PVP 23

Code: 23IT3603

III B.Tech - II Semester - Regular Examinations – APRIL 2026

MACHINE LEARNING
(INFORMATION TECHNOLOGY)

Duration: 3 hours Max. Marks: 70

Note: 1. This question paper contains two Parts A and B.
2. Part-A contains 10 short answer questions. Each Question carries 2 Marks.
3. Part-B contains 5 essay questions with an internal choice from each unit. Each Question carries 10 marks.
4. All parts of Question paper must be answered in one place.

BL – Blooms Level CO – Course Outcome

PART – A

		BL	CO
1.a)	State any two real-world applications of machine learning.	L2	CO1
b)	Define feature engineering and write the need of it.	L2	CO1
c)	List any two non-metric similarity functions.	L2	CO1
d)	How do you compute accuracy in a classification model?	L2	CO1
e)	State Bayes' rule in probabilistic learning.	L2	CO1
f)	Mention two impurity measures used in decision tree algorithms.	L2	CO1
g)	What is the role of a perceptron in a classification task?	L2	CO1
h)	What is class-conditional independence in Naive Bayes?	L2	CO1
i)	Define divisive clustering.	L2	CO1
j)	What is the role of matrix factorization in clustering?	L2	CO1

PART – B

			BL	CO	Max. Marks
		UNIT-I			
2		Explain supervised, unsupervised and reinforcement learning paradigms with suitable examples.	L2	CO1	10 M
		OR			
3		Describe each stage of the machine learning process from data acquisition to prediction.	L2	CO1	10 M
		UNIT-II			
4	a)	Discuss different metrics used to evaluate classification models for the given confusion matrix: TP=80, FP=20, TN=10, FN=10. Calculate precision, recall and F1-score.	L3	CO2	5 M
	b)	Apply the K-Nearest Neighbor algorithm with $k = 3$ to classify the test point (3,3) using the dataset {(1,1) - A, (2,2) - A, (4,4) - B, (5,5) - B}.	L3	CO2	5 M
		OR			
5	a)	Apply KNN regression to a sample dataset and compute the predicted value for a new point.	L3	CO2	5 M
	b)	Explain different proximity and distance measures used in nearest neighbor algorithms with suitable examples.	L3	CO2	5 M
		UNIT-III			
6	a)	Explain how decision trees perform classification using impurity measures such as entropy.	L2	CO1	3 M

<table>
<tr><td></td><td>b)</td><td>Analyse the following data set and apply the Naïve Bayes classifier to predict “if a new customer (Age Group = Young, Income = Medium) will buy the product or not”.
<table>
<tr><th>Customer</th><th>Age Group</th><th>Income</th><th>Buys Product?</th></tr>
<tr><td>1</td><td>Young</td><td>High</td><td>No</td></tr>
<tr><td>2</td><td>Young</td><td>Medium</td><td>Yes</td></tr>
<tr><td>3</td><td>Middle-Aged</td><td>High</td><td>Yes</td></tr>
<tr><td>4</td><td>Senior</td><td>Low</td><td>No</td></tr>
<tr><td>5</td><td>Young</td><td>Low</td><td>Yes</td></tr>
<tr><td>6</td><td>Middle-Aged</td><td>Medium</td><td>Yes</td></tr>
<tr><td>7</td><td>Senior</td><td>Medium</td><td>No</td></tr>
<tr><td>8</td><td>Middle-Aged</td><td>High</td><td>Yes</td></tr>
</table></td><td>L4</td><td>CO4</td><td>7 M</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td>7</td><td>a)</td><td>Explain the bias-variance trade-off in decision tree models with suitable examples.</td><td>L2</td><td>CO1</td><td>5 M</td></tr>
<tr><td></td><td>b)</td><td>Discuss the working principle of the Naïve Bayes Classifier, its assumptions, advantages, limitations and provide an example scenario where it performs well compared to other classification algorithms.</td><td>L4</td><td>CO4</td><td>5 M</td></tr>
<tr><td colspan="6">UNIT-IV</td></tr>
<tr><td>8</td><td>a)</td><td>Apply perceptron learning algorithm to AND and OR Boolean functions.</td><td>L3</td><td>CO2</td><td>5 M</td></tr>
<tr><td></td><td>b)</td><td>Explain the architecture of a Multi-Layer Perceptron and describe how back propagation updates weights during training.</td><td>L3</td><td>CO2</td><td>5 M</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td>9</td><td>a)</td><td>Illustrate the working of expectation and maximization based clustering.</td><td>L3</td><td>CO2</td><td>5 M</td></tr>
<tr><td></td><td>b)</td><td>Explain and Differentiate between single-layer perceptron and Multilayer Perceptron.</td><td>L3</td><td>CO2</td><td>5 M</td></tr>
</table>

