	Yes
65	No
68	Yes
69	Yes
70	Yes
71	No
72	No
72	Yes
75	Yes
75	Yes
80	No
81	Yes
83	Yes
85	No

Temp.< 64.5 $\Delta I=0.048$

Temp.< 66.5 $\Delta I=0.010$

Temp.< 70.5 $\Delta I=0.045$

Temp.< 73.5 $\Delta I=0.001$

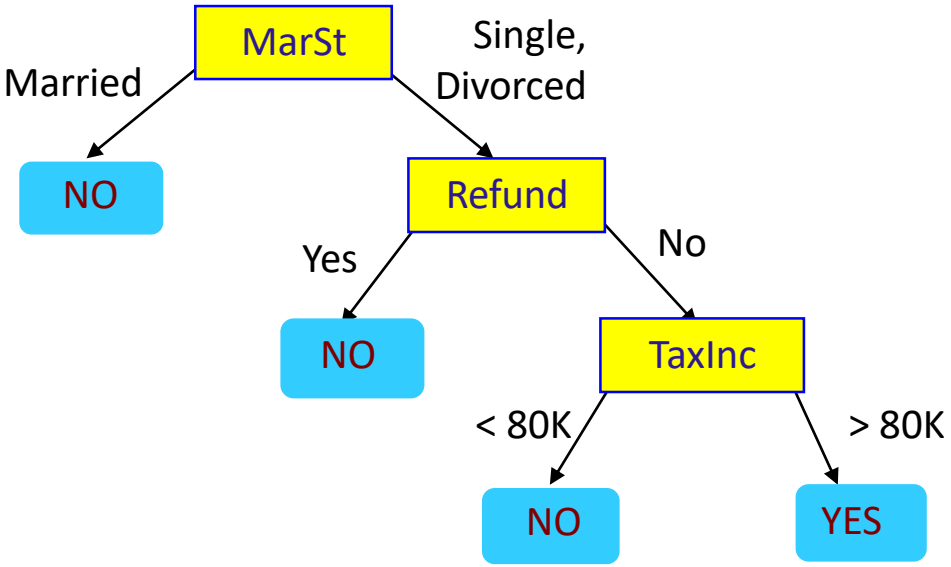
Temp.< 77.5 $\Delta I=0.025$

Temp.< 80.5 $\Delta I=0.000$

Temp.< 84 $\Delta I=0.113$

Example 3: Decision Tree

Tid	Refund	Marital Status	Taxable Income	Cheat
	categorical	*categorical*	*continuous*	*class*
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



There could be more than one tree that fits the same data!

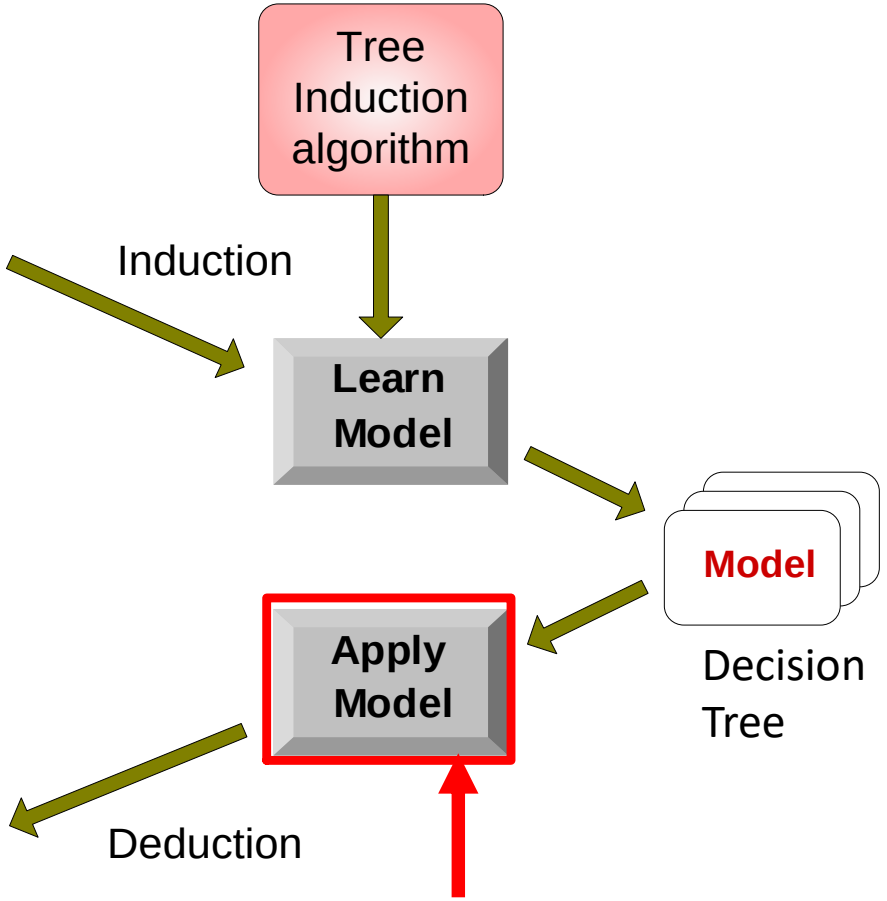
Decision Tree Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

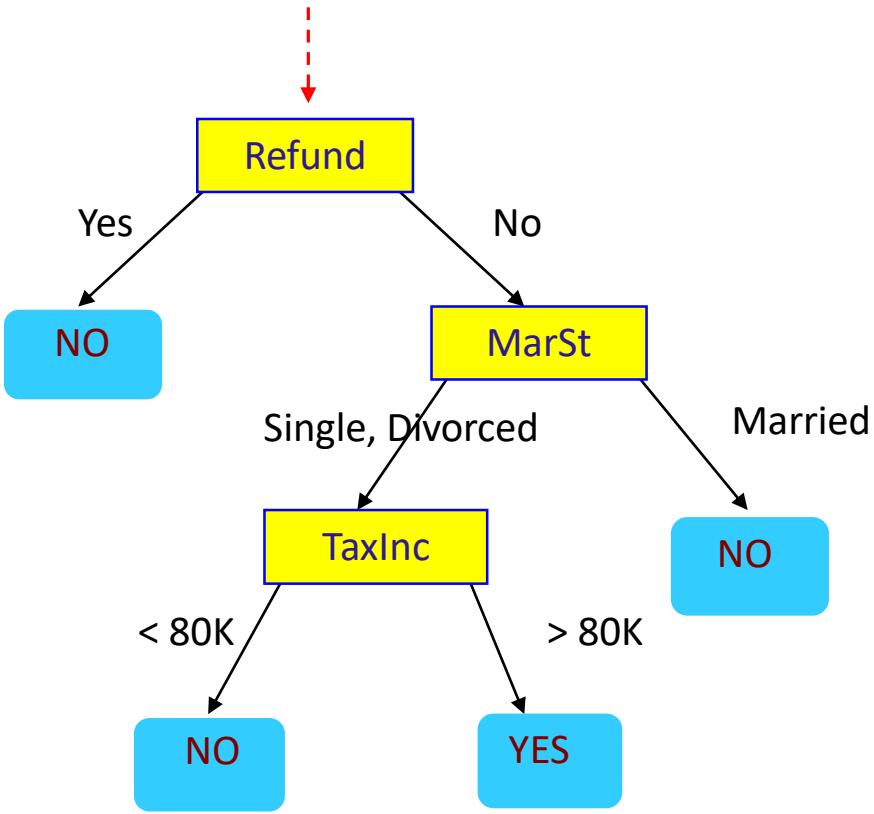
Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



Apply Model to Test Data

Start from the root of tree.



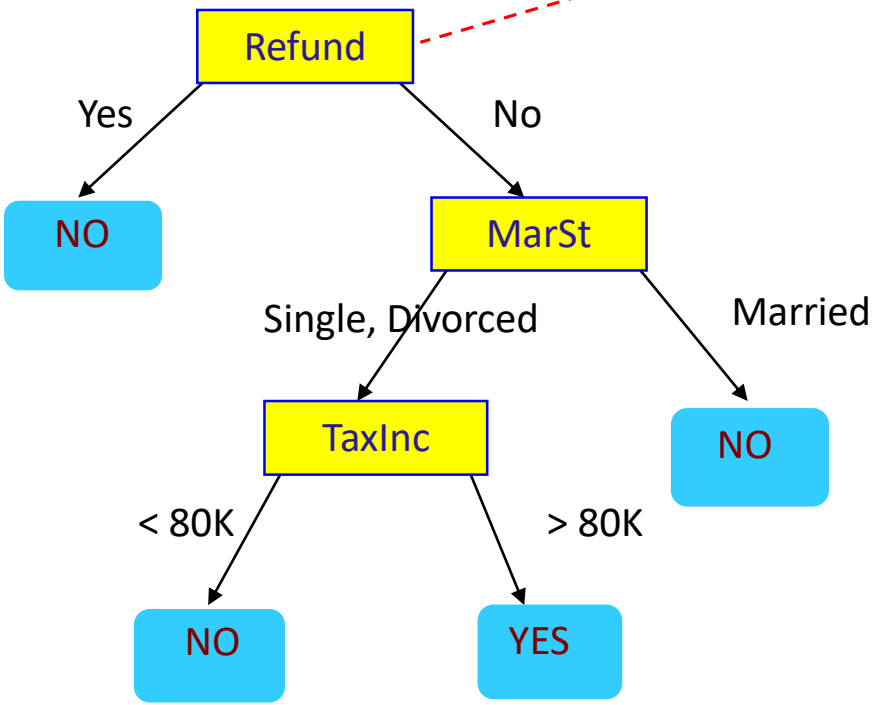
Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

Apply Model to Test Data

Test Data

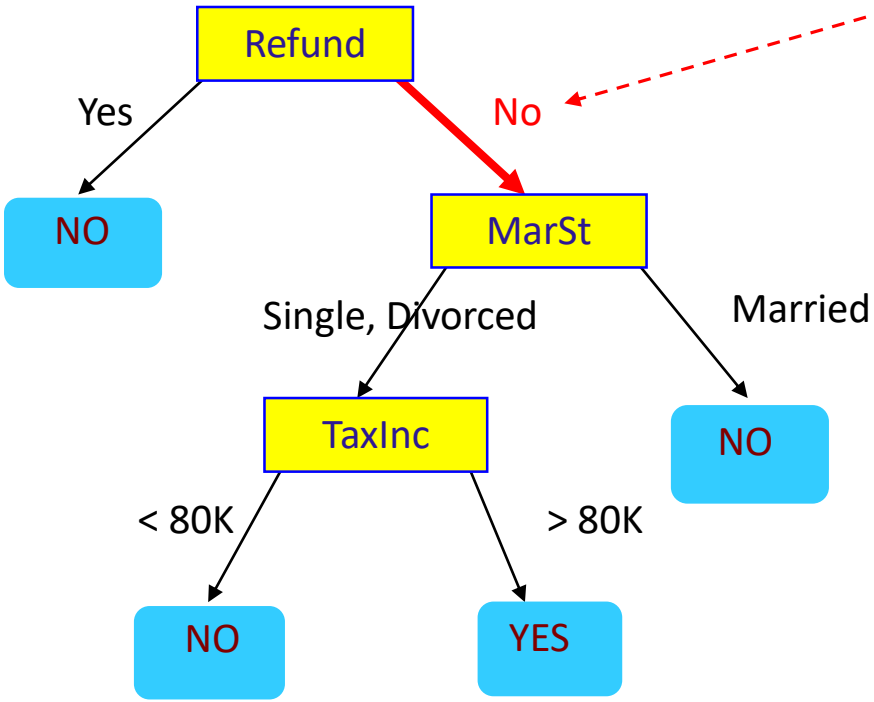
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Apply Model to Test Data

Test Data

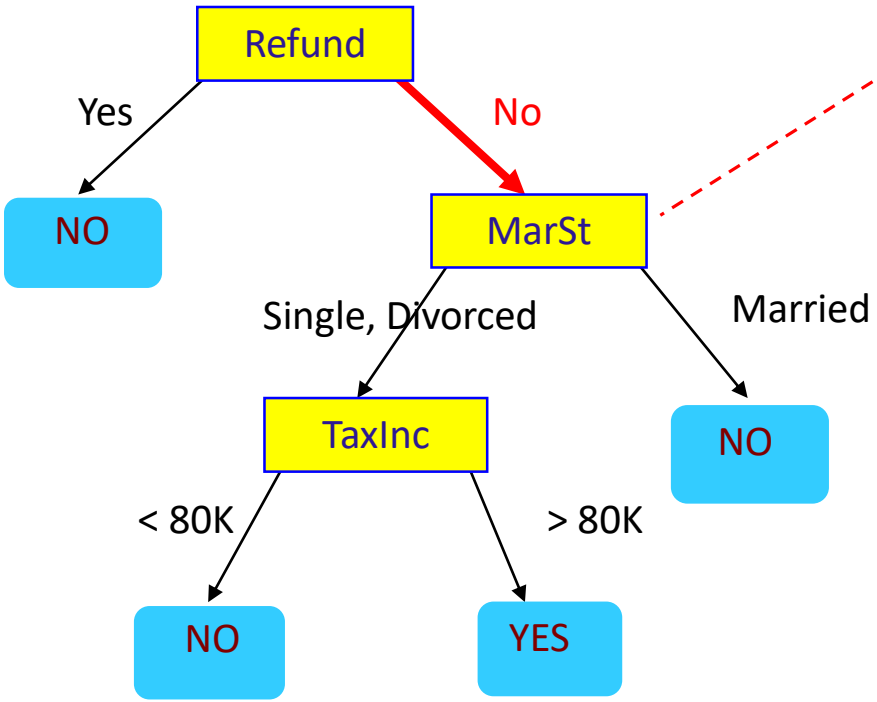
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Apply Model to Test Data

Test Data

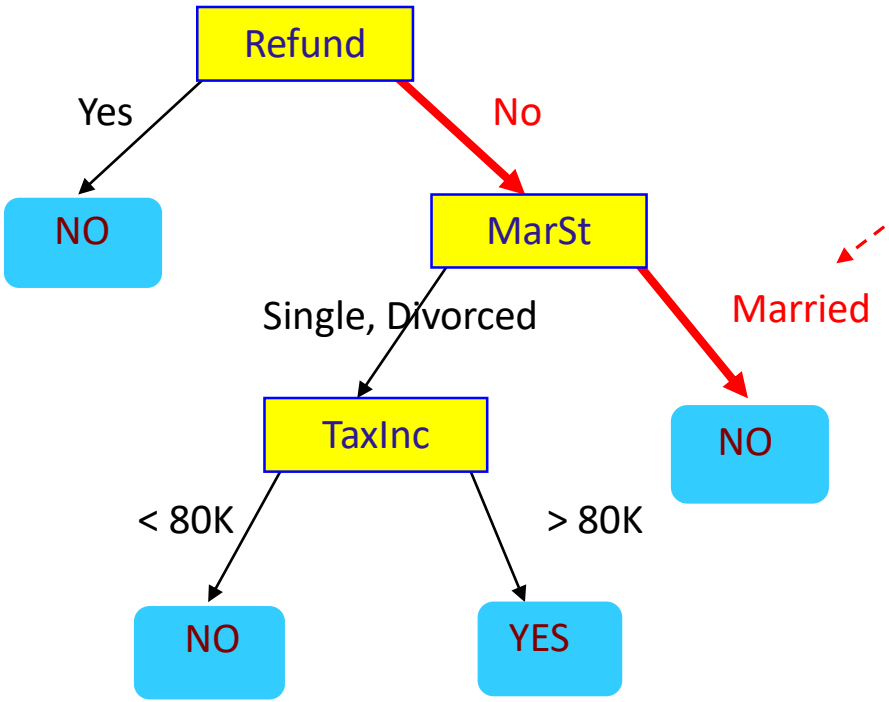
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



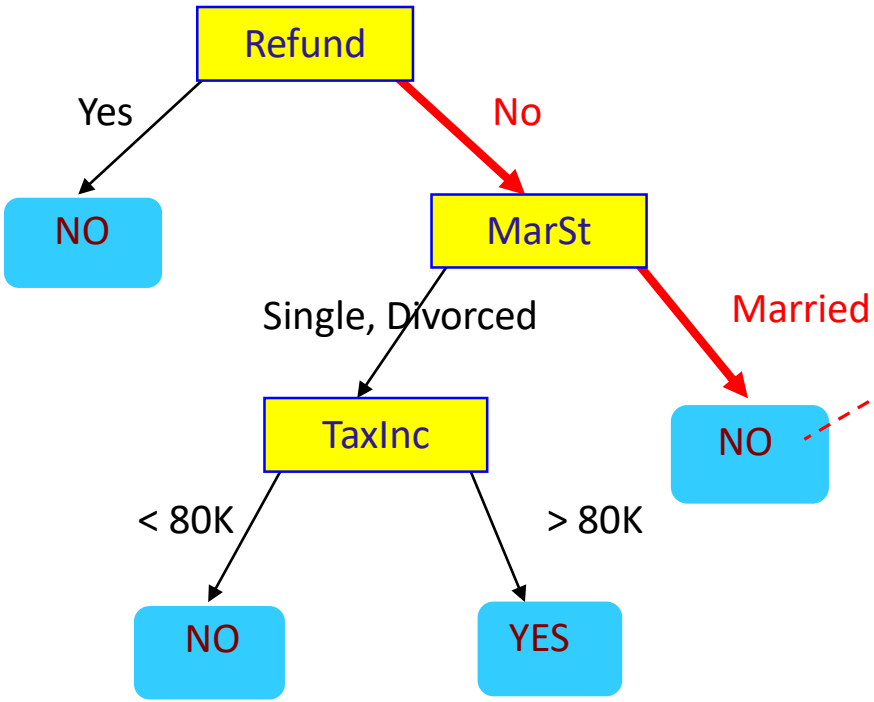
Apply Model to Test Data

Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Apply Model to Test Data



Test Data

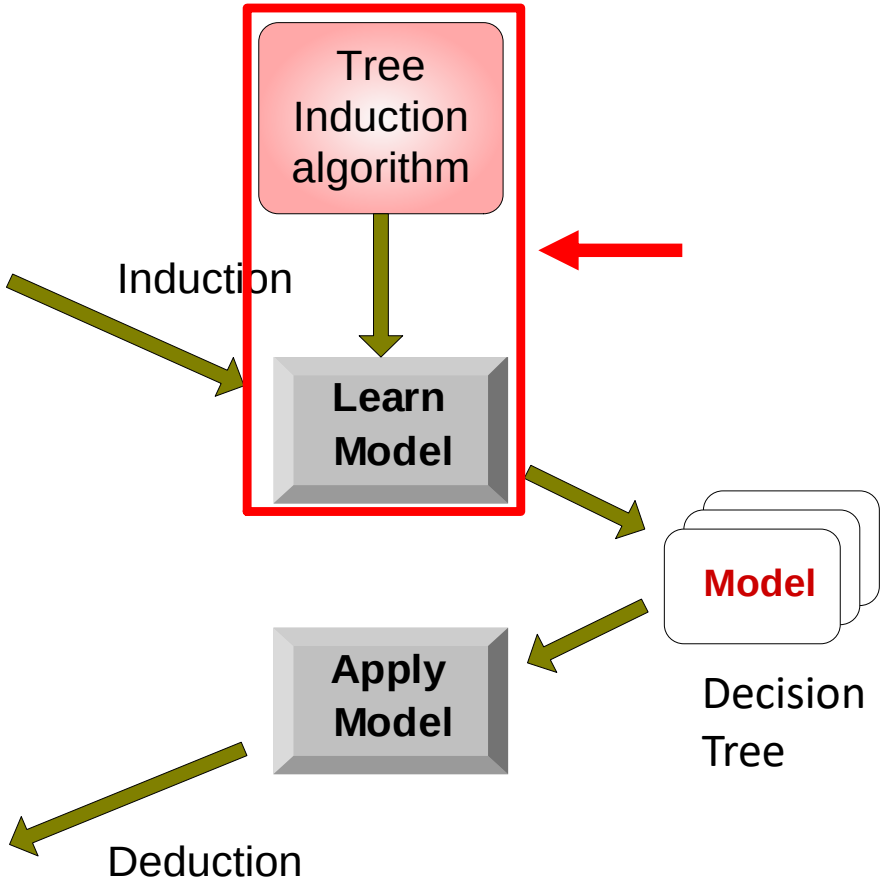
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

Assign Cheat to “No”

Decision Tree Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set



Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set

Advantages of using ID3

Understandable prediction rules are created from the training data.
Builds the fastest tree.
Builds a short tree.
Only need to test enough attributes until all data is classified.
Finding leaf nodes enables test data to be pruned, reducing number of tests.
Whole dataset is searched to create tree.

Disadvantages of using ID3

Only one attribute at a time is tested for making a decision.
Classifying continuous data may be computationally expensive, as many trees must be generated to see where to break the continuum.
Data may be over-fitted or over-classified, if a small sample is tested.

Comparing Attribute Selection Measures

The three measures, in general, return good results but

- Information gain:
  - biased towards multivalued attributes
- Gain ratio:
  - tends to prefer unbalanced splits in which one partition is much smaller than the others
- Gini index:
  - biased to multivalued attributes
  - has difficulty when # of classes is large
  - tends to favor tests that result in equal-sized partitions and purity in both partitions

Other Attribute Selection Measures

CHAID: a popular decision tree algorithm, measure based on χ^2 test for independence

C-SEP: performs better than info. gain and gini index in certain cases

G-statistics: has a close approximation to χ^2 distribution

MDL (Minimal Description Length) principle (i.e., the simplest solution is preferred):

The best tree as the one that requires the fewest # of bits to both (1) encode the tree, and (2) encode the exceptions to the tree

Multivariate splits (partition based on multiple variable combinations)

CART: finds multivariate splits based on a linear comb. of attrs.

Which attribute selection measure is the best?

