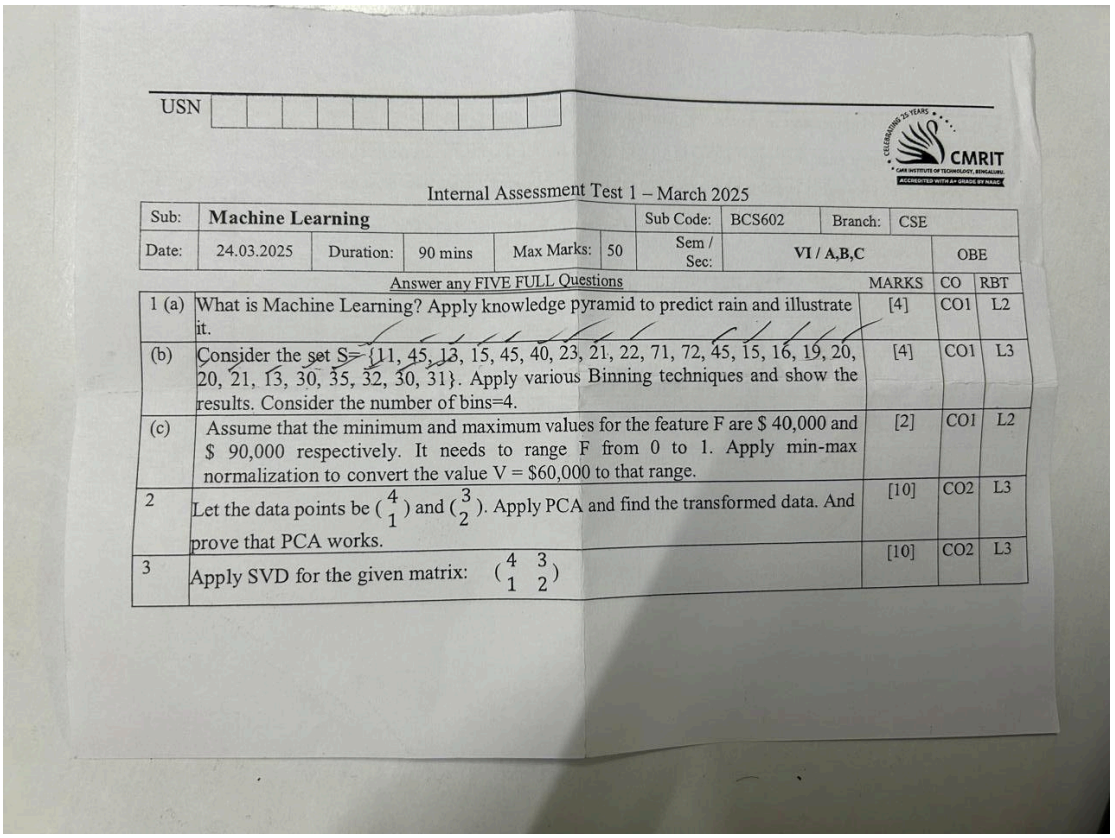
Sub:	Machine Learning					Sub Code:	BCS 602	Branch:	CSE
Date:	24/03/2025	Duration:	90 mins	Max Marks:	50	Sem / Sec:	6 A,B,C		



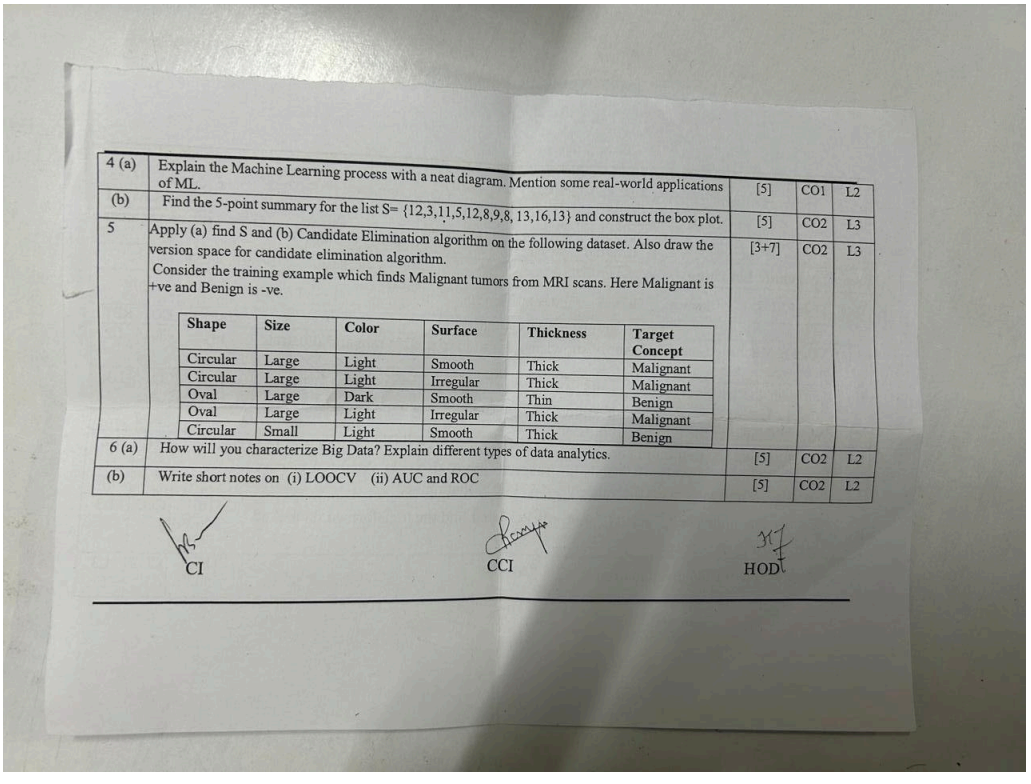
USN

Internal Assessment Test 1 – March 2025

Sub:	Machine Learning					Sub Code:	BCS602	Branch:	CSE	
Date:	24.03.2025	Duration:	90 mins	Max Marks:	50	Sem / Sec:	VI / A,B,C		OBE	

Answer any FIVE FULL Questions

		MARKS	CO	RBT
1 (a)	What is Machine Learning? Apply knowledge pyramid to predict rain and illustrate it.	[4]	CO1	L2
(b)	Consider the set S= {11, 45, 13, 15, 45, 40, 23, 21, 22, 71, 72, 45, 15, 16, 19, 20, 20, 21, 13, 30, 35, 32, 30, 31}. Apply various Binning techniques and show the results. Consider the number of bins=4.	[4]	CO1	L3
(c)	Assume that the minimum and maximum values for the feature F are $ 40,000 and $ 90,000 respectively. It needs to range F from 0 to 1. Apply min-max normalization to convert the value V = $60,000 to that range.	[2]	CO1	L2
2	Let the data points be $\binom{4}{1}$ and $\binom{3}{2}$. Apply PCA and find the transformed data. And prove that PCA works.	[10]	CO2	L3
3	Apply SVD for the given matrix: $\begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}$	[10]	CO2	L3



4 (a)	Explain the Machine Learning process with a neat diagram. Mention some real-world applications of ML.	[5]	CO1	L2
(b)	Find the 5-point summary for the list S= {12,3,11,5,12,8,9,8, 13,16,13} and construct the box plot.	[5]	CO2	L3
5	Apply (a) find S and (b) Candidate Elimination algorithm on the following dataset. Also draw the version space for candidate elimination algorithm. Consider the training example which finds Malignant tumors from MRI scans. Here Malignant is +ve and Benign is -ve.	[3+7]	CO2	L3

Shape	Size	Color	Surface	Thickness	Target Concept
Circular	Large	Light	Smooth	Thick	Malignant
Circular	Large	Light	Irregular	Thick	Malignant
Oval	Large	Dark	Smooth	Thin	Benign
Oval	Large	Light	Irregular	Thick	Malignant
Circular	Small	Light	Smooth	Thick	Benign

