GAYATRI VIDYA PARISHAD COLLEGE OF ENGINEERING FOR WOMEN

(AUTONOMOUS)

(Affiliated to Andhra University, Visakhapatnam)

I M. Tech II Semester Regular Examinations, August- 2026

Machine Learning- 24CS21RC05

(Common to CSE, CSE(DS)

……..

Key of Valuation

UNIT-I

1. **A. Explain the relationship between Artificial Intelligence, Machine Learning, and Deep Learning. Give suitable examples to show how they are related.**

Artificial Intelligence (AI) is the broad field of developing machines that can perform tasks requiring human intelligence, such as reasoning, decision-making, problem-solving, and language understanding.

Machine Learning (ML) is a subset of AI in which machines learn patterns from data and make predictions or decisions without being explicitly programmed for every situation.

Deep Learning (DL) is a subset of ML that uses multi-layered artificial neural networks to learn complex patterns from large amounts of data.

Relationship

AI ⟶ ML ⟶ DL

- AI is the broadest concept.
- ML is one approach used to achieve AI.
- DL is an advanced ML technique.

Example: In a self-driving car:

- AI: Makes decisions such as stopping, turning, and changing lanes.
- ML: Learns driving patterns from data.
- DL: Identifies pedestrians, vehicles, traffic signs, and lanes from camera images.

Thus, DL is a part of ML, and ML is a part of AI.

1 b. What are the major benefits of Machine Learning systems?

Major benefits of Machine Learning include:

1. Automation: Automates repetitive decision-making and prediction tasks.
2. Learning from data: Systems improve their performance as more data becomes available.

3. Better predictions: ML can identify patterns that may not be obvious to humans.
4. Handling large datasets: It can process and analyze huge volumes of data efficiently.
5. Adaptability: Models can be retrained when data or conditions change.
6. Personalization: Used to provide personalized recommendations and services.
7. Reduced human effort: Reduces manual analysis and repetitive work.
8. **Applications across domains:** Used in healthcare, finance, education, transportation, security, and entertainment.

Example: Netflix and other recommendation systems use machine learning to suggest content based on user behavior.

OR

2. a. Explain the three main types of Machine Learning?

The three main types are:

1. Supervised Learning

In supervised learning, the model is trained using **labeled data**, where both input and expected output are known.

Example: Predicting whether an email is spam or not spam.

Common algorithms:

- Linear Regression
- Logistic Regression
- Decision Trees
- k-NN
- Naïve Bayes

2. Unsupervised Learning

In unsupervised learning, the model receives **unlabeled data** and attempts to discover hidden patterns or groups.

Example: Grouping customers according to their purchasing behavior.

Common techniques:

- Clustering
- K-Means
- Hierarchical Clustering
- Principal Component Analysis (PCA)

3. Reinforcement Learning

In reinforcement learning, an **agent interacts with an environment** and learns by receiving rewards or penalties.

Example: A robot learning to navigate a room or an AI learning to play a game.

2 b. Differentiate between batch learning and online learning

Batch Learning

In **batch learning**, the model is trained using the entire available dataset at once. The model is updated periodically when new training data is collected.

Advantages:

- Simple to implement.
- Can provide stable models.
- Suitable when data changes slowly.

Disadvantage: Retraining can be computationally expensive when the dataset is very large.

Online Learning

In **online learning**, the model is trained incrementally using individual data instances or small batches of data.

Advantages:

- Can adapt quickly to changing data.
- Suitable for continuously arriving data.
- Requires less memory at a time.

Disadvantage: The model can be affected by noisy or poor-quality incoming data.

Batch Learning	**Online Learning**
Uses entire dataset	Uses individual samples/small batches
Model updated periodically	Model updated continuously
Higher computational requirement during retraining	Lower memory requirement
Suitable for static data	Suitable for continuously changing data
Example: periodic sales prediction model	Example: real-time recommendation system

UNIT-II

3. a. What is regularization in Machine Learning? [7M]

Regularization is a technique used to reduce **overfitting** by adding a penalty to the model for having overly complex parameters.

A model that is too complex may perform very well on training data but poorly on unseen data. Regularization encourages the model to learn simpler patterns and improves generalization.

Main types

1. L1 Regularization (Lasso)
Adds a penalty based on the absolute values of model parameters. It can make some parameters exactly zero, effectively performing feature selection.

2. L2 Regularization (Ridge)
Adds a penalty based on the squared values of model parameters. It discourages very large parameter values.

Example

Suppose a model has many features and learns unnecessary patterns from the training data. Applying regularization reduces the influence of unnecessary features and helps the model perform better on new data.