Most give good results, none is significantly superior than others

Overfitting and Tree Pruning

Overfitting: An induced tree may overfit the training data

- Too many branches, some may reflect anomalies due to noise or outliers
- Poor accuracy for unseen samples

Two approaches to avoid overfitting

- Prepruning: Halt tree construction early—do not split a node if this would result in the goodness measure falling below a threshold
  - Difficult to choose an appropriate threshold
- Postpruning: Remove branches from a “fully grown” tree—get a sequence of progressively pruned trees
  - Use a set of data different from the training data to decide which is the “best pruned tree”

Enhancements to Basic Decision Tree Induction

Allow for continuous-valued attributes

- Dynamically define new discrete-valued attributes that partition the continuous attribute value into a discrete set of intervals

Handle missing attribute values

- Assign the most common value of the attribute
- Assign probability to each of the possible values

Attribute construction

- Create new attributes based on existing ones that are sparsely represented
- This reduces fragmentation, repetition, and replication

Scalable Decision Tree Induction Methods

SLIQ (EDBT'96 — Mehta et al.)

Builds an index for each attribute and only class list and the current attribute list reside in memory

SPRINT (VLDB'96 — J. Shafer et al.)

Constructs an attribute list data structure

PUBLIC (VLDB'98 — Rastogi & Shim)

Integrates tree splitting and tree pruning: stop growing the tree earlier

RainForest (VLDB'98 — Gehrke, Ramakrishnan & Ganti)

Builds an AVC-list (attribute, value, class label)

BOAT (PODS'99 — Gehrke, Ganti, Ramakrishnan & Loh)

Uses bootstrapping to create several small samples

Motivation

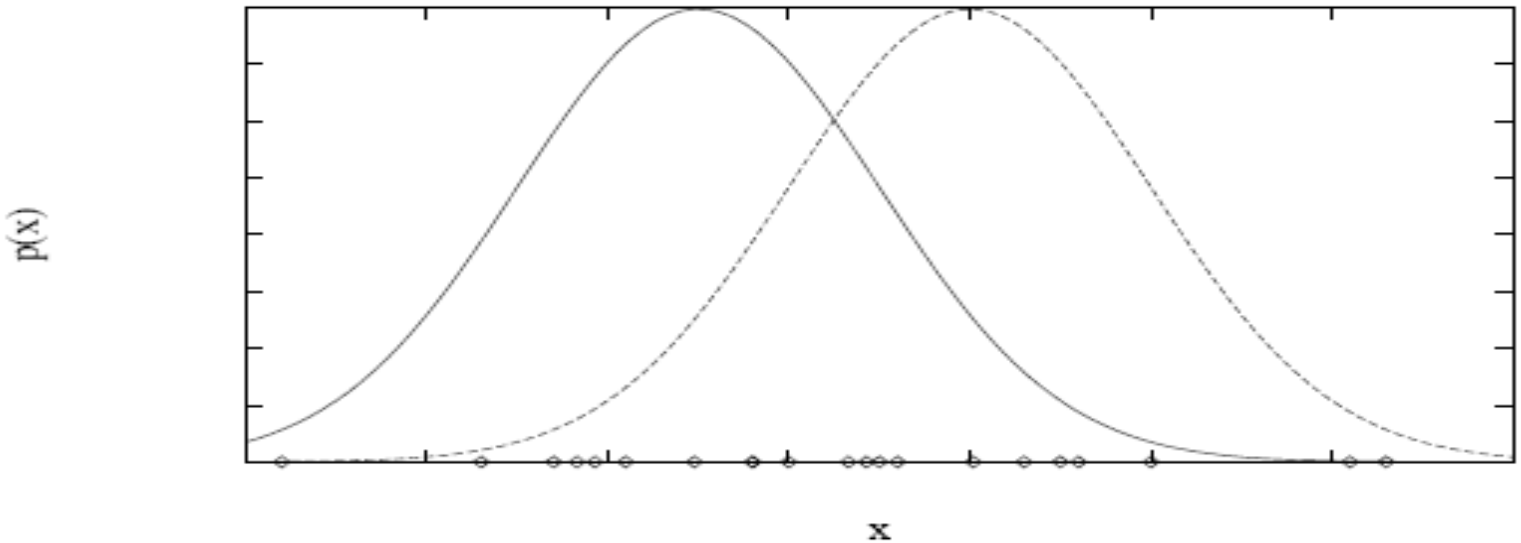
- In many practical learning settings, only a subset of the relevant instance features might be observable.
- For example, among many *Storm, Lightning, Thunder, ForestFire, Campfire,* and *BusTourGroup* have been observed. (In BBN example)
- If some variable is sometimes observed and sometimes not, then we can use the cases for which it has been observed to learn to predict its values when it is not.
- Many approaches have been proposed to handle the problem of learning in the presence of unobserved variables.
- EM algorithm (Dempster et al. 1977), a widely used approach to learning in the presence of unobserved variables.
- The EM algorithm can be used
  - even for variables whose value is never directly observed,
  - provided the general form of the probability distribution governing these variables is known.

Estimating Means of k Gaussians

- To simplify our discussion, we consider the special case
  - where the selection of the single Normal distribution at each step is based on choosing each with uniform probability,
  - where each of the k Normal distributions has the same variance σ^2, known value.
- The learning task is to output a hypothesis $h = (\mu_1, \ldots, \mu_k)$ that describes the means of each of the k distributions.
- We would like to find a maximum likelihood hypothesis for these means; that is, a hypothesis h that maximizes $p(D \mid h)$.

$$\mu_{ML} = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{m} (x_i - \mu)^2$$

Estimating Means of k Gaussians



Each instance x generated by

1. Choosing one of the k Gaussians with uniform probability
2. Generating an instance at random according to that Gaussian

EM for Estimation k Means

Given:

- Instances from X generated by mixture of k Gaussian distributions
- Unknown means $\langle \mu_1, \ldots, \mu_k \rangle$ of the k Gaussians
- Don't know which instance x_i was generated by which Gaussian

Determine:

- Maximum likelihood estimates of $\langle \mu_1, \ldots, \mu_k \rangle$

Think of full description of each instance as $y_i = \langle x_i, z_{i1}, z_{i2} \rangle$, where

- z_{ij} is 1 if x_i generated by jth Gaussian
- x_i observable
- z_{ij} unobservable

EM Algorithm: Pick random initial $h = \langle \mu_1, \mu_2 \rangle$, then iterate

E step: Calculate the expected value $E[z_{ij}]$ of each hidden variable z_{ij}, assuming the current hypothesis $h = \langle \mu_1, \mu_2 \rangle$ holds.

$$E[z_{ij}] = \frac{p(x = x_i | \mu = \mu_j)}{\sum_{n=1}^{2} p(x = x_i | \mu = \mu_n)}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_j)^2}}{\sum_{n=1}^{2} e^{-\frac{1}{2\sigma^2}(x_i - \mu_n)^2}}$$

M step: Calculate a new maximum likelihood hypothesis $h' = \langle \mu_1', \mu_2' \rangle$, assuming the value taken on by each hidden variable z_{ij} is its expected value $E[z_{ij}]$ calculated above. Replace $h = \langle \mu_1, \mu_2 \rangle$ by $h' = \langle \mu_1', \mu_2' \rangle$.

$$\mu_j \leftarrow \frac{\sum_{i=1}^{m} E[z_{ij}] \ x_i}{\sum_{i=1}^{m} E[z_{ij}]}$$

EM Algorithm

Converges to local maximum likelihood h
and provides estimates of hidden variables z_{ij}

In fact, local maximum in $E[\ln P(Y|h)]$

- Y is complete (observable plus unobservable variables) data
- Expected value is taken over possible values of unobserved variables in Y

General EM Problem

Given:

- Observed data $X = \{x_1, \ldots, x_m\}$
- Unobserved data $Z = \{z_1, \ldots, z_m\}$
- Parameterized probability distribution $P(Y|h)$, where
  - $Y = \{y_1, \ldots, y_m\}$ is the full data $y_i = x_i \cup z_i$
  - h are the parameters

Determine:

- h that (locally) maximizes $E[\ln P(Y|h)]$

General EM Method

Define likelihood function $Q(h'|h)$ which calculates $Y = X \cup Z$ using observed X and current parameters h to estimate Z

$$Q(h'|h) \leftarrow E[\ln P(Y|h')|h, X]$$

EM Algorithm:

Estimation (E) step: Calculate $Q(h'|h)$ using the current hypothesis h and the observed data X to estimate the probability distribution over Y.

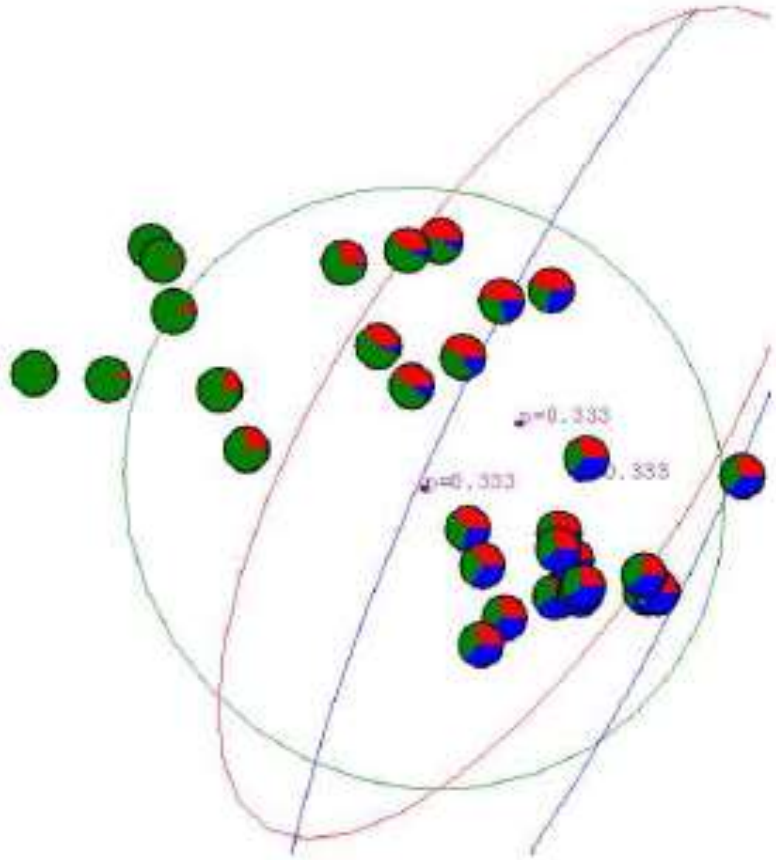
$$Q(h'|h) \leftarrow E[\ln P(Y|h')|h, X]$$

Maximization (M) step: Replace hypothesis h by the hypothesis h' that maximizes this Q function.

$$h \leftarrow \underset{h'}{\operatorname{argmax}}\, Q(h'|h)$$

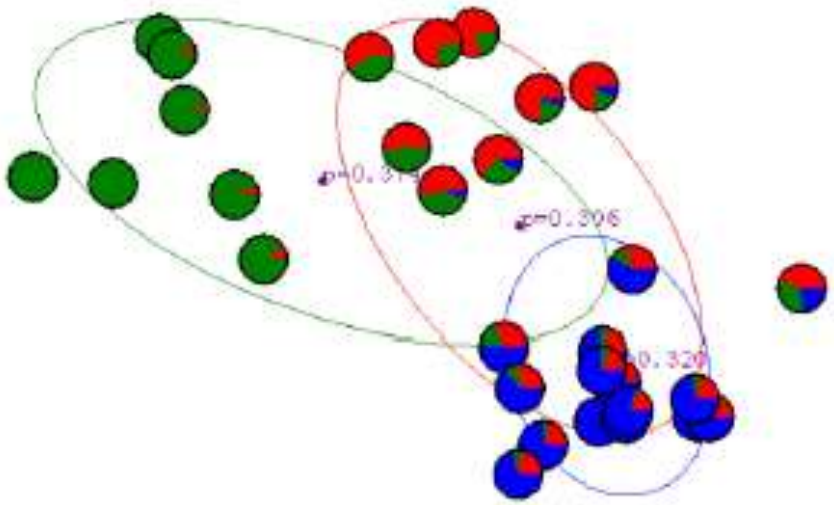
Example

Gaussian Mixture Example: Start

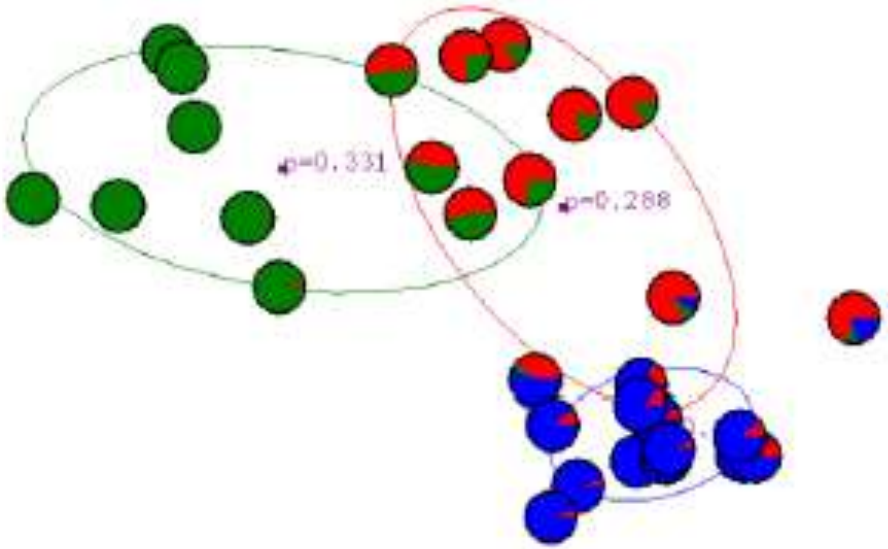


Advance apologies: in Black and White this example will be incomprehensible

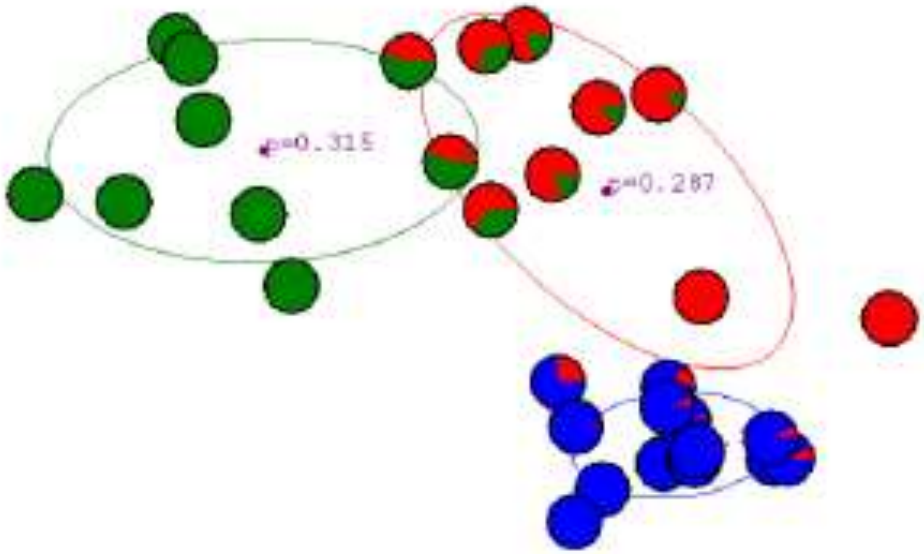
After 2nd iteration



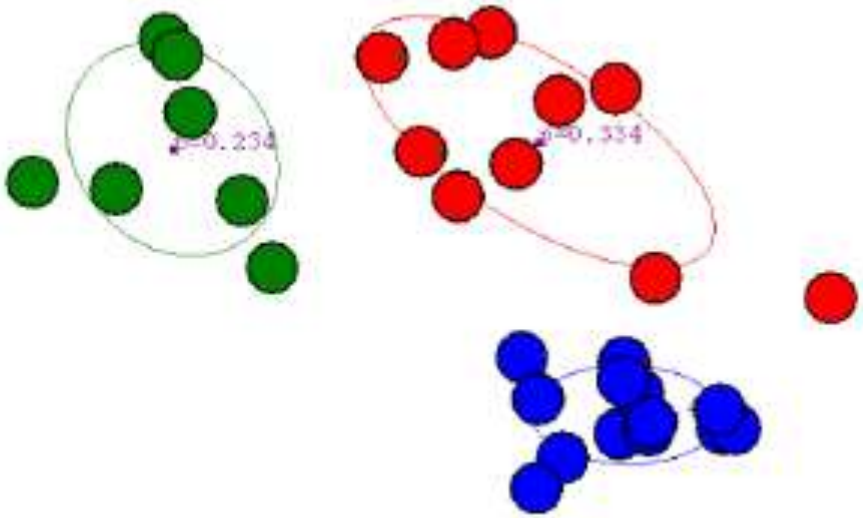
After 4th iteration



After 6th iteration



After 20th iteration



Thank You

Sol:- $\left(\frac{2}{5}\right) \log_2\left(\frac{2}{5}\right) + \frac{3}{5} \log_2\left(\frac{3}{5}\right)$

$= \underline{\underline{0.97}}$

(ii) [1 1] (1+1 = 2)

$= \frac{1}{2} \log_2 0.5 + \frac{1}{2} \log_2 0.5$

$= \underline{\underline{1}}$

(iii) [2 0] 2+0 = 2

$= \frac{2}{2} \log_2 \frac{2}{0} + \frac{0}{2} \log_2 \frac{0}{2}$

$= \underline{\underline{0}}$

[2 0]
= 0 (always)
[1 1]
= 1 (always)

(2) Consider the below dataset find the entropy of information gain of each attribute:

Student	Location	Target
1. Rama	College	Read +
2. Bhima	Home	Skips −
3. Rama	Office	Skips −
4. Bhima	Office	Read +

Sol:-

Student

Rama: [1 1] (+ −) = 1

Bhima: [1 1] (+ −) = 1

Location

College [1 0] = 0

Home [0 1] = 0

Office [1 1] = 1

3. Consider the below dataset draw the decision tree, using ID3 Algorithm.

	Student	First Last Year	Male	Workshard	sleep	Target (S) First Tris Year
1	Richard	Yes	Yes	No	Yes	Yes
2	Alan	Yes	Yes	Yes	No	Yes
3	Alison	No	No	Yes	No	Yes
4	Jeff	No	Yes	No	Yes	No
5	Gail	Yes	No	Yes	Yes	Yes
6	Simon	No	Yes	Yes	Yes	No

Soln:

Step 1: Find the entropy of the target S

$E(S) = [Yes \quad No]$

$= [4 \quad 2]$

$= \frac{4}{6} \log_2 \frac{4}{6} + \frac{2}{6} \log_2 \frac{2}{6}$

$= 0.918$

Step 2: Find the information gain of all the attributes and choose the attribute having the

highest information gain.

Information Gain $I_G = E(S) - \sum_V \frac{|S_v|}{V} E(S_v)$

$V(\text{Instance}) = 6$

First Last Year

Yes: $[3^+ \quad 0^-]$ $= \underline{\underline{0}}$

No: $[1^+ \quad 2^-]$ $= \frac{1}{3}\log_2 \frac{1}{3} + \frac{2}{3}\log_2 \frac{2}{3}$

$= 0.918$

$= 0.918 - \left[\frac{3}{6} \times 0 + \frac{3}{6} \times 0.918\right]$

$= \underline{\underline{0.459}}$

Male

Yes: $[2^+ \quad 2^-]$ $= \underline{\underline{1}}$

No: $[2^+ \quad 0^-]$ $= \underline{\underline{0}}$

$= 0.918 - \left[\frac{4}{6} \times 1 + \frac{2}{6} \times 0\right]$

$= \underline{\underline{0.251}}$

Workhard

Yes: $[3^+ \quad 1^-]$

$= \frac{3}{4}\log \frac{3}{4} + \frac{1}{4}\log_2 \frac{1}{4}$

$= \underline{\underline{0.8112}}$

No: $[1^+ \quad 1^-]$ $= \underline{\underline{1}}$

$I_G = 0.918 - \left[\frac{4}{6} \times 0.8112 + \frac{2}{6} \times 1\right]$

$= \underline{\underline{0.043}}$

Sleep

Yes: $[2^- \quad 2]$ $= \underline{\underline{1}}$

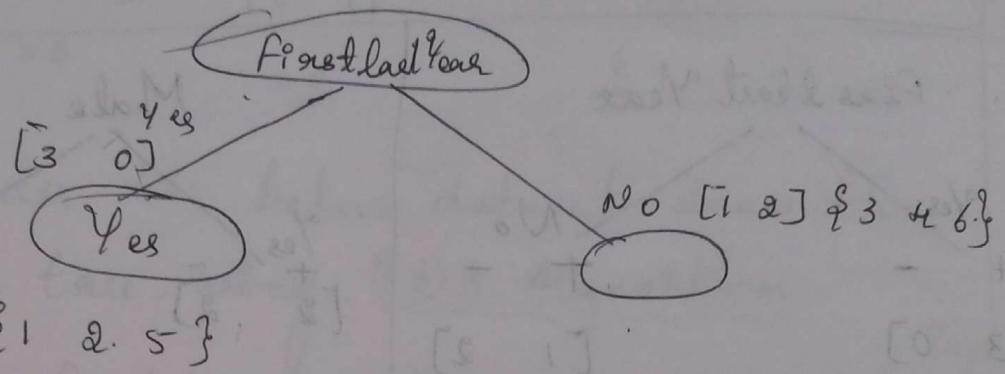
No: $[2 \quad 0]$ $= \underline{\underline{0}}$

$I_G = 0.918 - \left[\frac{4}{6} \times 1\right]$

$= \underline{\underline{0.251}}$

Step 3 :- Choose the attribute with the highest information gain.