6 (a)	How will you characterize Big Data? Explain different types of data analytics.	[5]	CO2	L2
(b)	Write short notes on (i) LOOCV (ii) AUC and ROC	[5]	CO2	L2

CI CCI HOD

Answer 1 Machine Learning (ML) is a subset of artificial intelligence (AI) that enables systems to learn from data and make predictions or decisions without being explicitly programmed. It involves training models on datasets to recognize patterns and improve performance over time. Common ML applications include image recognition, speech processing, recommendation systems, and predictive analytics.
(Definition 2 marks , Diagram 1 marks and explanation knowledge pyramid with respect to rain 1 marks)

Applying the Knowledge Pyramid to Predict Rain

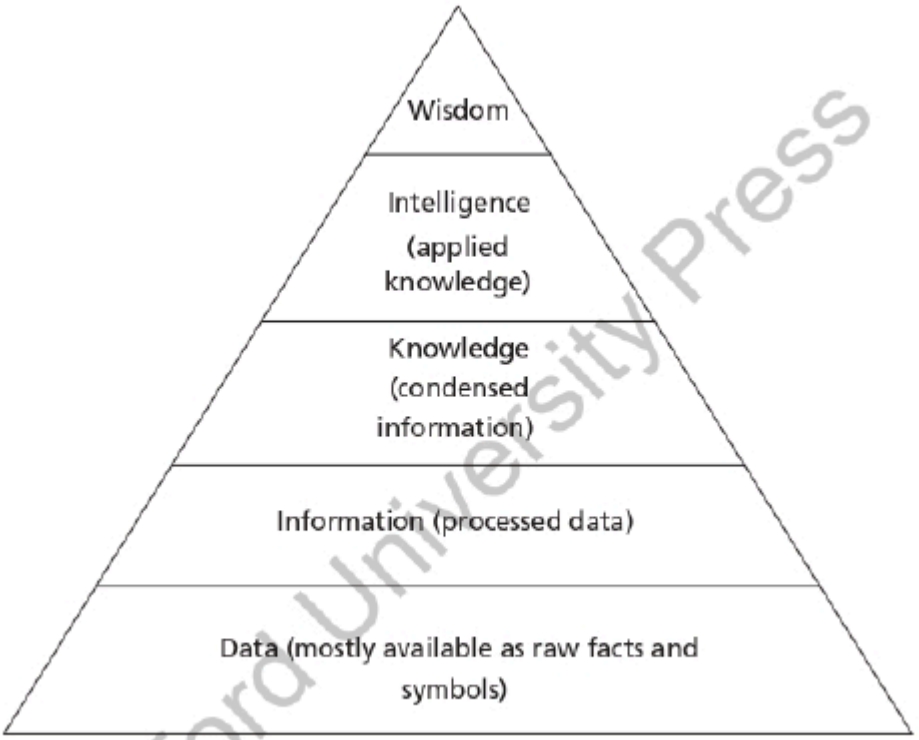


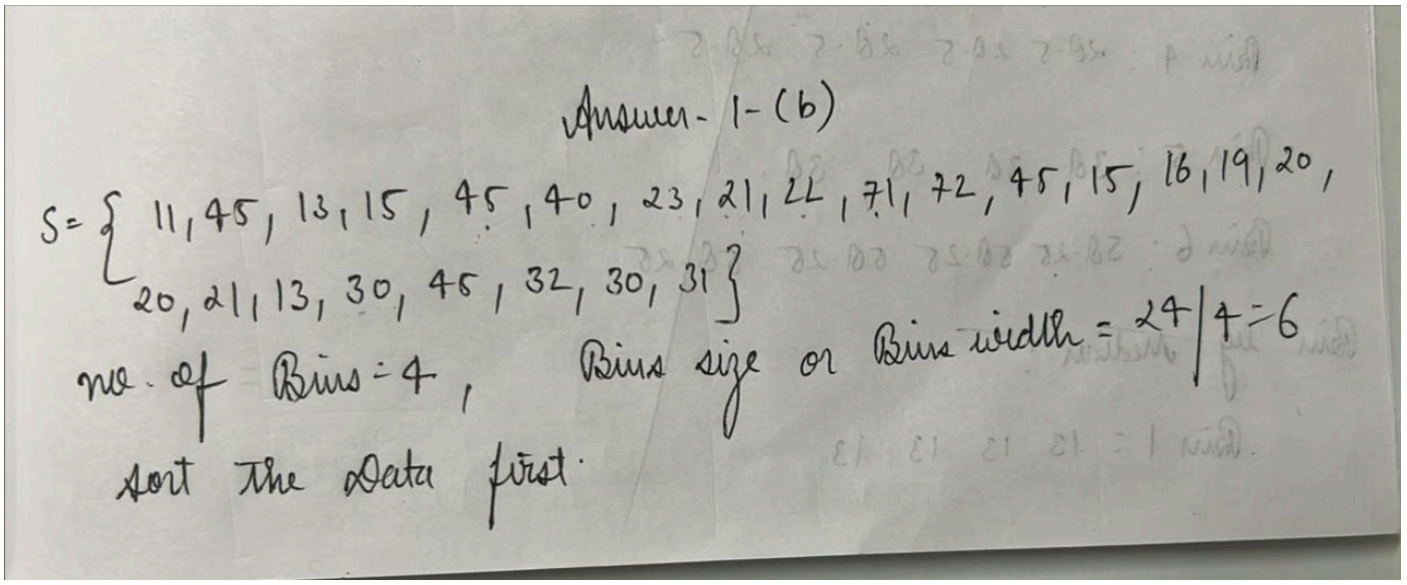
Figure 1.1: The Knowledge Pyramid

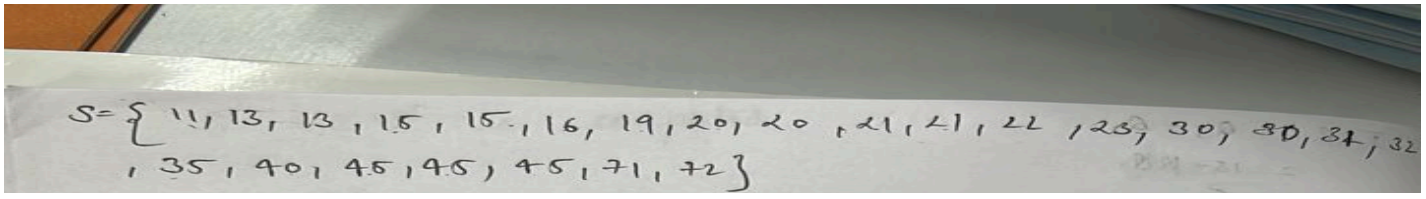
Now, let's apply this to predicting rain:

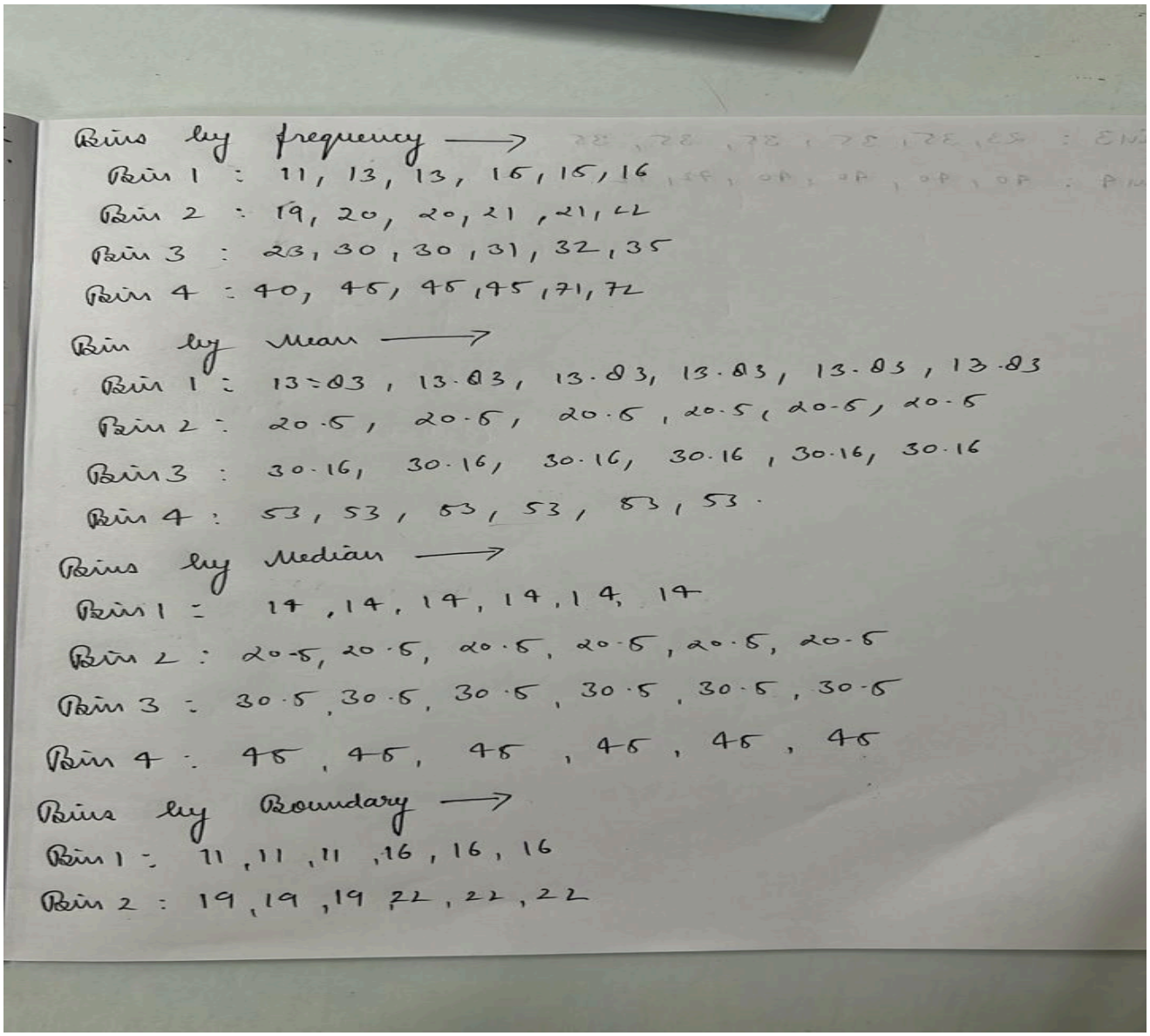
Level	Description	Example in Rain Prediction
Data	Raw facts and observations.	Temperature, humidity, wind speed, air pressure readings from sensors.
Information	Processed data with meaning.	Identifying patterns: High humidity and low pressure are often associated with rain.

Knowledge	Condensed information—understanding relationships.	ML models trained on past weather data recognize that a drop in pressure combined with high humidity increases the probability of rain.
Intelligence	Applying knowledge to make predictions.	A machine learning model predicts an **80% chance of rain** based on detected weather conditions.
Wisdom	Making informed decisions based on intelligence.	The system recommends carrying an umbrella or adjusting travel plans due to expected rainfall.

Answer 1-b **All four binning methods consist of 1 marks each**





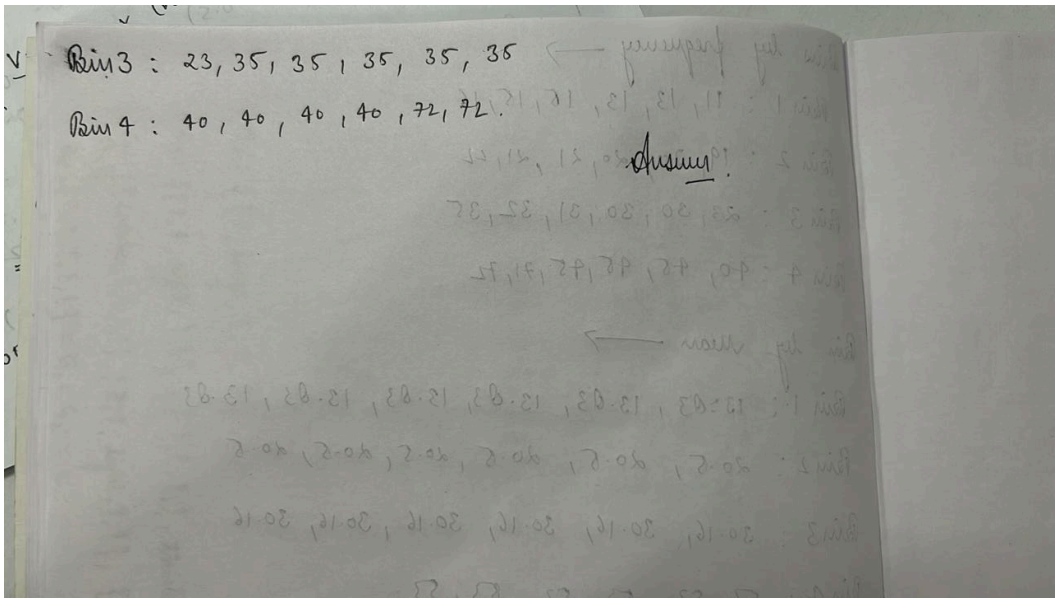


Bins by frequency →
Bin 1 : 11, 13, 13, 15, 15, 16
Bin 2 : 19, 20, 20, 21, 21, 22
Bin 3 : 23, 30, 30, 31, 32, 35
Bin 4 : 40, 45, 45, 45, 71, 72

Bin by Mean →
Bin 1 : 13.03, 13.03, 13.03, 13.03, 13.03, 13.03
Bin 2 : 20.5, 20.5, 20.5, 20.5, 20.5, 20.5
Bin 3 : 30.16, 30.16, 30.16, 30.16, 30.16, 30.16
Bin 4 : 53, 53, 53, 53, 53, 53.

Bins by Median →
Bin 1 : 14, 14, 14, 14, 14, 14
Bin 2 : 20.5, 20.5, 20.5, 20.5, 20.5, 20.5
Bin 3 : 30.5, 30.5, 30.5, 30.5, 30.5, 30.5
Bin 4 : 45, 45, 45, 45, 45, 45