Purpose:
Reduce overfitting → Improve generalization → Improve performance on unseen data.

3 b. Explain the k-Nearest Neighbours classification algorithm.

k-Nearest Neighbours (k-NN) is a supervised machine-learning algorithm used mainly for **classification** and also for regression.

The basic idea is that a new data point is classified according to the classes of its **k nearest training examples**.

Steps

1. Choose the value of **k**.
2. Calculate the distance between the new point and training data points.
3. Select the **k nearest points**.
4. Determine the majority class among these neighbours.
5. Assign that class to the new data point.

A common distance measure is **Euclidean distance**.

Example

Suppose we want to classify a fruit as an **apple or orange** based on weight and size.

If $k = 3$, and among the three closest fruits:

- 2 are apples
- 1 is an orange

Then k-NN classifies the new fruit as an **apple**.

Advantages

- Simple and easy to understand.
- No complex training process.
- Useful for classification problems.

Limitations

- Prediction can be slow for large datasets.
- Sensitive to the choice of k.
- Sensitive to feature scaling.
- Can be affected by irrelevant features.

OR

4. a. Compare k-NN and Naïve Bayes classification algorithms. [7M]

Feature	**k-NN**	**Naïve Bayes**
Type	Supervised learning	Supervised learning
Approach	Instance-based	Probabilistic
Main principle	Uses nearest data points	Uses Bayes' theorem
Training	Very little training	Builds probability estimates
Prediction	Based on neighbouring points	Based on calculated probabilities
Speed	Prediction can be slower	Usually fast
Data requirement	Sensitive to distance and scaling	Works well with many feature types
Example	Image classification	Spam email classification

4 b. Explain accuracy, precision, recall, F1-score, and confusion matrix with suitable examples.

These are commonly used evaluation measures for classification models.

Confusion Matrix

A confusion matrix summarizes the predictions of a classification model.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

- **TP:** Correctly predicted positive.

- **TN:** Correctly predicted negative.
- **FP:** Negative incorrectly predicted as positive.
- **FN:** Positive incorrectly predicted as negative.

Example

Consider a disease prediction system:

- TP = 80
- TN = 90
- FP = 10
- FN = 20

1. Accuracy

Accuracy represents the proportion of **all predictions that are correct**.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

For the example:

$$Accuracy = \frac{80 + 90}{80 + 90 + 10 + 20} = \frac{170}{200} = 85\%$$

2. Precision

Precision tells us, among all instances predicted as positive, **how many are actually positive**.

$$Precision = \frac{TP}{TP + FP}$$
$$Precision = \frac{80}{80 + 10} = 88.89\%$$

High precision means fewer false positives.

3. Recall

Recall tells us, among all actual positive instances, **how many were correctly identified**.

$$Recall = \frac{TP}{TP + FN}$$
$$Recall = \frac{80}{80 + 20} = 80\%$$

High recall means fewer false negatives.

4. F1-Score

F1-score is the **harmonic mean of precision and recall**.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

For the example:

UNIT-III

5. a. Explain the basic concept of a separating hyperplane, support vectors, and the objective of SVM classification.

Support Vector Machine (SVM) is a supervised learning algorithm mainly used for classification.

Separating Hyperplane

A **hyperplane** is a decision boundary that separates data points belonging to different classes.

For a linear SVM, the hyperplane can be represented as:

$$w^T x + b = 0$$

where $w$is the weight vector and $b$is the bias.

Support Vectors

Support vectors are the data points closest to the separating hyperplane. They determine the position and orientation of the optimal hyperplane.

Objective of SVM

The main objective is to find the hyperplane that:

- Correctly separates the classes.
- Maximizes the **margin**, i.e., the distance between the hyperplane and the nearest data points.
- Provides better generalization to unseen data.

Thus, SVM attempts to find the **maximum-margin decision boundary**.

5 b. Compare LDA, linear SVM, nonlinear SVM, and SVM regression.

Feature	**LDA**	**Linear SVM**	**Nonlinear SVM**	**SVM Regression**
Purpose	Classification/dimensionality reduction	Classification	Nonlinear classification	Regression
Boundary	Linear	Linear	Nonlinear	Regression function
Main idea	Maximizes class separation	Maximizes margin	Uses kernel functions	Fits data within an ε-insensitive margin

Feature	LDA	Linear SVM	Nonlinear SVM	SVM Regression
Data	Assumes class distributions	Linearly separable/approximately separable	Nonlinearly separable	Continuous target
Example	Classifying documents	Spam classification	Image classification	House-price prediction

OR

6. a. How does a linear SVM separate data belonging to two different classes?

A **linear SVM** separates two classes using a hyperplane.