<table>
<tr><td colspan="6">UNIT-V</td></tr>
<tr><td rowspan="2">10</td><td>a)</td><td>Perform one iteration of K-Means clustering for points {(2,2), (3,3), (7,7), (8,8)} assuming initial centroids (2,2) and (7,7).</td><td>L3</td><td>CO3</td><td>5 M</td></tr>
<tr><td>b)</td><td>Explain spectral clustering and describe how it differs from traditional clustering methods.</td><td>L3</td><td>CO3</td><td>5 M</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td rowspan="2">11</td><td>a)</td><td>Using K-Means clustering on the following dataset with K = 2, initialize the centroids as (15, 39) and (20, 81). Perform one iteration by computing the Euclidean distance of each data point from the centroids and assigning each point to the nearest centroid. Show the new cluster assignments after this iteration.
<table>
<tr><th>CID</th><th>Income</th><th>Score</th></tr>
<tr><td>1</td><td>15</td><td>39</td></tr>
<tr><td>2</td><td>20</td><td>81</td></tr>
<tr><td>3</td><td>30</td><td>20</td></tr>
<tr><td>4</td><td>35</td><td>60</td></tr>
<tr><td>5</td><td>40</td><td>65</td></tr>
</table></td><td>L3</td><td>CO3</td><td>5 M</td></tr>
<tr><td>b)</td><td>Illustrate the working of agglomerative hierarchical clustering using a small dataset and show cluster merging at each step.</td><td>L3</td><td>CO3</td><td>5 M</td></tr>
</table>

PVP23

Code: 23IT3603

III B. Tech- II Semester- Regular Examinations - APRIL 2026

MACHINE LEARNING
(INFORMATION TECHNOLOGY)

Duration: 3 hours Max.Marks:70

Note: 1. This question paper contains two Parts A and B.
2. Part-A contains 10 short answer questions. Each Question carries 2 Marks.
3. Part – B contains 5 essay questions with an internal choice from each unit. Each Question carries 10 marks.
4. All parts of Question paper must be answered in one place.

BL- Blooms Level CO- Course Outcome

Scheme of Evaluation

PART - A

		BL	CO
1. a)	State any two real-world applications of machine learning. Any two applications of machine learning ----- 2M	L2	CO1
b)	Define feature engineering and write the need of it. Definition ------ 1M Need ---------- 1M	L2	CO1
c)	List any two non-metric similarity functions. Any two non-metric similarity functions ---------- 2M	L2	CO1
d)	How do you compute accuracy in a classification model? Accuracy formula --------- 2M	L2	CO1
e)	State Bayes' rule in probabilistic learning. Bayes' rule ---------- 2M	L2	CO1
f)	Mention two impurity measures used in decision tree algorithms. Any two impurity measures ------ 2M	L2	CO1
g)	What is the role of a perceptron in a classification task? Role of a perceptron --------- 2M	L2	CO1
h)	What is class-conditional independence in Naïve Bayes? Class-conditional independence in Naïve Bayes -------- 2M	L2	CO1
i)	Define divisive clustering. Divisive clustering ---------- 2M	L2	CO1
j)	What is the role of matrix factorization in clustering? Role of matrix factorization -------- 2M	L2	CO1

PART B

			BL	CO	Max. Marks
		UNIT - I			
2		Explain supervised, unsupervised and reinforcement learning paradigms with suitable examples. Supervised learning ---------- 4M Unsupervised learning ---------- 4M Reinforcement learning ---------- 2M	L2	CO1	10M
		OR			
3		Describe each stage of the machine learning process from data acquisition to prediction. Stage of the machine learning process ---------- 10M	L2	CO1	10M
		UNIT - II			
4	a)	Discuss different metrics used to evaluate classification models for the given confusion matrix: TP=80, FP=20, TN=10, FN=10. Calculate precision, recall and F1-score. Precision -------- 2M Recall ---------- 2M F1-score ---------- 1M	L3	CO2	5M
	b)	Apply the K-Nearest Neighbor algorithm with k = 3 to classify the test point (3,3) using the dataset {(1,1) - A, (2,2) – A, (4,4) – B, (5,5) - B}. Solution --------- 5M	L3	CO2	5M
		OR			
5	a)	Apply KNN regression to a sample dataset and compute the predicted value for a new point. KNN regression ---------- 5M	L3	CO2	5M

<table>
<tr><td></td><td>b)</td><td>Explain different proximity and distance measures used in nearest neighbor algorithms with suitable examples.

Proximity and distance measures ---------- 5M</td><td>L3</td><td>CO2</td><td>5M</td></tr>
<tr><td colspan="6">UNIT - III</td></tr>
<tr><td rowspan="2">6</td><td>a)</td><td>Explain how decision trees perform classification using impurity measures such as entropy.

Entropy -------- 3M</td><td>L2</td><td>CO1</td><td>3M</td></tr>
<tr><td>b)</td><td>Analyse the following dataset and apply the Naïve Bayes classifier to predict “if a new customer (Age Group = Young, Income = Medium) will buy the product or not”.
<table>
<tr><th>Customer</th><th>Age Group</th><th>Income</th><th>Buys Product?</th></tr>
<tr><td>1</td><td>Young</td><td>High</td><td>No</td></tr>
<tr><td>2</td><td>Young</td><td>Medium</td><td>Yes</td></tr>
<tr><td>3</td><td>Middle-Aged</td><td>High</td><td>Yes</td></tr>
<tr><td>4</td><td>Senior</td><td>Low</td><td>No</td></tr>
<tr><td>5</td><td>Young</td><td>Low</td><td>Yes</td></tr>
<tr><td>6</td><td>Middle-Aged</td><td>Medium</td><td>Yes</td></tr>
<tr><td>7</td><td>Senior</td><td>Medium</td><td>No</td></tr>
<tr><td>8</td><td>Middle-Aged</td><td>High</td><td>Yes</td></tr>
</table>
Solution ---------- 7M</td><td>L4</td><td>CO4</td><td>7M</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td rowspan="2">7</td><td>a)</td><td>Explain the bias-variance trade-off in decision tree models with suitable examples.

bias-variance trade-off -------- 5M</td><td>L2</td><td>CO1</td><td>5M</td></tr>
<tr><td>b)</td><td>Discuss the working principle of the Naïve Bayes Classifier, its assumptions, advantages, limitations and provide an example scenario where it performs well compared to other classification algorithms.

