

# CBCS SCHEME

USN

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

BCS602

## Sixth Semester B.E./B.Tech. Degree Examination, June/July 2025 Machine Learning

Time: 3 hrs.

Max. Marks: 100

*Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.  
2. M : Marks , L: Bloom's level , C: Course outcomes.*

Module - 1				M	L	C																													
<b>Q.1</b>	a.	State Tom Mitchell's definition of machine learning. List and explain the challenges of machine learning.	7	L1	CO1																														
	b.	List and explain the visualization aids available for univariate data analysis with example for each.	7	L2	CO1																														
	c.	For the patients age list {12, 14, 19, 22, 24, 26, 28, 31, 34}. Find the IQR.	6	L3	CO1																														
<b>OR</b>																																			
<b>Q.2</b>	a.	Explain in detail the machine learning process with a neat diagram.	7	L2	CO1																														
	b.	Explain data preprocessing with measures to solve the problem of missing data.	7	L2	CO1																														
	c.	Find the 5-point summary of the list {13, 11, 2, 3, 4, 8, 9} and plot the box plot for the same.	6	L3	CO1																														
<b>Module - 2</b>																																			
<b>Q.3</b>	a.	Let the data points be $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 7 \end{pmatrix}$ . Apply Principal Component Analysis (PCA) and find the transformed data.	10	L3	CO1																														
	b.	Apply candidate elimination algorithm on the dataset given in Table Q.3(b) to obtain the complete version space. <div style="text-align: center; margin-top: 5px;">Table Q.3(b)</div> <table border="1" style="width: 100%; margin-top: 5px; border-collapse: collapse;"> <thead> <tr> <th>CGPA</th> <th>Interactiveness</th> <th>Practical knowledge</th> <th>Communication skills</th> <th>Logical thinking</th> <th>Job offer</th> </tr> </thead> <tbody> <tr> <td>≥ 9</td> <td>Yes</td> <td>Excellent</td> <td>Good</td> <td>Fast</td> <td>YES</td> </tr> <tr> <td>≥ 9</td> <td>Yes</td> <td>Good</td> <td>Good</td> <td>Fast</td> <td>YES</td> </tr> <tr> <td>≥ 8</td> <td>No</td> <td>Good</td> <td>Good</td> <td>Fast</td> <td>NO</td> </tr> <tr> <td>≥ 9</td> <td>Yes</td> <td>Good</td> <td>Good</td> <td>Slow</td> <td>YES</td> </tr> </tbody> </table>	CGPA	Interactiveness	Practical knowledge	Communication skills	Logical thinking	Job offer	≥ 9	Yes	Excellent	Good	Fast	YES	≥ 9	Yes	Good	Good	Fast	YES	≥ 8	No	Good	Good	Fast	NO	≥ 9	Yes	Good	Good	Slow	YES	10	L3	CO2
CGPA	Interactiveness	Practical knowledge	Communication skills	Logical thinking	Job offer																														
≥ 9	Yes	Excellent	Good	Fast	YES																														
≥ 9	Yes	Good	Good	Fast	YES																														
≥ 8	No	Good	Good	Fast	NO																														
≥ 9	Yes	Good	Good	Slow	YES																														
<b>OR</b>																																			
<b>Q.4</b>	a.	Find Singular Value Decomposition (SVD) of the matrix $A = \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix}$ .	10	L3	CO2																														
1 of 3																																			

- b.** Write Find-S algorithm. Apply the algorithm to obtain the hypothesis for the dataset given in the Table Q.4(b).

Table Q.4(b)

Sky	Air 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

**Module – 3**

- Q.5 a.** Apply K-nearest neighbor algorithm, for the dataset given in Table Q.5(a). Given a test instance (6.1, 40, 5), use the training set to classify the test instance. Choose K = 3.

Table Q.5(a)

CGPA	Assessment	Project submitted	Result
9.2	85	8	PASS
8	80	7	PASS
8.5	81	8	PASS
6	45	5	FAIL
6.5	50	4	FAIL
5.8	38	5	FAIL

- b.** Explain types of regression methods and limitations of regression methods.

- c.** Explain the structure of a decision tree and write the procedure to construct a decision the using ID3 algorithm.

**OR**

- Q.6 a.** Write the nearest-centroid classifier algorithm. Apply the same to predict the class for the given test instance (6, 5) using the training dataset given in Table Q.6(a)

X	Y	Class
3	1	A
5	2	A
4	3	A
7	6	B
6	7	B
8	5	B

Table Q.6(a)

- b.** Distinguish between  
 i) Regression and correlation  
 ii) Regression and causation  
 iii) Linearity and non-linearity relationships.

- c.** Explain the advantages and disadvantages of decision tree. Write the general algorithm for decision tree.

## Module – 4

**Q.7 a.** Using Naïve bayes classifier classify the new data (Red, SUV, Domestic) using the training dataset given in Table Q.7(a).