The decision boundary is:

$$w^T x + b = 0$$

The SVM creates two parallel margin boundaries:

$$w^T x + b = 1$$

and

$$w^T x + b = -1$$

The data points lying closest to these boundaries are called **support vectors**.

The SVM chooses the hyperplane that **maximizes the margin** between the two classes.

For a new data point:

- If $w^T x + b > 0$, it is assigned to one class.
- If $w^T x + b < 0$, it is assigned to the other class.

Therefore, a linear SVM separates classes by finding the **optimal maximum-margin hyperplane**.

6 b. Describe basic principle, working process of LDA and how it can be used for classification.

Linear Discriminant Analysis (LDA) is a supervised technique used for **classification and dimensionality reduction**.

Basic Principle

LDA tries to find a projection that:

- Maximizes the distance between different classes.
- Minimizes the variation within the same class.

Working Process

1. Calculate the **mean vector** for each class.
2. Calculate the **within-class scatter**.
3. Calculate the **between-class scatter**.
4. Find the projection that maximizes between-class separation relative to within-class variation.
5. Project the data onto the selected discriminant direction.
6. Classify new data based on its projected position.

Example

Suppose students are classified as **Pass** or **Fail** based on:

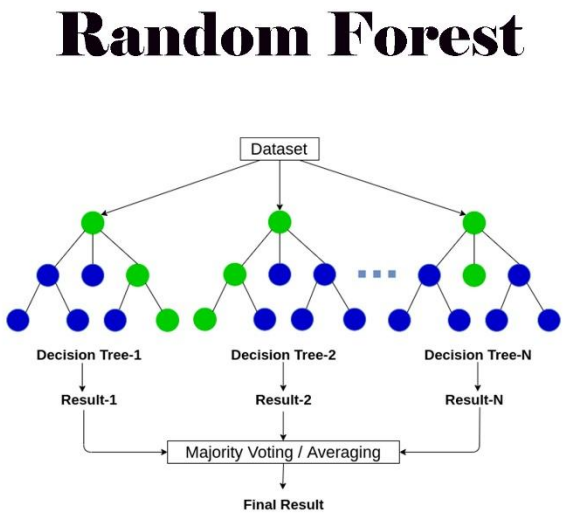
- Study hours
- Attendance

LDA finds a linear combination of these features that provides maximum separation between Pass and Fail students.

UNIT-IV

7. a. Describe how Random Forests are constructed using decision trees and discuss its advantages over a single decision tree.

Random Forest is an ensemble-learning method that combines multiple decision trees to produce a more accurate and stable prediction.



Construction

1. Select different **bootstrap samples** from the training dataset.

2. Build a decision tree for each sample.
3. At each split, consider a **random subset of features**.
4. Grow multiple decision trees.
5. Combine the predictions of all trees:
    - **Classification:** Majority voting.
    - **Regression:** Average of predictions.

Advantages over a Single Decision Tree

- **Higher accuracy:** Combines multiple trees.
- **Reduced overfitting:** Less dependent on one tree.
- **Better stability:** Less sensitive to changes in training data.
- **Handles many features:** Can work with large feature sets.
- **Useful for classification and regression.**
- Can estimate **feature importance**.

Thus, Random Forest generally provides more robust predictions than an individual decision tree.

7 b. Discuss the major strengths and weaknesses of the Decision Tree approach

Strengths

1. **Easy to understand:** The tree structure is simple to interpret.
2. **Easy to visualize:** Decisions can be represented as branches.
3. **Handles numerical and categorical data.**
4. **Little preprocessing:** Usually does not require feature scaling.
5. **Captures nonlinear relationships.**
6. **Useful for both classification and regression.**

Weaknesses

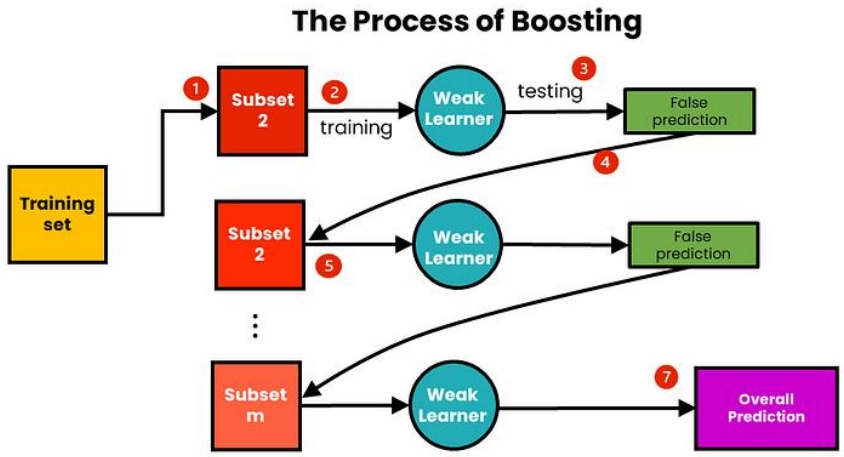
1. **Overfitting:** A deep tree may memorize training data.
2. **Unstable:** Small changes in data can produce a different tree.
3. **Biased splits:** Some splitting criteria may favor features with many possible values.
4. **Can become complex:** Large trees are difficult to interpret.
5. **Lower generalization:** A single tree may be less accurate than ensemble methods.