Naïve Bayes Classifier -------- 5M</td><td>L4</td><td>CO4</td><td>5M</td></tr>
<tr><td colspan="6">UNIT - IV</td></tr>
</table>

8	a)	Apply perceptron learning algorithm to AND and OR Boolean functions. AND ---------- 3M OR ------- 2M	L3	CO2	5M
	b)	Explain the architecture of a Multi-Layer Perceptron and describe how back propagation updates weights during training. Multi-Layer Perceptron ------- 5M	L3	CO2	5M
		OR			
9	a)	Illustrate the working of expectation and maximization based clustering. EM Clustering ---------- 5M	L3	CO2	5M
	b)	Explain and Differentiate between single-layer perceptron and Multilayer Perceptron. Differences ---------- 5M	L3	CO2	5M
		UNIT - V			
10	a)	Perform one iteration of K-Means clustering for points {(2,2), (3,3), (7,7), (8,8)} assuming initial centroids (2,2) and (7,7). Solution ---------- 5M	L3	CO3	5M
	b)	Explain spectral clustering and describe how it differs from traditional clustering methods. Spectral clustering -------- 3M Differences from traditional clustering methods ---------- 2M	L3	CO3	5M
		OR			

<table>
<tr>
<td>11</td>
<td>a)</td>
<td>Using K-Means clustering on the following dataset with K = 2, initialize the centroids as (15, 39) and (20, 81). Perform one iteration by computing the Euclidean distance of each data point from the centroids and assigning each point to the nearest centroid. Show the new cluster assignments after this iteration.
<table>
<tr><th>CID</th><th>Income</th><th>Score</th></tr>
<tr><td>1</td><td>15</td><td>39</td></tr>
<tr><td>2</td><td>20</td><td>81</td></tr>
<tr><td>3</td><td>30</td><td>20</td></tr>
<tr><td>4</td><td>35</td><td>60</td></tr>
<tr><td>5</td><td>40</td><td>65</td></tr>
</table>
Solution ---------- 5M</td>
<td>L3</td>
<td>CO3</td>
<td>5M</td>
</tr>
<tr>
<td></td>
<td>b)</td>
<td>Illustrate the working of agglomerative hierarchical clustering using a small dataset and show cluster merging at each step.

Agglomerative hierarchical clustering ---------- 5M</td>
<td>L3</td>
<td>CO3</td>
<td>5M</td>
</tr>
</table>

Code: 23IT3603

PVP23

III B. Tech – II Semester – Regular Examinations - APRIL 2026

MACHINE LEARNING
(INFORMATION TECHNOLOGY)

Duration: 3 hours Max. Marks: 70

Note:
1. This question paper contains two Parts A and B.
2. Part-A contains 10 short answer questions. Each Question carries 2 Marks.
3. Part – B contains 5 essay questions with an internal choice from each unit. Each Question carries 10 marks.
4. All parts of Question paper must be answered in one place.

BL – Blooms Level CO – Course Outcome

Scheme of Evaluation

PART – A

1.

a) State any two real-world applications of machine learning. **[L2] [CO1]**

Speech Recognition
Traffic prediction
Product recommendations
Self-driving cars etc.

b) Define feature engineering and write the need of it. **[L2] [CO1]**

Feature engineering is the process of transforming raw data into meaningful input variables (features) that improve the performance of a machine learning model.
Feature engineering is crucial because machine learning models don't understand raw data directly—they learn from patterns in features.

c) List any two non-metric similarity functions. **[L2] [CO1]**

Jaccard Similarity
Cosine Similarity

d) How do you compute accuracy in a classification model? **[L2] [CO1]**

$$\text{Accuracy} = \frac{\text{correct classifications}}{\text{total classifications}}$$

e) State Bayes' rule in probabilistic learning. **[L2] [CO1]**

$$P(C_i \mid A) - \frac{P(A \mid C_i)\,P(C_i)}{P(A)}$$

f) Mention two impurity measures used in decision tree algorithms. **[L2] [CO1]**

1. Entropy
2. Information Gain
3. Gini Index
4. Misclassification error
5. Mean Squared Error (MSE) *(for Regression Trees)*

g) What is the role of a perceptron in a classification task? **[L2] [CO1]**

The Perceptron model operates in a step-by-step process that involves computing the weighted sum of inputs, applying an activation function, and classifying the output.

h) What is class-conditional independence in Naïve Bayes? [L2] [CO1]

In Naive Bayes classifier, class-conditional independence means that all features are assumed to be independent of each other given the class label.

i) Define divisive clustering. [L2] [CO1]

Divisive clustering follows a top–down hierarchical approach. In this method, the clustering process starts with all data points grouped together in a single cluster.

j) What is the role of matrix factorization in clustering? [L2] [CO1]

Matrix Factorization is a technique used in clustering and data analysis to decompose a large data matrix into smaller matrices. This decomposition helps in discovering hidden patterns, structures, and relationships in the data.