Table Q.7(a)

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

10 L3 CO4

**b.** Explain the simple model of an artificial neuron along with the artificial neural network structure.

10 L2 CO4

## OR

**Q.8 a.** Explain Bayes theorem, Maximum A Posteriori (MAP) hypothesis and Maximum Likelihood (ML) hypothesis in detail.

10 L2 CO4

**b.** Explain different activation functions used in artificial neural network.

10 L2 CO4

## Module – 5

**Q.9 a.** Consider the following set of data given in Table Q.9(a). Cluster it using K-means algorithm with initial value of objects 2 and 5 with the coordinate values (4, 6) and (12, 4) as initial seeds.

Table Q.9(a)

Objects	X-coordinate	Y-coordinate
1	2	4
2	4	6
3	6	8
4	10	4
5	12	4

10 L3 CO5

**b.** Explain the various components of reinforcement learning.

10 L2 CO5

## OR

**Q.10 a.** Find the Manhattan and Chebyshev distance if the coordinates of the objects are (0, 3) and (5, 8).

4 L3 CO5

**b.** Explain the mean shift clustering algorithm.

6 L2 CO5

**c.** List and explain the  
i) Characteristics of reinforcement learning  
ii) Challenges of reinforcement learning  
iii) Applications of reinforcement learning

10 L3 CO5

\*\*\*\*\*

# Machine Learning

(BCS 602)

Q1 (a) Tom Mitchell's definition:

2M

A computer program is said to learn from experience  $E$ , with respect to task  $T$  and some performance measure  $P$ , if its performance on  $T$  measured by  $P$  improves with experience ' $E$ '.

Challenges of Machine Learning:

- ① problems
- ② Huge data
- ③ High computation power
- ④ Complexity of the algorithms
- ⑤ Bias/Variance.

5M

Problems: Machine learning can deal with the well-posed problems where specifications are complete & available. Computers cannot solve ill-posed problems.

Huge data: Availability of quality data is a challenge

High Computation Power: GPU & TPU are required to execute machine learning (ML) algorithms.

Complexity of algorithms: Selection, describing, application & comparison of machine learning algorithms is necessary task.

Bias/Variance: Overfitting & underfitting are great challenges in machine learning algorithms.

Q1 (b)

Visualization aids are:-

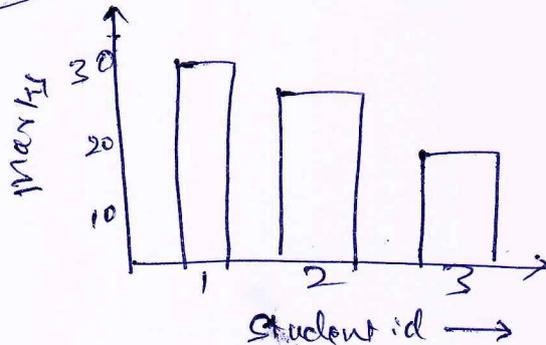
- ① Bar chart
- ② Pie chart
- ③ Histogram
- ④ Dot plots.

7

1 M

Bar Chart: Displays the frequency distribution of variables. Illustrates discrete data.

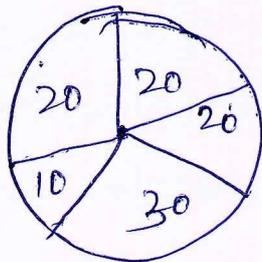
e.g.



2 M

Pie Chart: Illustrates univariate data.

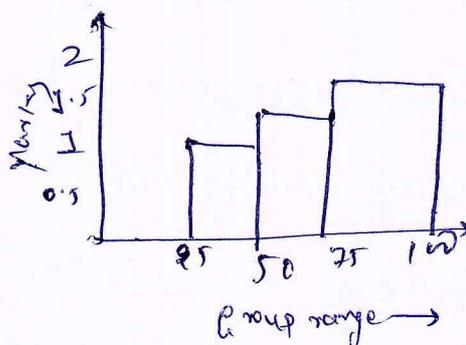
e.g.



2 M

Histogram: Plays important role in data mining for showing frequency distributions.

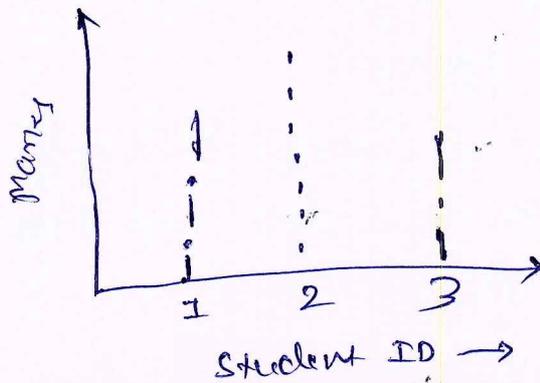
e.g.



conveys useful information like nature of data & its mode.

Dot plot : Similar to bar charts. They are clustered as compared to bar chart. (less 2M)

Fig.



