

Explain PCA and LDA

Principal Component Analysis (PCA)

1. What it is: PCA is a dimensionality reduction technique.
2. Purpose: Reduces data to fewer dimensions while preserving as much information as possible.
3. Usage: Common in image processing, noise reduction, and visualization.

Linear Discriminant Analysis (LDA)

1. What it is: LDA is a classification and dimensionality reduction technique.
2. Purpose: Helps separate data by classes, improving classification.
3. Usage: Often used in face recognition and other classification tasks with labeled data.

What is Regression

1. Regression is a type of machine learning task focused on predicting a continuous outcome or numerical value based on input data.
2. It aims to find relationships between input features and a target variable so that predictions can be made for new data.
3. In simple terms, regression helps us estimate values or trends based on patterns learned from historical data.

What is R2 and RMSE

R^2 (R-squared): It shows how well the model's predictions match the actual values. It ranges from 0 to 1, where 1 means perfect predictions and 0 means no correlation.

RMSE (Root Mean Squared Error): RMSE (Root Mean Squared Error) tells us how much the model's predictions differ from the actual values, on average. A lower RMSE means the model's predictions are closer to the real values, while a higher RMSE means they are further apart.

Ridge & Lasso Regression

Ridge Regression: It adds a penalty to the model for large coefficients (weights), helping to reduce overfitting. The penalty term is the sum of the squares of the coefficients, making the model simpler.

Lasso Regression: It also adds a penalty, but it's the sum of the absolute values of the coefficients. Lasso can set some coefficients to zero, effectively removing certain features, which makes the model simpler and easier to interpret.

SVM

SVM (Support Vector Machine) is a supervised machine learning algorithm used for classification and regression.

It works by finding the best line (or hyperplane in higher dimensions) that separates different classes of data points.

K-Nearest Neighbour

1. K-Nearest Neighbour is one of the only Machine Learning algorithms based totally on supervised learning approach.
2. It is used for classification and regression.
3. It works by comparing a new data point to the points in the dataset and checking which are most similar.
4. Imagine you have a group of apples and oranges, and you want to classify a new fruit. The KNN algorithm looks at its closest neighbors (fruits) and checks if they are mostly apples or oranges. If most are apples, it classifies the new fruit as an apple.

What is Clustering and K-Means

Clustering: It's a technique used to group similar items together. The goal is to find patterns or structures in data by grouping items that are similar to each other.

K-Means: It's a popular clustering algorithm. It divides data into 'K' clusters by finding the average (centroid) of each group. Each data point is assigned to the nearest centroid. The algorithm repeats until the clusters are stable.

K-Mediod Algorithm

The K-Medoids algorithm is a clustering method similar to K-Means, but instead of using the mean of points as the center of a cluster, it uses an actual data point (called a medoid) as the center.

Silhouette Method

The Silhouette Method is used to evaluate the quality of clusters in clustering algorithms.

It measures how similar an object is to its own cluster compared to other clusters.

The Silhouette Score ranges from -1 to +1:

A score close to +1 means the object is well clustered (similar to its own group and far from other groups).

A score close to 0 means the object is on the boundary between two clusters.

A score close to -1 indicates that the object might be in the wrong cluster.

What is elbow method

1. The Elbow Method is used to find the best number of clusters (K) for K-Means clustering.
2. You plot the inertia (the sum of squared distances between data points and their cluster centroids) for different values of K.
3. As K increases, inertia decreases.
4. The "elbow" is the point where the decrease in inertia starts slowing down. This point suggests the optimal number of clusters.

Random Forest Classifier model

1. The Random Forest Classifier is an ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting.
2. Each tree makes a prediction, and the class that gets the most votes from all trees is the final prediction.

Adaptive Boosting

Adaptive Boosting (AdaBoost) is a machine learning algorithm that combines multiple weak models to create a stronger, more accurate model.

Gradient Tree Boosting (GBM)

Gradient Tree Boosting (or Gradient Boosting Machine, GBM) is a powerful machine learning technique that builds a strong model by combining many smaller models, typically decision trees.

XGBoost

XGBoost (Extreme Gradient Boosting) is an advanced version of gradient boosting, designed to be faster and more efficient, especially on large datasets.

Reinforcement Learning

1. Reinforcement Learning (RL) is a type of machine learning where an agent learns by interacting with its environment.
2. The agent performs actions, and in return, it gets feedback in the form of rewards or penalties.
3. It's like learning by trial and error, where the agent improves its strategy based on past experiences