PART - B

UNIT – 1

2. Explain supervised, unsupervised and reinforcement learning paradigms with suitable examples. [L2] [CO1]

Supervised Learning: -

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output. In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y). In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

Unsupervised Learning: -

As the name suggests, unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data. It can be compared to learning which takes place in the human brain while learning new things. It can be defined as:

Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.

Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

Reinforcement Learning: -

Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance. Agent gets rewarded for each good action and gets punished for each bad action; hence the goal of reinforcement learning agent is to maximize the rewards. In reinforcement learning, there is no labelled data like supervised learning, and agents learn from their experiences only. The reinforcement learning process is similar to a human being; for example, a child learns various things by experiences in his day-to-day life.

OR

3. **Describe each stage of the machine learning process from data acquisition to prediction.** **[L2] [CO1]**

STAGES OF MACHINE LEARNING: -

Building a machine learning system involves a number of steps. Note the emphasis on data in the form of training, validation and test data. Typically, the available data is split into training, validation and test data. Training data is used in model learning or training and validation data is used to tune the ML model. Test data is used to examine how the learnt model is performing. The components of an ML system are

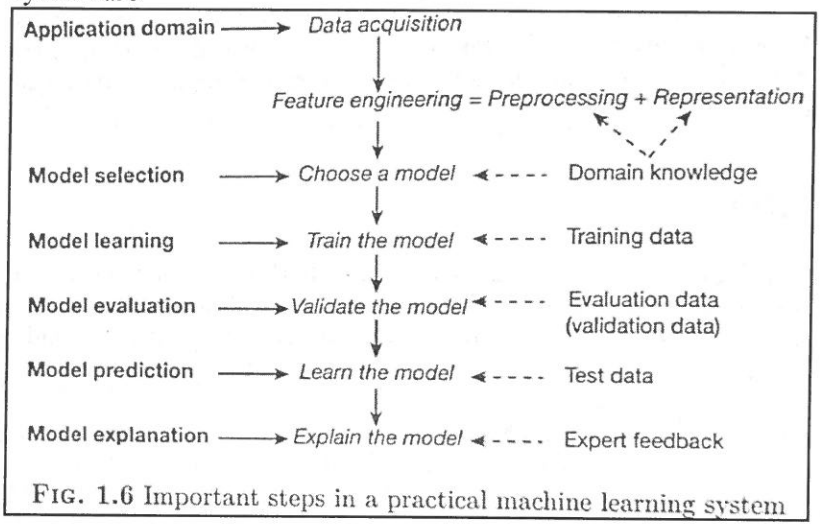


FIG. 1.6 Important steps in a practical machine learning system

Data Acquisition: -

This depends on the domain of the application. For example, to distinguish between adults and children, measurements of their height or weight are adequate; however, to distinguish between normal and COVID-19-infected humans, their body temperature and chest congestion may be more important than their height or weight. Typically, data collection is carried out before feature engineering.

Feature Engineering: -

This step involves a combination of **data preprocessing** and **data representation.**

Data preprocessing: -

In several practical applications, the raw data available needs to be updated before it can be used by an ML model. The common problems encountered with raw data are missing values, different ranges for different variables and the presence of outliers.

Data Representation: -

Representation is an important step in building ML models. In the process, it deals with both feature selection and feature extraction and introduces different categories of dimensionality reduction. It is often stated in DL literature that feature engineering is important in ML, but not in DL because DL systems have automatic representation learning capability.

Model Selection: -

Model selection in machine learning refers to the process of choosing the most appropriate model for a given dataset so that it performs well not only on training data but also on unseen data. Model selection is the process of choosing an appropriate machine learning model based on the nature of data and application requirements.

Model Learning: -
This step depends on the size and type of the training data. In practice, a subset of the labelled data is used as training data for learning the model and another subset is used for model validation or model evaluation. Some models are transparent, such as decision trees, which use easy-to-understand rules for decision making.

Model Evaluation: -
This step is also called model validation. This step requires specifically earmarked data called validation data. It is possible that the ML model works well on the training data; then we say that the model is well trained.

Model Prediction: -
This step deals with testing the model that is learnt and validated. It is used for prediction because both classification and regression tasks are predictive tasks. This step employs the test data set earmarked for the purpose. In the real world, the model is used for prediction as new patterns keep coming in.

Model Explanation: -
This step is important to explain to an expert or a manager why a decision was taken by the ML model. This will help in explicit or implicit feedback from the user to further improve the model. Explanation had an important role earlier in expert systems and other AI systems. However, explanation has become very important in the era of DL. This is because DL systems typically employ neural networks that are relatively opaque. So, their functioning cannot be easily explained at a level of detail that can be appreciated by the domain expert/ user.