Q1  
(C)

Age = { 12, 14, 19, 22, 24, 26, 28, 31, 34 }

Median = 24

2M

$$Q_{0.25} = \frac{12 + 14 + 19 + 22}{4} = 16.5$$

$$Q_{0.75} = \frac{26 + 28 + 31 + 34}{4} = 29.5$$

2M

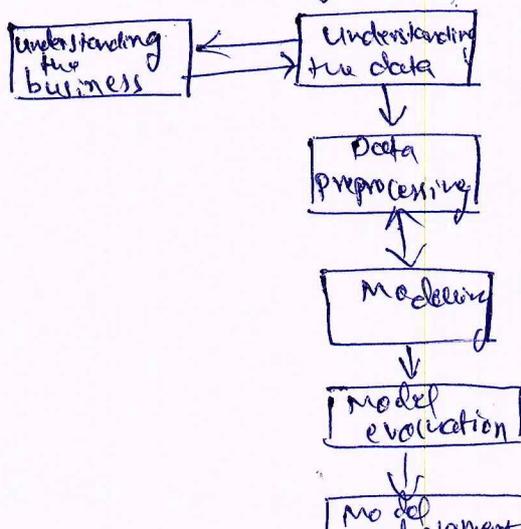
$$\begin{aligned} IQR &= Q_{0.75} - Q_{0.25} \\ &= 29.5 - 16.5 \end{aligned}$$

$$\boxed{IQR = 13}$$

2M

Q2

(a) Machine Learning process



2M

Understanding the data: It involves understanding the objectives and requirements of the business organizations.

2) Understanding the data: It involves data collection, characteristics of data, hypothesis formulation & matching of patterns to selected hypothesis. 5M

3) Preparation of data: It involves producing the final dataset by cleaning the raw data and preparation of data for the data mining process.

4) Modelling - This involves application of data mining algorithm for the data to obtain a model or pattern.

5) Evaluate - It involves evaluation of data mining results using statistical analysis & visualization methods.

6) Deployment - It involves deployment of results of the data mining algorithm to improve existing process.

Q2

(b) Data preprocessing improves the quality of the 3M data mining techniques. The raw data must be processed to give accurate results. The process of detection & removal of errors in data is called data cleaning. Data wrangling means making data processable for machine learning algorithms.

Measures to solve the problem of missing data.

① Ignore the tuple 4M

② Fill in the values manually.

③ Use global constant

④ Fill by attribute value.

⑤ Use Attribute mean      ⑥ Use most possible value.

$$Q. 2 (c) \quad \{13, 11, 2, 3, 4, 8, 9\} \Rightarrow \{2, 3, 8, 11, 13\}$$

Five-Point Summary:

① Minimum = 2

Median = 8

4M

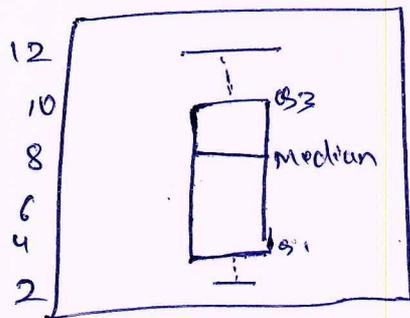
② Maximum = 13

③  $Q_1 = 3$

④  $Q_2 = 8$

⑤  $Q_3 = 11$

Box Plot



2M  
 $\{\text{min}, Q_1, \text{Median}, Q_3, \text{Max}\}$

Q. 3  
 (a)

- Data points =  $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$  &  $\begin{pmatrix} 1 \\ 7 \end{pmatrix}$ . Apply PCA.  
soln

Mean vectors can be computed as

$$\mu = \left( \frac{2+1}{2}, \frac{6+7}{2} \right) = \begin{pmatrix} 1.5 \\ 6.5 \end{pmatrix}$$

2M

$$x_1 = \begin{pmatrix} 2 - 1.5 \\ 6 - 6.5 \end{pmatrix} = \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix}$$

$$x_2 = \begin{pmatrix} 1 - 1.5 \\ 7 - 6.5 \end{pmatrix} = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}$$

Covariance =

$$m_1 = \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix} \begin{pmatrix} 0.5 & -0.5 \end{pmatrix}$$

$$= \begin{pmatrix} 0.25 & -0.25 \\ -0.25 & 0.25 \end{pmatrix}$$

$$m_2 = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \begin{pmatrix} -0.5 & 0.5 \end{pmatrix}$$

$$= \begin{pmatrix} 0.25 & -0.25 \\ -0.25 & 0.25 \end{pmatrix}$$

$$C = \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$$

2M

Eigen values of  $C = \lambda_1 = 1, \lambda_2 = 0$

Eigen vectors  $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$  &  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

2M

$$\text{Matrix } A = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$A^T = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

2M

$$\text{Normalized } A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$A A^T = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\therefore \text{Transformed matrix is } y = A X(x-m) \quad 2M$$
$$= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}$$

Q3  
(b)

Candidate Elimination Algorithm

$$S_0 < \phi \phi \phi \phi \phi >$$

$$S_1 < \rightarrow a, \text{yes, excellent, Good, Fast} >$$

$$S_2 < \rightarrow a, \text{yes, ? , Good, Fast} >$$

4M

$$S_3 < \rightarrow a, \text{yes, ? , Good, Fast} >$$

$$S_4 < \rightarrow a, \text{yes, ? , Good, ?} >$$

$$G_0 < \rightarrow a, \text{yes, ? , ? , ?} > < \rightarrow a, ? , ? , \text{Good} > < ? , \text{yes, ? , Good} >$$

