Machine Learning Internal Question Paper Scheme and Solution

USN		CMRIT

Internal Assessment Test 1 - Mar 2026

Sub:	**Machine Learning**				Sub Code:	**BCS602**	Branch:	**ISE**
Date:	**03/03/2026**	Duration:	**90 min**	Max Marks: **50**	Sem/Sec:	**VI / A, B & C**		**OBE**

	Answer any FIVE FULL Questions	MARKS	CO	RBT
1a	Tom Mitchell's definition of machine learning. List out the factors that drive the popularity of machine learning.	[04]	CO1	L1
1b	Explain Supervised, Unsupervised, and semi-supervised Learning with examples	[06]	CO1	L2
2a	Explain in detail the machine learning process model with the process flow diagram	[05]	CO1	L2
2b	Explain Big Data Analytics and Types of Analytics with examples	[05]	CO1	L2
3	Let the data points be $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 7 \end{pmatrix}$ Apply PCA and find the transformed data. Apply reverse and prove that PCA Works.	[10]	CO2	L3
4a	Write the Find-S algorithm with a solved example for the same, and specify the limitations of Find-S.	[07]	CO2	L3
4b	Do the stem and Leaf plot for the following English Marks {38, 53, 63, 82, 88, 83, 85,92}	[03]	CO2	L3
5a	Apply the Candidate Elimination algorithm for the training dataset below, which consists of 6 instances. Show the updated **S and G boundaries** after processing every instance. (table below)	[08]	CO2	L3

Sky	**AirTemp**	**Humidity**	**Wind**	**Water**	**Forecast**	**EnjoySport**
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

		MARKS	CO	RBT
5b	Sorted data for price(in dollars): 2, 6, 7, 9, 13, 20, 21, 24, 30. Apply various Binning Techniques	[02]	CO2	L3
6a	Write the definition for the K-Nearest Neighbour (KNN) algorithm. Consider the following student performance training dataset: (table below) Using the K-Nearest Neighbour (KNN) algorithm with K = 3, classify the following test instance: (5.8,90,8)	[08]	CO3	L3

S.No	**CGPA**	**Assessment**	**Project**	**Result**
1	9	88	9	Pass
2	8.3	75	8	Pass
3	7.9	73	7	Pass
4	5.6	45	5	Fail
5	7	53	6	Fail
6	6.6	55	5	Fail
7	8.6	82	8	Pass
8	5.8	48	4	Fail

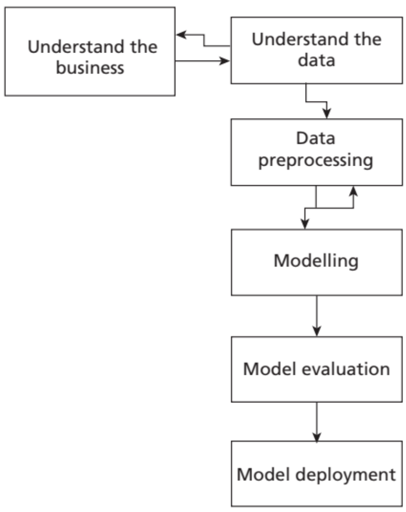
		MARKS	CO	RBT
6b	Write the Nearest centroid classifier algorithm	[02]	CO3	L1

CI CCI HOD

USN										

Internal Assessment Test 1 - Mar 2026								
Sub:	**Machine Learning**				Sub Code:	**BCS602**	Branch:	**ISE**
Date:	**03/03/2026**	Duration:	**90 min**	Max Marks: **50**	Sem/Sec:	**VI / A, B & C**		**OBE**