8. a. Explain boosting and stacking techniques in Ensemble Learning.

Boosting

Boosting is an ensemble technique that builds models **sequentially**, where each new model attempts to correct the errors made by previous models.

Process



Models are combined to produce a stronger learner.

Examples:

- AdaBoost
- Gradient Boosting
- XGBoost

Example: If the first classifier incorrectly classifies some students, the next classifier gives greater attention to those difficult cases.

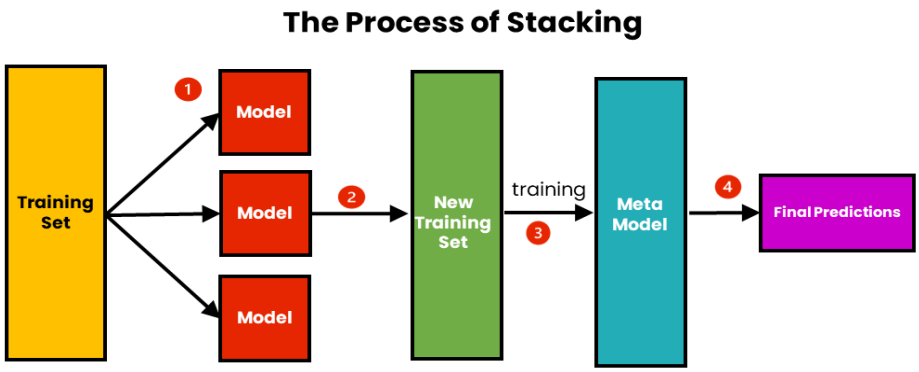
Advantage

Boosting can significantly improve prediction accuracy.

Stacking

Stacking, or stacked generalization, combines predictions from **different types of machine-learning models** using another model called a **meta-model**.

Process



For example:

- Model 1: Decision Tree

- Model 2: SVM
- Model 3: Logistic Regression
- Meta-model: Logistic Regression

The meta-model learns how to combine the predictions of the base models.

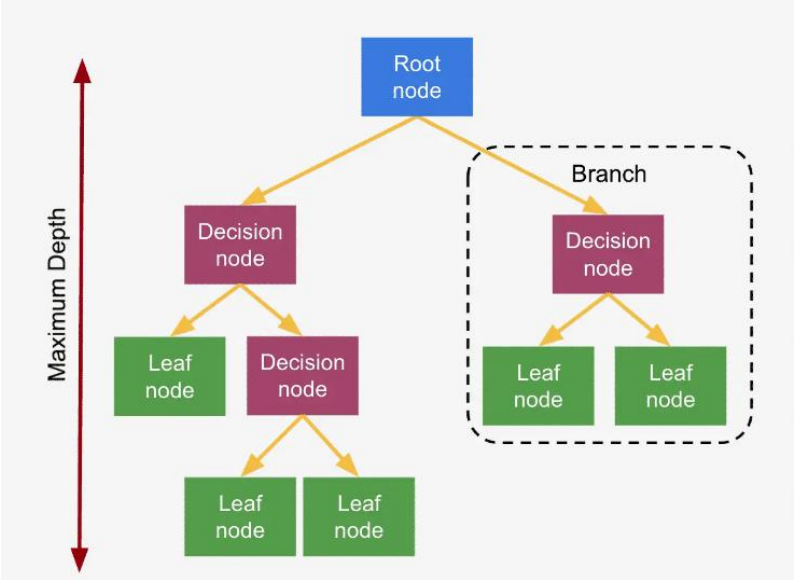
Difference

Boosting: Models are usually built sequentially to correct previous errors.

Stacking: Different models are combined and their outputs are given to a meta-model.

8 b. Explain the steps involved in constructing a decision tree using ID3 and describe how the be attribute is selected for splitting.

D3 (Iterative Dichotomiser 3) is a decision-tree algorithm that uses **Information Gain** to select the best attribute for splitting.