3M

$$G_4 < \rightarrow a, ? , ? , ? , ? > < ? , \text{yes} ? , ? , ? >$$

$$G_3 < \rightarrow a, ? , ? , ? , ? > < ? , \text{yes} ? , ? , ? > < ? , ? , ? , ? >$$

3M

$$G_2 < ? , ? , ? , ? , ? >$$

$$G_1 < ? , ? , ? , ? , ? >$$

$$G_0 < ? , ? , ? , ? , ? >$$

Q4

(a) Find SVD of  $A = \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix}$

Soln

$$AA^T = \begin{pmatrix} 1 & 2 \\ 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 2 & 9 \end{pmatrix} = \begin{pmatrix} 5 & 22 \\ 22 & 97 \end{pmatrix}$$

2M

Eigen values of  $AA^T = 0.0098$  &  $101.9902$

$$\text{Eigen vectors } u_1 = \begin{pmatrix} 0.2268 \\ 1 \end{pmatrix} \quad u_2 = \begin{pmatrix} -4.4086 \\ 1 \end{pmatrix}$$

$$\text{Normalized vectors } u_1 = \begin{pmatrix} 0.2212 \\ 0.9752 \end{pmatrix} \quad u_2 = \begin{pmatrix} -0.9752 \\ 0.2212 \end{pmatrix}$$

$$U = u_1 u_2 = \begin{pmatrix} 0.2212 & -0.9752 \\ 0.9752 & 0.2212 \end{pmatrix}$$

2M

$$A^T A = \begin{pmatrix} 17 & 37 \\ 38 & 35 \end{pmatrix}$$

$$\text{Eigen values} = 10.99, 0.0098$$

2M

$$\text{Eigen vectors} = \begin{pmatrix} 0.0047 \\ 1 \end{pmatrix} \text{ \& } \begin{pmatrix} -2.236 \\ 1 \end{pmatrix}$$

Normalized vectors

$$v_1 = \begin{pmatrix} 0.0082 \\ 0.9129 \end{pmatrix} \quad v_2 = \begin{pmatrix} 0.9129 \\ 0.0082 \end{pmatrix}$$

2M

$$V = [v_1 \ v_2] = \begin{pmatrix} 0.0081 & 0.9129 \\ 0.9129 & 0.0082 \end{pmatrix}$$

$$S = \begin{pmatrix} 10.99 & 0 \\ 0 & 0.0098 \end{pmatrix}$$

2M

Q4(b)

Initial hypothesis

$$h_0 = \langle \phi \ \phi \ \phi \ \phi \ \phi \ \phi \rangle$$

6M

$$h_1 = \langle \text{sunny, warm, } \overset{\text{normal,}}{\text{strong}}, \text{ warm, same} \rangle \text{ true}$$

$$h_2 = \langle \text{sunny, warm, ? , strong, warm, same} \rangle \text{ true}$$

true

$$h_3 = h_2$$

$$h_4 = \langle \text{sunny, warm, ? , strong, ? , ?} \rangle \text{ true}$$

FIND-S Algorithm

Input: Positive instances in the training dataset

Output: Hypothesis 'h'

1. Initialize 'h' to the most specific hypothesis.  $\phi M$

$$h = \langle \phi, \phi, \phi, \dots \rangle$$

2. Generalize the initial hypothesis for first the instance.

3. For each subsequent instances:

IF it is a positive instance,

check for each attribute value in the instance with the hypothesis 'h'.

IF the attribute value is the same as the hypothesis value, do nothing.

ELSE if the attribute value is different than the hypothesis value, change it to '?' in 'h'.

ELSE if it is a negative instance,

Ignore it.

Q5

(a) APPLY KNN, TEST instance (6.1, 40, 5)  $K=3$

STEP 1: calculate Euclidean distance between test  $\phi M$  instance & each of the training instance.

CRPA	Assessment	Proj submitted	Result	Distance
9.2	85	8	Pass	45.20
8	80	7	Pass	40.09
8.5	81	8	Pass	41.179
6	45	5	Fail	5.01
6.5	50	4	Fail	10.05
8.8	38	5	Fail	2.022