	Answer any FIVE FULL Questions and Answers	MARKS	CO	RBT
1a	Tom Mitchell's definition of machine learning. List out the factors that drive the popularity of machine learning. **Ans:** A computer program is said to learn from experience (E) with respect to some task (T) and performance measure (P), if its performance at task T, as measured by P, improves with experience E. Factors driving popularity of ML: 1. High volume of available data to manage: Big companies such as Facebook, Twitter, and YouTube generate huge amount of data that grows at a phenomenal rate. It is estimated that the data approximately gets doubled every year. 2. Second reason is that the cost of storage has reduced. The hardware cost has also dropped. Therefore, it is easier now to capture, process, store, distribute, and transmit the digital information. 3. Third reason for popularity of machine learning is the availability of complex algorithms now. Especially with the advent of deep learning, many algorithms are available for machine learning. With the popularity and ready adaption of machine learning by business organizations, it has become a dominant technology trend now.	[04]	CO1	L1
1b	Explain Supervised, Unsupervised, and semi-supervised Learning with examples **Ans:** Supervised Learning: Uses labeled data. Explain the example: Email spam classification. • Supervised algorithms use labelled dataset. As the name suggests, there is a supervisor or teacher component in supervised learning. • A supervisor provides labelled data so that the model is constructed and generates test data. • In supervised learning algorithms, learning takes place in two stages. • In layman terms, during the first stage, the teacher communicates the information to the student that the student is supposed to master. • The student receives the information and understands it. • During this stage, the teacher has no knowledge of whether the information is grasped by the student. Unsupervised Learning: Uses unlabeled data. Explain the example: Customer segmentation using clustering. • The second kind of learning is by self-instruction. • There are no supervisor or teacher components. • In the absence of a supervisor or teacher, self-instruction is the most common kind of learning process. • This process of self-instruction is based on the concept of trial and error.	[06]	CO1	L2

	• Here, the program is **<u>supplied with objects, but no labels are defined</u>**. • The algorithm itself observes the examples and recognizes patterns based on the principles of grouping. • Grouping is done in ways that similar objects form the same group. • <u>Cluster analysis and Dimensional reduction</u> algorithms are examples of unsupervised algorithms. Semi-Supervised Learning: Combination of labeled and unlabeled data. Explain the example: Image classification with few labeled samples. • There are circumstances where the dataset has a huge collection of <u>unlabelled data and some labelled data</u>. • Labelling is a costly process and difficult to perform by the humans. • Semi-supervised algorithms use **<u>unlabelled data by assigning a pseudo-label.</u>** • Then, the labelled and pseudo-labelled dataset can be combined.			
2a	Explain in detail the machine learning process model with the process flow diagram Ans:  • 1. Understanding the business – This step involves understanding the objectives and requirements of the business organization. Generally, a single data mining algorithm is enough for giving the solution. This step also involves the formulation of the problem statement for the data mining process. • 2. Understanding the data – It involves the steps like data collection, study of the characteristics of the data, formulation of hypothesis, and matching of patterns to the selected hypothesis. • 3. Preparation of data – This step involves producing the final dataset by cleaning the raw data and preparation of data for the data mining process. The missing values may cause problems during both training and testing phases. Missing data forces classifiers to produce inaccurate results. This is a perennial problem for the classification models. Hence, suitable strategies should be adopted to handle the missing data • 4. Modelling – This step plays a role in the application of data mining algorithm for the data to obtain a model or pattern.	[05]	CO1	L2