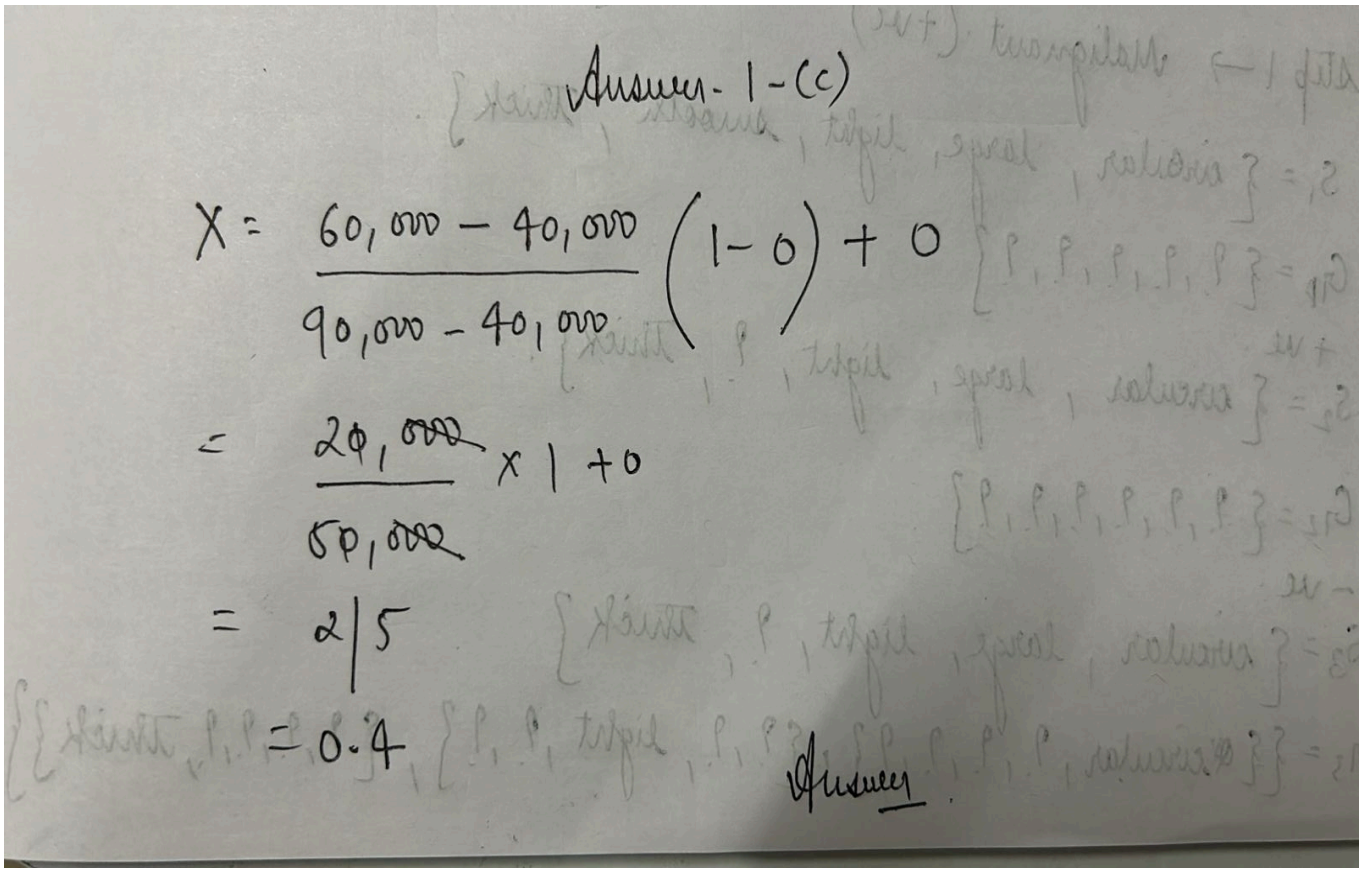
Bins by Boundary →
Bin 1 : 11, 11, 11, 16, 16, 16
Bin 2 : 19, 19, 19, 22, 22, 22



Bin 3 : 23, 35, 35, 35, 35, 35
Bin 4 : 40, 40, 40, 40, 72, 72.

Answer

Answer 1-c: Correct Answer 2 marks else 0



Answer 2: Eigen Values 2 marks, Eigen Vectors 2 marks, Feature Vector 2 marks , Transform Features 2 marks and Inverse 2 marks.

Answer. 2.

$\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$

step 1 $\longrightarrow$ $m = \begin{bmatrix} 4+3/2 \\ 1+2/2 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 1.5 \end{bmatrix}$

step 2 $\longrightarrow$ Mean vector : $(x - m)$

$$m_1 = \begin{bmatrix} 4 - 3.5 \\ 1 - 1.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}$$

$$m_2 = \begin{bmatrix} 3 - 3.5 \\ 2 - 1.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}$$

$\Longrightarrow$ $M = [m_1 \quad m_2] = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$

step 3 $\longrightarrow$ covariance Matrix :

$$C = \sum_{i=1}^{n} m_i m_i^T$$

$$C_1 = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 0.25 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix} \begin{bmatrix} -0.5 & 0.5 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.25 & -0.25 \\ -0.25 & 0.25 \end{bmatrix}$$

$C = C_1 + C_2$

$$= \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$$

step 4. Eigen Values :

$= \lambda^2 - T(A) + |A| = 0$

$\Rightarrow \lambda^2 - \lambda = 0$

$\Rightarrow \lambda(\lambda - 1) = 0$

$\lambda = 0, 1$

step 5 : sort eigen Values.

$\lambda = 1, 0$

step 6 : eigen vectors

$$(A - \lambda I)\bar{x} = 0$$

$$\begin{bmatrix} 0.5 - \lambda & -0.5 \\ -0.5 & 0.5 - \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

for $\lambda = 1$

$$\begin{bmatrix} -0.5 & -0.5 \\ -0.5 & -0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\Rightarrow -0.5x_1 - 0.5x_2 = 0$ — i

$-0.5x_1 - 0.5x_2 = 0$ — ii

Taking eqn — i

$-0.5x_1 = +0.5x_2$ or $0.5x_1 = -0.5x_2$

Divide by -0.5 or Divide by -0.5

$\frac{x_1}{1} = \frac{x_2}{-1}$ or $\frac{x_1}{-1} = \frac{x_2}{1}$

$e_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ or $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ both are correct.

for $\lambda = 0$

$$\begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$0.5x_1 - 0.5x_2 = 0$ — i

$-0.5x_1 + 0.5x_2 = 0$ — ii

Taking eqn ii

$0.5x_2 = 0.5x_1$

Divide by 0.5

$\frac{x_2}{1} = \frac{x_1}{1}$

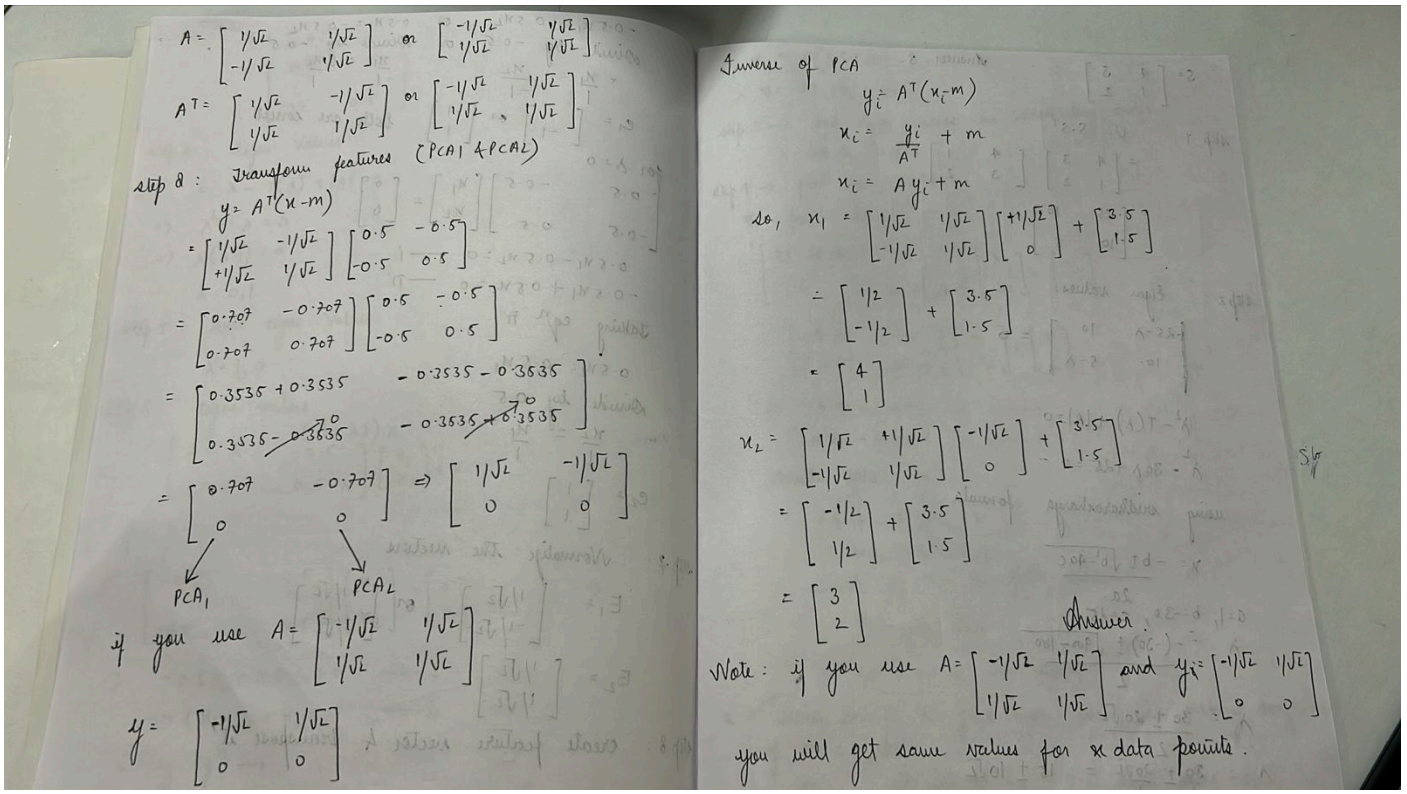
$e_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