I_G (FirstlastYear) = 0.459



Step 4 :- Consider Sub-dataset i.e 3 4 6

Student	Male	Workhard	Sleep	FirstAtYear
3 Alison	No	Yes	No	Yes
4 Jef	Yes	No	Yes	No
6 Simon	Yes	Yes	Yes	No

V = 3 S = [1 2]

E(S) = 0.918

male

Yes: [0 2] (+ −) = 0

No: [1 0] (+ −) = 0

$I_G = 0.918 - \left[\frac{2}{3} \times 0 + \frac{1}{3} \times 0\right]$

= 0.918

Workhard

Yes: [1 1] = 1

No: [0 1] = 0

$= 0.918 - \left[\frac{2}{3} \times 1\right]$

= 0.251

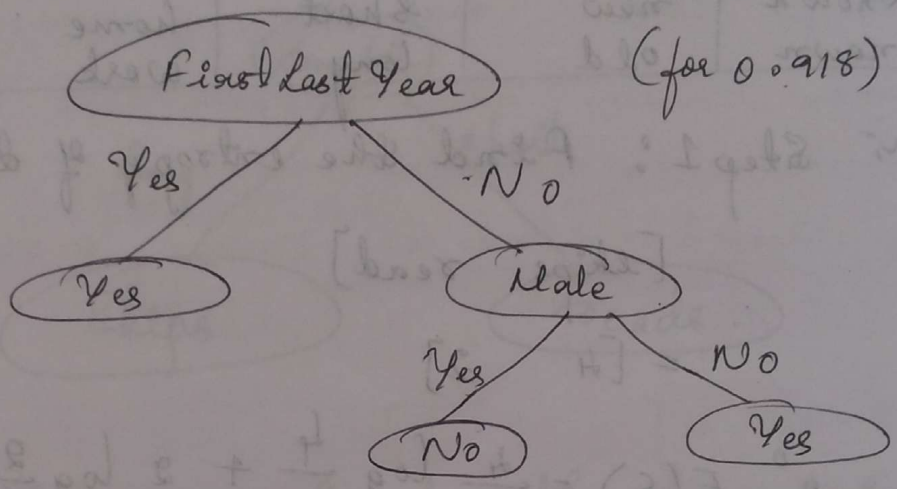
Sleep

Yes → [0+ 2−] = 0

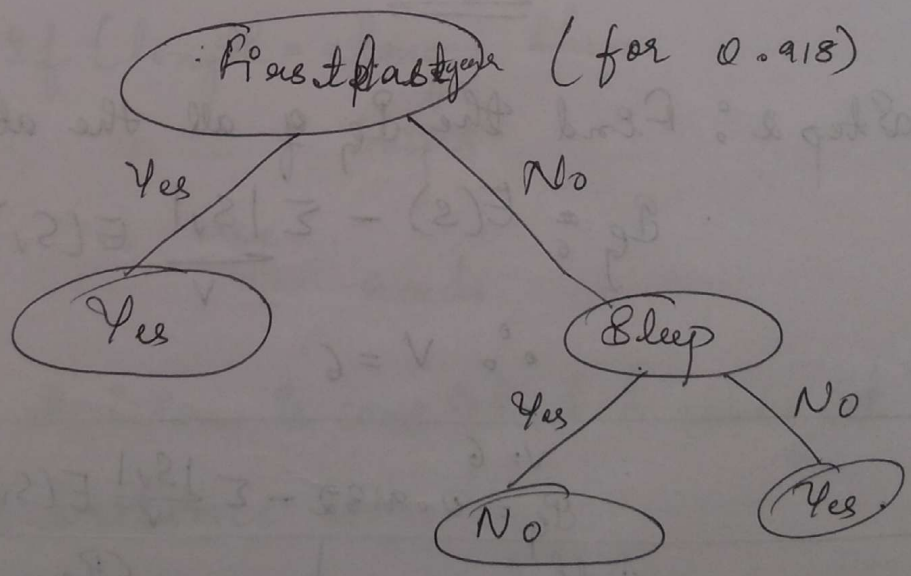
No → [1 0] = 0

= 0.918

Conclusion :-



(for 0.918)



(for 0.918)

(2) Consider the dataset draw the decision tree using ID3 algorithm.

Author	Thread	Length	Where	Target
1. Known	new	Long	Home	skips +
2. Unknown	new	short	Work	read -
3. Unknown	old	Long	Work	skips +
4. known	old	Long	home	skips +
5. known	new	short	home	reads -
6. known	old	long	work	skips +

Sol: Step 1: Find the entropy of dataset

[skips read]

$= [4 \quad 2]$

$$\therefore E(S) = \frac{4}{6}\log_2\frac{4}{6} + \frac{2}{6}\log_2\frac{2}{6}$$

$$= \underline{\underline{0.9182}}$$

Step 2: Find the I_G of all the attributes

$$I_G = E(S) - \sum_V \frac{|S_V|}{V} E(S_V)$$

$$\therefore V = 6$$

$V = 6$

$$I_G = 0.9182 - \sum \frac{|S_V|}{V} E(S_V)$$

Author

[3 1] [1 1]

$= \frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}$

$= \underline{\underline{0.0442}}$

Thread

[1 2] [3 0]

$= \underline{0.9182}$ $= \underline{\underline{0}}$

$= 0.9182$ $I_G =$

$= 0.4589$

Length

Long: [4 0] $= 0$

Short: [0 2] $= 0$

$\therefore I_G = 0.9182$ (✓)

I_G is more

Where

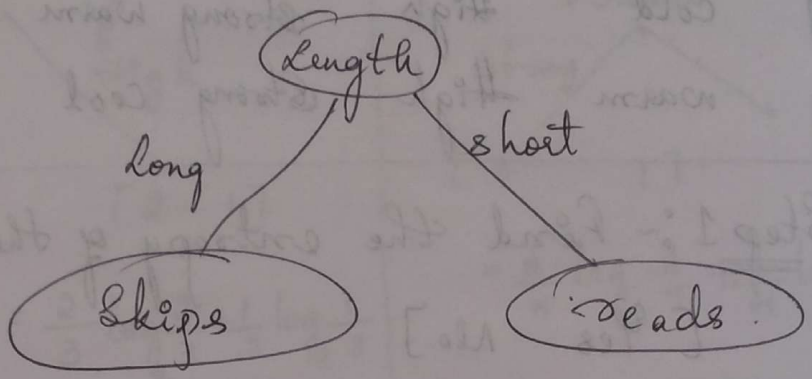
Home: [2 1] $= 0.9182$

Work: [2 1] $= 0.9182$

$= 0.9182 - \left[\frac{3}{6} \times 2 \times 0.9182\right]$

$= 0$

Step 3 :-



Step 4 :-

The conjunctive expression for the above decision tree is

if (length == long) then
skips
else
reads.

Step 5 :-

The above decision is consistent because it classifies 6 instance correctly.

3. Consider the dataset, draw decision tree using ID3, give the conjunctive expression, (boolean), state whether the decision tree is consistent or not.

Sky	Temp	Humidity	Wind	Water	Forecast	Enjoy sport
Sunny	warm	Normal	Strong	warm	Same	Yes
Sunny	warm	High	Strong	warm	Same	Yes
Rainy	cold	High	Strong	warm	change	No
Sunny	warm	High	Strong	cool	change	Yes

Sol:- Step 1:- Find the entropy of the dataset

[Yes No]

[3 1]

$$\therefore E(S) = \frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}$$

$$= 0.811$$

Step 2:- Find the IG of all the attributes

$$IG = E(S) - \sum_{V} \frac{|S_V|}{} E(S_V)$$

$$\therefore V = 4$$

$V = 4$

$$IG = 0.811 - \sum_{V} \frac{|S_V|}{} E(S_V)$$

Sky

Sunny: [3+ 0−] = 0

Rainy: [0+ 1−] = 0

$\therefore I_G = 0.811$

Temp

warm: [3 0] = 0

cold: [0 1] = 0

$\therefore I_G = 0.811$

Humidity

Normal: [1 0] = 0

High: [2 1] $= \frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3} = 0.9182$

$\therefore I_G = 0.811 - \left[\frac{1}{4}\times 0 + \frac{3}{4}\times 0.9182\right]$

$= 0.122$

Wind

Strong: [3 1] $= \frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4} = 0.811$

$I_G = 0.811 - \left[\frac{4}{4}\times 0.811\right]$

$= 0$

Water

warm: [2 1] $= 0.9182$

cool: [1 0] $= 0$

$I_G = 0.811 - \left[\frac{3}{4}\times 0.9182 + \frac{1}{4}\times 0\right]$

$= 0.122$

Forecast

Same: [2 0] $= 0$

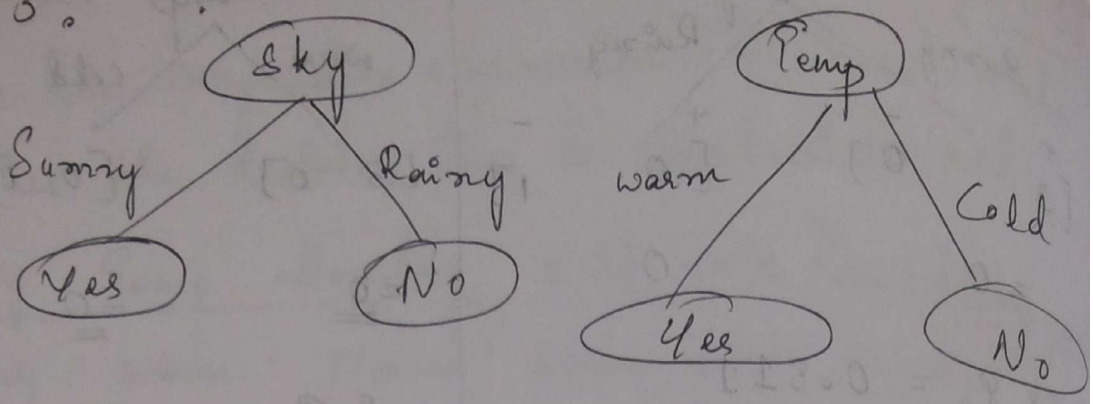
change: [1 1] $= 1$

$\therefore I_G = 0.811 - \left[\frac{2}{4}\log_2 0 + \frac{2}{4}\times 1\right]$

$0 + 0.5$

$= 0.311$

Step 3:



Step 4:- if (sky == Sunny) then Yes
else No

if (Temp == warm) Yes
else No

Step 5:- The above decision tree is consistent because it satifices 4 Instance Correctly

ID3 Algorithm:-

Algorithm of ID3 (Dataset, OutputAttributes, Attribut

// Dataset ← It is a training examples

// OutputAttributes ← It is a Attribute which contains the values predicted by the tree.

// Attribute ← A list of features.

Step 1:- Create a Root Node of the tree

1) a Node with +ve Label ← If all the examples are positive

2) a node with -ve label ← If all the examples are negative.

3) Most common value ← Attributes = ϕ

1. get Node A ⟵ The attribute with the highest information gain.

2. for every attributes/features
   1. ~~for~~ Add a new branch below the Root
   2. get a subset examples for each attribute value.

3. if the Subset Examples == ϕ
   then return most common outputAttribute

else

ID3(SubsetExamples, outputAttributes, Attribute-{Node})

end

Step 4: return root.

Problems:-

1 Consider the dataset and find the or give the Conjuntive expression and classify below instance

Rain	mild	Normal	weak	?

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Sol:- Find the entropy of entire dataset Target

$E(S) = [\text{Sunny} \quad \text{Overcast} \quad \text{Rain}]$ } wrong

$= [2 \quad 3] \quad [4 \quad 0] \quad [3 \quad 2]$

Only you need to take Yes and No (count)

[Yes No]

[9 5]

$$\therefore E(S) = \frac{9}{14} \log_2 \frac{9}{14} + \frac{5}{14} \log_2 \frac{5}{14}$$

$$= 0.940$$

Step 2 :- Find the I_G of all the attributes

$$I_G : E(S) - \sum_{V} \frac{|S_V|}{} E(S_V)$$

$\therefore V = 14$

---

(1) Outlook

Sunny, Overcast, Rain

[2 3] [4 0] [3 2]

$= 0.970$ $= 0$ $= 0.970$

$$I_G : 0.940 - \left[\frac{5}{14} \times 0.970 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.970\right]$$

$$= 0.940 - [0.3464 + 0 + 0.3464]$$

$= 0.247$ (✓) (Highest information gain)

---

(2) Temperature

Hot, Mild, Cool

[2 2] [4 2] [3 1]

$= 1$ $= 0.918$ $= 0.811$

$$I_G : 0.940 - \left[\frac{4}{14} \times 1 + \frac{6}{14} \times 0.918 + \frac{4}{14} \times 0.811\right]$$

$= 0.029$

---

(4) Humidity

High, Normal

[3 4] [6 1]

$= 0.985$ $= 0.591$

$$I_G = 0.940 - \left[\frac{7}{14} \times 0.985 + \frac{7}{14} \times 0.591\right]$$

0.152

(4) Wind

weak: [6 2] $= 0.811$

strong: [3 3] $= 1$

$$I_G = 0.940 - \left[\frac{8}{14} \times 0.811 + \frac{6}{14} \times 1\right]$$

$$= 0.048$$

Step3:

Outlook

[Yes, No] → Sunny: [2, 3] $\{D_1, D_2, D_8, D_9, D_{11}\}$

[Yes No] → Overcast: [4 0] $\{D_3, D_7, D_{12}, D_{13}\}$

[Yes No] → Rain: [3 2] $\{D_4, D_5, D_6, D_{10}, D_{14}\}$

Step 4:- Consider Subset dataset i,e D_1 D_2 D_8 D_9 D_{11}

Day	Temperature	Humidity	Wind	Play Tennis
D_1	Hot	High	Weak	No
D_2	Hot	High	Strong	No
D_8	Mild	High	Weak	No
D_9	Cool	Normal	Weak	Yes
D_{11}	Mild	Normal	Strong	Yes

Step 1:- Find the entropy of entire dataset.

[Yes No]

$\therefore E(S) = [2 \quad 3]$

$$= \frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5} = 0.970$$

Step 2:- Find the I_G for the attributes

$$I_G = E(S) - \sum_V \frac{|S_V|}{} E(S_V)$$

$\boxed{\therefore V = 5}$

Temperature

Hot: [0 2] $= 0$

Mild: [1 1] $= 1$

Cool: [1 0] $= 0$

$$I_g = 0.970 - \left[\frac{2}{5}\times 0 + \frac{2}{5}\times 1 + 0\right]$$

$$= 0.57$$

Humidity

High: [0 3] $= 0$

Normal: [2 0] $= 0$

$$I_g = 0.970 - \left[\frac{3}{5}\times 0 + \frac{2}{5}\times 0\right]$$

$I_g = 0.970$ ✓ (Highest)

Wind

Weak: [1 2] $= 0.918$

Strong: [1 1] $= 1$

$$I_g = 0.970 - \left[\frac{3}{5}\times 0.918 + \frac{2}{5}\times 1\right]$$

$$= 0.019$$

Step 3:- Humidity

High: [0 3] { No }

Normal: [2 0] { D_9, D_{11} } Yes

Step 4: Consider subset dataset 4, 5, 6, 10, 14.

Day	Temperature	Wind	Play tennis
D_4	Mild	Weak	Yes
D_5	Cool	Weak	Yes
D_6	Cool	Strong	No
D_{10}	Mild	Weak	Yes
D_{14}	Mild	Strong	No

Step 1: Find the entropy

$\therefore E(S) = [3\ 2]$

$$= \frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5} = 0.970$$

$\therefore V = 5$

Temperature

Mild: [2 1] $= 0.918$

Cool: [1 1] $= 1$

$I_g = 0.019$

Wind

Weak: [3 0] $= 0$

Strong: [0 2] $= 0$

$I_g = 0.970$ (Highest)

Step 3:- Wind

Weak: Yes

Strong: No

Consider below Dataset, Find the

1. Entropy of the target.
2. Draw the D. T.

	Student	FirstLastyear	Male	workshard	sleep	Firstthisyear
1	Richard	yes	yes	no	yes	yes
2	Alan	yes	yes	yes	no	yes
3	Alison	no	no	yes	no	yes
4	Jeff	no	yes	no	yes	$\boxed{\text{no}}$
5	Gail	yes	no	yes	yes	yes
6	Simon.	no	yes	yes	yes.	$\boxed{\text{no}}$

Step1:- Find the entropy of the target, S

$$E(S) = [\text{yes}, \text{NO}]$$

$$= [4, 2]$$

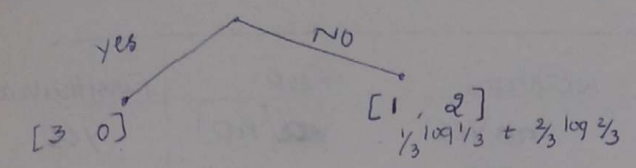
$$\therefore = \frac{4}{6}\log_2 4/6 + 2/6 \log_2 2/6$$

$$= 0.918$$

Step 2 :- Find the IG of all the attributes. & choose the attribute ~~with~~ having highest IG.

$v = 6$ $\quad S - \sum \frac{|v|}{v} \times \text{entropy}$

1. First Last year

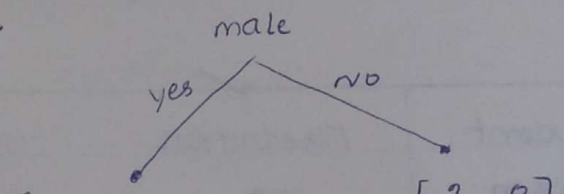


$E = 0$ $\quad E = 0.918$

$\therefore = 0.918 - \left[\frac{3}{6} \times 0 + \frac{3}{6} \times 0.918\right]$

$\boxed{I_G = 0.459}$

2.

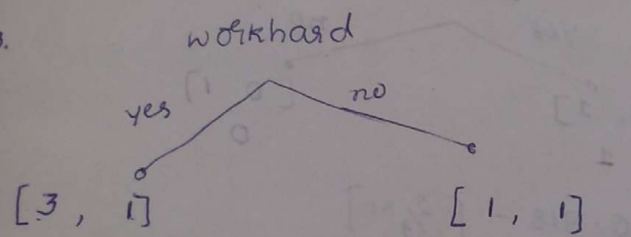


$E = 1$ $\quad E = 0$

$\therefore = 0.918 - \left[\frac{4}{6} \times 1 + 0\right]$

$\boxed{= 0.251}$

3.

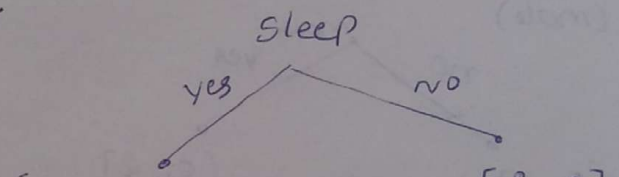


$= E = 0.8112$ $\quad E = 1$

$= 0.918 - \left[\frac{4}{6} \times 0.8112 + \frac{2}{6} \times 1\right]$

$\boxed{= 0.043}$

4.

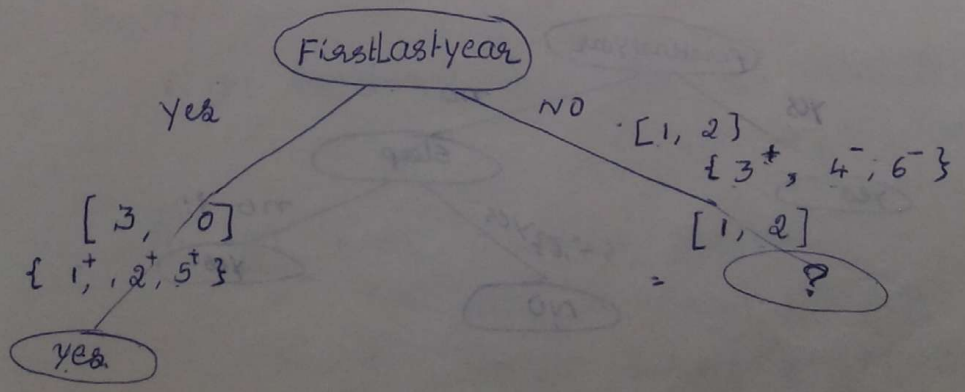


$E = 1$ $\quad E = 0$

$\therefore = 0.918 - \left[\frac{4}{6} \times 1\right]$

$\boxed{= 0.251}$

Step 3 :- The attribute with Highest IG is, IG(FirstLastyear) = 0.459



Step4 :- consider the 4 instances.

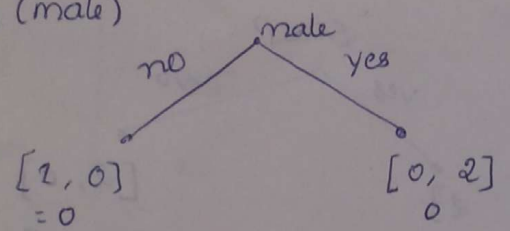
$\{3^+, 4^-, 6\}$

	student	FirstLastYear	Male	workhard	Sleep	FirstThisYear
3	Alison	no	no ~~yes~~	~~no~~ yes	~~yes~~ no	yes
4	Jeff	no	yes	no	yes	no
6	Simon	no	yes	yes	yes	no

$S = [1, 2]$ $\therefore$ Entropy$(S) = 0.918$

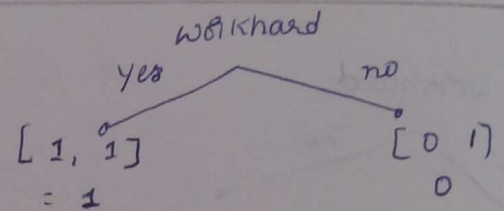
$V = 3$

IG (male)



$\therefore IG = 0.918 - [1/3 \times 0 + 2/3 \times 0]$

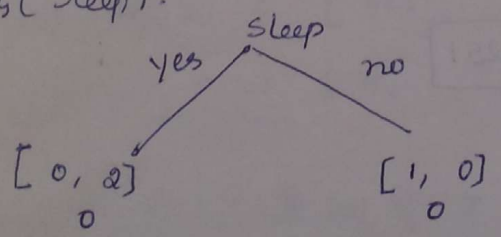
$\boxed{IG = 0.918}$



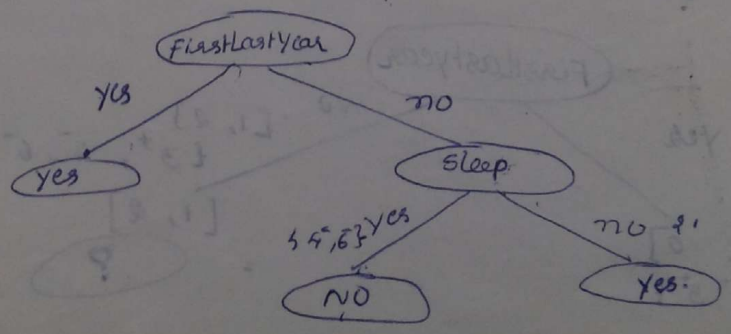
$= 0.918 - [2/3 \times 1]$

$= 0.251$

IG ('Sleep):



$\boxed{IG = 0.918.}$



problems :

1. Give a decision tree to represent the below Boolean Function.

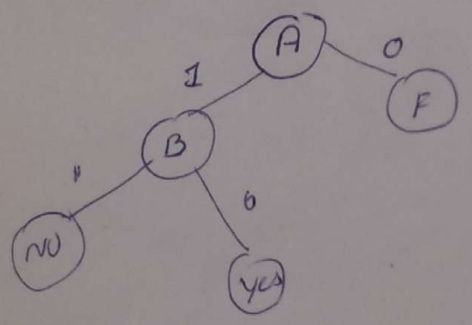
1. $A \wedge \sim B$.

check A

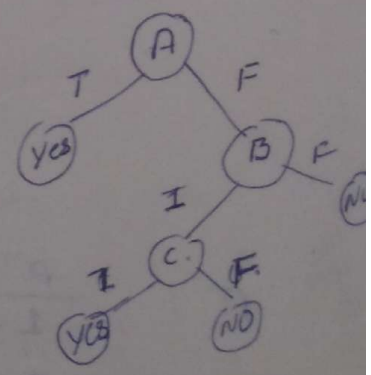
if A == 1 : check B

if B = 1 classify NO

if B = 0 class yes.



