USN | | | | | | | | | | |

Internal Assessment Test 1 – March 2025

Sub:	**Machine Learning**					Sub Code:	**BCS602**	Branch:	**AInDS**
Date:		Duration:	**90 minutes**	Max Marks:	**50**	Sem	**VI**		**OBE**

		Answer any FIVE Questions	**MARKS**	**CO**	**RBT**
1	a	Solve by using Find-S Algorithm S.No. | Fever | Cough | Throat_Pain | Body_Pain | Covid-19 I1. | N | Y | Y | Y | Positive I2. | Y | Y | Y | Y | Positive I3. | Y | N | Y | Y | Positive I4. | N | Y | N | N | Negative I5. | Y | Y | Y | N | Positive I6. | N | N | N | Y | Negative I7. | N | N | N | N | Negative	[5]	CO2	L3
	b	What are measures of central tendency explain with example	[5]	CO1	L1
2		How you will design a Learning System? Explain with example	[10]	CO2	L2
3		For a given dataset S={5,10,15,20,25,30} of marks. Find 5point summary and IQR value and draw boxplot	[10]	CO1	L2
4	a	Explain different types of Machine Learning and different Applications in the field of machine learning	[5]	CO1	L1
	b	Explain Descriptive statistics with respect to Data Types	[5]	CO1	L1
5		Apply PCA for following matrix and prove that it works M= 4 3 1 2	[10]	CO2	L1
6		Find Covariance and Correlation Coefficients for following two sets of data X: 1 2 6 12 Y: 8 12 18 22	[10]	CO2	L2

CI **CCI** **HOD**

Q 1 a) Solve by using Find-S Algorithm

S.No.	Fever	Cough	Throat_Pain	Body_Pain	Covid-19
I1.	N	Y	Y	Y	Positive
I2.	Y	Y	Y	Y	Positive
I3.	Y	N	Y	Y	Positive
I4.	N	Y	N	N	Negative
I5.	Y	Y	Y	N	Positive
I6.	N	N	N	Y	Negative
I7.	N	N	N	N	Negative

Answer:

Q 1 b) What are measures of central tendency explain with example

Answer:

One cannot remember all the data. Therefore, a condensation or summary of the data is necessary. This makes the data analysis easy and simple. One such summary is called central tendency. Thus, central tendency can explain the characteristics of data and that further helps in comparison. Mass data have tendency to concentrate at certain values, normally in the central location. It is called measure of central tendency (or averages). This represents the first order of measures. Popular measures are mean, median and mode.

1. **Mean** – Arithmetic average (or mean) is a measure of central tendency that represents the 'center' of the dataset. This is the commonest measure used in our daily conversation such as average income or average traffic. It can be found by adding all the data and dividing the sum by the number of observations. Mathematically, the *average* of all the values in the sample (population) is denoted as $\bar{x}$. Let $x_1, x_2, \cdots, x_N$ be a set of 'N' values or observations, then the arithmetic mean is given as:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (2.3)$$

For example, the mean of the three numbers 10, 20, and 30 is $\frac{10 + 20 + 30}{3} = \frac{60}{3} = 20$

2. **Median** – The middle value in the distribution is called median. If the total number of items in the distribution is odd, then the middle value is called median. If the numbers are even, then the average value of two items in the centre is the median. It can be observed that the median is the value where x_i is divided into two equal halves, with half of the values being lower than the median and half higher than the median. A median class is that class where $(N/2)^{th}$ item is present.

In the continuous case, the median is given by the formula:

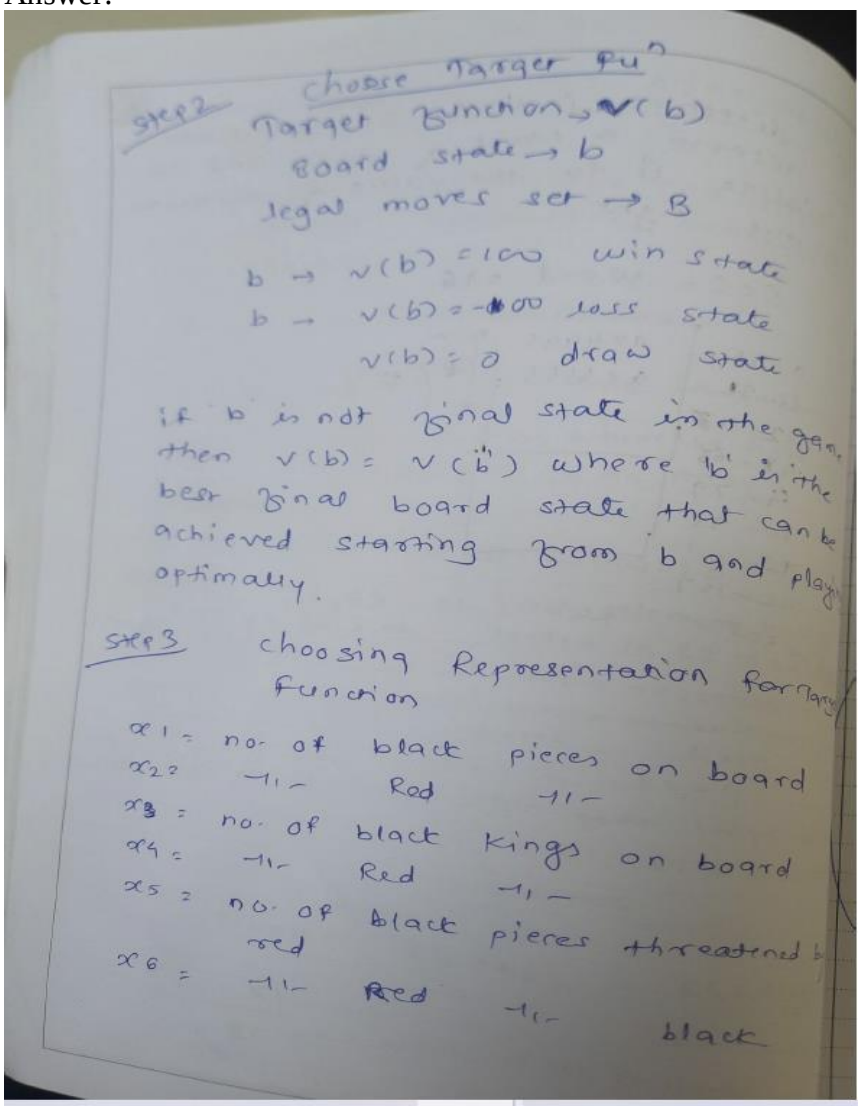
$$\text{Median} = L_1 + \frac{\frac{N}{2} - cf}{f} \times i \qquad (2.7)$$

Median class is that class where $N/2^{th}$ item is present. Here, i is the class interval of the median class and L_1 is the lower limit of median class, f is the frequency of the median class, and cf is the cumulative frequency of all classes preceding median.

3. **Mode** – Mode is the value that occurs more frequently in the dataset. In other words, the value that has the highest frequency is called mode. Mode is only for discrete data and is not applicable for continuous data as there are no repeated values in continuous data.

The procedure for finding the mode is to calculate the frequencies for all the values in the data, and mode is the value (or values) with the highest frequency. Normally, the dataset is classified as unimodal, bimodal and trimodal with modes 1, 2 and 3, respectively.

Q 2) How you will design a Learning System? Explain with example
Answer:



$\hat{V}(b) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5 + w_6 x_6$

where w_1 to w_6 are coefficients or weights chosen by learning algorithm

w_0 = additive constant to board value.

Step 4 choosing function Approximation algorithm

① Estimating training values
② Adjusting the weights

In this step we need a set of training examples.

It is a ordered pair $(b, V(b))$

eg. following training example describes a board state b in which black has won the game for which target fun^n value $V_{train}(b) = +100$