step 7 : Normalize the vectors

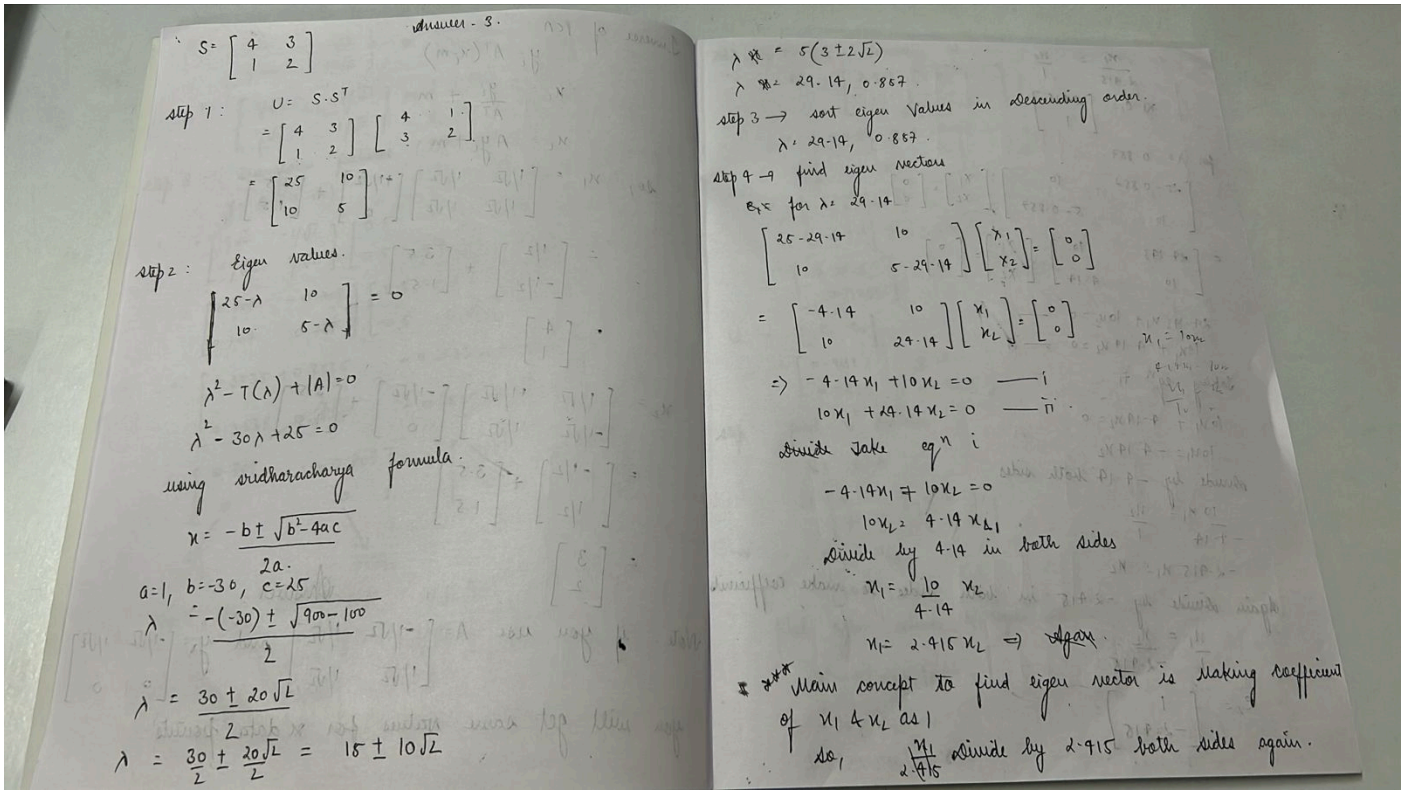
$E_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ or $\begin{bmatrix} +1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

$E_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

step 8 : create feature vector & transpose it



Answer 3: Eigen Values for both matrixes 2 marks, Eigen Vectors for both matrixes 2 marks, U Matrix 2 marks , V Transpose 2 mark and Sigma∑ 2 marks



$$\frac{x_1}{2.915} = \frac{x_2}{1}$$

$x_1 e_1 = \begin{bmatrix} 2.915 \\ 1 \end{bmatrix}$

for $\lambda = 0.857$

$$\begin{bmatrix} 25 - 0.857 & 10 \\ 10 & 5 - 0.857 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 24.143 & 10 \\ 10 & 4.14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$24.143 x_1 + 10 x_2 = 0$ — i

$10 x_1 + 4.14 x_2 = 0$ — ii

Taking eqn ii.

$10 x_1 + 4.14 x_2 = 0$

$10 x_1 = -4.14 x_2$

divide by -4.14 both sides

$$\frac{10 x_1}{-4.14} = \frac{x_2}{1}$$

$-2.915 x_1 = x_2$

again divide by -2.415 in both sides to make coefficient

$$\frac{x_1}{1} = \frac{x_2}{-2.915}$$

$e_2 = \begin{bmatrix} 1 \\ -2.915 \end{bmatrix}$

step step 5- Ex Normalize it

$$e_1 = \begin{bmatrix} 2.915 / \sqrt{2.915^2 + 1^2} \\ 1 / \sqrt{2.915^2 + 1^2} \end{bmatrix}$$

$$= \begin{bmatrix} 0.9239 \\ 0.3825 \end{bmatrix}$$

$$e_2 = \begin{bmatrix} 1 / \sqrt{1^2 + (2.915)^2} \\ -2.915 / \sqrt{1^2 + (-2.915)^2} \end{bmatrix}$$

$$= \begin{bmatrix} 0.3025 \\ -0.9239 \end{bmatrix}$$

$U = \begin{bmatrix} e_1 & e_2 \end{bmatrix}$

$$= \begin{bmatrix} 0.9239 & 0.3825 \\ 0.3825 & -0.9239 \end{bmatrix}$$

step: find V^T Matrix

$V = A^T A$

$$= \begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 17 & 14 \\ 14 & 13 \end{bmatrix}$$

find eigen values & eigen vectors

$\lambda^2 - T(\lambda) + |A| = 0$

$\lambda^2 - 30\lambda + 25 = 0$

$\lambda = 29.14, 0.857$

for $\lambda = 29.14$

$$\begin{bmatrix} 17 - 29.14 & 14 \\ 14 & 13 - 29.14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -12.14 & 14 \\ 14 & 16.14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$-12.14 x_1 + 14 x_2 = 0$ — i