STEP 2: Sort the distances in ascending order & select first 3 neighbors

Instance	Euclidean Distance	Result
6	2.022	Fail
4	5.001	Fail
5	10.05	Fail

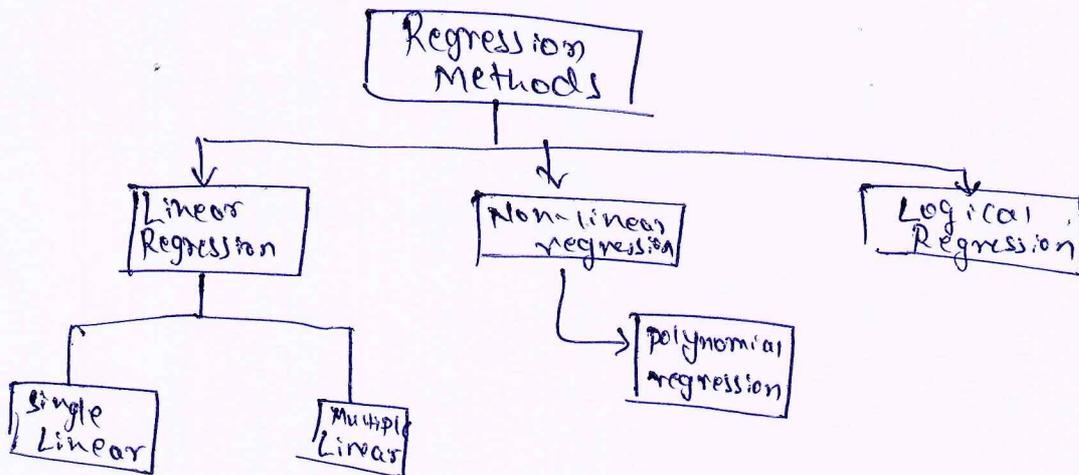
2M

Step 3: Predict the class of the test instance by majority voting.

The class for test instance is 'Fail'.

Q 5  
(b)

### Types of Regression Methods



Linear Regression: A line is fitted upon given data for finding the linear relationship between one independent and one dependent variable.

Multiple Regression: A line is fitted for finding linear relationship between two or more independent variables & one dependent variable.

Polynomial Regression: Model two or more independent variables and one dependent variable.

Logistic Regression: used for predicting categorical variables that involve one or more independent variables & one dependent variable.

## Limitations of Regression Method

3M

1. Outliers - outliers bias the outcome of regression model.
2. Number of cases - Ratio of independent & dependent variable should be at least 20:1
3. Missing data - missing data in training data can make the model unfit for the sampled data.
4. Multicollinearity: If explanatory variables are highly correlated, the regression is vulnerable to bias.

### Q 5(c) Structure of Decision Tree

A decision tree has root node, internal node, branches & leaf nodes.

• topmost node is root node.

• Internal nodes are the test nodes - Decision nodes  
• the branches are labelled as per the outcomes of the test condition

• The leaf node represent the labels or the outcomes of the decision path.

2M



— Root node



— Decision node



— Leaf node.

## Algorithm to Construct Decision tree Using ID3 SM

- ① Compute Entropy info for the whole training dataset based on the target attribute.

$$\text{Entropy\_Info}(P) = -(p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_n \log_2 p_n)$$

- ② Compute Entropy\_Info & Information\_Gain for each of the attribute in the training dataset.

$$\text{Information\_Gain}(A) = \text{Entropy\_Info}(T) - \text{Entropy\_Info}(T_A)$$

$$\text{Entropy\_Info}(T_A) = \sum_{i=1}^n \frac{|A_i|}{|T|} \times \text{Entropy\_Info}(A_i)$$

- ③ Choose the attribute for which entropy is minimum and therefore the gain is maximum as the best split attribute.
- ④ The best split attribute is placed as the root node.
- ⑤ The root node is branched into subtrees with each subtree as an outcome of the test condition of the root node attribute.
- ⑥ Recursively apply the same operation for the subset of the training set with the remaining attributes until a leaf node is derived or no more training instances are available in the subset.

## Q 6(a) Nearest Centroid Classifier Algorithm:

Inputs: Training dataset  $T$ , Distance metric  $d$ , Test instance  $t$

Output: Predicted class or category.

Q 6  
(b)

### i) Regression

- ① Used to predict one variable from another
- ② Dependent & Independent variables are defined
- ③  $y = ax + b$

### ii) Regression

- ① Shows statistical relationship
- ② Cannot prove cause; only association

### iii) Linearity

- ① Straight line
- ② Constant rate of change
- ③ eg. Linear regression

### Advantages of Decision tree.

- 1. Easy to model & interpret
- 2. Simple to understand
- 3. Input attributes can be discrete/continuous
- 4. Quick to train

Q 6  
(c)

### Correlation 2M

- ① Used to measure strength & direction of relationship.
- ② No dependent & independent variables
- ③ Number between  $-1$  &  $+1$

### Causation 2M

- ① Shows cause and effect relations
- ② Requires experiments, theory

### NonLinearity 2M

- ① Curve
- ② Varied Rate of Change
- ③ eg. Polynomial, exponential.

3M

- ① Compute the mean/Centroid of each class 2M
- ② Compute the distance between the test instance & mean/centroid of each class.
- ③ Predict the class by choosing the class with smaller distance.