Steps in ID3

1. Start with the complete training dataset.
2. Calculate the **entropy** of the target attribute.
3. Calculate the Information Gain for each candidate attribute.
4. Select the attribute with the **highest Information Gain** as the decision node.
5. Divide the dataset according to the selected attribute.
6. Repeat the process for each resulting subset.
7. Stop when:
    - All examples belong to the same class, or
    - No useful attributes remain.

Entropy

Entropy measures the impurity or uncertainty in a dataset:

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

Information Gain

Information Gain measures how much an attribute reduces uncertainty:

$$IG(S, A) = Entropy(S) - \sum_v \frac{|\ S_v\ |}{|\ S\ |} Entropy(S_v)$$

The attribute with the **highest Information Gain** is selected as the best attribute for splitting.

UNIT-V

9. a. Discuss feature selection and feature extraction with suitable examples.

Feature Selection:
Feature selection is the process of selecting the **most relevant features** from the original dataset while removing irrelevant or redundant features.

Methods:

- Filter methods – Correlation, Chi-square
- Wrapper methods – Forward/Backward selection
- Embedded methods – Lasso, Decision Trees

Example:
For predicting house prices, features like **area, location, number of rooms** may be selected, while an irrelevant feature like **house ID** can be removed.

Feature Extraction:
Feature extraction transforms the original features into a **new, smaller set of meaningful features**.

Example:
PCA converts 10 original features into 2 or 3 principal components while preserving most of the important information.

Difference:

- Selection → chooses existing features.
- Extraction → creates new features from existing ones.

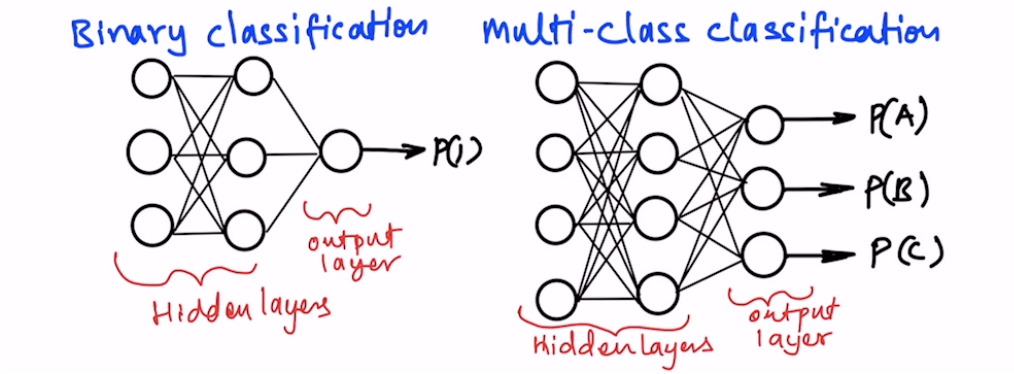
9 b. Differentiate between Incremental PCA, Randomized PCA, and Kernel PCA.

Feature	Incremental PCA	Randomized PCA	Kernel PCA
Basic idea	Processes data in small batches	Uses randomization for faster computation	Uses kernel functions
Suitable for	Very large datasets	Large/high-dimensional datasets	Nonlinear data
Memory	Low memory requirement	Relatively efficient	Higher computational cost
Data type	Mainly linear	Linear	Nonlinear
Example	Processing large datasets batch-by-batch	Fast dimensionality reduction	Image/complex pattern analysis

OR

10. a. Explain how Artificial Neural Networks can be used to solve classification problems

An **Artificial Neural Network (ANN)** is a computational model consisting of interconnected neurons arranged in **input, hidden, and output layers**.



Steps for classification:

1. Input features are given to the input layer.
2. Neurons calculate weighted sums and apply activation functions.
3. Hidden layers learn important patterns from the data.
4. The output layer produces class probabilities or class labels.
5. The network is trained by minimizing the classification error using **backpropagation and gradient descent**.

Example:
An ANN can classify emails as **spam or not spam** based on features such as words, sender information, and email structure.

10 b. Explain the basic steps involved in PCA.

Principal Component Analysis (PCA) is a dimensionality-reduction technique that converts correlated features into a smaller number of uncorrelated **principal components**.

Steps:

1. **Standardize the data** – Bring features to a common scale.
2. **Calculate covariance matrix** – Determine relationships between features.
3. **Calculate eigenvalues and eigenvectors** – Find important directions.
4. **Rank principal components** – Arrange them according to eigenvalues.
5. **Select components** – Choose the components explaining most of the variance.
6. **Transform the data** – Project the original data onto the selected components.

Example:
If a dataset has **10 features**, PCA may reduce it to **2 principal components** while retaining most of the useful information.