2. $[A \vee [B \wedge C]]$



check A :

if A = 1 : classif yes

else "

check B :

if B = 1 : check C

if C = 1 : classif yes.

if C = 0 : classif NO

if B = 0 : classif NO

3. $A \oplus B$:

$= A\bar{B} + \bar{A}B.$

$\underbrace{(A \wedge \sim B)}_{(T)} \vee \underbrace{(\sim A \wedge B)}$

check A :

if $A = 1$: check B

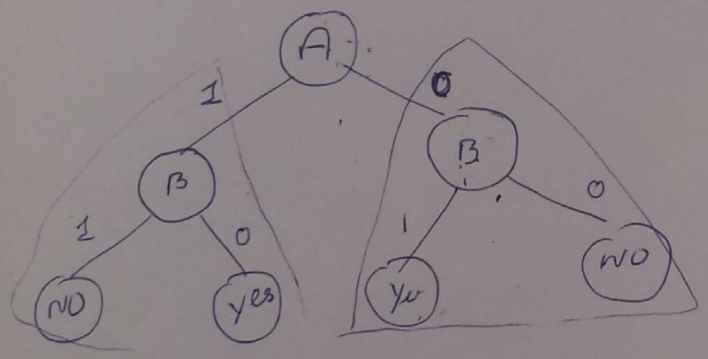
if $B = 1$: classify NO

if $B = 0$: classify yes

if $A = 0$: check B

: if $B = 1$: class yes

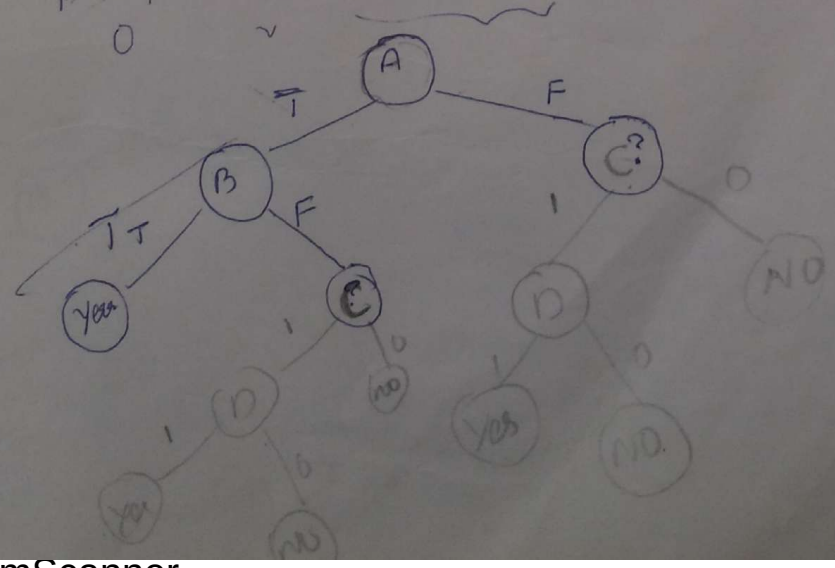
if $B = 0$: class NO



4. $[A \wedge B] \vee [C \wedge D]$

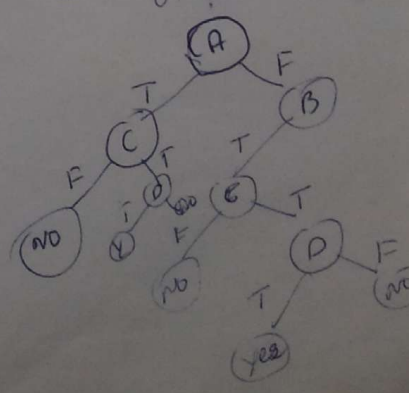
$F \wedge F$

0



$[A \vee B] \wedge (C \wedge D)$

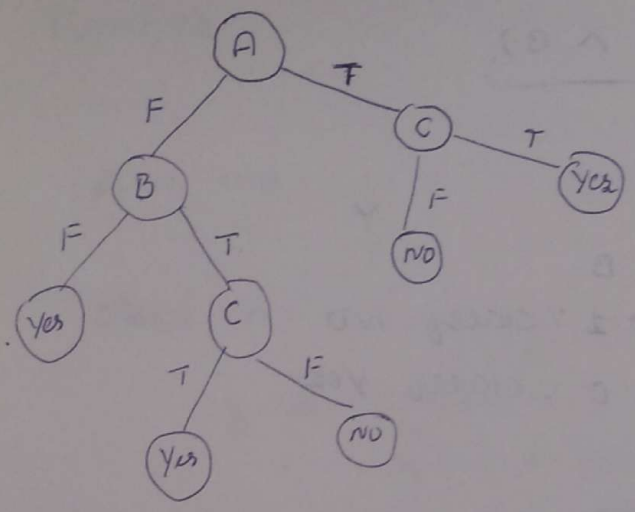
$0 + 1 \wedge 1 = 1$



5. $(\sim A \wedge \sim B) \vee C$

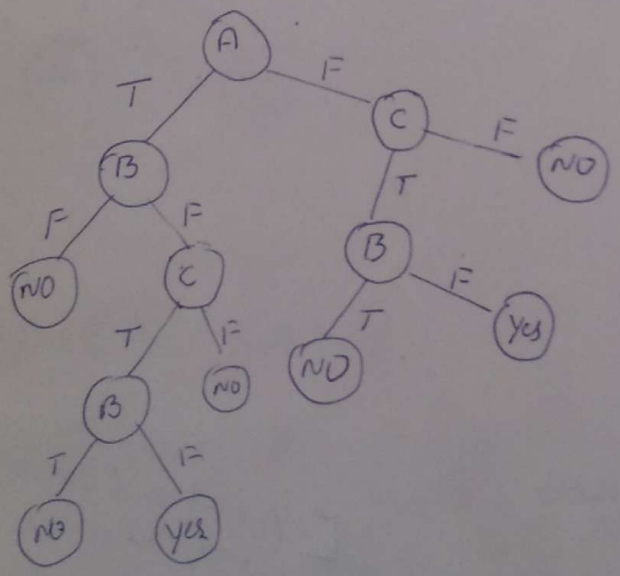
T ∧ F

F



6. $(A \wedge B) \vee (C \wedge \sim B)$

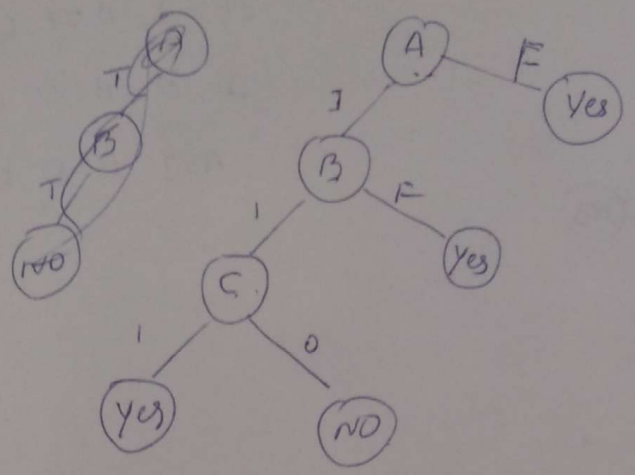
T F. F



7. $\sim(A \wedge B) \vee C$

$A=0,\ B=0,\ C=1$

$A=1 \quad B=1 \quad C=0$

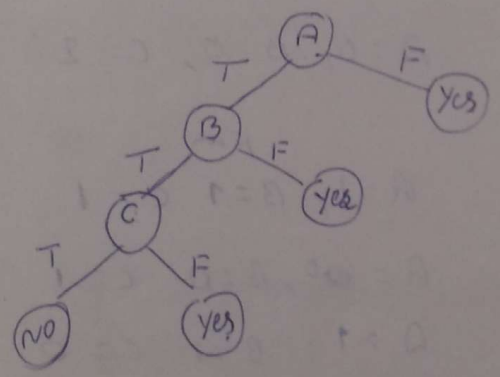


8. $[\sim(A \wedge B) \vee \sim(B \wedge C)]$

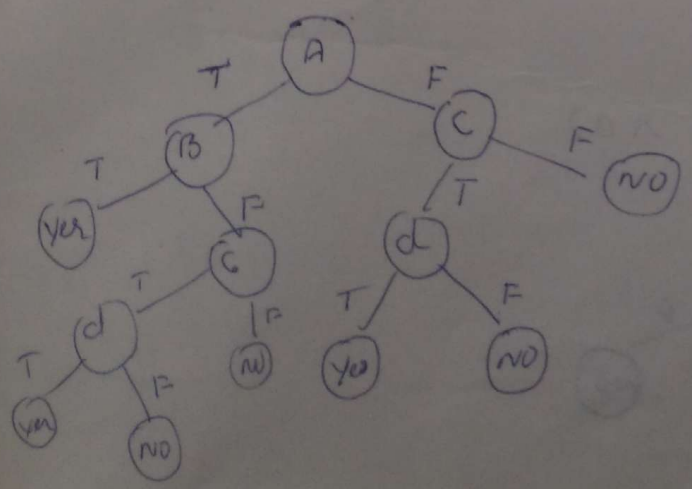
$A=0,\ B=0,\ C=1$

$1 \quad B=1 \quad C=0$

=



9. $[(A \wedge B) \vee (C \wedge d)]$



10. $[A \wedge \sim(B \wedge C)] \vee B$

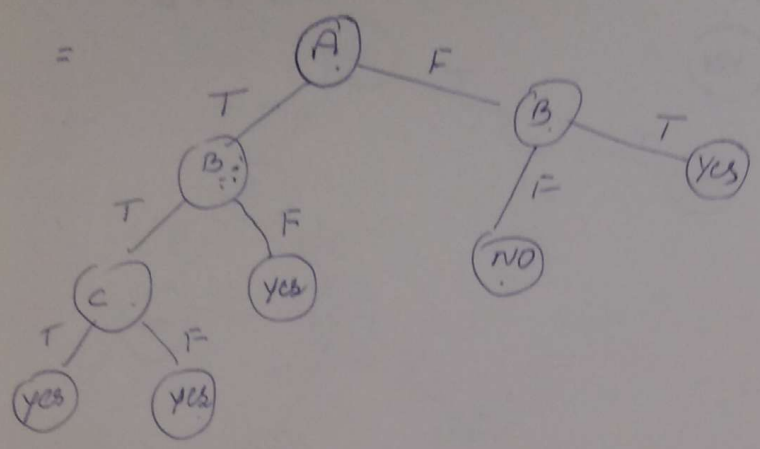
$T \wedge \sim(T \wedge F)$

$A = 1, B = 1, C = 0$

$A = 0 \quad B = 0 \quad C = 1$

$A = 1 \quad B = 0 \quad C = 1$

=

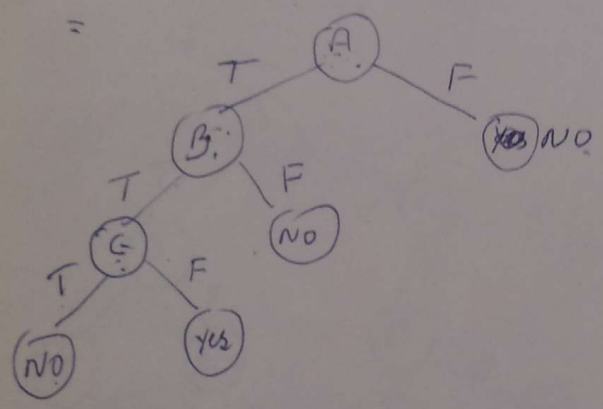


11. $\sim[\sim(A \wedge B) \vee (B \wedge C)]$

$F + F$

$(A \wedge \cancel{\sim B}) \wedge \sim(B \wedge C)$

=



$A = 0, \quad B = 0, \quad C = 1$

False

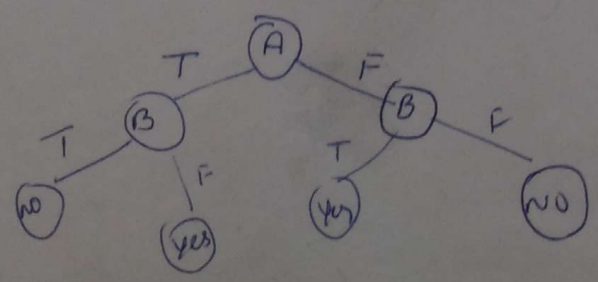
$A = 1, \quad B = 1 \quad C = 1$

$A = 0, \quad B = 0 \quad C = 1$

$A = 1 \quad B = 1 \quad C = 0$

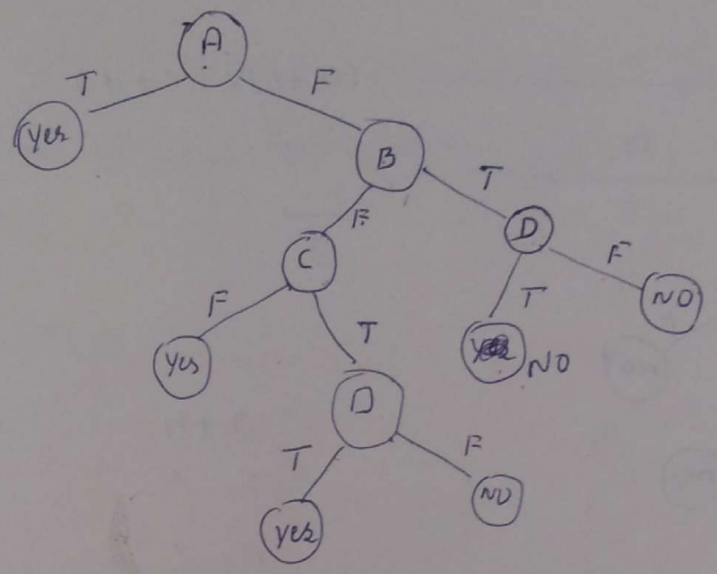
12. $A \oplus B$. $\quad A B' + \bar{A} B$

$= (A \wedge \sim B) \vee (\sim A \wedge B)$.



13. $[(A \vee (\sim B \wedge \sim C) \vee (D \wedge \sim B)]$

:

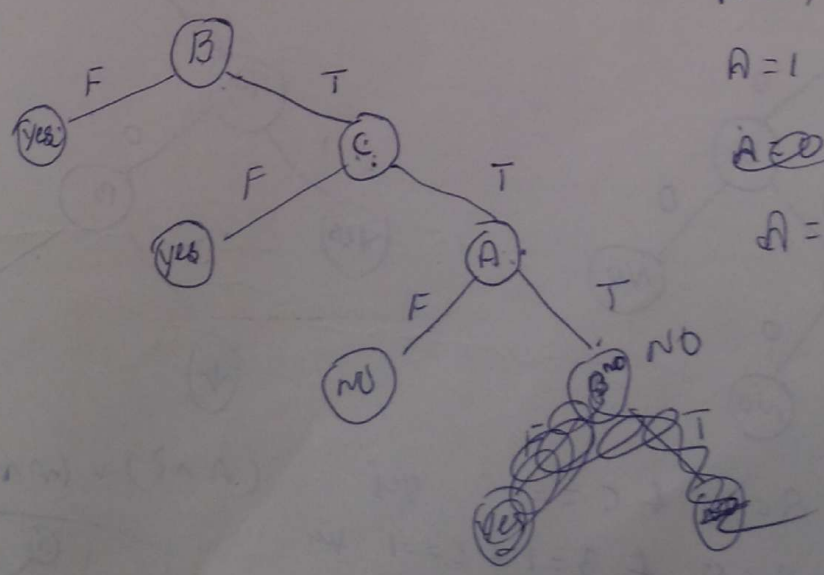


A = 1, B = 1, C = 0, D = 0 => yes

A = 0, B = 1, C = 1, D = 1 => False.

A = 0, B = 0 C = 0 D = 1 => yes

14. $\underset{T}{\sim B} \vee \sim C \vee (A \wedge \sim B)$



A = 1, B = 0, C = 1 => yes

A = 1 B = 1 C = 1 => NO

~~A = 0 B = 1 C = 1~~

A = 0, B = 1, C = 1 = NO

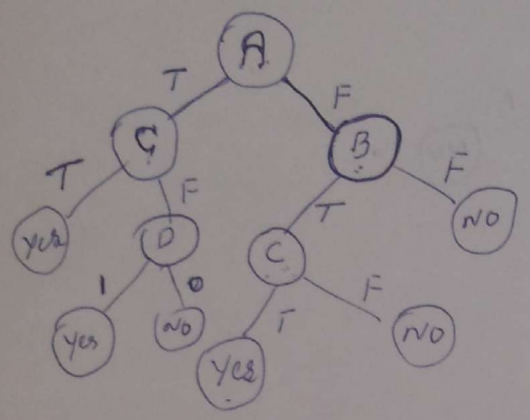
15. $[(A \vee B) \wedge (C \vee D)]$

1 . 0

$(a+b) \cdot (c+d) =$

$\underbrace{(a+b)}_{F} = F$

$\underbrace{\quad}_{T} \quad ? = ?$

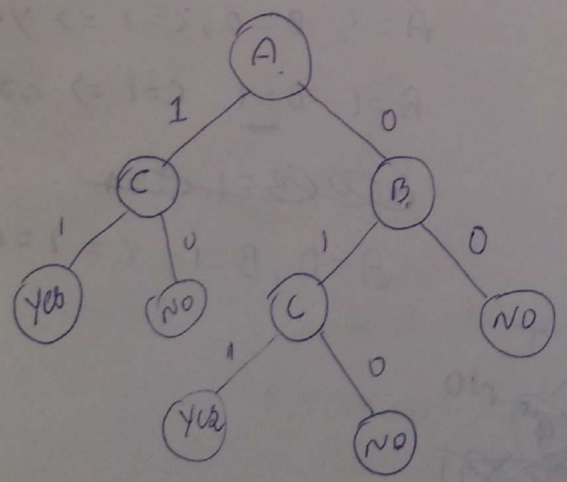


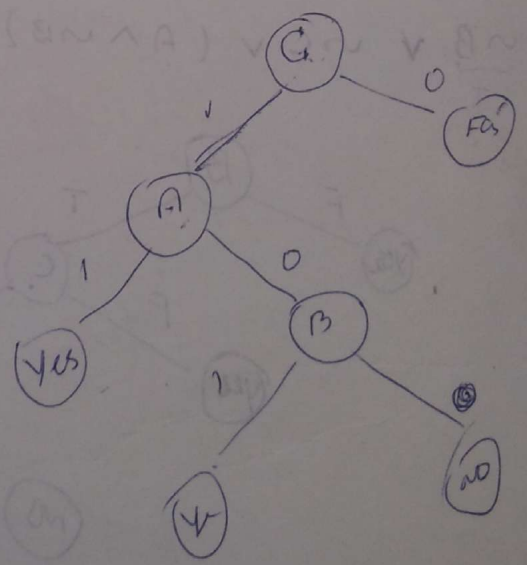
	a+b
T	×
F	?

A=0, B=1, C=0, D=1 = yes

A=1 B=0 C=1 D=0 = yes.

16. $(A \vee B) \wedge C$ or $C \wedge (A \vee B)$



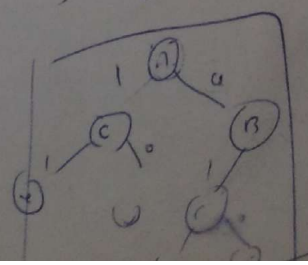


if (A==1 & C==1) yes $(A \wedge C) \vee (\neg A \wedge B \wedge C)$

if (A=0 & B=1 & C==1 yes

else

NO



Dataset 1 :-

Name	Over 6ft	Eye Color	Hair Style	Sex
Morgan	Yes	Blue	Long	Female
Bob	No	Brown	None	Male
Vincent	Yes	Brown	Short	Male
Amanda	No	Brown	Short	Female
Reid	No	Blue	Short	Male
Lauren	No	Purple	Long	Female
Elisa	Yes	Brown	Long	Female

Classify using naive bayes for the unseen instance

(i) (yes, Blue, None) as Male or Female

(ii)

Name	Over 6ft	Eye Color	Hairstyle	Sex
Amanda	No	Brown	Short	?

Sol: Step 1 : Using Naive Bayes Classifier

Male : P(Male) . P(Amanda|Male) P(No|Male) .

P(Brown|Male) . P(Short|Male) .

Female : P(Female) . P(Amanda|female) . P(No|female)

P(Brown|Female) . P(Short|Female) .