X	Y	Class
3	1	A
5	2	A
4	3	A
7	6	B
6	7	B
8	5	B

- ① Compute mean/centroid of each class 2M

$$\text{Centroid of class 'A'} = \frac{(3+5+4, 1+2+3)}{3} = \frac{(12, 6)}{3} = (4, 2)$$

$$\text{Centroid of class 'B'} = \frac{(7+6+8, 6+7+5)}{3} = \frac{(21, 18)}{3} = (7, 6)$$

Test instance (6, 5)

- ② Calculate Euclidean distance between test instance (6, 5) & each of the centroid. 3M

$$\text{Euclidean dist} [(6, 5), (4, 2)] = \sqrt{(6-4)^2 + (5-2)^2} = \sqrt{13} = 3.6$$

$$\text{Euclidean dist} [(6, 5), (7, 6)] = \sqrt{(6-7)^2 + (5-6)^2} = \sqrt{2} = 1.414$$

The test instance has smaller distance to class 'B', hence test instance belongs to 'B'.

~~Q.11~~

## Disadvantages of Decision Trees

UM

- ① It is difficult to determine how deeply a decision tree can be grown, or when to stop growing it.
- ② A complex decision tree may also be over-fitting with the training data
- ③ Decision tree learning is not well suited for classifying multiple output classes.

## General Algorithm for Decision trees

1. Find the best attribute from the training dataset using an attribute selection measure & place it at the root of the tree.
2. Split the training dataset into subsets based on the outcomes of the test attribute & each subset in a branch contains the data instances with same value for selected test attribute.
3. Repeat steps 1 & 2 on each subset until we end up on leaf nodes in all branches of tree.
4. Splitting process is recursive until stopping criterion is reached.

Q7  
(a)

Frequency table of color

	Yes	No
Red	3	2
Yellow	2	3

Likelihood table of color 2M

	$P(\text{Yes})$	$P(\text{No})$
Red	$3/5$	$2/5$
Yellow	$2/5$	$3/5$

Frequency table of 'type'

	Yes	No
Sports	4	2
SUV	1	3

Likelihood Table of 'type'

	Yes	No
Sports	$4/5$	$2/5$
SUV	$1/5$	$3/5$

2M

Frequency table of 'origin'

	Yes	No
Domestic	2	3
Imported	3	2

Likelihood Table of 'origin'

	Yes	No
Domestic	$2/5$	$3/5$
Imported	$3/5$	$2/5$

2

Test data = (Red, SUV, Domestic) = x

$$\begin{aligned}
 P(\text{Yes} | x) &= P(\text{Red} | \text{Yes}) \times P(\text{SUV} | \text{Yes}) \times P(\text{Domestic} | \text{Yes}) \times P(\text{Yes}) \\
 &= \frac{3}{5} \times \frac{1}{5} \times \frac{2}{5} \times 1 \\
 &= 0.048
 \end{aligned}$$

1M

1M

$$\begin{aligned}
 P(\text{No} | x) &= P(\text{Red} | \text{No}) \times P(\text{SUV} | \text{No}) \times P(\text{Domestic} | \text{No}) \times P(\text{No}) \\
 &= \frac{2}{5} \times \frac{3}{5} \times \frac{3}{5} \times 1 \\
 &= 0.144
 \end{aligned}$$

2M

Since  $0.144 > 0.048$ , the test instance can be classified as 'No'

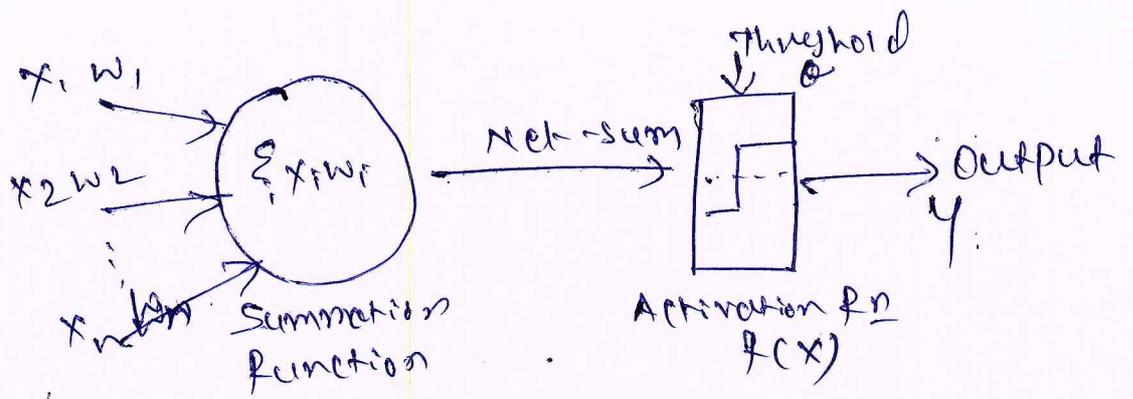
Q 7b)

Simple model of Artificial Neuron:

5M

- ① It receives weighted inputs from other neurons
- ② It operates with a threshold function/activation fn.

The received inputs are computed as a weighted sum which is given to the activation fn & if the sum exceeds the threshold value the neuron gets fired.