UNIT - II

4. a) Discuss different metrics used to evaluate classification models for the given confusion matrix: TP=80, FP=20, TN=10, FN=10. Calculate precision, recall and F1-score. [L3] [CO2]
Solution: -

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$= \frac{80}{80 + 20} = \frac{80}{100} = 0.8$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$= \frac{80}{80 + 10} = \frac{80}{90} \approx 0.89$$

$$\text{F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

$$= \frac{2 \times (0.8 \times 0.89)}{0.8 + 0.89}$$

$$= \frac{2 \times 0.712}{1.69} = \frac{1.424}{1.69} \approx 0.84$$

b) Apply the K-Nearest Neighbor algorithm with k = 3 to classify the test point (3,3) using the dataset {(1,1) - A, (2,2) – A, (4,4) – B, (5,5) - B} [L3] [CO2]
Solution: -
Given k=3

Calculate distance Metric
We use **Euclidean distance**:

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

Test point is (3,3)

Training Instance	(x_1, x_2)	Class	Distance from t_1	Rank (k = 3)
I_1	(1,1)	A	$\sqrt{[(3-1)^2 + (3-1)^2]} = \sqrt{[4+4]}$ = 2.83	3
I_2	(2,2)	A	$\sqrt{[(3-2)^2 + (3-2)^2]} = \sqrt{[1+1]}$ = 1.41	1
I_3	(4,4)	B	$\sqrt{[(3-4)^2 + (3-4)^2]} = \sqrt{[1+1]}$ = 1.41	2
I_4	(5,5)	B	$\sqrt{[(3-5)^2 + (3-5)^2]} = \sqrt{[4+4]}$ = 2.83	–

Majority Voting
Class A → 2
Class B → 1
The test point **(3,3)** is classified as **Class A**

OR

5. a) Apply KNN regression to a sample dataset and compute the predicted value for a new point. [L3] [CO2]

Solution: -
KNN regression is a non-parametric method used for predicting continuous values. The core idea is to predict the target value for a new data point by averaging the target values of the K nearest neighbors in the feature space.

Sample dataset: -
Given the following training dataset, where each data point has two input features (x_1,x_2) and a continuous target value y. Compute the class label for test instance Q = (2, 1) with K = 2.

Training data:

(x_1, x_2)	y
(1,1)	10
(2,2)	20
(3,3)	30
(6,6)	60

Solution: -
Step 1: Given k=2
Step 2: Choose Distance Metric
We use **Euclidean distance**:

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

Step 3: Finding Nearest Neighbors
Query point: Q = (2, 1)
Distance metric: Euclidean distance

Training Point	y	Distance from Q(2,1)	Nearest (K=2)
(1,1)	10	$\sqrt{(2-1)^2 + (1-1)^2} = 1.0$	✓
(2,2)	20	$\sqrt{(2-2)^2 + (1-2)^2} = 1.0$	✓
(3,3)	30	$\sqrt{(2-3)^2 + (1-3)^2} \approx 2.24$	✗
(6,6)	60	$\sqrt{(2-6)^2 + (1-6)^2} \approx 6.40$	✗

Step 4: Predict Output (KNN Regression)
Nearest distances:
(1,1) → y = 10
(2,2) → y = 20
For KNN Regression, prediction is the **average of target values**:

$$\hat{y} = \frac{10 + 20}{2} = 15$$

Predicted output for query point $Q = (2, 1)$: $\boxed{y = 15}$

b) Explain different proximity and distance measures used in nearest neighbor algorithms with suitable examples. **[L3] [CO2]**

Answer: -
Proximity measures are used to quantify the degree of similarity or dissimilarity between two or more pattern vectors. These pattern vectors can represent documents, images or even entire audio or video files. Proximity measures are often used by ML algorithms to compare and classify or group or make predictions using these patterns. There are many types of proximity measures and the popular ones are:

Euclidean distance: -
Euclidean distance measures the straight-line distance between two points in a multi-dimensional space. So, it is also called as the crow files distance.

Formula

For two points

$A(x_1, y_1)$ and $B(x_2, y_2)$:

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

Cosine Similarity: -
Cosine similarity measures the angle between two vectors, not their magnitude. It tells how similar the direction of two vectors is.

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

Jaccard Similarity: -

Jaccard Similarity is a measure used to find how similar two sets are by comparing the elements they have in common with the total unique elements present in both sets.

$$\text{Jaccard Similarity}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Hamming Distance: -

In machine learning, the hamming distance measures the similarity between two strings of the same length. Hamming distance measures, the number of positions at which two strings differ.

UNIT – III

6. a) **Explain how decision trees perform classification using impurity measures such as entropy.**

Answer: - **[L2] [CO1]**

Entropy: -

Entropy measures the uncertainty or randomness in a dataset. If all data belongs to one class → entropy is 0 (pure). If data is evenly mixed → entropy is high. Used in ID3 and C4.5

$$Entropy = -\sum p_i \log_2(p_i)$$

b) Analyse the following dataset and apply the Naïve Bayes classifier to predict "if a new customer (Age Group = Young, Income = Medium) will buy the product or not". **[L4] [CO4]**

Customer	Age Group	Income	Buys Product?
1	Young	High	No
2	Young	Medium	Yes
3	Middle-Aged	High	Yes
4	Senior	Low	No
5	Young	Low	Yes
6	Middle-Aged	Medium	Yes
7	Senior	Medium	No
8	Middle-Aged	High	Yes

Solution: -

For this you need:

P(buy=yes)=5/8

P(buy=no)=3/8

P(Young|buy=yes)=2/5

P(Young|buy=no)=1/3

P(Medium|buy=yes)=2/5

P(Medium|buy=no)=1/3

P(Buy=yes|Young,Medium)=P(Young|Buy=yes)P(Medium|Buy=yes)P(Buy=yes)= 2/5*2/5*5/8 = 20/200 =0.1

P(Buy=no|Young,Medium)=P(Young|Buy=no)P(Medium|Buy=no)P(Buy=no)=1/3*1/3*3/8 = 3/72 =0.0417

The classifier **predicts "Yes"**, the customer **will buy the product**.