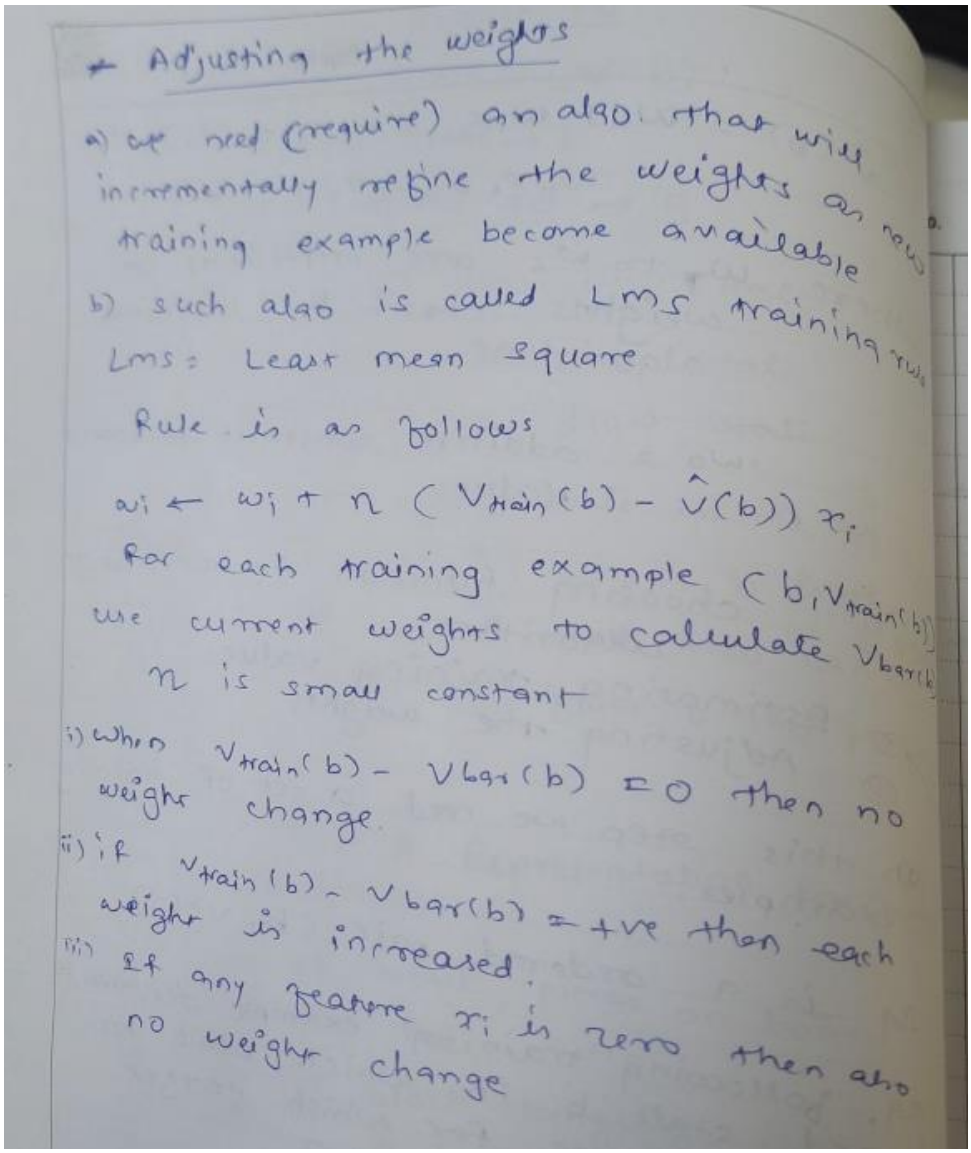
It is associated with values assigned to intermediate states

Rule is as follows

$V_{train}(b) \leftarrow \hat{V}(\text{Successor}(b))$ (intermediate state)