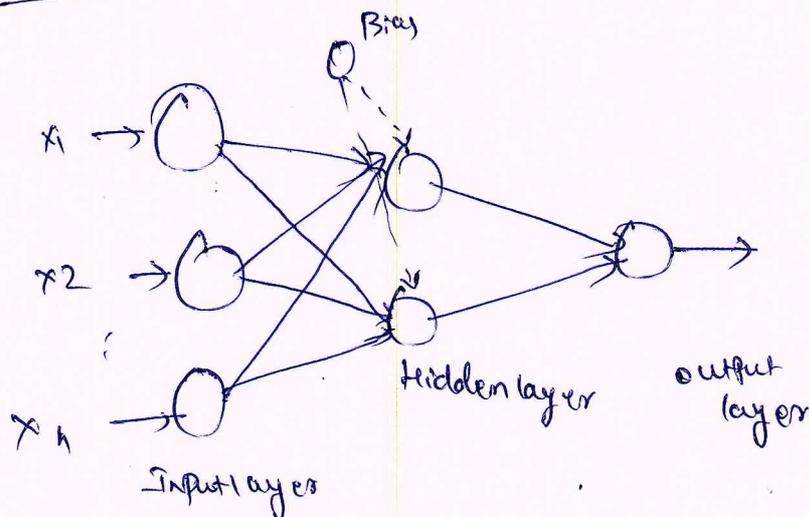
$$\text{Net-sum} = \sum_{i=1}^n x_i w_i$$

Activation function is a binary step function.

$$Y = \begin{cases} 1 & f(x) \geq 0 \\ 0 & f(x) < 0 \end{cases}$$

### Artificial Neural Network Structure

SM



- \* Artificial neural network (ANN) imitates brain which imitates some intelligence
- \* The input layer receives the information ( $x_1, x_2, \dots, x_n$ ) & passes it to nodes in the hidden layer.
- \* The edges connecting the nodes from the input layer to the hidden layer are associated with connection weights.
- \* If weighted sum of inputs to a neuron is above the threshold then the neuron fires.

28  
(a)

Bayes Theorem is used to select most probable hypothesis from data, considering both prior knowledge & posterior distributions.

$$P(\text{Hypothesis } h | \text{Evidence } E) = \frac{P(\text{Evidence } E | \text{Hypothesis } h) P(\text{Hypothesis } h)}{P(\text{Evidence } E)}$$

$P(\text{Hypothesis } h) \Rightarrow$  prior probability

$P(\text{Evidence } E | \text{Hypothesis } h) \Rightarrow$  likelihood probability of Evidence E after observing training data that the hypothesis h is correct.

$P(\text{Hypothesis } h | \text{Evidence } E) =$  posterior probability.

posterior probability  $\propto$  prior probability  $\times$  likelihood prob.

Maximum A Posteriori (MAP) hypothesis,  $h_{MAP}$  3M

Given a set of candidate hypotheses, the hypothesis which has the maximum value is maximum probable hypothesis

$$\begin{aligned} h_{MAP} &= \max_{h \in H} P(\text{Hypothesis } h | \text{Evidence } E) \\ &= \max_{h \in H} \frac{P(\text{Evidence } E | \text{Hypothesis } h) P(\text{Hypothesis } h)}{P(\text{Evidence } E)} \\ &= \max_{h \in H} P(\text{Evidence } E | \text{Hypothesis } h) \times P(\text{Hypothesis } h) \end{aligned}$$

Maximum Likelihood (ML) hypothesis  $h_{ML}$  3M

Given a set of candidate hypotheses, if every hypothesis is equally probable, only  $P(E|h)$  is used to find the most probable hypothesis.  
 $h_{ML} = \max_{h \in H} P(\text{Evidence } E | \text{Hypothesis } h)$

Activation functions

10M

① Identity Function

$$f(x) = x$$

$f(x)$  increases proportionally with value of  $x$ .

② Binary Step Function

$$f(x) = \begin{cases} 1 & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) < 0 \end{cases}$$

output is 0/1 based on the threshold value 0.

③ Bipolar Step Function

$$f(x) = \begin{cases} 1 & \text{if } (f(x) > 0) \\ -1 & \text{if } (f(x) < 0) \end{cases}$$

output is +1/-1 based on threshold value 0.

④ Sigmoidal Function

$$s(x) = \frac{1}{1 + e^{-x}}$$

output values are in range 0 & 1.

⑤ Bipolar Sigmoid Function

$$s(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$

outputs value between -1 & +1.

⑥ Ramp Function

$$f(x) = \begin{cases} 1 & \text{if } x > 1 \\ x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x < 0 \end{cases}$$

It is a linear transformation  $f_n$  whose upper & lower limits are fixed.

### 7. Tanh - Hyperbolic Tangens Function

$$\tanh(x) = \frac{2}{1+e^{-2x}} - 1$$

The output value ranges between  $-1$  &  $1$ .

### 8. ReLu - Rectified Linear unit

$$f(x) = \max(0, x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Outputs '0' for -ve input values & works like linear  $f_n$  for +ve input values.

### 9. Softmax

$$S(x_i) = \frac{e^{x_i}}{\sum_{j=0}^k e^{x_j}} \quad i = 0 \dots k$$

Calculates probability of each target class which ranges between  $0$  &  $1$ .