	• 5. Evaluate – This step involves the evaluation of the data mining results using statistical analysis and visualization methods. The performance of the classifier is determined by evaluating the accuracy of the classifier. The process of classification is a fuzzy issue. For example, classification of emails requires extensive domain knowledge and requires domain experts. Hence, performance of the classifier is very crucial. • 6. Deployment – This step involves the deployment of results of the data mining algorithm to improve the existing process or for a new situation.			
2b	Explain Big Data Analytics and Types of Analytics with examples Ans: Big Data Analytics refers to the process of analyzing extremely large and complex datasets to discover hidden patterns, correlations, trends, and useful insights that support better decision-making. With the growth of digital technologies, organizations generate massive volumes of structured and unstructured data from sources such as social media, sensors, transactions, and web applications. Big Data Analytics uses advanced tools and techniques to process this data efficiently. **Big Data Analytics and Types of Analytics** The primary aim of data analysis is to assist businesses and organizations in making better decisions. Data analytics involves examining data to extract useful information, identify patterns, and support decision-making. There are four main types of data analytics: 1. **Descriptive Analytics** Descriptive analytics focuses on describing the main features of collected data. It summarizes historical data to understand what has happened in the past. It uses statistical measures such as mean, median, percentage, and frequency to present data in a meaningful way. 2. **Diagnostic Analytics** Diagnostic analytics deals with the question "Why did it happen?". It is also known as causal analysis. This type of analytics identifies the causes and relationships behind events by examining patterns and correlations in data. 3. **Predictive Analytics** Predictive analytics deals with forecasting future outcomes based on historical data. It answers the question "What will happen in the future?". It uses statistical models and machine learning algorithms to identify patterns and predict future trends. 4. **Prescriptive Analytics** Prescriptive analytics suggests the best course of action to take. It goes beyond prediction and provides recommendations for decision-making. It helps organizations plan effectively and reduce risks by suggesting optimal solutions.	[05]	CO1	L2

3	Let the data points be $\begin{pmatrix}2\\6\end{pmatrix}$ and $\begin{pmatrix}1\\7\end{pmatrix}$ Apply PCA and find the transformed data. Apply reverse and prove that PCA Works. Ans: Principal Component Analysis Let the data points be $\begin{pmatrix}2\\6\end{pmatrix}$ and $\begin{pmatrix}1\\7\end{pmatrix}$. Apply PCA and find the transformed data. Solution: The mean vector can be calculated as, $\mu = \frac{x_1 + x_2}{2}$, $\mu = \begin{pmatrix}\frac{2+1}{2}\\ \frac{6+7}{2}\end{pmatrix} = \begin{pmatrix}1.5\\6.5\end{pmatrix}$ Centering the Data:- The mean must subtracted from the data to get the adjusted data, $x_1 = X_1 - \mu = \begin{pmatrix}2-1.5\\6-6.5\end{pmatrix} = \begin{pmatrix}0.5\\-0.5\end{pmatrix}$ $x_2 = X_2 - \mu = \begin{pmatrix}1-1.5\\7-6.5\end{pmatrix} = \begin{pmatrix}-0.5\\0.5\end{pmatrix}$ Compute the Covariance Matrix:- Centered data point is used to compute the covariance matrix, $m_1 = x_1 x_1^T = \begin{pmatrix}0.5\\-0.5\end{pmatrix}\begin{pmatrix}0.5 & -0.5\end{pmatrix}$ $= \begin{pmatrix}0.25 & -0.25\\-0.25 & 0.25\end{pmatrix}$	[10]	CO2	L3

$$m_2 = x_2 x_2^T = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \begin{pmatrix} -0.5 & 0.5 \end{pmatrix}$$

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

summing these Matrices,

$$C = m_1 + m_2 = \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$$

Compute Eigenvalues & Eigenvectors of the Covariance Matrix :-

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

Eigen Values → $\det(C - \lambda I) = 0$

substituting the $C - \lambda I$:

$$\begin{vmatrix} 0.5-\lambda & -0.5 \\ -0.5 & 0.5-\lambda \end{vmatrix} = 0$$

Expanding determinant,

$$(0.5-\lambda)(0.5-\lambda) - (-0.5)(-0.5) = 0$$

$$(0.5-\lambda)^2 - 0.25 = 0$$

$$(0.5-\lambda)^2 = 0.25 \quad \{ \text{Take sqrt on both sides}$$

$$0.5 - \lambda = \pm 0.5$$

solving for λ,

case 1 : $0.5 - \lambda = 0.5$

$\lambda = 0.5 - 0.5 = 0$

$\boxed{\lambda = 0}$

case 2 : $0.5 - \lambda = -0.5$

$\lambda = 0.5 + 0.5$

$\boxed{\lambda = 1}$ $\boxed{\text{Eigen values } \lambda_1 = 1, \lambda_2 = 0}$

