Answers to Sections A, B and C

SECTION - A

a) Differentiate between precision and recall:

Precision tells us how many of the predicted positive results are actually correct, while recall tells us how many of the actual positive results we correctly predicted.

b) Discuss Mean Absolute Error:

Mean Absolute Error (MAE) is the average of the absolute differences between actual and predicted values. It shows how far predictions are from actual values.

c) What is Correlation:

Correlation is a statistical measure that shows the relationship between two variables. It can be positive, negative, or zero.

d) Discuss various applications of clustering:

Clustering is used in customer segmentation, image segmentation, grouping documents, and anomaly detection.

e) Define Logistic Regression:

Logistic regression is used to predict binary outcomes (like yes/no). It uses a logistic function to give output between 0 and 1.

SECTION - B

Q2. Explain the need of confusion metrics. Differentiate between precision and recall:

Confusion metrics are needed to evaluate classification models. They show true/false positives and negatives. Precision focuses on correct positive predictions, while recall focuses on detecting all actual positives.

Q3. Why hyperplane is used in SVM:

In SVM, a hyperplane is the decision boundary that separates different classes of data. The best hyperplane maximizes the margin between the classes.

Q4. What are various distance metrics used in K-NN algo:

Common distance metrics in K-NN are Euclidean distance, Manhattan distance, and Minkowski distance. These help determine which neighbors are closest for making predictions.

SECTION - C

Q5. Discuss various methods of clustering.

Clustering is an unsupervised machine learning technique used to group similar data points into clusters based on patterns and similarities. One of the most commonly used methods is K-Means Clustering, where data is divided into K groups based on the nearest average value. Another method is Hierarchical Clustering, which builds a tree structure of clusters without requiring the number of clusters in advance. DBSCAN groups data based on density and is good at handling noise. Mean Shift shifts data points towards the region of highest density. These methods are used in marketing, image analysis, fraud detection, and more.

Q6. Describe decision tree and random forest tree.

A Decision Tree is a flowchart-like structure used for classification and regression. It splits data into branches based on feature conditions, and the leaves give the final result. It is simple and easy to understand but can overfit the data. Random Forest is an ensemble method that builds multiple decision trees using random subsets of data and features. It improves accuracy and reduces overfitting by taking the majority vote or average of the individual trees. These models are widely used in fields like medical diagnosis, loan approval, and fraud detection.