OR

7. **a) Explain the bias-variance trade-off in decision tree models with suitable examples. [L2] [CO1]**

Answer: -

A supervised Machine Learning model aims to train itself on the input variables(X) in such a way that the predicted values(Y) are as close to the actual values as possible. This difference between the actual values and predicted values is the error and it is used to evaluate the model. The error for any supervised Machine Learning algorithm comprises of 3 parts:

- Bias error
- Variance error
- The noise

What is Bias?

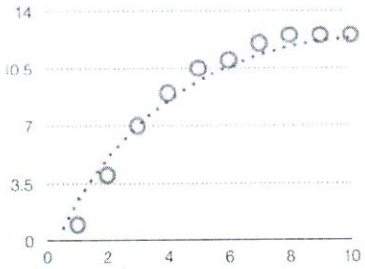
Bias is calculated as the difference between average prediction and actual value. In machine learning, bias (systematic error) occurs when a model makes incorrect assumptions about data. A model with high bias does not match well training data as well as test data. It leads to high errors in training and test data. While the model with low bias matches the training data well (high training accuracy or less error in training). It leads to low error in training data but high error in test data.

What is Variance?

Variance is a measure of the spread or dispersion of numbers in a given set of observations with respect to the mean. It basically measures how a set of numbers is spread out from the average. In statistics and probability, variance is defined as the expectation of the squared deviation of a random variable from the sample mean. In machine learning, variance is the variability of model prediction on different datasets. The variance shows how much model prediction varies when there is a slight variation in data. If model accuracies on training and test data vary greatly, the model has high variance.

Bias Variance Tradeoff : -

If the algorithm is too simple (hypothesis with linear equation) then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex (hypothesis with high degree equation) then it may be on high variance and low bias. In the latter condition, the new entries will not perform well. Well, there is something between both of these conditions, known as a Trade-off or Bias Variance Trade-off. An algorithm can't be more complex and less complex at the same time. For the graph, the perfect tradeoff will be like this.



b) Discuss the working principle of the Naïve Bayes Classifier, its assumptions, advantages, limitations and provide an example scenario where it performs well compared to other classification algorithms. [L4] [CO4]

Answer: -

Naïve Bayes Classifier: -

Naïve Bayes is a probabilistic classification algorithm based on Bayes Theorem with the assumption that all features are conditionally independent given the class. It is called *naïve*

because the independence assumption is usually not true in real life, but the method still works well.

Bayes Theorem

$$P(C_k \mid X) = \frac{P(C_k)\, P(X \mid C_k)}{P(X)}$$

For classification, denominator is ignored:

$$P(C_k \mid X) \propto P(C_k)\, P(X \mid C_k)$$

Advantages: -

1. Simple and fast
2. Works well with high-dimensional data
3. Requires less training data
4. Handles categorical data naturally
5. Performs well even with irrelevant features

Limitations: -

1. Strong independence assumption
2. Zero probability problem
3. Cannot capture feature interactions
4. Probability estimates may be inaccurate

UNIT - IV

8. a) Apply perceptron learning algorithm to AND and OR Boolean functions. [L3] [CO2]

Solution: -

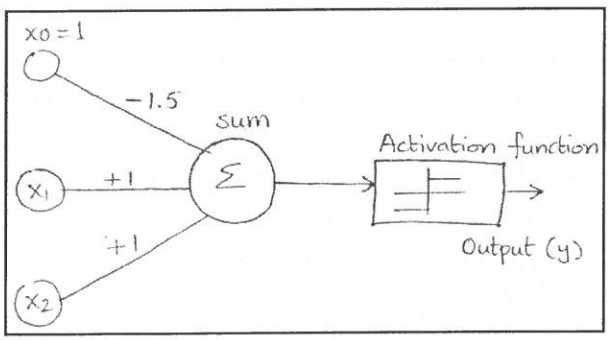
In a Boolean function, the inputs are binary and the output is 1 if the corresponding function value is true and 0 otherwise. Therefore, it can be seen as a two class classification problem. As an example, for learning to AND two inputs, the table of inputs and required outputs is given in table below.

AND function.