Eigen Vectors

For each eigenvalues, solve $(C - \lambda I)v = 0$

1. For $\lambda_1 = 1$:

$$\begin{pmatrix} 0.5-1 & -0.5 \\ -0.5 & 0.5-1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -0.5 & -0.5 \\ -0.5 & -0.5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

simplifies to,

$$-0.5v_1 - 0.5v_2 = 0$$

$$v_1 = -v_2$$

setting $v_2 = 1$, we get $v_1 = -1$

$$v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Normalizing a vector,

$$V_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

magnitude of the vector v is given by

$$\|v\| = \sqrt{(-1)^2 + (1)^2} = \sqrt{1+1} = \sqrt{2}$$

Normalize the vector, v divide each component by the magnitude,

$$\frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

2. For $\lambda_2 = 0$:

$$\begin{pmatrix} 0.5-0 & -0.5 \\ -0.5 & 0.5-0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

simplifies to,

$$0.5v_1 - 0.5v_2 = 0$$

$$v_1 = v_2$$

setting $v_2 = 1$, we get $v_1 = 1$, so the eigen vector is,

$$v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Normalizing a vector,

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}$$

After Normalization,

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

<u>checking the orthogonality of A</u> :-

A matrix is orthogonal if it's inverse is equal to its transpose :

$$A^{-1} = A^{T}$$

or equivalently $AA^{T} = I$

compute AA^{T}

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

$$= \begin{pmatrix}\frac{1}{2}+\frac{1}{2} & -\frac{1}{2}+\frac{1}{2}\\ -\frac{1}{2}+\frac{1}{2} & \frac{1}{2}+\frac{1}{2}\end{pmatrix} \Rightarrow \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = I$$

The transformation of the matrix using the equation,

$$y = A(x - m)$$

$x \rightarrow$ original datapoint
$m \rightarrow$ mean of the datapoint
$A \rightarrow$ Transformation Matrix
$x - m \rightarrow$ Mean-adjusted data.

$$y = A(x-m) = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \left\{ 0.5 \text{ as } \frac{1}{2} \text{ for convenience} \right\}$$

$$y = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}$$

This means that after the PCA transformation, the second row becomes all zeros, which indicates that the second principal component does not contribute to the variance of this dataset

	One can check the original matrix with the formula, $x = A^T y + m$ $= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1.5 \\ 6.5 \end{pmatrix}$ $= \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1.5 \\ 6.5 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 6 & 7 \end{pmatrix}$ So, the original is obtained without any loss of information.			

4a	Write the Find-S algorithm with a solved example for the same, and specify the limitations of Find-S.
	Ans:
	Initialize ‘h’ to the most specific hypothesis.
	h = ⟨φ, φ, φ, φ, φ, …⟩
	Generalize the initial hypothesis for the first positive instance.
	For each subsequent instance:
	• **If it is a positive instance:**
	    ○ **Check each attribute value in the instance with the hypothesis ‘h’.**
	    ○ **If the attribute value is the same as the hypothesis value, do nothing.**
	    ○ **If the attribute value is different from the hypothesis value, change it to ‘?’ in ‘h’.**
	• **If it is a negative instance:**
	    ○ **Ignore it.**
		[07]	CO2	L3

Generalization : Specific to General Learning

Ex:1) $h = \langle \phi\ \phi\ \phi\ \phi\ \phi\ \phi\ \phi\ \phi \rangle$

Read the first instance I1, to generalize the hypothesis h so that this positive instance can be classified by the hypothesis h1.

I1 : NO short yes NO NO Black NO Big yes (positive Instance)

h_1 = < NO short yes NO NO Black NO Big >

When reading the second Instance I2, it is a negative instance, so ignore it.

I2 : yes short NO NO NO Brown yes Medium NO (Negative Instance)

h_2 = < NO short yes NO NO Black NO Big >

when reading the third Instance I3, it is a positive instance so generalize h_2 to h_3 to accomodate it.