$14 x_1 + 16.14 x_2 = 0$ — ii

Taking eqn — i

$14 x_2 = 12.14 x_1$

$$\frac{14 x_2}{12.14} = \frac{x_1}{1}$$

$1.153 x_2 = x_1$

$$\frac{x_2}{1} = \frac{x_1}{1.153}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.153 \\ 1 \end{bmatrix}$$

for $\lambda = 0.857$

$$\begin{bmatrix} 17 - 0.857 & 14 \\ 14 & 13 - 0.857 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 16.143 & 14 \\ 14 & 12.14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$16.14 x_1 + 14 x_2 = 0$ — i

$14 x_1 + 12.14 x_2 = 0$ — ii

Taking eqn — ii

$14 x_1 = -12.14 x_2$

$$\frac{14 x_1}{-12.14} = \frac{x_2}{1}$$

$-1.153 x_1 = x_2$

$$\frac{x_1}{1} = \frac{x_2}{-1.153}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1.153 \end{bmatrix}$$

Normalize it

$$e_1 = \begin{bmatrix} 1.153 / \sqrt{1^2 + 1.153^2} \\ 1 / \sqrt{1^2 + 1.153^2} \end{bmatrix}$$

$$= \begin{bmatrix} 0.755 \\ 0.655 \end{bmatrix}$$

$$e_2 = \begin{bmatrix} 0.655 \\ -0.755 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.755 & 0.655 \\ 0.655 & -0.755 \end{bmatrix}$$

$$V^T = \begin{bmatrix} 0.755 & 0.655 \\ 0.655 & -0.755 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sqrt{29.14} & 0 \\ 0 & \sqrt{0.857} \end{bmatrix} = \begin{bmatrix} 5.39 & 0 \\ 0 & 0.925 \end{bmatrix}$$

$S = U \Sigma V^T$ is completed

Answer

Answer 4-a Machine Learning Process: Diagram 2 marks and Explanation 1 mark. Applications : 2 marks:

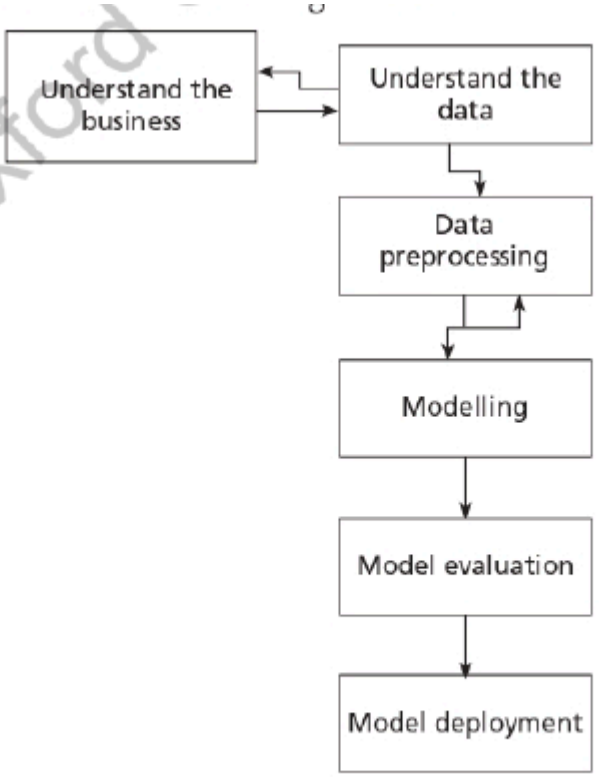
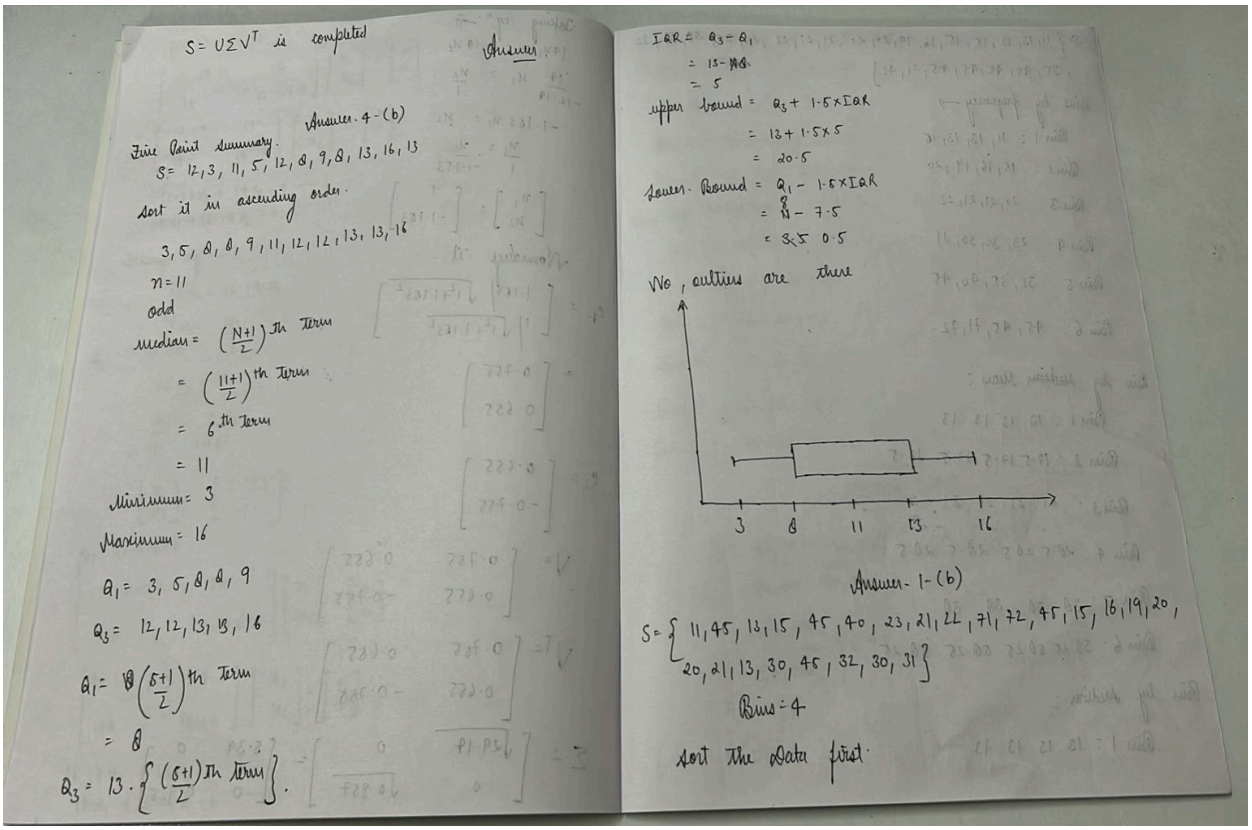


Figure 1.11: A Machine Learning/Data Mining Process

1. Sentiment analysis – This is an application of natural language processing (NLP) where the words of documents are converted to sentiments like happy, sad, and angry which are captured by emoticons effectively. For movie reviews or product reviews, five stars or one star are automatically attached using sentiment analysis programs.
2. Recommendation systems – These are systems that make personalized purchases possible. For example, Amazon recommends users to find related books or books bought by people who have the same taste like you, and Netflix suggests shows or related movies of your taste. The recommendation systems are based on machine learning.
3. Voice assistants – Products like Amazon Alexa, Microsoft Cortana, Apple Siri, and Google Assistant are all examples of voice assistants. They take speech commands and perform tasks. These chatbots are the result of machine learning technologies.
4. Technologies like Google Maps and those used by Uber are all examples of machine learning which offer to locate and navigate shortest paths to reduce time.

Answer 4-b Five Point Summary, 1 Marks, Outlier Identification 2 Marks and Boxplot: 2 marks Calculation mistake cut 1 marks.