Step 2: Construct the frequency table

$N_{male} = 3$ $\qquad$ $N_{female} = 4$

$P(N_{male}) = \frac{3}{7}$ $\qquad$ $P(N_{female}) = \frac{4}{7}$

$= 0.42$ $\qquad$ $= 0.57$

	Male	Female	$P(A_i \mid Male)$	$P(A_i \mid Female)$
Amanda	0	1	$0/3 = 0$	$1/4 = 0.25$
No	2	2	$2/3 = 0.66$	$2/4 = 0.5$
Brown	2	2	$2/3 = 0.66$	$2/4 = 0.5$
Short	2	1	$2/3 = 0.6$	$1/4 = 0.25$

Step 3: Apply the formula

Male: $0.428 \times 0 \times 0.6 \times 0.6 \times 0.6$

$= 0$

Female: $0.571 \times 0.25 \times 0.5 \times 0.5 \times 0.25$

$= 0.0089$

Step 4: Apply Normalization

Male: $\frac{0}{0 + 0.0089}$ $\qquad$ Female: $\frac{0.0089}{0 + 0.0089}$

$= 0\%$ $\qquad$ $= 1 = 100\%$

Dataset 2 :

Home Owner	Marital Status	Job Experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

Classify whether Mr. Bob is defaulted or not using NB classifier ?

(i) < No , Married , 3 , ? >

Sol- **Step 1 :** Using NB classifier.

Yes : $P(Yes) \cdot P(No|Yes) \cdot P(Married|Yes) \cdot P(3|Yes)$

No : $P(No) \cdot P(No|No) \cdot P(Married|No) \cdot P(3|No)$

Step 2 : Construct the frequency table

$N_{yes} = 3$ $\qquad N_{No} = 7 \qquad N = 10$

$P(N_{yes}) = 3/10 = \underline{0.3}$ $\qquad P(N_{No}) = 7/10 = \underline{0.7}$

	Yes	No	$P(A_i \mid Yes)$	$P(A_i \mid No)$
No	2	4	$2/3 = 0.66$	$4/7 = 0.57$
Married	1	4	$1/3 = 0.33$	$4/7 = 0.57$
3	1	2	$1/3 = 0.33$	$2/7 = 0.28$

Step 3: Yes : $3 \times 0.66 \times 0.33 \times 0.33$

$= 0.0162$

No : $7 \times 0.57 \times 0.57 \times 0.28$

$= 0.6368$

Step 4: Apply Normalization

$$Yes = \frac{0.0162}{0.0162 + 0.0636} \qquad No = \frac{0.0636}{0.0162 + 0.0636}$$

$$= \frac{0.0162}{0.0798} \qquad = \frac{0.0636}{0.0798}$$

$Yes = 0.20$ $\qquad = 0.79$

$Yes = 20\%$ $\qquad No = 79\%$

Dataset A :-

Name	Give Birth	Can fly	Live in Water	Have Legs	Class
Human	Yes	No	No	Yes	mammals
Python	No	No	No	No	non-mam
Salman	No	No	Yes	No	non-mam
Whale	Yes	No	Yes	No	mammals
frog	No	No	sometimes	Yes	non-mam
Komodo	No	No	No	Yes	non-mam
bat	Yes	Yes	No	Yes	mammals
Pigeon	No	Yes	No	Yes	non-mam
Cat	Yes	No	No	Yes	mammals
Leopard shark	Yes	No	Yes	Yes	non-mam
turtle	No	No	sometimes	No	non-mam
Penguin	No	No	sometimes	Yes	non-mam
Porcupine	Yes	No	No	Yes	mammals
eel	No	No	Yes	No	non-mam
Salamander	No	No	sometimes	Yes	non-mam
gila monster	No	No	No	Yes	non-mam
Platypus	No	No	No	Yes	mammals
owl	No	Yes	No	Yes	non-mam
dolphin	Yes	No	Yes	No	mammals
eagle	No.	Yes	No.	Yes	non-mam

Use NB classifier to classify the above instance. ?

Yes	No	Yes	No	?

Sol: **Step 1**: Using NB classifier

Mammals: P(Mammals) · P(Yes | Mammals) · P(No | Mammals) P(Yes | Mammals) · P(No | Mammals)

Non-mammals: P(Non-Mammals) · P(Yes | Non-mammals) · P(No | Non-mammals) · P(Yes | Non-mammals) · P(No | non-mammals)

Step 2: Construct the frequency table

$N_{mammals} = 7$ $\qquad$ $N_{non\text{-}mammals} = 13$

$P(N_{mammals}) = 7/20 = \underline{0.35}$ $\qquad$ $P(N_{non\text{-}mammals}) = \frac{13}{20} = \underline{0.65}$

	Mammals	Non-Mammals	P(Ai | Mammals)	P(Ai | Non-mammals)
Yes	6	1	6/7 = 0.85	1/13 = 0.07
No	6	10	6/7 = 0.85	10/13 = 0.76
Yes	2	3	2/7 = 0.28	3/13 = 0.23
No	2	4	2/7 = 0.28	4/13 = 0.30

Step 3 : Apply the formula

Mammals = $0.35 \times 0.85 \times 0.85 \times 0.28 \times 0.28$

$= 0.019$

Non-mammals = $0.65 \times 0.07 \times 0.76 \times 0.23 \times 0.30$

$= 0.0023$

Step 4 : Apply Normalization

Mammals = $\frac{0.019}{0.019 + 0.0023}$

$= 0.892$

Non-mammals = $\frac{0.0023}{0.019 + 0.0023}$

$= 0.107$

Mammals = 89%

Non-mammals = 10%

∴ The new instance is "mammals"

Mammals > Non-Mammals

Dataset A:

Example No	Colour	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

use Naive Bayes classifier using m estimate $m = 3$, $p = 0.5$ to classify the new instance.

(i) <Red, Domestic, SUV>

Sol: Step 1: Using Naive Bayes classifier

Yes: $P(Yes) \cdot P(Red|Yes) \cdot P(SUV|Yes) \cdot P(Domestic|Yes)$

No: $P(No) \cdot P(Red|No) \cdot P(SUV|No) \cdot P(Domestic|No)$

Step 2: Construct the frequency table

$N_{yes} = 5$

$N_{No} = 5$

$P(Yes) = \frac{5}{10} = 0.5$

$P(No) = \frac{5}{10} = 0.5$

	Yes	No	$P(A_i \mid Yes)$	$P(A_i \mid No)$
Red	3	2	$3/5 = 0.6$	$2/5 = 0.4$
SUV	1	3	$1/5 = 0.2$	$3/5 = 0.6$
Domestic	2	3	$2/5 = 0.4$	$3/5 = 0.6$

Step 3: Apply the formula

$$D_1 = 0.5 \times 0.6 \times 0.2 \times 0.4$$

$$= 0.024$$

$P_2 = 0.5 \times 0.4 \times 0.6 \times 0.6$

$= 0.072$

Step 4: Apply Normalization.

$Yes = \frac{0.024}{0.024 + 0.072}$ $\quad$ $No = \frac{0.072}{0.024 + 0.072}$

$= 0.25$ $\quad$ $= 0.75$

$= 25\%$ $\quad$ $= 75\%$

∴ The new instance is "No".

Yes < No

By: Darshan K.M & Yashanath N.L
Asst Prof, Dept of CSE
A.I.T, Chikmagalur.

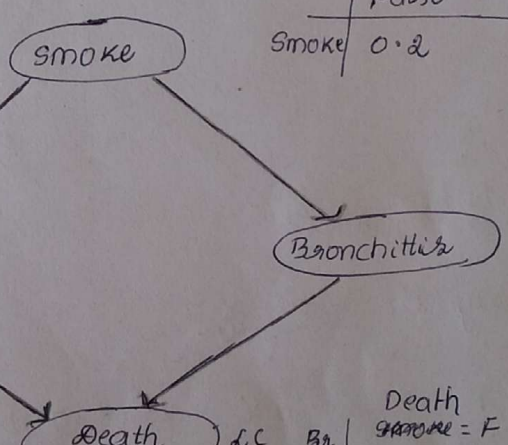
module- 4

Baysian Learning.

Defn:- "The bayesian learning is a probabilistic model which combines the past experience (prior) with the new data (Likelihood) which will calculate the probability for hypothesis (posterior)."

* The Baysian learning works on the principal of Joint distribution.
* It gives the probability of hypothesis.

eg:- consider the below baysian n/w and predict the proba of death of the patient



	False	True
Smoke	0.2	0.9

Smoke	LC = F	LC = T
F	0.8	0.1
T	0.1	0.8

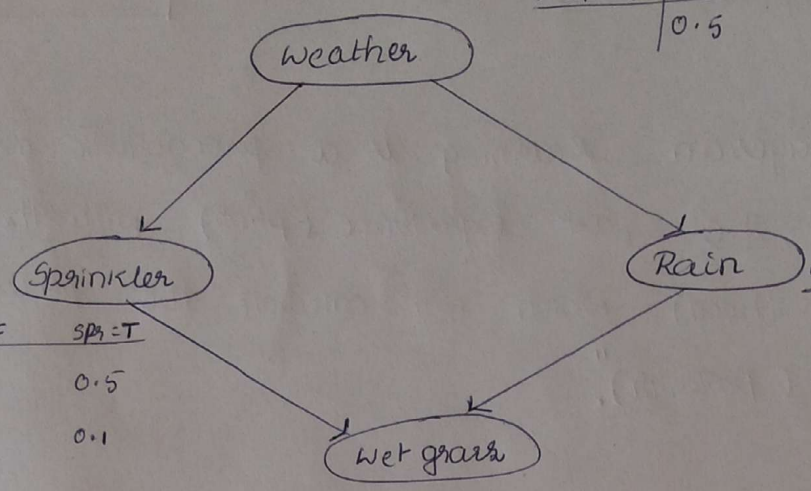
LC	Br	Death Smoke = F	Death Smoke = True
F	F	0.1	0.3
T	F	0.02	0.98
F	T	0.4	0.6
T	T	0.001	0.99

$= P(\text{Death} = T \mid LC = T, Br = F)$

$= 0.99$ ie. 99% of chances of death of a patient.

eg 2:- predict the probability of wetgrass of the below baysian network.

weather	False	True
	0.5	0.5



weather	Rain = F	Rain = T
F	0.8	0.2
T	0.2	0.8

weather	Spr = F	Spr = T
F	0.5	0.5
T	0.9	0.1

Spr	Rain	weather = F	weather = True
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

= P(weather = True | Sprinkler = True, Rain = False)

= 0.9

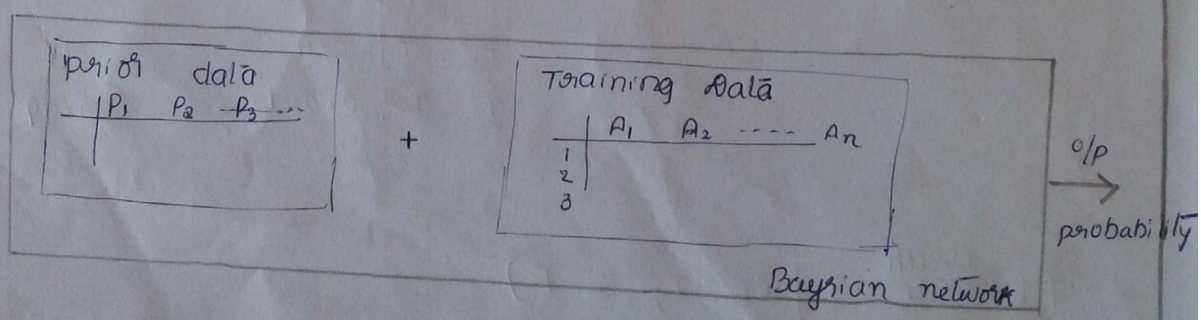
ie, 90 %

*** ~~Joint Distribution (JD)~~ :-

*** Features of Baysian Learning :-

1. The prior knowledge is combined with the training data to determine the final probability of hypothesis.

ie,



2. The Bayesian networks works on the principal of Priar ($P(x)$) and posterior probability :- $P(A|B)$.

3. The Bayesian methods always gives the probabilistic prediction

eg :- $P(\text{cancer}) = 0.93$, 93% of chance of cancer.

4. A new instance can be classified by combining the prediction of multiple hypothesis.

5. Bayesian methods gives standard optimal decision making over other methods (eg Find-S, CE)..

6. It provides a flexible approach where the prob of training data can be increased or decreased.

* * * Bayes Theorem :-

Defn :- " The Bayes theorem determines the best hypothesis H for a given observed or training data (D). along with initial or prior probability."

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)}$$

Likelihood

prior knowledge

Posterior proba for a given T.D.

Normalization constant.

where,

$$P(D) = P(D \mid \text{Attribute 1} = \text{True})\, P(\text{Attri 1}) + P(D \mid \text{Attribute} = \text{False}) \cdot P(\text{Attribut 2}).$$

1. Prior probability :-

It is the Probability of hypothesis of an attribute known in advance.

* These attributes are independent.

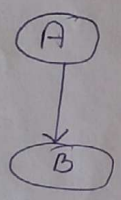
eg:- $P(\text{Smoke}) = 0.4$, $P(A) = 0.2$ etc

Posterior probability :-

The posterior probability is a conditional probability whose probability is found w.r.t other attributes.

* The Attributes are depend on other attributes.
* It is represented as $P(B \mid \text{parent}(B))$.

eg:- 1.



$\Rightarrow P(B \mid \text{parent}(B))$
$= P(B \mid A)$.

2.

$\Rightarrow P(\text{LungCancer} \mid \text{parent}(\text{LungCancer}))$
$= P(\text{LunCancer} \mid \text{Smoke})$.

Likelihood Function :-

The Function $P(D \mid h)$ is the Likelihood Function which assumes the probability of data is true.

*** The Maximum posterior hypothesis (MAP):

Defn :-

" It is the maximum of all the candidate hypothesis H for a given training data, D.

eg :-

consider the hypothesis space H

$$H = \{ h_1 , h_2 , h_3 \}$$

$P(h_1 | D) = 0.4$ $\quad$ $P(h_2 | D) = 0.3$ $\quad$ $P(h_3 | D) = 0.3$

$\therefore h_{map} = h_1$ $\quad$ [since, h_1 is a maximum candidate]

In general,

i.e. $$h_{map} \equiv \underset{h \in H}{\arg\max}\ P(h|D).$$

$$\equiv \underset{h \in H}{\arg\max}\ \frac{P(D|h) . P(h)}{P(D)}$$ ..... By Bayes Theorem

$$\boxed{h_{map} \equiv \underset{h \in H}{\arg\max}\ P(D|h) . P(h)}$$

$P(D)$ is neglected $\because$ It is a constant

*** Maximum Likelihood :-

The $P(D|h)$ is called as Likelihood of the data D w.r.t h, and its maximum called as "maximum Likelihood"

$$\boxed{h_{ml} \equiv \arg\max P(D|h)}$$

NOTE :-

1. What are the Features of Bayesian Learning.

2. Explain Bayes theorem.

Sol:- Defn + prior & posterior prba
+ MAP & Likelihood.

3. With an example explain the MAP hypothesis.

4. Given the Likelihood = 98%, prior hypo = 0.02 and $P(D) = 0.092$ Find the hypothesis using bayes theorem.

Sol:-

1. Likelihood, $P(D|h) = 0.98$
2. prior hypo $P(h) = 0.02$, $P(D) = 0.092$

$$\therefore P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)}$$

$$= \frac{0.98 \times 0.02}{0.092} = 0.0213$$

$$=$$

5. The test returns a correct positive result in only 98% of the cases in which disease is present and negative result in only 97% of the cases disease not present and 0.008 of the entire population have cancer. use bayes theorem, predict weather a patient have cancer or not.

Soln :-

given,

$P(\text{cancer}) = 0.008$

$P(\oplus \mid \text{cancer}) = 0.98$ $\qquad$ $P(- \mid \text{cancer}) = 0.02$ $\quad$ $(1-0.98)$

$P(\oplus \mid \text{notcancer}) = 0.03$ $\qquad$ $P(- \mid \text{notcancer}) = 0.97$

Step 1 :- The target is : { cancer , notCancer }.

Step 2 :- Find $P(\text{cancer} \mid \oplus)$, and $P(\text{notcancer} \mid \oplus)$.

Step 3 :- Bayes theorem.

$$P(\text{cancer} \mid \oplus) = \frac{P(\oplus \mid \text{cancer}) \cdot P(\text{cancer})}{P(D)}$$

$$P(D) = P(+ \mid \text{cancer})\, P(\text{cancer}) + P(- \mid \text{notcancer})\, P(\text{notcancer})$$

$$= 0.98 \times 0.008 + 0.03 \times \underset{(1-0.008)}{0.992} = 0.0376$$

$$= \frac{0.98 \times 0.008}{0.0376} = 0.21 \qquad (21\%)$$

$$P(\text{notcancer} \mid \oplus) = \frac{P(\oplus \mid \text{notcancer})\, P(\text{notcancer})}{P(D)}$$

$$= \frac{0.03 \times 0.992}{0.0376} = 0.79$$

Step 4 :- $h_{map} = \underset{}{\overset{\text{argmax}}{}} \{0.21, 0.79\}$

$= \{0.79\}$ ie 79% of patient do not have cancer.

Bayes Theorem and Concept Learning :-

Brute Force Bayes Concept Learning :-

This algo learns or understands the training data by finding the highest posterior probability as shown below.

Algorithm :- Brute Force Map Learning algo

Given an instance space x (Training Data)

- A hypothesis space H
- The target concept $C : x \rightarrow \{0,1\}$
- A sequence of instances,

$$\langle\langle x_1, d_1\rangle, \langle x_2, d_2\rangle \text{ ----- } \langle x_n\, d_n\rangle\rangle$$

Step 1 : For each hypothesis h in H, Find the posterior probability

$$\boxed{P(h|D) = \frac{P(D|h).P(h)}{P(D)}}$$

Step 2 : Output the highest ~~pro~~ posterior Probability h_{map}.

$$\boxed{h_{map} = \underset{h \in H}{\operatorname{argmax}}\ P(h|D)}$$

Remarks :-

1. It is impractical for huge dataSets
2. It is computationally infeasible ~~is It takes~~
3. It is less sensitive to inconsistent data.
4. It is not robust to noise

In order to address the above remarks,

The brute force learning algo specifies the boundary values for $\underbrace{P(h)}_{\text{prior prob}}$ and $\underbrace{P(D|h)}_{\text{likelihood}}$.

* The $P(h)$ & $P(D|h)$ must be choosen to be consistent with below assumption,

1. The training data D is noise free
2. The target concept C is present in H
3. No hypothesis is more probable than any other hypothesis.

$\therefore$

Assume,

$$P(h) = \frac{1}{|H|} \Rightarrow \text{Length of the Hypothesis space}$$

$$P(D|h) = \begin{cases} 1 & \text{if } d_i \text{ is consistent} \\ 0 & \text{otherwise. (inconsistent)} \end{cases}$$

Consider bayes formula, $P(h|D) = \dfrac{P(D|h) \cdot P(h)}{P(D)}$.

Step 1 :- Assume d_i is inconsistent then

$$\therefore P(h|D) = \frac{0 \cdot P(h)}{P(D)} = 0$$

Step 2 :- Assume d_i is consistent then

$$P(h|D) = \frac{1 \cdot \frac{1}{|H|}}{P(D)} \Rightarrow \frac{\frac{1}{|H|}}{\frac{V.S_{H,D}}{|H|}} \Rightarrow \frac{1}{|V.S_{H,D}|}$$

Finally $P(h|D) = \begin{cases} \dfrac{1}{|VS_{H,D}|} & \text{if } d_i \text{ consistent} \\ 0 & \text{otherwise.} \end{cases}$

*** -: MAP hypothesis and consistent Learner algorithm :-

or

The evolution of posterior probability

or

The Relationship b/w MAP hyp & consistent Algo: (Find-S, CE)
(Difference)

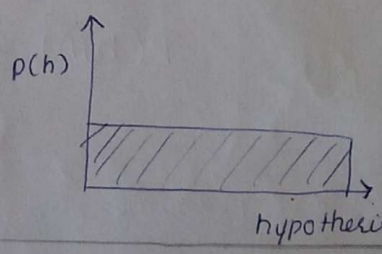
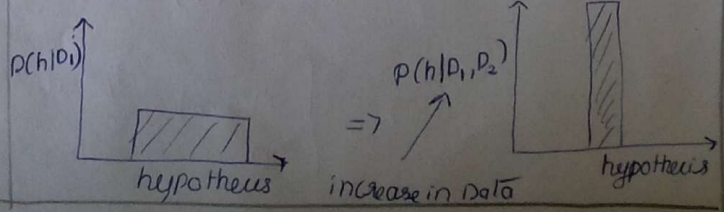
Any learning algo is said to be consistent if the algorithm commits zero error over the training examples.

*. Every consistent algo (Find-S, candidate elimination) always outputs a map hypothesis.

Assumption,

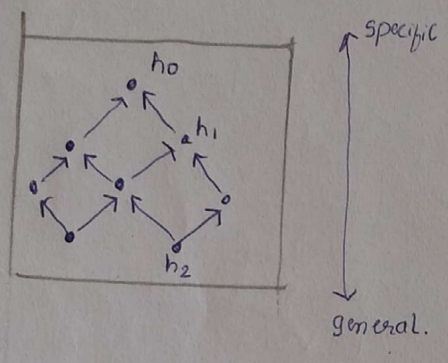
1. A uniform prior probability distribution over H.
2. The Training data is noise-Free.

Relationship :-

Consistent algo (Find-S, CE)	Evolution of MAP hypothesis
1. Initially, the hypothesis space H is assumed to be specific. <∅, ∅, ∅, ∅>	1. There is a uniform distribution of probability over H. 
2. As, the T.D increases, the specific hypo moves towards general hypo <?, ?, High, ?>	2. As the T.D increases, the posterior probability increases. 

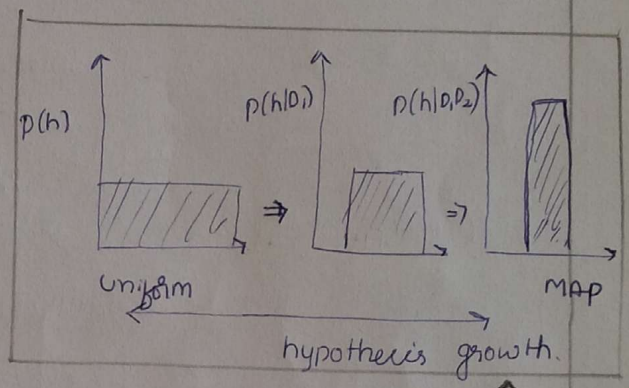
3. The general hypo commits zero error for all positive & negative instance

4. the hypothesis grows from specific to general vice-versa.