Q 9  
a)

Choose the object 2 & 5 with coordinate values.

4M

Cluster 1	Cluster 2
(4, 6)	(12, 4)
Centroid 1 (4, 6)	Centroid 2 (12, 4)

Iteration 1: Compare all the data points with the centroid and assign to the nearest ~~right~~ sample.

$$\text{Dist}(1, \text{centroid 1}) = \sqrt{(2-4)^2 + (11-6)^2} = \sqrt{8}$$

$$\text{Dist}(1, \text{centroid 2}) = \sqrt{(2-12)^2 + (11-4)^2} = \sqrt{100} = 10$$

Cluster Table after Iteration 1:

Cluster 1	Cluster 2
(4, 6)	(10, 4)
(2, 4)	(12, 4)
(6, 8)	.
Centroid 1 (4, 6)	Centroid 2 (11, 4)

4M

Iteration 2:

Cluster 1	Cluster 2
(4, 6)	(10, 4)
(2, 4)	(12, 4)
(6, 8)	.
Centroid 1 (4, 6)	Centroid (11, 4)

4M

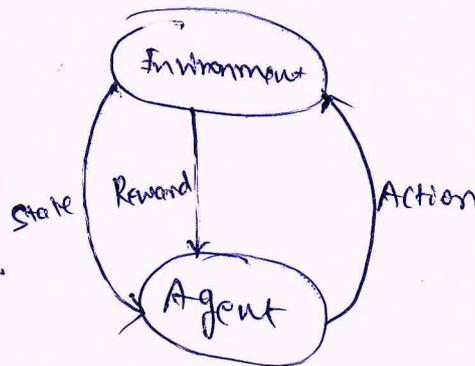
There is no change in the cluster table. Therefore, the K-Means algorithm terminates with two clusters with data points.

Q 9  
(b)

### Components of Reinforcement Learning

#### Environment & Agent

Environment describes the state/state variables as state.



An agent is an autonomous body that looks at the environment & takes an action.

#### States & Actions

The input for reinforced learning is called state & the output is action.

1M

Rewards: The system performance is measured by reward system. A reward is something that the environment produces to the agent. 2M

### Policies

The role of algorithms is called a policy in reinforcement learning and policy maps states to actions. 2M

Deterministic policy:

$$a = \pi(s)$$

Stochastic Policy:

$$\pi(a|s) = \Pr(a_t = a | s_t = s)$$

It is the probability of taking an action 'a' while in state 's'.

A better policy is the one that has better cumulative expected reward over other policies.

Q10

(a) Manhattan distance  $(x_i, x_j) = \sum_{k=1}^n |x_{ik} - x_{jk}|$

$$= |(0-5) + (3-8)|$$

$$= 10$$

2M

Chebyshev distance  $(x_i, x_j) = \max_k |x_{ik} - x_{jk}|$

$$= \max\{|0-5|, |3-8|\}$$

$$= \max\{5, 5\}$$

$$= 5$$

2M

Q 10  
(b)

## Mean-shift Clustering Algorithm

6M

- Step 1: Design a window
- Step 2: Place the window on a set of data points.
- Step 3: Compute the mean for all the points that come under the window.
- Step 4: Move the center of the window to the mean computed in step 3.

$$V_s = \frac{1}{K} \sum_{x_i \in S_k} (x_i - x)$$

$K \Rightarrow$  number of points

$S_k \Rightarrow$  Distance from data points  $x_i$  to centroid of kernel within radius of sphere.

Centroid update,  $x = x + V_s$ .

Steps: Repeat steps 3-4 for convergence. Once convergence is achieved, no further points can be accommodated.

Q 10  
(c)

## Characteristics of Reinforcement Learning 4M

- ① Sequential decision making - From start to goal step there is a sequence of decisions.
- ② Delayed Feedback: one must spend many moves to get final success/failure. Feedback in terms of reward is often delayed.
- ③ The agent actions are interdependent as any actions affect the subsequent actions.

Time related: All actions are associated with time stamps inherently as all actions are ordered as per the timeline inherently.

## Challenges of Reinforcement Learning 3M

- ① Reward design is a big challenge as in many games, determining the rewards & its value is challenging.
- ② Absence of a model is a challenge - simulation must be done to gather experience.
- ③ Partial observability of states - Many states are fully observable. In some scenarios partial observability exists.
- ④ Time consuming operations - more state spaces and possible actions may complicate the scenarios, resulting in more time consumption.
- ⑤ Complexity - more complexity involved in designing of reinforcement algorithms.

## Applications of Reinforcement Learning

① Industrial automation

② Resource management application to allocate resource

③ traffic light controller to reduce congestion of traffic.

④ Personalized recommendation system like news

⑤ Bidding for advertisement.

*[Signature]*

*[Signature]*

*[Signature]*  
HOD  
Computer Science & Engineering  
KLS Vishwanathrao Deshpande  
Institute of Technology, Navi Mumbai

*[Signature]*  
17/2/24