$S = U\Sigma V^T$ is completed

Answer

Answer. 4-(b)

Five Point summary.

S = 12, 3, 11, 5, 12, 8, 9, 8, 13, 16, 13

sort it in ascending order.

3, 5, 8, 8, 9, 11, 12, 12, 13, 13, 16

n = 11

odd

median = $\left(\frac{N+1}{2}\right)$ th Term

= $\left(\frac{11+1}{2}\right)$ th Term

= 6th Term

= 11

Minimum = 3

Maximum = 16

Q_1 = 3, 5, 8, 8, 9

Q_3 = 12, 12, 13, 13, 16

$Q_1 = 8\left(\frac{5+1}{2}\right)$ th Term

= 8

$Q_3 = 13 \cdot \left\{\left(\frac{5+1}{2}\right)$ th Term$\right\}$.

IQR = $Q_3 - Q_1$

= 13 - 8

= 5

upper bound = $Q_3 + 1.5 \times IQR$

= 13 + 1.5 × 5

= 20.5

lower Bound = $Q_1 - 1.5 \times IQR$

= 8 - 7.5

= 3.5 0.5

No, outliers are there

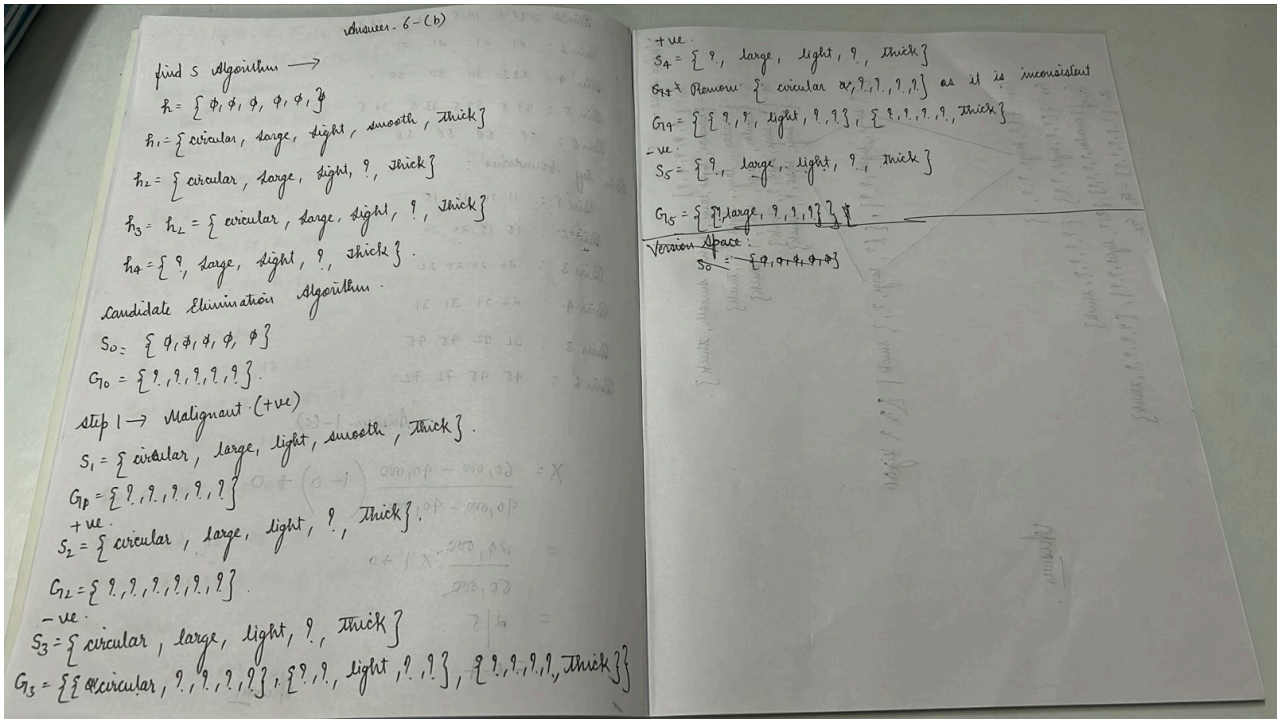
Answer. 1-(b)

S = { 11, 45, 13, 15, 45, 40, 23, 21, 22, 71, 72, 45, 15, 16, 19, 20, 20, 21, 13, 30, 45, 32, 30, 31 }

Bins = 4

sort the data first.

Answer 5-a Find S Algorithm: 3 marks and if any hypothesis is wrong cut 1 marks. Candidate Elimination: 5 marks and version space 2 marks



Answer. 6-(b)

find S Algorithm →

h = { ϕ, ϕ, ϕ, ϕ, ϕ, ϕ }

h_1 = { circular, large, light, smooth, thick }

h_2 = { circular, large, light, ?, thick }

$h_3 = h_4$ = { circular, large, light, ?, thick }

h_4 = { ?, large, light, ?, thick }.

Candidate Elimination Algorithm.

S_0 = { ϕ, ϕ, ϕ, ϕ, ϕ, ϕ }

G_0 = { ?, ?, ?, ?, ?, ? }.

step 1 → Malignant (+ve)

S_1 = { circular, large, light, smooth, thick }.

G_1 = { ?, ?, ?, ?, ?, ? }

+ve.

S_2 = { circular, large, light, ?, thick }.

G_2 = { ?, ?, ?, ?, ?, ? }.

-ve.

S_3 = { circular, large, light, ?, thick }

G_3 = { { circular, ?, ?, ?, ? }, { ?, ?, light, ?, ? }, { ?, ?, ?, ?, thick } }

+ve.

S_4 = { ?, large, light, ?, thick }

G_4 = Remove { circular, ?, ?, ?, ? } as it is inconsistent

G_4 = { { ?, ?, light, ?, ? }, { ?, ?, ?, ?, thick } }

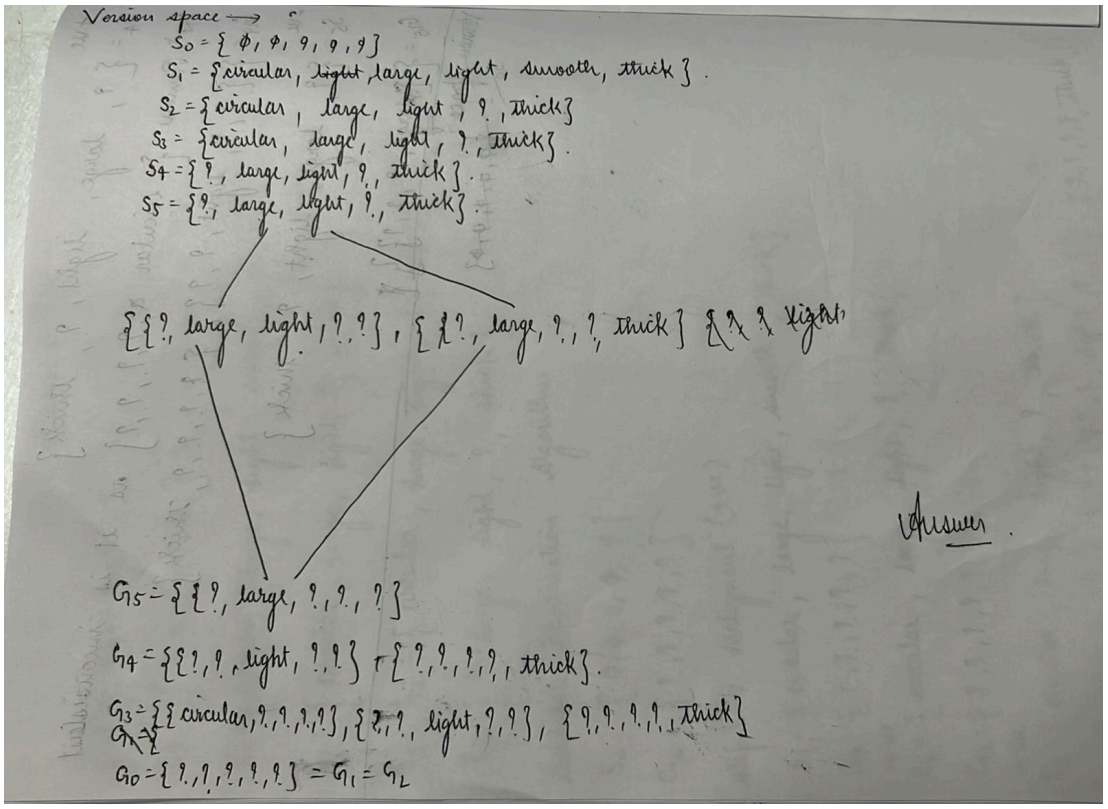
-ve.