I3 : NO short yes NO NO Black NO Medium yes (Positive Instance)

h_2, I_3 h_3 = < NO short yes NO NO Black NO ? >

Ignore I4, since it is a negative instance

I4 : NO Long NO yes yes White NO Medium
NO (Negative Instance)

same h3 h4 = < NO short yes NO NO Black NO ? >

when reading the fifth Instance I5, h4 is further generalized to h5.

I5 : NO short yes yes yes Black NO Big
yes (positive Instance)

I5, h4 h5 = < NO short yes ? ? Black NO ? >

The generated h5 hypothesis can classify any subsequent positive instance to true.

4b Do the stem and Leaf plot for the following English Marks {38, 53, 63, 82, 88, 83, 85,92} — [03] CO2 L3

Ans: Stem | Leaf 38, 53, 63, 82, 83, 85, 88, 92

-- Stem | Leaf------------

3 | 8
5 | 3
6 | 3
8 | 2 3 5 8
9 | 2

5a Apply the Candidate Elimination algorithm for the training dataset below, which consists of 6 instances. Show the updated **S and G boundaries** after processing every instance. — [08] CO2 L3

Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
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

Ans: The **Candidate Elimination algorithm** works by maintaining two boundaries: the **Specific boundary (S)** and the **General boundary (G)**. Initially, S is set to the most specific hypothesis ⟨Ø, Ø, Ø, Ø, Ø, Ø⟩ and G is set to the most general hypothesis ⟨?, ?, ?, ?, ?, ?⟩.

For the first training instance (Sunny, Warm, Normal, Strong, Warm, Same) with output **Yes**, since it is a positive example, the specific boundary S is updated by replacing all Ø with the corresponding attribute values. Thus, S becomes ⟨Sunny, Warm, Normal, Strong, Warm, Same⟩, while G remains unchanged.

	For the second positive instance (Sunny, Warm, High, Strong, Warm, Same), S is compared attribute-wise. Since the Humidity value differs (Normal and High), it is generalized to “?”. All other attributes remain the same. Hence, S becomes ⟨Sunny, Warm, ?, Strong, Warm, Same⟩ and G is still ⟨?, ?, ?, ?, ?, ?⟩. The third instance (Rainy, Cold, High, Strong, Warm, Change) is a **negative example**. In this case, the general boundary G must be specialized so that it does not cover this negative instance but remains consistent with S. The minimal specializations produce hypotheses such as ⟨Sunny, ?, ?, ?, ?, ?⟩, ⟨?, Warm, ?, ?, ?, ?⟩, and ⟨?, ?, ?, ?, ?, Same⟩. The specific boundary S remains unchanged. The fourth instance (Sunny, Warm, High, Strong, Cool, Change) is positive. The specific boundary S is updated again. Since the Water attribute (Warm and Cool) and Forecast attribute (Same and Change) differ, they are generalized to “?”. Thus, S becomes ⟨Sunny, Warm, ?, Strong, ?, ?⟩. From G, any hypothesis inconsistent with this positive example is removed, so ⟨?, ?, ?, ?, ?, Same⟩ is eliminated. Finally, the version space is bounded by: Specific boundary S = ⟨Sunny, Warm, ?, Strong, ?, ?⟩ General boundary G = {⟨Sunny, ?, ?, ?, ?, ?⟩, ⟨?, Warm, ?, ?, ?, ?⟩}.			
5b	Sorted data for price(in dollars): 2, 6, 7, 9, 13, 20, 21, 24, 30. Apply various Binning Techniques Ans: 2, 6, 7, 9, 13, 20, 21, 24, 30 Number of observations (n) = 9 **Create Bins** • **Bin 1:** 2, 6, 7 • **Bin 2:** 9, 13, 20 • **Bin 3:** 21, 24, 30 Calculate mean of each bin and replace values with mean. • **Bin 1 Mean:** (2+6+7)/3 = 15/3 = 5 → 5, 5, 5 • **Bin 2 Mean:** (9+13+20)/3 = 42/3 = 14 → 14, 14, 14 • **Bin 3 Mean:** (21+24+30)/3 = 75/3 = 25 → 25, 25, 25 • **Bin 1 Median:** 6 → 6, 6, 6 • **Bin 2 Median:** 13 → 13, 13, 13 • **Bin 3 Median:** 24 **Bin 1 Boundaries:** 2 and 7 → 2, 7, 7 • **Bin 2 Boundaries:** 9 and 20 → 9, 9, 20 • **Bin 3 Boundaries:** 21 and 30 → 21, 21, 30 • → 24, 24, 24	[02]	CO2	L3
6a	Write the definition for the K-Nearest Neighbour (KNN) algorithm. Consider the following student performance training dataset:			