3. The posterior probability becomes zero for the inconsistent hypothesis

4. The hypothesis grows from uniform distribution to maximum posterior probability. vice-versa



∴ Every consistent algorithm outputs a MAP hypothesis.
(Find-S, CE)

Note :->

1. Explain Brute-Force Map Learning algorithm.

2. Define
   1. consistent Learner.
   2. Inconsistent Learner

**3. Explain how Bayesian Framework characterizes the behaviour of consistent Learning algorithm.

4. Justify, how Find-S & CE outputs Map hypothesis.

5. With a neat diagram, Explain the growth (evolution) of posterior probability. (MAP hypothesis).

Maximum Likelihood and Least Square Error hypo

or

Learning a real valued target Function.

Suppose, the dataset contains the continuous real valued target attribute ie, it can have any value b/w ~~the~~ 0 & 1 as shown below.

A_1	A_2	A_3	target
-	-	-	0.1
-	-	-	0.2
-	-	-	0.01
			0.4
			⋮

∴ The Learning algo must minimize the square error & outputo the maximum likelihood.

consider, the above dataset of the form,

$$\langle x_i, d_i \rangle \; ; \; d_i = f(x_i) + e$$

target Func

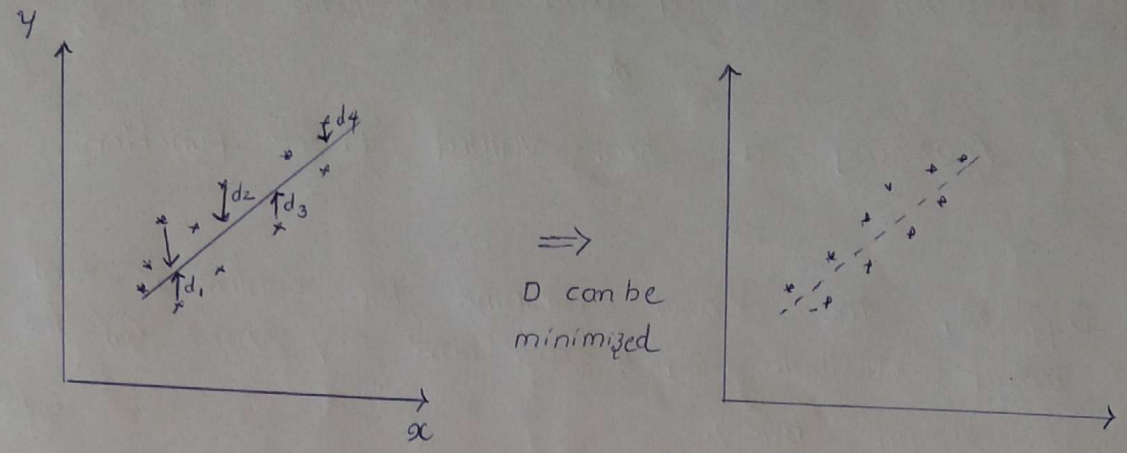
random variable represents error.

Inorder to minimize the error, there are two cases,

case 1 :- The data are mutually Linearly dependent.

Suppose, the data sets are linearly dependent then, the Linear Function fit the line by passing through the data points.

* the error can be minimized by Finding the vertical distance b/w the line & the data points.



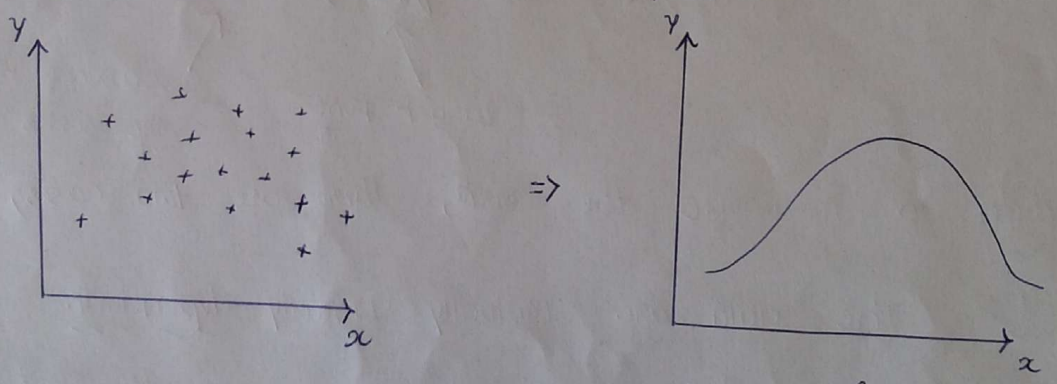
$$D = d_1^2 + d_2^2 + \cdots d_n^2$$

—— : The solid line indicates, the Linear Function with more error

----- : The dotted line indicates, the minimum squared error (D). And increases the maximum likelihood

~~Case~~ Case 2:- Data are mutually independent & non-linear:

suppose, if the data are arranged non-linearly then error can be minimized by using normal distribution (a smooth, bell shaped curve).



gaussian Normal distribution.

The gaussian normal distribution Formula is,

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(d_i - \mu)^2}$$

where,

σ^2 : variance

μ = mean Function

Consider, the maximum likelihood

$$= \arg\max P(D|h)$$

$$\equiv \arg\max \prod_{i=1}^{n} P(D_i|h_i) \quad \text{... for } n \text{ attributes}$$

$$\equiv \arg\max \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(d_i - h_i)^2}$$

--> for each hypo

Take Log. ie ($\log ab = \log a + \log b$).

$$\equiv \arg\max \sum \ln \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}} - \frac{1}{2\sigma^2}(d_i - h_i)^2$$

$$\equiv \arg\max \sum - \cancel{\frac{1}{2\sigma^2}} (d_i - h_i)^2 \quad \text{... } \because \text{ does not contain } h.$$

$$\equiv \arg\max \sum -(d_i - h_i)^2$$

Take Inverse.

$$\boxed{\therefore h_{ML} \equiv \arg\min \sum (d_i - h_i)^2}$$

error term

$\therefore$ The above eqn. says that, the square error $(d_i - h_i)^2$ can be minimized by using normal distribution. The maximum likelihood can be increased.

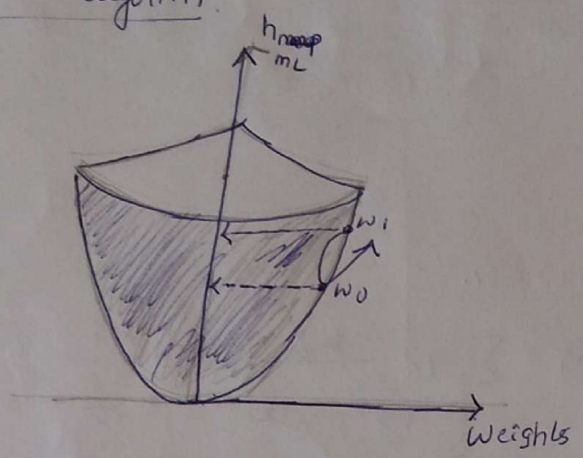
Remark :

1. It considers only noise in the target data
2. It doesnot considers noise in the attributes of training data.

Gradient Search to maximize likelihood in N.N :-

In a baysian neural network, to maximize the Likelihood of a hypothesis by updating the weights.

* The weights are updated by using gradient Ascent search algorith.



$$w_{JK} \leftarrow w_{JK} + \Delta w_{JK}$$

$$\Delta w_{JK} = \alpha \sum_{i=1}^{n} \underbrace{h(x_i)(1-h(x_i))}_{\text{Derivative of Sigmoid Func}} \cdot \underbrace{(d_i - h(x_i))}_{\text{Error diff b/w observed \& training}} \cdot \underbrace{x_{iJ}}_{\text{Inputs}}$$

α → Learning Rate

***

Minimum Description Length Principle :→

The MDL principle is used to interprete the shortest explanation for the observed data.

consider,

$$h_{map} = \arg\max P(D|h) \cdot P(h)$$

Take $\log_2$.

$$= \arg\max \log_2 P(D|h) + \log_2 P(h) \quad \cdots \log(pq) = \log p + \log q$$

& minimize the negative quantity,

$$= \arg\min -\log_2 P(D|h) - \log_2 P(h).$$

$-\log_2 P(h)$: The Length of h under optimal hypothesis space H.

$-\log_2 P(D|h)$: The Length of training data for a given hypothesis.

In general,

$$\boxed{h_{MDL} \equiv \underset{h \in H}{\operatorname{argMin}} \; L_{C_1}(h) + L_{C_2}(D|h)}$$

NOTE :-

1. Explain how to minimize the Least square error in bayesian learning.

2. Show how the normal distribution Formula (gaussian) minimizes the squared Error.

3. Explain MDL principle.

4. Explain how to maximize the Likelihood of a real valued ~~Fun~~ target Function.

Naive Bayes Classifier :-

It is a simple probabilistic classifier based on bayes theorem.

* The every feature is assumed to be conditionally independent.

* The baysian approach is,

$$\boxed{V_{N.B} \equiv \text{argMax } P(v_J) \prod P(a_i | v_J)}$$

where, $P(v_J)$: The probability of target. class
ie

$P(a_i | v_J)$: The probability of attributes w.r.t to target. ie, $\frac{n_c}{n}$ $\frac{(\text{no of target class})}{\text{Total Instance}}$.

$\prod$: The Joint multiplication of conditional probability.

Problems :-

1. Consider the below patient data set and classify the instance [Name = Rama RunNose = NO headche = mild Fever = yes Disease = ?] using Naive Bayes. classifier.

	Name	RumingNose	headche	Fever	Disease
1	Rama	NO	mild	yes	D_2
2	Rama	yes	no	no	D_1
3	Rama	NO	strong	yes	D_1
4	Sita	yes	mild	yes	D_1
5	Sita	no	no	no	D_2
6	Sita	yes	strong	yes	D_1
7	Sita	yes	strong	no	D_2
8	Rama	yes	mild	yes	D_1

Step 1 : By using Naive Bayes Classifier,

$D_1 = P(D_1) \cdot P(Rama \mid D_1) \cdot P(no \mid D_1) \cdot P(mild \mid D_1) \cdot P(yes \mid D_1)$
$=$

$D_2 = P(D_2) \cdot P(Rama \mid D_2) \cdot P(no \mid D_2) \cdot P(mild \mid D_2) \cdot P(yes \mid D_2)$

Step 2 :- Construct the Frequency Table.

$N_{D_1} = 5$ $N_{D_2} = 3$, $N = 8$

$P(D_1) = 5/8 = 0.625$ $P(D_2) = 3/8 = 0.375$

	Attribute value	D_1	D_2	$P(\text{Attribute value} \mid D_1)$	$P(\text{Attribute} \mid D_2)$
A_1	Rama	3	1	$= 3/5 = 0.6$	$1/3 = 0.333$
A_2	No	1	2	$1/5 = 0.2$	$2/3 = 0.666$
A_3	mild	2	1	$2/5 = 0.4$	$1/3 = 0.333$
A_4	yes	4	1	$4/5 = 0.8$	$1/3 = 0.333$

Step 3 :- Apply the Formula.

$D_1 = 0.625 \times 0.6 \times 0.2 \times 0.4 \times 0.8$
$= 0.024$

$D_2 = 0.375 \times 0.333 \times 0.666 \times 0.333 \times 0.333$
$= 0.0092$

Step 4 :- Apply Normalization

$D_1 = \dfrac{0.024}{(0.024 + 0.0092} = 0.72$ ie 72%

$D_2 = \dfrac{0.0092}{0.024 + 0.0092} = 0.27$ ie 27%

Step 5 : Since $D_1 > D_2$, the new instance is Classified as,

$\boxed{\therefore \text{ Disease } = D_1}$ [72% chance of Disease D_1].

Problem No:- 2

Kitchen Data Set

problem No:-3

playTennis DataSet

} Refer class Notes.

Estimating the Probability :→

The Naive Bayes classifier, becomes 0 (zero) whenever any one conditional probability is zero.

* which leads to poor estimation.

∴ The probability of Naive Bayes classifier must be $\neq 0$

i.e., $= P(v_J) \prod P(a_i | v_J) \neq 0$

eg:-

Name	Target
Rama	yes
Rama	yes
Sita	yes
Sita	No

i. $P(yes) \cdot P(Rama|yes)$

$= 3/4 \cdot 2/3 \Rightarrow \neq$ zero

ii. $P(no) \cdot P(Rama|No)$

$= 1/4 \cdot 0$

$= 0 \Rightarrow$ Leads to poor estimation

∴ The new probability estimation formula is

$$K = \frac{n_c + mp}{n + m} \quad \text{S.T } K \neq 0$$

where, n_c = no of target class.

m = A weight or a any constant

n = Total no of instance

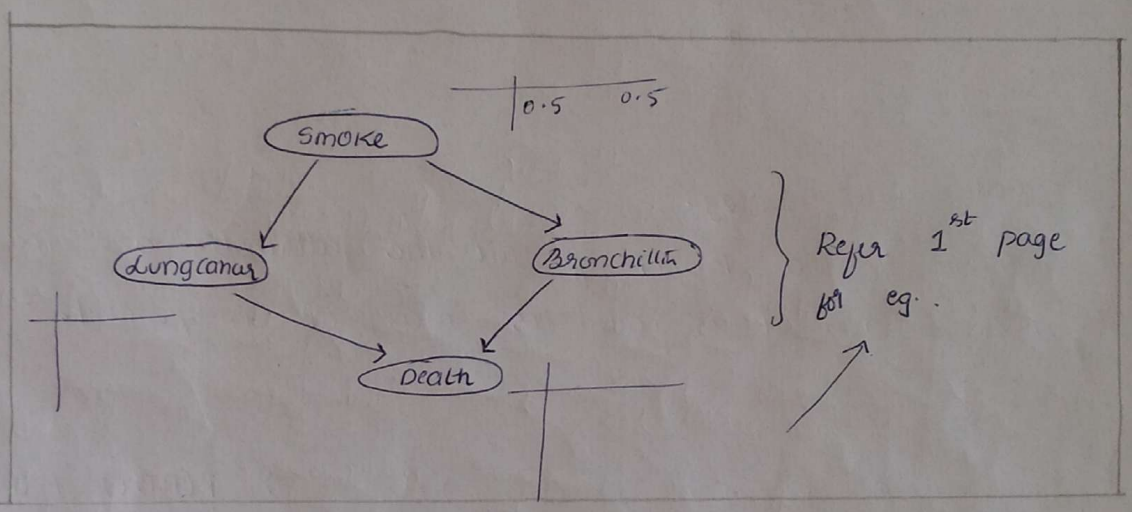
p = prior estimate.

Baysian Belief Network :-

"The Baysian belief network is a probabilistic graphical ~~model~~ model that represents a set of variables and their probabilistic dependencies".

* It is a directed acyclic graph whose node represents a variables and arc represents the conditional probability."

eg:-



Representation :-

"The Baysian belief network represents the joint probability distribution for a set of variables".

ie.,

In general,

$$P(v_1, v_2, v_3 \cdots v_n) = \prod P(v_i \mid parent(v_i))$$

variables — Joint probability. — Conditional prob of v_i

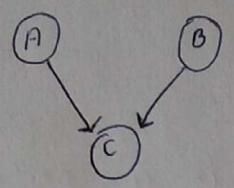
where,

$parent(v_i)$: A set of parent nodes of v_i

Baysian Belief N/w

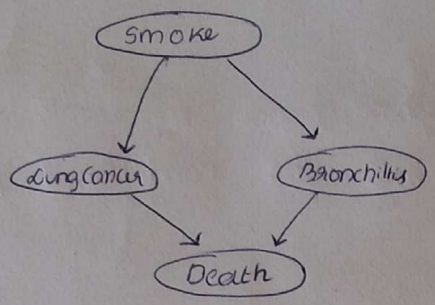
Representation

eg:- 1.



$\Rightarrow P(A,B,C) = P(A) \cdot P(B) \cdot P(C \mid A,B)$.

2.



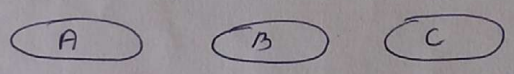
$\Rightarrow P(S, L, B, D)$

$= P(S) \cdot P(L \mid S) \cdot P(B \mid S) \cdot P(D \mid L, B)$

marginal Independence:-

Defn:- In a BBN, if all the attributes are independent ie then it is called as marginal independent.

eg:-



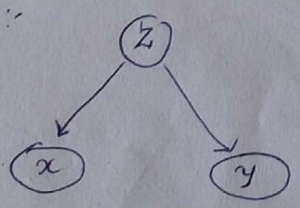
$\Rightarrow P(A,B,C) = P(A) \cdot P(B) \cdot P(C)$

Bayson Belief N/w

Conditionally Independent:-

Defn:- Let x, y, z be the 3 discrete valued variable, then x & y are conditionally independent given z. as shown below

eg:-



$\Rightarrow P(x,y,z) = P(z) \cdot P(x \mid z) \cdot P(y \mid z)$

*, x & y are conditionally indep

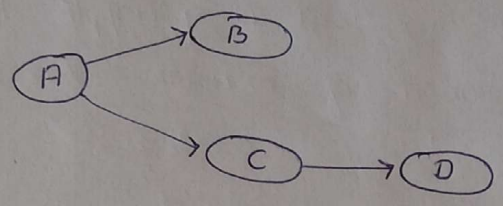
Problems :-

1. Consider the Full Joint probability Function,

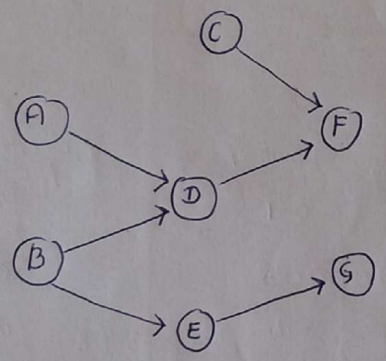
$$P(A,B,C,D) = P(D|C) \cdot P(C|A) \cdot P(B|A) \cdot P(A)$$

Draw the corresponding Baysian n/w.

Soln :-



2. Given the baysian n/w, ~~give~~ write the full Joint ~~o~~ prob. representation.



Soln :-

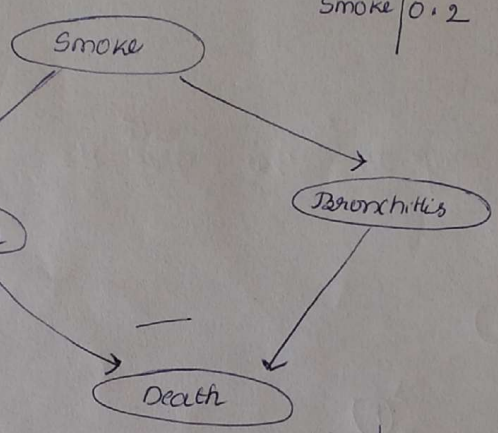
$$P(A,B,C,D,E,F,G) = P(A) \cdot P(B) \cdot P(C) \cdot P(F|C,D) \cdot P(D|A,B) \cdot P(E|B) \cdot P(G|E).$$

Baysian Inference :-

The Baysian n/w can be used infer (predict) the probability distribution of target variable given the observed values of other variables.

* The B.N, can also be used to compute the probability distribution of subset of Network variables.

eg:-



Smoke	F	T
	0.2	0.9

Smoke	Br = F	Br = T
F	0.3	0.5
T	0.6	0.8

Smoke	LC = F	LC = T
F	0.8	0.1
T	0.1	0.8

LC	Br	Death = F	Death = T
F	F	0.1	0.3
F	T	0.02	0.98
T	F	0.4	0.6
T	T	0.001	0.99

* From the above B.N having the probability distribution of the observed variables, we can infer, target & subset of n/w information as shown below.

1. $P(Death = T \mid LC = T, Br = T) = 0.99$ ie 99% } Target
2. $P(LC = T \mid Smoke = F) = 0.1$ ie 10%
3. $P(Br = F \mid Smoke = T) = 0.5$ ie 50% } Subsets of a N/w.

NOTE :

1. Explain Naive Bayes Classifier.

Soln :- Defn + Formula + Estimating the probability.

2. ~~Ex~~ With a neat diagram explain the bayesian belief n/w.

soln :-

1. Defn +
2. Representation
3. eg

3. write a short note on

i. marginal independence ii. conditionally independence.

4. problems on converting from bayesian n/w to probability distr.

Refer pageno 22

5. Briefly explain how the bayesian n/w can be used to predict the target value. or subset of a N/w.

Soln :- Bayesian inference

(Refer page No :- 23)

Advantages and disadvantages of Baysian Learning

Advantages :-

1. It provides a standard way to combine the prior inf. with data

2. It obeys the Likelihood principle.

3. It provides interpretable answers such as 95% of ther is a cancer

4. It provides a convenient setting for wide range of models such as hierarchical models.

Disadvantages :-

1. It does not tell how to select the prior data

2. It produces the posterior probability that are heavily influenced by prior probability

3. It is computationally infeasible for the large data set.

Darshan L.M & Taranath N.L
Asst prof, Dept of CSE
AIT, Chikmagalur

:- Evaluating Hypothesis :-

The hypothesis (eg: version space, Decision tree, etc) ~~which~~ are found by many concept learning algorithm such as Find-S, candidate elimination algo, ANN etc, are to be evaluated by estimating the true error of the hypothesis.

* The true error can be estimated by using stastical ~~math~~ operators such as mean, variance and standard dev.

* There are many models which evaluates the performance of the hypothesis such as,
    1. Binomial Distribution
    2. Normal Distribution
    3. Paired-t test & K-nearest ~~neighbor~~ neighbor algorithm etc

Motivation & Difficulties in estimating the error :-

There are many factors which makes the evaluation very difficult such as,

1. Limited Data Set :-
   The Limited size of the dataset affects the accuracy of the learned hypothesis.
   eg:- A dataset of new disease.