S_5 = { ?, large, light, ?, thick }

G_5 = { { ?, large, ?, ?, ? } }

Version Space:

S_0 = { ?, ?, ?, ?, ? }



Answer 6-a: All types of Big Data : 2 marks and Different types 3 marks

1. Volume – Since there is a reduction in the cost of storing devices, there has been a tremendous growth of data. Small traditional data is measured in terms of gigabytes (GB) and terabytes (TB), but Big Data is measured in terms of petabytes (PB) and exabytes (EB). One exabyte is 1 million terabytes.
2. Velocity – The fast arrival speed of data and its increase in data volume is noted as velocity. The availability of IoT devices and Internet power ensures that the data is arriving at a faster rate. Velocity helps to understand the relative growth of big data and its accessibility by users, systems and applications.
3. Variety – The variety of Big Data includes:
   - Form – There are many forms of data. Data types range from text, graph, audio, video, to maps. There can be composite data too, where one media can have many other sources of data, for example, a video can have an audio song.
   - Function – These are data from various sources like human conversations, transaction records, and old archive data.
   - Source of data – This is the third aspect of variety. There are many sources of data. Broadly, the data source can be classified as open/public data, social media data and multimodal data. These are discussed in Section 2.3.1 of this chapter.

   Some of the other forms of Vs that are often quoted in the literature as characteristics of Big data are:
4. Veracity of data – Veracity of data deals with aspects like conformity to the facts, truthfulness, believability, and confidence in data. There may be many sources of error such as technical errors, typographical errors, and human errors. So, veracity is one of the most important aspects of data.
5. Validity – Validity is the accuracy of the data for taking decisions or for any other goals that are needed by the given problem.
6. Value – Value is the characteristic of big data that indicates the value of the information that is extracted from the data and its influence on the decisions that are taken based on it.

There are four types of data analytics:

1. Descriptive analytics
2. Diagnostic analytics
3. Predictive analytics
4. Prescriptive analytics

Answer 6-b LOOCV: 1 mark figure and 1 mark explanation

AUC Curve: 1 mark for correct explanation and ROC Curve Explanation: 1 marks figure and 1 mark for explanation/

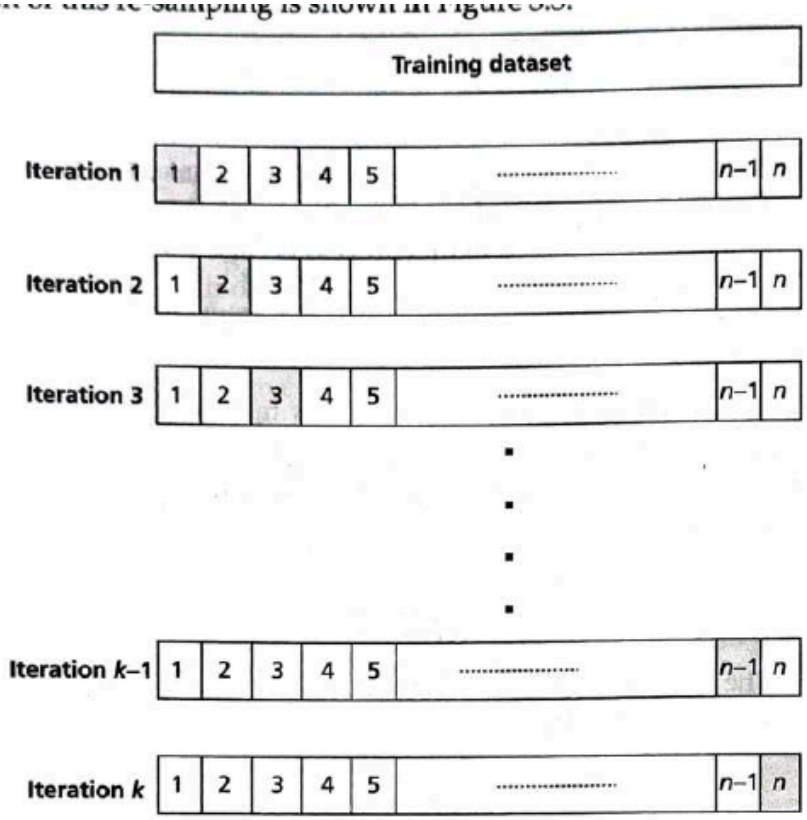


Figure 3.5: Illustration of Leave-One-Out Cross-Validation

Visual Classifier Performance Receiver Operating Characteristic (ROC) curve and Precision-Recall curves indicate the performance of classifiers visually. ROC curves are visual means of checking the accuracy and comparison of classifiers. ROC is a plot of sensitivity (True Positive Rate) and the 1-specificity (False Positive Rate) for a given model.

A sample ROC curve is shown in Figure 3.6, where results of five classifiers are given. A is the ROC of an average classifier. The ideal classifier is E where the area under curve is 1.0. Theoretically, it can range from 0.9 to 1. The rest of the classifiers B, C, D are categorized based on area under curve as good, better and still better based on the area under curve values.

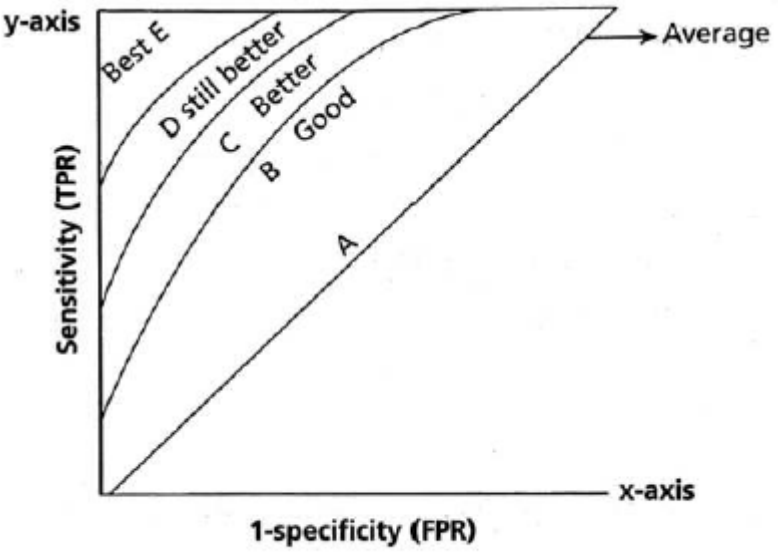


Figure 3.6: A Sample ROC Curve

Instead of predicting the label of a classifier, one can predict the probabilities of the model. Probabilities allow some better evaluation by functions that are called scoring functions or scoring rules. The area under curve (AUC) is one such score that can be used for classifier model evaluation. The integrated AUC is a measure of the model across threshold values.

AUC indicates the accuracy of the model. A model is perfect if it has area under ROC curve as one. The AUC score 0 of a model indicates the wrong model. The approximate area under precision-recall curve also indicates the power of the model across thresholds.