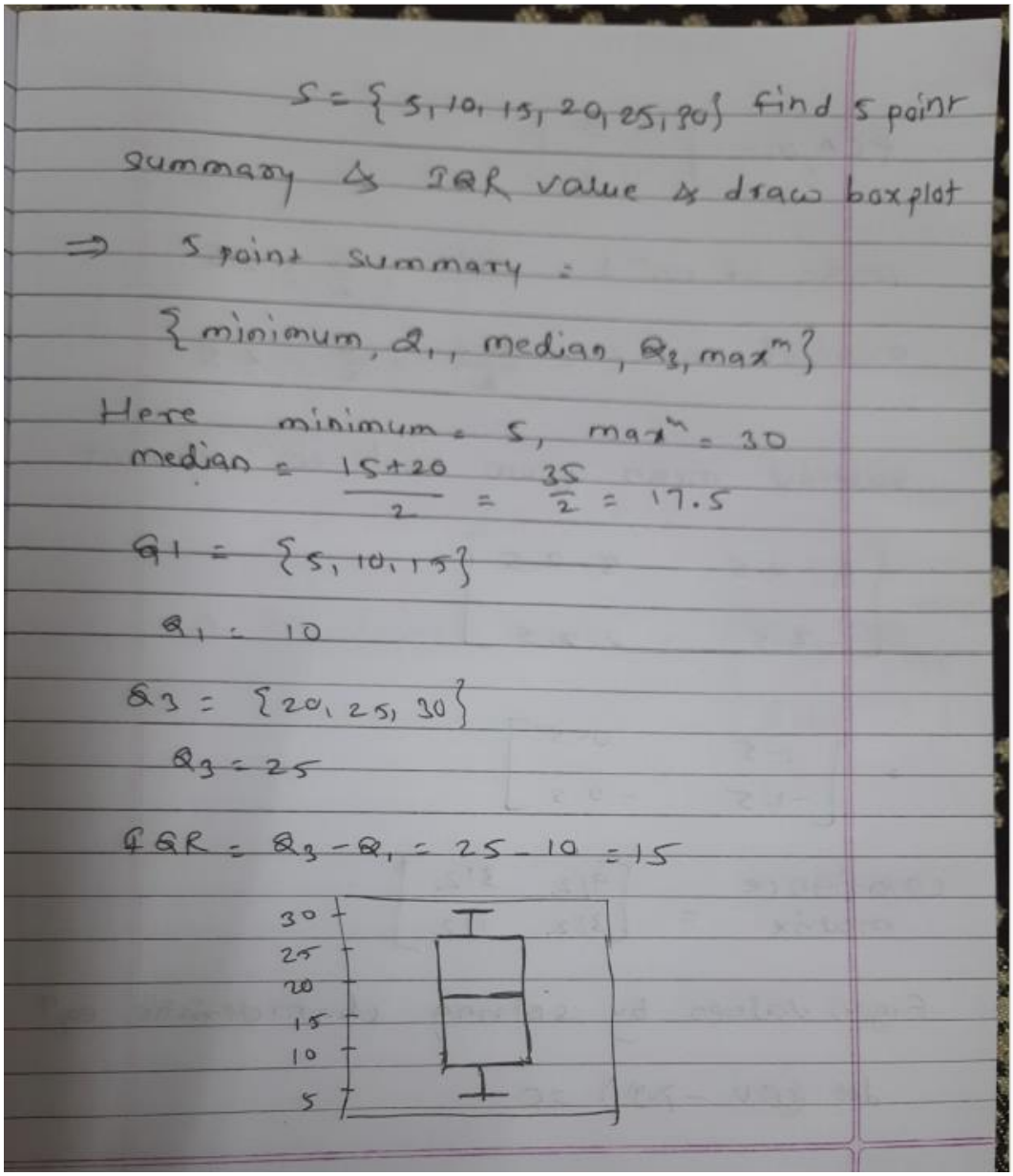
where $\hat{V}$ = current approximation to

Successor(b) = next board state



Q 3) For a given dataset S={5,10,15,20,25,30} of marks. Find 5point summary and IQR value and draw boxplot

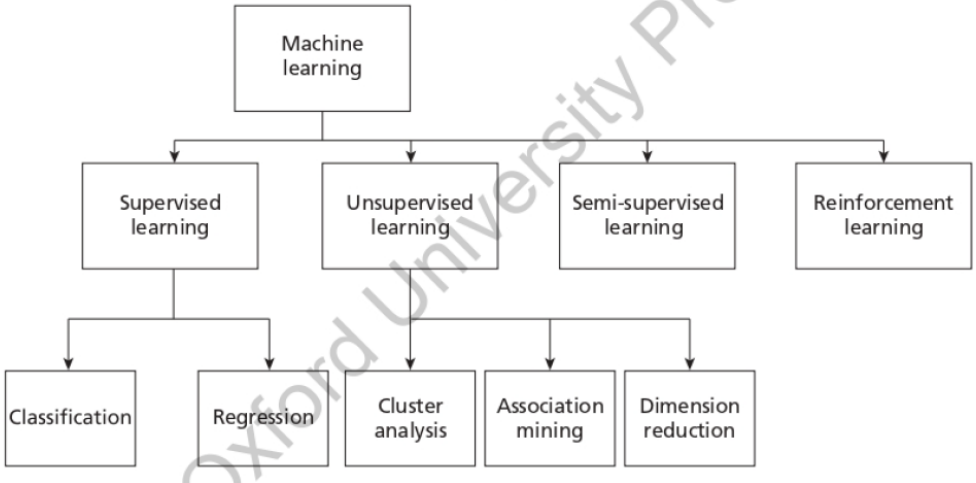
Answer:



Q 4 a) Explain different types of Machine Learning and different Applications in the field of machine learning

Answer:

What does the word 'learn' mean? Learning, like adaptation, occurs as the result of interaction of the program with its environment. It can be compared with the interaction between a teacher and a student. There are four types of machine learning as shown in Figure 1.5.