S.No	CGPA	Assessment	Project	Result
1	9	88	9	Pass
2	8.3	75	8	Pass
3	7.9	73	7	Pass
4	5.6	45	5	Fail
5	7	53	6	Fail
6	6.6	55	5	Fail
7	8.6	82	8	Pass
8	5.8	48	4	Fail

K-Nearest Neighbour (KNN) is a **non-parametric, instance-based learning algorithm** used for both **classification and regression**. It works by finding the **K closest training samples** (neighbors) to a given test instance based on a distance measure (such as Euclidean distance).

- In **classification**, the test instance is assigned to the class that is most common among its K nearest neighbors (majority vote).
- In **regression**, the output is typically the average (or weighted average) of the values of its K nearest neighbors.

Using the K-Nearest Neighbour (KNN) algorithm with K = 3, classify the following test instance: (5.8,90,8)

[08] CO3 L3

Ans:

K = 3

We calculate Euclidean Distance:

Euclidean Distance Formula (2D)

$$d(p,q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

General Euclidean Distance Formula (n-dimensions)

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Step 1: Compute Distances

S.No	(CGPA, Assess, Project)	Result	Distance from (5.8,90,8)
1	(9,88,9)	Pass	$\sqrt{[(9-5.8)^2 + (88-90)^2 + (9-8)^2]} = \sqrt{(10.24+4+}$ 3.90
2	(8.3,75,8)	Pass	$\sqrt{(6.25+225+0)}=\sqrt{231.25} \approx 15.21$
3	(7.9,73,7)	Pass	$\sqrt{(4.41+289+1)}=\sqrt{294.41} \approx 17.15$
4	(5.6,45,5)	Fail	$\sqrt{(0.04+2025+9)}=\sqrt{2034.04} \approx 45.10$
5	(7,53,6)	Fail	$\sqrt{(1.44+1369+4)}=\sqrt{1374.44} \approx 37.07$
6	(6.6,55,5)	Fail	$\sqrt{(0.64+1225+9)}=\sqrt{1234.64} \approx 35.13$
7	(8.6,82,8)	Pass	$\sqrt{(7.84+64+0)}=\sqrt{71.84} \approx 8.47$

	8 (5.8,48,4) Fail $\sqrt{(0+1764+16)}=\sqrt{1780} \approx 42.19$			
	Step 2: Select 3 Nearest Neighbours (K=3) Smallest distances: 1. 3.90 → Pass (S.No 1) 2. 8.47 → Pass (S.No 7) 3. 15.21 → Pass (S.No 2)			
	Step 3: Majority Voting Among 3 nearest neighbors: Pass = 3 Fail = 0			
6b	Write the Nearest centroid classifier algorithm **Ans:** separable. **Algorithm 4.3: Nearest Centroid Classifier** **Inputs:** Training dataset *T*, Distance metric *d*, Test instance *t* **Output:** Predicted class or category 1. Compute the mean/centroid of each class. 2. Compute the distance between the test instance and mean/centroid of each class (Euclidean Distance). 3. Predict the class by choosing the class with the smaller distance.	[02]	CO3	L1