2. Overfitting :-
   The learned hypothesis are to be evaluated to avoid overfitting of data.
   eg:- post-pruning method in Decision-tree.

3. Bias in estimate :-

The Learned hypothesis may be biased (error) for the unobersved or Fututure Traming examples.

∴ The hypothesis must be (estimated) tested on Training examples choosen randomly. So that the hypothesis may be unbiased. (Zero Error)

4. Variance in estimate :-

Even though, the obtained hypothesis is unbiased (No error) but still there will be variance from true accuracy.

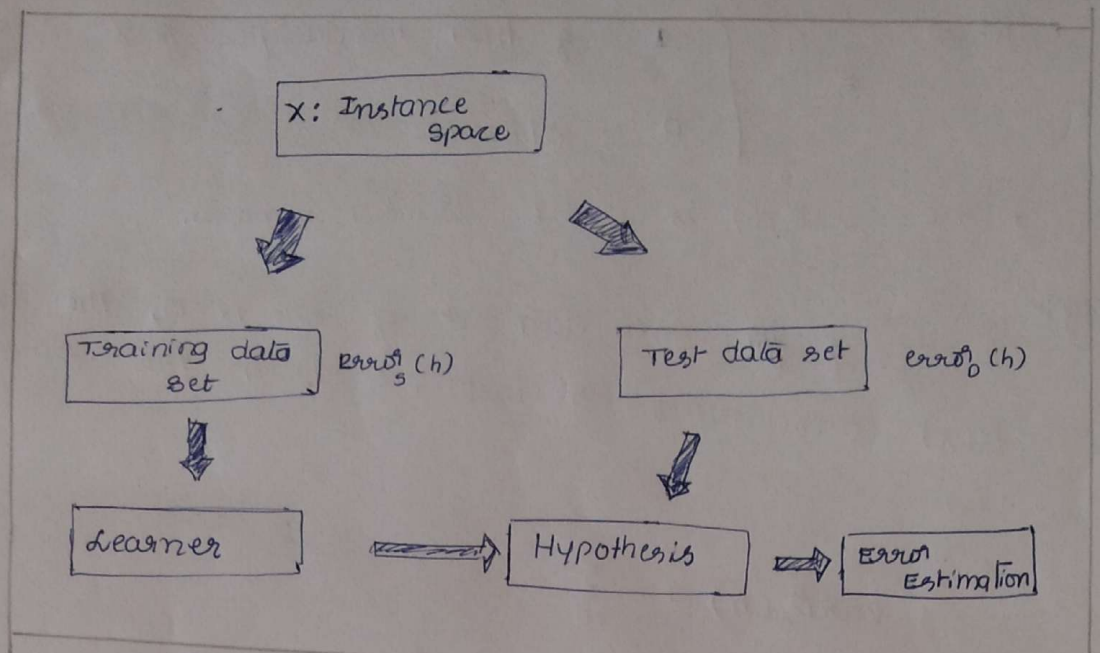
* The variance is used in estimating the confidence interval.

Estimating Hypothesis Accuracy :- [configuring the algorithm].

Inorder, to estimate the accuracy of the hypothesis for the unobserved or Future Training datasets. the machine learning algorithms are trained with the below configuration

* The general setup is :

1. X : The instance space [The data set]
2. D : The probability distribution of Training dataset.
3. Learner : The algorithm which understands the data set.
4. Hypothesis : The hypothesis are tested with the test data set & the accuracy is estimated for the Future dataset as shown below.



The steps to estimate the error.

With the above setup, the error ~~estimation~~ is estimated to find the accuracy & probable error. of the algorithm.

1. The instance space x are choosen randomly or distributed as training & testing data sets.

2. The Learner (Find-S, CE, ANN etc) understands the target using training dataset and generate the hypothesis

3. The hypothesis accuracy is estimated by using test data set.

- It is estimated by ~~usin~~ Finding
  1. Sample Error 2. True Error 3. Confidence Interval.

Sample Error :- [Mean Error or Expected value] :-

Defn :- "The sample error of a hypothesis w.r.t target function f of N samples is given by

$$error_s(h) = \frac{1}{N} \sum \delta(f(x), h(x))$$

target func. hypothesis

where,

$$\delta = \begin{cases} 1 & \text{if } h(x) \text{ misclassifies } x \\ 0 & \text{otherwise.} \end{cases}$$

* The $error_S(h)$ is w.r.t training examples.

eg:-

Suppose, if for given data set of size $N = 10$, the hypothesis $h(x)$ misclassifies 5 instances then the sample error is,

$$error_S(h) = \frac{5}{10} = 0.5$$

2. True Error:

Defn:- The true error of a hypothesis w.r.t target function f over the entire unknown D distribution is given by,

$$error_D(h) = \underset{x \in D}{\text{Probability}} \left[f(x) \neq h(x) \right]$$

* The true error is w.r.t to test ~~data~~ or Future data set

Remark:-

1. It is very difficult to arrive at one particular conclusion
2. The true error is estimated by using the confidence interval.

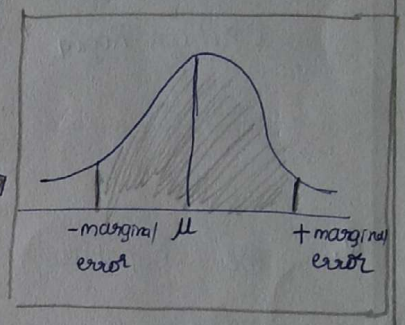
3. Confidence intervals for Discrete valued hypothesis:-

The confidence interval for given hypothesis h is measured interms of ~~percentage~~ N percentage ie, 95%, 89% etc. over a given samples of data.

* The true error ie, $error_D(h)$ is estimated by using confidence interval.

ie:,

$$\boxed{error_D(h) = \mu \pm \text{marginal error}}$$



mean error ($error_S(h)$.

$$= error_S(h) \pm Z_n \sqrt{\frac{error_S(h)\,(1 - error_S(h))}{n}}$$

where, Z_n = A constant value for N% confidence

eg:- For 95% $Z_n = 1.96$

90% $Z_n = 1.64$ etc.

Problem:-

A patient dataset contains 40 samples and the hypothesis 'h' of some learning algo misclassifies 12 instances then find

1. Sample error
2. True error,

assume 95% confidence interval, $Z_n = 1.96$.

Soln:-

1. Sample Error :- $error_S(h) = \frac{12}{40} = 0.30$

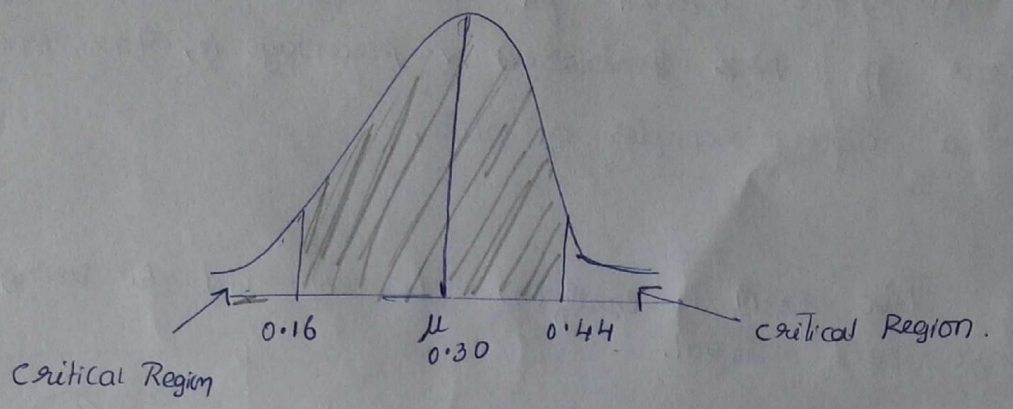
2. True error : $error_D(h) = error_S(h) \pm Z_n \sqrt{\frac{error_S(h) \cdot (1 - error_S(h))}{n}}$

$$= 0.30 \pm 1.96 \sqrt{\frac{0.30 \cdot (1 - 0.3)}{40}}$$

$$= 0.30 \pm 0.14.$$

The range is $[0.44, 0.16]$

$\therefore$



---

2. The hypothesis h, commits 28 success out of 70 independent instances. compute 90% confidence interval to true error for where $Z_n = 1.64$

Sol:-

Total instance $N = 70$

Success Rate $= 28$

$\therefore$ error $= 70 - 28$

$= 42$ instances are misclassified by h.

$\therefore$ 1. $Error_S(h) = \frac{42}{70} = 0.6$

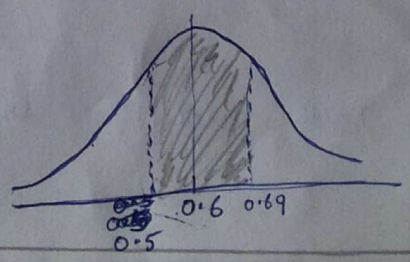
2. True error, $error_D(h) = 0.6 \pm 1.64\sqrt{\frac{0.6(1-0.6)}{70}}$

$= 0.6 \pm 1.64 \times 0.0585$

$= 0.6 \pm 0.0986$

Interval Range is $= [0.69, 0.5]$

3. the confidence interval,



NOTE - 1

*** 1. Explain the difficulties in evaluating the hypothesis.?

2. write a short note on

1. Bias 2. unbiased 3. variance.

*** 3. Explain the confidence interval for discrete hypothesis.

4. Explain the configuration of estimating the error of a learning algorithm.

5. Explain the steps (general setup) to estimate the accuracy of the hypothesis.

Refer page NO

*** 6. List and explain the statistical operators to estimate the error of a hypothesis h.

Soln:- explain → sample error, ~~mean~~ True error & confidence Interval

---

Basics of Sample Theory:-

The basics are,

1. Random Variable :-

It is the probabilistic outcome of a given experiment.

eg:- tossing a coin., $P(h) = 0.5$

2. probability distribution:-

It is the probability that y will take on the value y_i.

eg:- $Pr(\text{head} = 0.21)$.

3. Mean value of Expected outcome:-

Refer page NO:- 3

Variance :- It describes the width of the distribution about its mean value.

$$var(y) = E(y - \mu_y)^2.$$

$$\text{standard dev} = \sqrt{var(y)}.$$

The Binomial Distribution :-

Defn :- "The Binomial Distribution is a probabilistic model where the task 'r' is observed over a series of N independent task".

*. It is given by,

$$\boxed{P(X = r) = {}^{n}C_{r} \; p^{r} (1-p)^{n-r}}$$

* If the instances X of a data set are selected independently then the true error is estimated according to binomial distribution. is

ie,

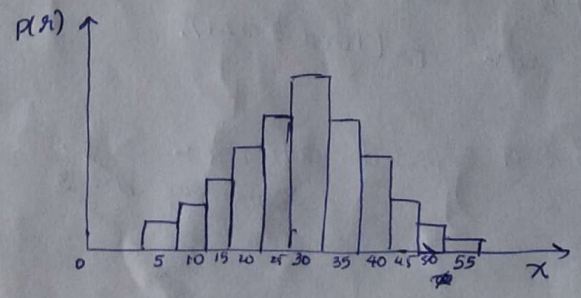
$$\boxed{P(X = r) = {}^{n}C_{r} \; error_S(h)^{r} (1 - error_S(h))^{n-r}}.$$

where, $r \rightarrow$ no of positive classification.

$n \rightarrow$ An identical independent task.

$error_S(h) \rightarrow$ Sample error (r/n).

* The binomial distribution for a given task 'r' is shown as,



Properties :-

1. Mean value (μ) :-

Defn :- "Let y be the random variable which takes the ~~positive~~ possible values $y_1, y_2, y_3 \cdots y_R$

then mean is given by

$$\boxed{E(y) = \sum_{i=0}^{R} y_i \cdot P}$$

probability value.

2. Variance :-

Defn :- "Let y be the random variable, then its dispersion is given by,

$$var(y) = n \cdot p \cdot (1-p)$$

$$std(y) = \sqrt{var(y)}$$

$$= \sqrt{n \cdot p(1-p)}.$$

3. Estimators :-

The estimators such as $error_S(h)$ & $error_D(h)$ ~~is~~ are used to estimate the accuracy of the hypothesis.

$$error_S(h) = \frac{r}{n}$$

$$= \frac{\text{Total misclassified instance}}{\text{Total no of instance}}$$

$$error_D(h) = P$$

= probability value of misclassifying a instance

4. Estimating bias :-

Defn :- "Let $y_1, y_2 \cdots y_n$ be the random variable for a parameter P is,

$$= E[y] - P$$

$$= \begin{cases} 0 & \text{unbiased [no error]} \\ \neq 0 & \text{biased [error]} \end{cases}$$

5. The standard deviation in $error_S(h)$:-

Defn:- The S.D for $error_S(h)$ is given by,

$$\sigma_{error_S(h)} = \sqrt{\frac{P(1-P)}{n}}$$

$$= \sqrt{\frac{error_S(h)\,(1 - error_S(h))}{n}}$$

the samples are choosen independently.

Problems:-

1. A dataset of $n=10$ samples and the sample error i.e., $error_S(h) = 0.4$. Find the probability for $x=7$. use Binomial distribution. Find variance & S.D

Soln:-

1. $P(x=r) = {}^{n}C_{r}\; error_S(h)^{r}\,(1 - error_S(h))^{n-r}$

$$\therefore P(x=7) = {}^{10}C_{7}\,(0.4)^{7}\,(1-0.4)^{10-7}$$

$$= \underline{\underline{0.04}}$$

2. variance $= n.P.(1-P)$

$$= 10 \cdot error_S(h)\,(1 - error_S(h))$$

$$= 10 \times 0.4 \times 0.6 = 2.4$$

3. standard dev: $= \sqrt{var}$

$$= \sqrt{2.4} = 1.54.$$

2. use Binomial distribution, Find the probability value b/w 1 to 5. Assume $p = 2/3$

1. Find the expected value & mean

2. Find variance & S.D.

Soln :- (1) The pr b/w 1 to 5

$n = 5$,

$$P(x=0) = {}^5C_0 (0.6)^0 (1-0.6)^{5-0} = 0.004$$

$$P(x=1) = {}^5C_1 (0.6)^1 (1-0.6)^{5-1} = 0.041$$

$$P(x=2) = {}^5C_2 \, 0.6^2 (1-0.6)^{5-2} = 0.1646$$

$$P(x=3) = \quad = 0.3292$$

$$P(x=4) = \quad = 0.3292$$

$$P(x=5) = \quad = 0.13$$

2. Mean value

Interval Y	0	1	2	3	4	5
proba P	0.004	0.041	0.1646	0.3292	0.3292	0.13

$$= \sum_{i=0}^{n} Y_i \cdot P$$

$$\therefore = (0 \times 0.004 + 1 \times 0.041 + 2 \times 0.1646 + 3 \times 0.3292 + 4 \times 0.3292 + 5 \times 0.13)$$

$$= 3.3$$

3. variance & S.D :-

$$var = n \cdot P(1-P)$$
$$= 5 \cdot 0.6 (1-0.6) = 1.2$$

$$S.D = \sqrt{var} = \sqrt{1.2}$$
$$= 1.09$$

3. Suppose hypothesis h commits $r = 300$ errors on a sample data of size $n = 1000$ randomly drawn test examples. Find Standard deviation in $error_S(h)$?

Soln :-

Since, the instances are drawn randomly, ∴ it is a Binomial distribution.

$$\text{The } S.D = \sqrt{\frac{error_S(h)\,(1 - error_S(h))}{n}}$$

$$\therefore \quad error_S(h) = \frac{r}{n} = \frac{300}{1000} = 0.3$$

$$S.D = \sqrt{\frac{0.3\,(1-0.3)}{1000}}$$

$$= 0.0144$$

4. The hypothesis h, of some learning algo gives the mean error $error_S(h) = 0.4$ over 100 samples. Find the ~~error~~ total no of instances which misclassified by h?

Sol :-

$$error_S(h) = \frac{r}{n} = \frac{\text{Total no of misclassification}}{\text{Total no instances}}$$

$$\therefore \quad 0.4 = \frac{r}{100}$$

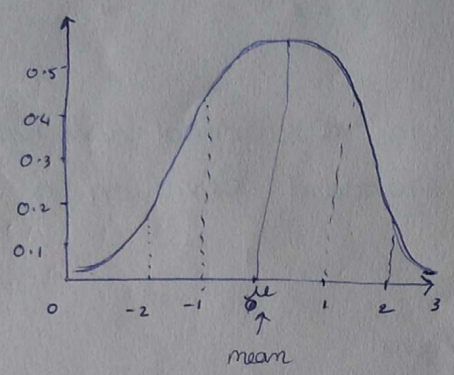
$\therefore r = 40$ instances out of 100 are misclassified by h.

The Normal or gaussian distribution :-

The normal distribution is a bell shaped distribution function defined by the probability density Function. (PDF).

* The PDF will fall in the interval (a, b)

$$\int_a^b P(x) \cdot dx.$$



where, $$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

* The Normal distribution is determined by σ & μ.
* The normal distribution is symmetric about μ (mean).
* The ~~ta~~ total area of the curve is 1
* The expected value & mean is,

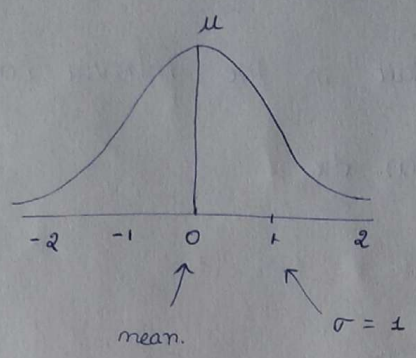
$$E(x) = \mu.$$

* The curve attains the maximum value at, $\frac{1}{\sqrt{2\pi\sigma}}$
* The S.D $\sigma = \sqrt{\frac{(x - \text{mean})^2}{n-1}}$
* If a random variable Y follows the normal distribution then its value lies b/w,

$$\boxed{x = \mu \pm z_n \sigma} \qquad \therefore z = \frac{x-\mu}{\sigma}$$

Special cases of normal distribution :- [properties]

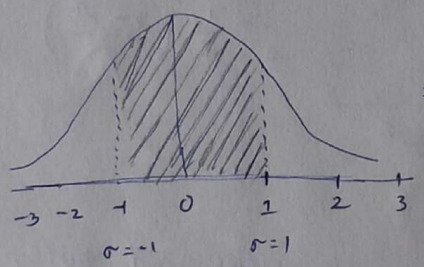
1. The normal distribution is said to be standard if the mean $\mu = 0$ & standard deviation of $\sigma = 1$.



2. The N% of the values in a normal distribution will lie within 1, 2 & 3 standard deviation of the mean.

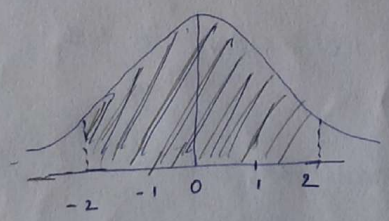
ie,

i. 68% confidence interval.



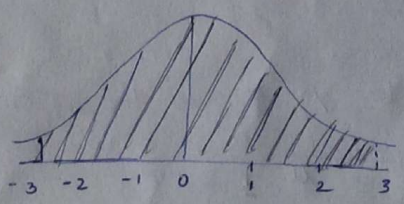
⇒ $\mu \pm 1$ standard dev line

ii. 95% confidence interval.



⇒ $\mu \pm 2$ standard dev line

iii. 99% confidence interval,



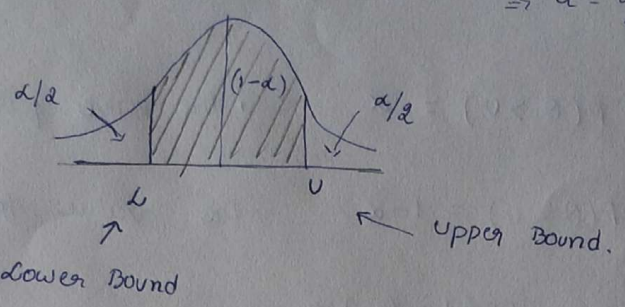
⇒ $\mu \pm 3$. standard dev line

Two sided and one-sided Bounds of Normal distribution:

I. Two sided Bounds :-

- The two sided confidence interval is given by (L, U) and used to estimate the confidence interval w.r.t Mean & it is given by,

$\Rightarrow \alpha = \alpha/2 + \alpha/2.$



i). $\alpha/2$ is the unshaded region at Left & Right of the mean.

ii). $(1-\alpha)$: the shaded region where new probability value falls in.

iii. The confidence interval is $100(1-\alpha)\%$

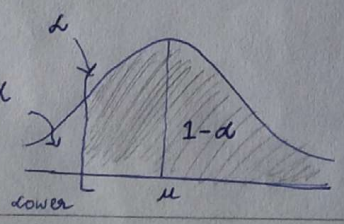
eg:-

If $\alpha = 0.05$ then $C.I = 100(1-0.05)$
$= 95\%$

II. One Sided Bounds :-

Lower sided Bounds:-

The Lower sided bound confidence interval for a Parameter θ is given by,



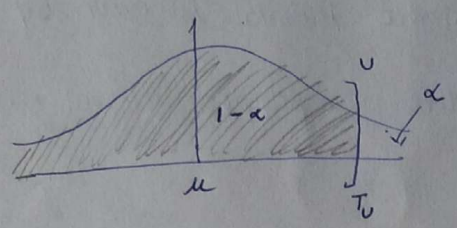
$\Rightarrow \therefore P(\theta > L) =$ then θ falls in $(1-\alpha)$ reg region

$P(\theta < L) = \alpha$

$C.I = 100(1-\alpha/2).$

ii. Upper Bounds :-

The upper bounded confidence interval for a parameter θ is given by.



then if,

$P(\theta > U) = \alpha$ ie, the value falls in α region

$P(\theta < U) = 1-\alpha$ The value falls in $1-\alpha$ region

$CI = 100(1-\alpha/2)\%$

Problem:-

1. Consider the values $x = 10.5, 20.5, 30.2, 22.2, 9.5$.

Find, 1. Mean & S.d of Normal distr

2. Find the value of Z (S.d Line) for $x_1 = 29.1$, $x_2 = 33.5$

Sol :-

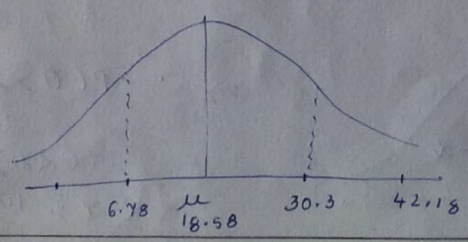
1. mean $= \mu = \dfrac{10.5 + 20.5 + 30.2 + 22.2 + 9.5}{5} = 18.58$

2. S.D $\sigma = \sqrt{\dfrac{\Sigma (x-\bar{x})^2}{n-1}}$

$= \sqrt{\dfrac{(10.5-18.58)^2 + (20.5-18.58)^2 + (30.2-18.58)^2 + (22.2-18.58)^2 + (9.5-18.58)^2}{4}}$

$= 11.8.$

3. The C.I.



4. The S.d ~~line~~ line for $x_1 = 29.1$

$$Z = \frac{x - \bar{x}}{\sigma}$$

$$= \frac{29.1 - 18.58}{11.8}$$

$$= 0.89 \approx < 1$$

$\therefore$ It is less than 1 $\therefore$ the point falls with in first S.D line

$\therefore$ it is 68% confidence interval.

for $x_2 = 33.5$,

$$Z = \frac{33.5 - 18.58}{11.8}$$

$$= 1.26 \approx > 1$$

$\therefore$ It is greater than 1 $\therefore$ The point x_2 falls in Second S.D Line

$\therefore$ It is 95% C.I.