MACHINE LEARNING APPLICATIONS

ıine Learning technologies are used widely now in different domains. Machine learning appli-ıs are everywhere! One encounters many machine learning applications in the day-to-day ;ome applications are listed below:

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.

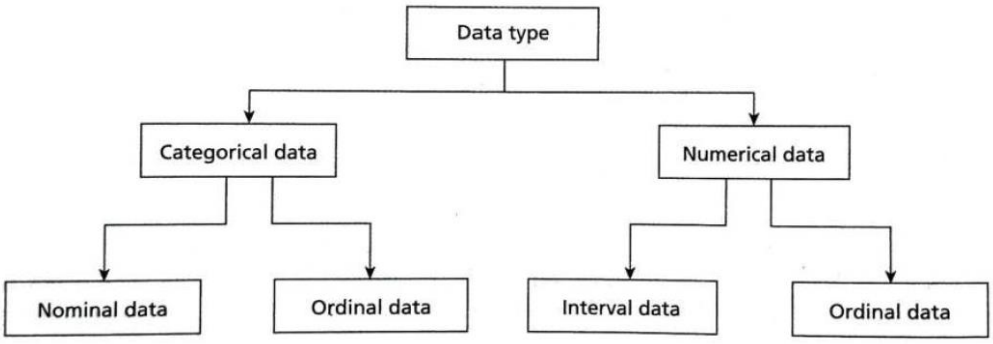
Q 4 b) Explain Descriptive statistics with respect to Data Types

Answer:

Descriptive statistics is a branch of statistics that does dataset summarization. It is used to summarize and describe data. Descriptive statistics are just descriptive and do not go beyond that. In other words, descriptive statistics do not bother too much about machine learning algorithms and its functioning.

Data visualization is a branch of study that is useful for investigating the given data. Mainly, the plots are useful to explain and present data to customers.

Descriptive analytics and data visualization techniques help to understand the nature of the data, which further helps to determine the kinds of machine learning or data mining tasks that can be applied to the data. This step is often known as Exploratory Data Analysis (EDA). The focus of EDA is to understand the given data and to prepare it for machine learning algorithms. EDA includes descriptive statistics and data visualization.



Categorical or Qualitative Data The categorical data can be divided into two types. They are nominal type and ordinal type.

- Nominal Data – In Table 2.2, patient ID is nominal data. Nominal data are symbols and cannot be processed like a number. For example, the average of a patient ID does not make any statistical sense. Nominal data type provides only information but has no ordering among data. Only operations like (=, ≠) are meaningful for these data. For example, the patient ID can be checked for equality and nothing else.
- Ordinal Data – It provides enough information and has natural order. For example, Fever = {Low, Medium, High} is an ordinal data. Certainly, low is less than medium and medium is less than high, irrespective of the value. Any transformation can be applied to these data to get a new value.

Numeric or Qualitative Data It can be divided into two categories. They are interval type and ratio type.

- Interval Data – Interval data is a numeric data for which the differences between values are meaningful. For example, there is a difference between 30 degree and 40 degree. Only the permissible operations are + and –.
- Ratio Data – For ratio data, both differences and ratio are meaningful. The difference between the ratio and interval data is the position of zero in the scale. For example, take the Centigrade-Fahrenheit conversion. The zeroes of both scales do not match. Hence, these are interval data.

Another way of classifying the data is to classify it as:

1. Discrete value data
2. Continuous data

Discrete Data This kind of data is recorded as integers. For example, the responses of the survey can be discrete data. Employee identification number such as 10001 is discrete data.

Continuous Data It can be fitted into a range and includes decimal point. For example, age is a continuous data. Though age appears to be discrete data, one may be 12.5 years old and it makes sense. Patient height and weight are all continuous data.

Q 5) Apply PCA for following matrix and prove that it works

M= 4 3
1 2

Answer:

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

mean of col^{n} 1 = $\frac{4+1}{2} = \frac{5}{2} = 2.5$

mean of col^{n} 2 = $\frac{3+2}{2} = \frac{5}{2} = 2.5$

Subtract mean from each col^{n} element

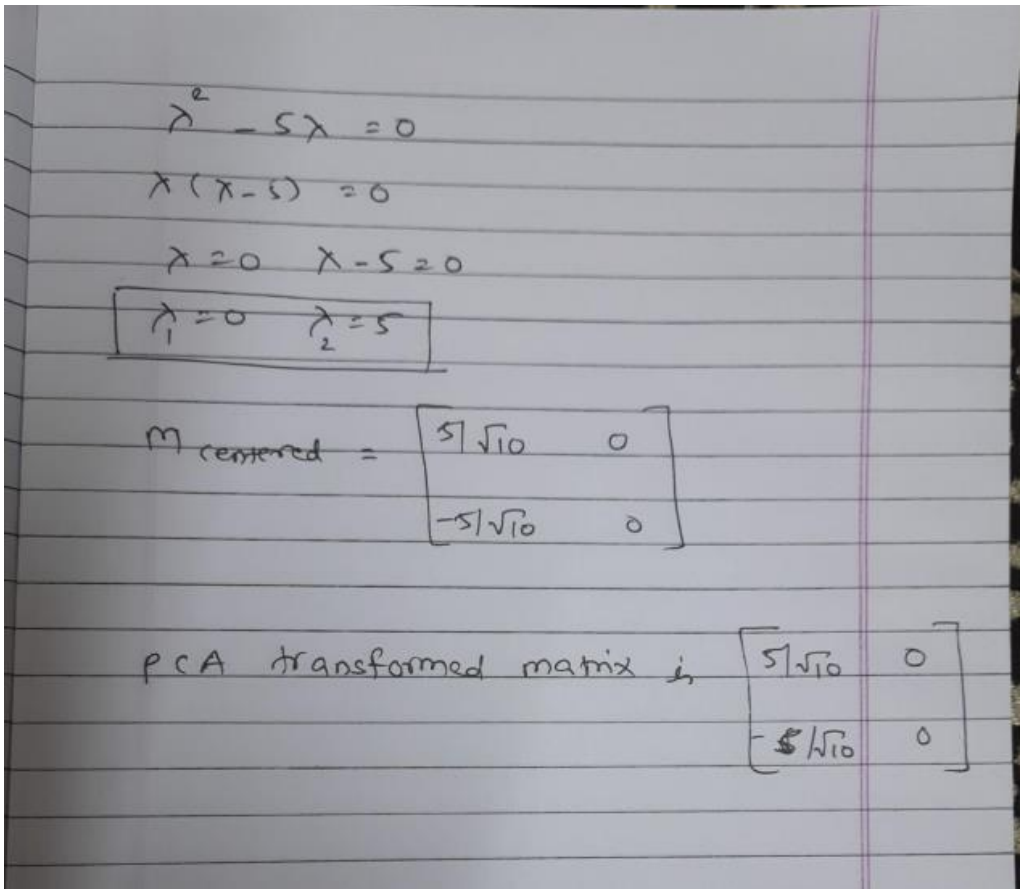
$$\begin{bmatrix} 4-2.5 & 3-2.5 \\ 1-2.5 & 2.2.5 \end{bmatrix}$$

$$= \begin{bmatrix} 1.5 & 0.5 \\ -1.5 & -0.5 \end{bmatrix}$$

covariance matrix = $\begin{bmatrix} 9/2 & 3/2 \\ 3/2 & 1/2 \end{bmatrix}$

Eigen values by solving characteristics eq^{n}

$$\det(COV - \lambda I) = 0$$



Q 6) Find Covariance and Correlation Coefficients for following two sets of data

X: 1 2 6 12

Y: 8 12 18 22

Answer:

Find covariance & correlation coefficients for following two sets

X: 1 2 6 12
Y: 8 12 18 22

$\bar{x}$ = mean (x) = $\frac{1+2+6+12}{4} = \frac{21}{4} = 5.25$

$\bar{y}$ = mean (Y) = $\frac{8+12+18+22}{4} = \frac{60}{4} = 15$

$$\text{cov.} = \frac{(1-5.25)(8-15) + (2-5.25)(12-15) + (6-5.25)(18-15) + (12-5.25)(22-15)}{4}$$

$$= \frac{(-4.25)(-7) + (-3.25)(-3) + (0.75)(3) + (6.75)(7)}{4}$$

$$= \frac{29.75 + 9.75 + 2.25 + 47.25}{4}$$

$$= \frac{89}{4} = 22.25 \quad \text{OR} \quad \frac{89}{n-1} = \frac{89}{3} = 29.66 \cong 30$$

correlation coefficient $r = \frac{cov(x,y)}{\sigma_x \sigma_y}$

std dev x = 4.9916

std dev y = 6.2182

$r = \frac{22.25}{4.99 \times 6.21}$ | $r_2 = \frac{29.67}{4.99 \times 6.21}$

$r = 0.716$ $r = 0.96$