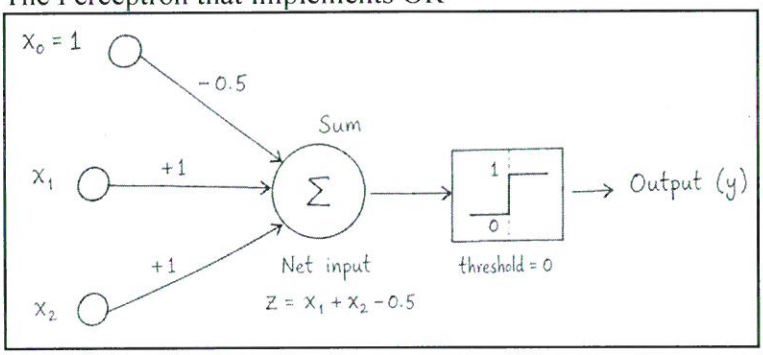
x_1	x_2	r
0	0	0
0	1	0
1	0	0
1	1	1

The Perceptron that implements AND from

$$y' = \sum_{i=1}^{n} w_i x_i + w_0$$



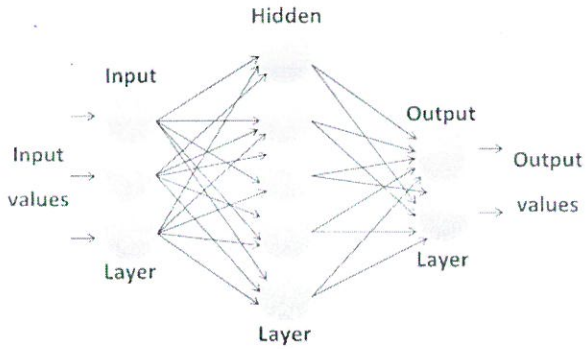
The Perceptron that implements OR



b) Explain the architecture of a Multi-Layer Perceptron and describe how back propagation updates weights during training. **[L3] [CO2]**

Answer: -

A **Multilayer Perceptron (MLP) classifier** is a type of **feedforward artificial neural network** used in machine learning for classification tasks. It can learn **complex, non-linear relationships** in data, distinguishing it from simpler models like a single-layer perceptron which can only handle linearly separable data.



Architecture of MLP: -

An MLP consists of multiple layers of interconnected nodes (neurons) organized into three main parts:

Input Layer: - The input layer is the first layer of a Multilayer Perceptron (MLP) and is responsible for receiving the raw input data from the dataset. Each neuron in this layer corresponds to one feature or attribute of the input. This layer does not perform any computation or transformation; it simply passes the input values directly to the next layer (hidden layer).

Hidden Layers: - Hidden layers are the intermediate layers between the input and output layers and are responsible for performing the main computations in the network. An MLP can have one or more hidden layers, and each layer consists of several neurons. Every neuron receives inputs from the previous layer, multiplies them by weights, adds a bias, and then applies an activation function such as ReLU, Sigmoid, or Tanh.

Output Layer: - The output layer is the final layer of the MLP that produces the result or prediction. The number of neurons in this layer depends on the type of problem being solved.

Back propagation: The network updates its weights and biases to minimize the loss using optimization techniques like Gradient Descent.

Simple Weight Update Formula

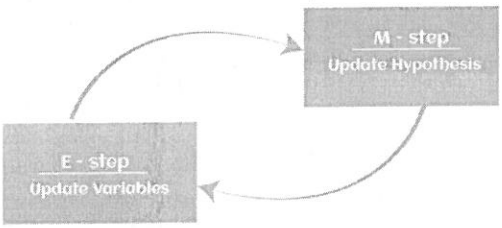
$$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w}$$

OR

9. **a) Illustrate the working of expectation and maximization based clustering.** **[L3] [CO2]**

Answer: -

The EM algorithm is the combination of various unsupervised ML algorithms, such as the k-means clustering algorithm. Being an iterative approach, it consists of two modes. In the first mode, we estimate the missing or latent variables. Hence it is referred to as the Expectation/estimation step (E-step). Further, the other mode is used to optimize the parameters of the models so that it can explain the data more clearly. The second mode is known as the maximization-step or M-step.



Expectation step (E - step): It involves the estimation (guess) of all missing values in the dataset so that after completing this step, there should not be any missing value.

Maximization step (M - step): This step involves the use of estimated data in the E-step and updating the parameters.

Repeat E-step and M-step until the convergence of the values occurs.

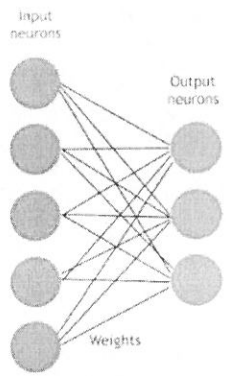
b) Explain and Differentiate between single-layer perceptron and Multilayer Perceptron.

Answer: - **[L3] [CO2]**

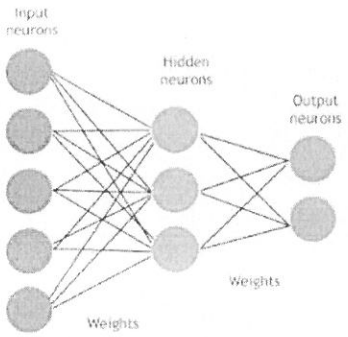
Perceptron models are basic building blocks of neural networks and are mainly classified based on their structure and application. These are as follows:

Single-layer Perceptron Model: - A Single Layer Perceptron (SLP) consists of only an input layer and an output layer, without any hidden layers, and is capable of solving only linearly separable problems such as simple logical operations.