] Refer Page No. 14 ($\mu \pm 2$).

***

The general approach for deriving the confidence-interval:-

To derieve the C.I the steps are,

1. Identify the population parameter 'p' of the training dataset.

2. Define the estimator y such that gives variance & unbiased estimator.

3. Determine the probability distribution which includes mean & variance such as binomial & normal distribution etc.

   1. $^{n}C_{r}\, p^{r}(1-p)^{n-r}$ 2. $\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

4. Determine the N% C.I by finding lower & upper bounds of normal distribution.

Central Limit theorem :-

* The central Limit theorem simplifies the derivation of confidence interval.

* Defn :- "
A set of independent, identical random variable $Y_1, Y_2 \ldots Y_n$ ~~with by any dist~~ by any probability distribution, with mean μ & variance σ^2.

$$\therefore z = \frac{\bar{Y}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

As $n \to \infty$, ie as n approaches to ∞ any distribution becomes the normal distribution

Differences in error of two hypothees:-

Consider the hypothesis H_1 & H_2 which is tested on the sample S_1 & S_2 containing n_1 & n_2 randomly drawn ~~ex~~ examples, then,

i. Difference in true error of H_1 & H_2 is,

$$\boxed{d = error_D(h_1) - error_D(h_2)}$$

ii. Difference in sample error of H_1 & H_2 is.

$$\hat{d} = error_S(h_1) - error_S(h_2).$$

iii. The N% confidence interval for H_1 & H_2 is,

$$= \hat{d} + z_n \sqrt{\frac{error_S(h_1)(1-error_S(h_1))}{n_1} + \frac{error_S(h_2)(1-error_S(h_2))}{n_2}}$$

iv. The variance is, $\sigma_d^2 = \frac{error_S(h_1)(1-error_S(h_1))}{n_1} + \frac{error_S(h_2)(1-error_S(h_2))}{n_2}$

NOTE :-

1. List & explain the basics of sampling theory

2. Explain the Binomial distribution ~~along~~ and its properties.

3. Explain the normal or gaussian distribution with an ex.

4. Briefly explain the bounds of normal distribution.

5. Define Central Limit theorem. & Explain the appr to derive confidence interval.

6. Explain the procedure in estimating the difference in error of two hypothesis.

*** Comparing the learning methods :-

The hypothesis of two learning algo can be compared statistically by

1. Paired - t test
2. K- nearest neighbor.

1. Paired-t-test :-

The paired t test is used to compute the mean of two samples where one sample paired with other sample.

* The observation is done before & after on the same sub

eg :- Sugar level before & after food.

Before Breakfast	After Breakfast
22.5	100.5
30.5	102.5
⋮	⋮

Algorithm:-

1. Let X, Y be the data Before ~~obs~~ and after observation

(Before) X	(After) Y
-	-
-	-
-	-
	-
	-

2. Test the null hypothesis h_0, where the true mean difference is 0

i. calculate the diff.

$$d_i = y_i - x_i$$

ii. calculate the mean difference,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} d_i$$

iii. ~~th~~ calculate the standard deviation,

$$S.D = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (d_i - \bar{y})^2}$$

iv. calculate the t-statistic value,

$$\text{hypothesis} = \bar{y} \pm t_{n-1} \, S.d.$$

Problem :-

Consider before and after values of some observation and perform paired t test, Find.

1. The degrees of Freedom.

2. The 95% C.I for alternate hypo, ($t_{n-1} = 3.182$)

3. The t - statistic value for null hypothesis.

Solution

	X Before	Y After
1.	6.01	5.23
2	2.28	1.21
3	1.51	1.40
4.	2.12	1.38

Soln :-

1. The degrees of freedom is

$= n - 1$

$= 4 - 1 = \underline{3}$

2. The Sample mean.

X	Y	$y_i - x_i$
6.01	5.23	-0.78
2.28	1.21	-1.07
1.51	1.40	-0.11
2.12	1.38	-0.74

$$\text{mean} = \frac{-0.78 - 1.07 - 0.11 - 0.74}{4} = -0.675$$

3. The s.d. $= \sqrt{\frac{1}{n(n-1)} \Sigma (y_i - \text{mean})^2}$

$$= \sqrt{\frac{(-0.675+0.78)^2 + (-0.675+1.07)^2 + (-0.675+0.11)^2 + (-0.675+0.74)^2}{4 \cdot 3}}$$

$= 0.2$

4. The 95% CI is, mean $\pm t_{n-1} \sigma$

$= -0.68 \pm 3.182 (0.2)$

$= -0.68 \pm 0.63 \quad = [-1.3164 \quad -0.0436]$.

5. The t-statistic value for null hypothesis, $h_1 = 0$

$$h_1 = \bar{y} \pm t_{n-1} \cdot s.d$$

$$t_{n-1} = \frac{\bar{y}}{s.d} = \frac{-0.68}{0.2}$$

$$= -3.4$$

2. Consider the sample value Find paired t-test, Find

1. degrees of freedom
2. 95% CI & Assume $t_{n-1} = 2.776$
3. The t-statistic value for $h_1 = 0$.

n	X Before	Y After
1	0	5
2	2	8
3	1	8
4	2	5
5	1	7

⇒ Refer Class Notes.

K-nearest algorithm :-

- It is instance based Learning method where it assumes all the instances as points in 'n'-dimensional space.
- This algo finds the distances b/w the two points

$$d(x_i, x_j) = \sqrt{\sum (a(x_i) - a(x_j))^2}$$

Algorithm :-

Step 1 :- Store the training examples to a list

Step 2 :- If the target function is discrete values then

$$\boxed{\text{return} \leftarrow \max \sum_{i=1}^{n} \delta(v, f(x_i))}$$

else

if the target function is continuous then

$$\boxed{\text{return} \leftarrow f(x) = \frac{\sum f(x)}{n}}$$ i.e mean value

Problem :-

Consider the training dataset as shown below

A	B	target
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Classify the new instance using K nearest classifier.

A = 3	B = 7	Class = ?

Soln :-

A	B	target	distance	Rank
7	7	Bad	$\sqrt{(3-7)^2+(7-7)^2} = 4$	3
7	4	Bad	$\sqrt{(3-7)^2+(7-4)^2} = 5$	4
3	4	Good	$\sqrt{0+(7-4)^2} = 3$	1
1	4	Good	$\sqrt{2^2+3^2} = 3.6$	2

if K = 1 [1-nearest neighbor].

~~Consider~~ choose least first Rank ie. 1

∴

A = 3	B = 7	Class = Good

if K=2 [2 - nearest neighbor]

choose least 2 Rank. ie, 1 , 2
(3 3.6)

A = 3	B = 7	Class = Good

if K=3 [3 - nearest neighbor].

choose least 3 Ranks [1 , 2 , 3]
3 3.6 4
Good Good Bad

A = 3	B = 7	Class = good.

BGS Institute of Technology

VTU

Bengaluru – Hassan National Highway (NH-75), Nagamangala Taluk, Mandya District, B G Nagar, Karnataka 571448

Ph: 08234 - 288418, Fax:288419

Email: principalbgsit@rediffmail.com, Web:www.bgsit.ac.in

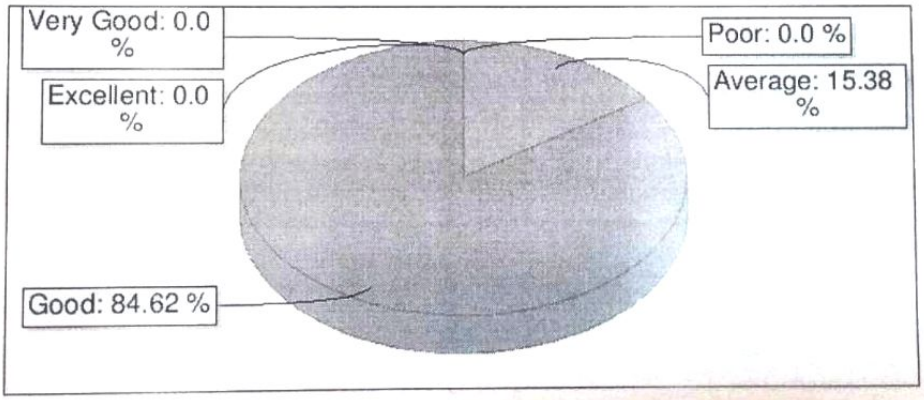
Student Feedback On Faculty 2020-21

Batch : BE , 2017-2021
Staff Name : Mr Prasanna Kumar M J
Subject Code : 17CS73
Subject Name : MACHINE LEARNING

Department : Computer Science and Engineering,Information Science and Engineering
Semester 7
Date : 12 Feb 2021

No	Questions	Poor	Average	Good	Very Good	Excellent	Feedback Percentage	Average Score (5)
		1	2	3	4	5		
	Time Sense							
1	Teacher conducts the classes regularly	0	2	11	0	0	56.9	2.8
2	Syllabus of this course is completed in time	0	2	11	0	0	56.9	2.8
3	Assignments, class tests, quizzes and seminars were conducted as per schedule	0	2	11	0	0	56.9	2.8
4	Alternate arrangements were made during his/her absence	0	2	11	0	0	56.9	2.8
	Class Control/Management							
1	The faculty is effective in controlling and conducting the class	0	2	11	0	0	56.9	2.8
2	The faculty invites student participation.	0	2	11	0	0	56.9	2.8
3	The faculty rightfully addresses inappropriate behaviour of students	0	2	11	0	0	56.9	2.8
4	The faculty has a tendency of inviting opinion and questions on subject matter from students	0	2	11	0	0	56.9	2.8
5	The faculty enhances learning by judicious reinforcement mechanism	0	2	11	0	0	56.9	2.8
	Subject Command							
1	The faculty focused on the defined syllabus	0	2	11	0	0	56.9	2.8
2	The faculty conducted and involved students in classroom discussions	0	2	11	0	0	56.9	2.8
3	The faculty had good communication skills	0	2	11	0	0	56.9	2.8
4	The lectures were well structured	0	2	11	0	0	56.9	2.8
5	The faculty related the subject to real life applications of concepts	0	2	11	0	0	56.9	2.8
6	The faculty referred to latest developments in the fields	0	2	11	0	0	56.9	2.8
	Use of Teaching Aid							
1	The faculty used different teaching aids like PPT's, Blackboard, Overhead Projectors etc	0	2	11	0	0	56.9	2.8
2	The blackboard/whiteboard work was clear in terms of legibility, visibility and structure	0	2	11	0	0	56.9	2.8

3	The faculty used different teaching methods in conducting the class. (Example group discussion, seminars, student presentations, etc.)	0	2	11	0	0	56.9	2.8
4	The faculty shared and discussed the answers to class tests or sessional tests	0	2	11	0	0	56.9	2.8
5	The faculty allowed the review of answer scripts of class tests	0	2	11	0	0	56.9	2.8
6	The faculty made sure all students are able to understand him/her.	0	2	11	0	0	56.9	2.8
	Helping Attitude							
1	The faculty has a helping attitude towards varied academic interests of students.	0	2	11	0	0	56.9	2.8
2	The faculty helps students gain access to material not readily available in text books, through e-resources, e-journals, reference books etc	0	2	11	0	0	56.9	2.8
3	The faculty has helps students facing physical, emotional and learning challenges.	0	2	11	0	0	56.9	2.8
4	The faculty's approach is towards development of professional skills among students	0	2	11	0	0	56.9	2.8
5	The faculty helps students in realizing career goals	0	2	11	0	0	56.9	2.8
6	The faculty helps students in realizing their strengths and development needs	0	2	11	0	0	56.9	2.8
	Total Count	**0**	**54**	**297**	**0**	**0**	**56.9**	**2.8**



Comments
Good
Mode rate
Nice teaching sir
Nice teaching sir
Nice teaching sir
Nice teaching sir
Nice teaching sir
1
.
Good
Good

B. K. Raghu
HOD
Dept. of Computer Science & Engg.
B.G.S. Institute of Technology,
B.G. Nagar - 571 448

BGS Institute of Technology

VTU

Bengaluru – Hassan National Highway (NH-75), Nagamangala
Taluk, Mandya District, B G Nagar, Karnataka 571448
Ph: 08234 - 288418, Fax:288419
Email: principalbgsit@rediffmail.com, Web:www.bgsit.ac.in

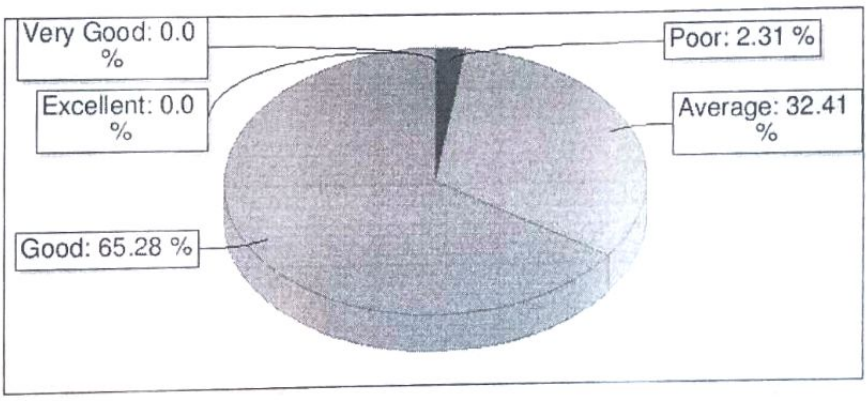
Student Feedback On Faculty 2020-21

Batch : BE , 2017-2021
Staff Name : Mr Prasanna Kumar M J
Subject Code : 17CS73
Subject Name : MACHINE LEARNING

Department : Computer Science and Engineering,Information Science and Engineering
Semester 7 , Sec : A
Date : 12 Feb 2021

No	Questions	Poor	Average	Good	Very Good	Excellent	Feedback Percentage	Average Score (5)
		1	2	3	4	5		
	Time Sense							
1	Teacher conducts the classes regularly	0	4	12	0	0	55	2.8
2	Syllabus of this course is completed in time	1	4	11	0	0	52.5	2.6
3	Assignments, class tests, quizzes and seminars were conducted as per schedule	1	5	10	0	0	51.2	2.6
4	Alternate arrangements were made during his/her absence	1	6	9	0	0	50	2.5
	Class Control/Management							
1	The faculty is effective in controlling and conducting the class	0	4	12	0	0	55	2.8
2	The faculty invites student participation.	0	5	11	0	0	53.8	2.7
3	The faculty rightfully addresses inappropriate behaviour of students	0	5	11	0	0	53.8	2.7
4	The faculty has a tendency of inviting opinion and questions on subject matter from students	0	5	11	0	0	53.8	2.7
5	The faculty enhances learning by judicious reinforcement mechanism	0	6	10	0	0	52.5	2.6
	Subject Command							
1	The faculty focused on the defined syllabus	0	6	10	0	0	52.5	2.6
2	The faculty conducted and involved students in classroom discussions	0	6	10	0	0	52.5	2.6
3	The faculty had good communication skills	0	7	9	0	0	51.2	2.6
4	The lectures were well structured	0	6	10	0	0	52.5	2.6
5	The faculty related the subject to real life applications of concepts	0	8	8	0	0	50	2.5
6	The faculty referred to latest developments in the fields	1	5	10	0	0	51.2	2.6
	Use of Teaching Aid							
1	The faculty used different teaching aids like PPT's, Blackboard, Overhead Projectors etc	1	5	10	0	0	51.2	2.6
2	The blackboard/whiteboard work was clear in terms of legibility, visibility and structure	1	4	11	0	0	52.5	2.6

3	The faculty used different teaching methods in conducting the class (Example group discussion, seminars, student presentations, etc.)	1	5	10	0	0	51.2	2.6
4	The faculty shared and discussed the answers to class tests or sessional tests	1	5	10	0	0	51.2	2.6
5	The faculty allowed the review of answer scripts of class tests	1	5	10	0	0	51.2	2.6
6	The faculty made sure all students are able to understand him/her.	1	6	9	0	0	50	2.5
	Helping Attitude							
1	The faculty has a helping attitude towards varied academic interests of students.	0	4	12	0	0	55	2.8
2	The faculty helps students gain access to material not readily available in text books, through e-resources, e-journals, reference books etc	0	5	11	0	0	53.8	2.7
3	The faculty has helps students facing physical, emotional and learning challenges.	0	4	12	0	0	55	2.8
4	The faculty's approach is towards development of professional skills among students	0	6	10	0	0	52.5	2.6
5	The faculty helps students in realizing career goals	0	5	11	0	0	53.8	2.7
6	The faculty helps students in realizing their strengths and development needs	0	4	12	0	0	55	2.8
	Total Count	**10**	**140**	**282**	**0**	**0**	**52.6**	**2.65**



Comments
.
Good
Teaching is good
Good tecahing
They teach us very well
Teaching was good
A Good Lecturer
.
.
Good
Good

- Do the content beyond syllabus for the next batch students

B. K. Raghav
HOD
Dept. of Computer Science & Engg.
B.G.S. Institute of Technology

BGS Institute of Technology

Department of Computer science & Engineering

INTERNAL AUDITING

DATE: 12/01/2021

Name of the Faculty:	M. J. Prasanna Kumar
Designation:	Asst professor
Subject Name with Code	Machine Learning 17CS73 and machine Learning Lab 17CS76

SL. No.	Contents	(2020-21) ODD Theory	Lab
1	Faculty Profile	✓	✓
2	Vision and Mission of the Institute	✓	✓
3	Vision and Mission of the Department	✓	✓
4	Department PEO's and PSOs,	✓	✓
5	Course Outcome	✓	✓
6	Mapping of COs and POs, PEOs, PSOs	✓	✓
7	Assessment Tools and Procedure for Assessment of Cos (IA Test, Quiz, Surprise test, Assignment, University Examination)	✓	✓
8	Previous University Question Papers	✓	X
9	COE of Institute and COE of the Department (COE= Calendar of Events)	✓	✓
10	Time Table (Class and Individual)	✓	✓
11	Course Plan (Syllabus Copy along with CO and hours)	✓	✓
12	List of Text and Reference Books	✓	✓
13	Lesson Plan	✓	✓
14	Batch wise Assignments Batchwise assignment needs to be given	X	X
15	Students Roll Call with phone numbers (Procter Details batch wise)	X	X
16	Report of Guest Lectures Needs to conduct guest lecture	X	X
17	Notes	✓	X
18	Question Bank	✓	✓
19	FEED Back Report (Mid of the semester & End of the Semester)	✓	✓
20	Communications with Faculty and Students	✓	✓
21	Academic Diary	✓	✓
22	Course end survey	✓	✓

Signature Of External Auditor

Signature Of Academic Incharge

B. K. R. HOD
Dept. of Computer Science & Engg.
B.G.S. Institute of Technology
B.G. Nagar - 571 448
Mandya Dist.

Signature Of HoD

Signature Of principal
Principal
BGSIT

||Jai Sri Gurudev||

BGS Institute of Technology

Department of Computer Science and Engineering

Academic year 2020-21 (ODD / EVEN)	
Name of the Faculty with Designation	M.J. PRASANNA KUMARA
Course Name with code	Machine Learning 17CS73

Feed Back Report			
No. of Students participated	29	Overall Feedback	100%

Course End Survey						
CO's	CO.1	CO.2	CO.3	CO.4	CO.5	CO.6
Av. Rating	2.6	2.8	2.86	2.64	2.64	

CO Attainment						
CO's	CO.1	CO.2	CO.3	CO.4	CO.5	CO.6
Attainment	2.19	2.3	2.26	2.37	2.37	

PO / PSO Attainment														
PO/PSO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
Attainment	1.53	1.53	1.53	0.76		0.752			0.76			1.07	1.53	1.53

Analysis of CO, PO/PSO Attainment [Review of attainment (course attainment)]

Try improve the attainment level which is not mapped for the next batch.

B. K. Raghu

HOD

Dept. of Computer Science & Engg.
B.G.S. Institute of Technology,
B.G. Nagar - 571 448
Nagamangala Tq. Mandya Dist
Karnataka (INDIA)

||Jai Sri Gurudev||

BGS Institute of Technology

Department of Computer Science and Engineering

Result Analysis CIE				
	Test-1	Test-2	Test-3	IA Final
22 (≥76%)	70	77	86	86
12-22 (≥41% ≤75%)	12	06		
12 (≤40%)	–			
Total No of Students	86	86	86	86
	4 AB	03	–	

Action taken for Slow learners:

Test-1

Test-2

Result Analysis SEE					
Course name with Code	Total Appeared	FCD	FC	Pass %	Failed
Machine Learning 17CS73	85	35	34	100%	–

Remarks: Performance is good & satisfactory

Faculty

B. K. Raghu

HOD

Dept. of Computer Science & Engg.
B.G.S. Institute of Technology