Multi-layer Perceptron model: - In contrast, a Multi-Layer Perceptron (MLP) contains one or more hidden layers between input and output, allowing it to learn complex and non-linear patterns using advanced activation functions.



Single layer perceptron



Multi-layer perceptron

UNIT - V

10. a) Perform one iteration of K-Means clustering for points {(2,2), (3,3), (7,7), (8,8)} assuming initial centroids (2,2) and (7,7). [L3] [CO3]

Solution: -

Data points: (2,2), (3,3), (7,7), (8,8)
Initial centroids:
C1 = (2,2)
C2 = (7,7)
Distance metric: Euclidean

Final Result after One Iteration

Cluster 1: (2,2), (3,3) → New centroid = (2.5, 2.5)
Cluster 2: (7,7), (8,8) → New centroid = (7.5, 7.5)

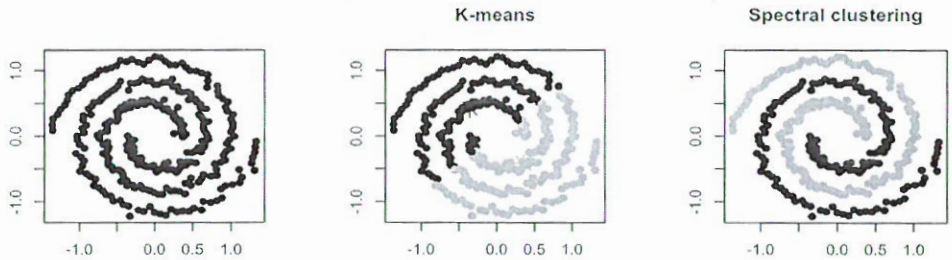
b) Explain spectral clustering and describe how it differs from traditional clustering methods.
Answer: - [L3] [CO3]

Spectral clustering: -

As the number of dimensions increases, a distance-based similarity measure converges to a constant value between any given examples. Reduce dimensionality either by using PCA on the feature data, or by using "spectral clustering" to modify the clustering algorithm.

In compactness, the points are closer to each other and are compact towards the cluster center. Distance is used as a measure to compute closeness. There are different types of distance metrics that are in use. A few of them are Euclidean distance, Manhattan distance, Minkowski distance, and Hamming distance. K-means algorithm uses the compactness approach.

In connectivity, the points in a cluster are either immediately next to each other (epsilon distance) or connected. Even if the distance is less, they are not put in the same cluster. Spectral clustering is one of the techniques to follow this approach.



OR

11. a) Using K-Means clustering on the following dataset with K = 2, initialize the centroids as (15, 39) and (20, 81). Perform one iteration by computing the Euclidean distance of each data point from the centroids and assigning each point to the nearest centroid. Show the new cluster assignments after this iteration. [L3] [CO3]

CID	Income	Score
1	15	39
2	20	81
3	30	20
4	35	60
5	40	65

Solution: -
Data points: (15,39), (20,81), (30,20), (35,60), (40,65)
Initial centroids:
C1 = (15,39)
C2 = (20,81)
Distance metric: Euclidean

Final Result after One Iteration
Cluster 1: (15,39), (30,20) → New centroid = (22.5, 29.5)
Cluster 2: (20,81), (35,60), (40,65) → New centroid = (31.67, 68.67)

b) Illustrate the working of agglomerative hierarchical clustering using a small dataset and show cluster merging at each step. [L3] [CO3]
Answer: -
Agglomerative clustering is a hierarchical clustering algorithm that starts with each data point as its own cluster and iteratively merges the closest clusters until a stopping criterion is reached. It is a bottom-up approach that produces a dendrogram, which is a tree-like diagram that shows the hierarchical relationship between the clusters.
Examples: -
Apply Agglomerative Hierarchical Clustering to the given data points and show how the clusters are formed step-by-step

Point	Coordinates
A	(7,8)
B	(1,2)
C	(8,9)
D	(2,3)
E	(9,10)

Solution: -
Given Data Points
A (7,8), B (1,2), C (8,9), D (2,3), E (9,10)

Step 1: Initial Clusters
Each point is its own cluster:
{A}, {B}, {C}, {D}, {E}

Step 2: Calculate Pairwise Distances (Euclidean): -
The distance between every pair of clusters is calculated using a distance measure such as Euclidean distance. Since each cluster has only one point at this stage, distances are computed between individual data points. Consider the closest distance.

Step 3: First Merge (Closest Points)
Merge closest pair:
{A, C} (or {B, D} — tie, choose any)

Clusters now:
{A,C}, {B}, {D}, {E}

Step 4: Second Merge

Next smallest distance:
Merge {C} with {E} → effectively merge {A, C} with E

Clusters:
{A, C, E}, {B}, {D}

Step 5: Third Merge

Merge closest pair:
{B, D}
Clusters:
{A, C, E}, {B, D}

Step 6: Final Merge

Merge the two remaining clusters:
{A, C, E} and {B, D}

Final cluster:
{A,B,C,D,E}

Final Clusters (if cut at 2 clusters)

- Cluster 1: {A, C, E}
- Cluster 2: {B, D}

The final clusters are obtained by stopping the merging process at a certain level or by cutting the dendrogram. This determines the number of meaningful groups in the dataset.