Assignment 1:

Linear regression : Linear regression is a statistical method used to model the relationship between a dependent variable (y) and one or more independent variables (x). It assumes that the relationship between the dependent and independent variables is linear, meaning that it can be represented by a straight line. The goal of linear regression is to find the equation of this line, which can then be used to predict the value of the dependent variable for new values of the independent variable.

Assignment 2:

K-nearest neighbors (KNN) is a non-parametric, instance-based learning algorithm used for classification and regression. It works by classifying new data points based on the majority class or mean value of their K nearest neighbors in the training set. The distance between data points is typically calculated using a distance metric such as Euclidean distance or Manhattan distance.

1. Calculate the distance between the new data point and all data points in the training set.
2. Sort the data points in the training set by their distance to the new data point.
3. Select the K nearest neighbors from the sorted list.

Choosing the Value of K

The value of K is a critical parameter in KNN, and it can significantly impact the performance of the algorithm. A small value of K may lead to overfitting, while a large value of K may lead to underfitting. There is no single best value of K, and it typically needs to be determined through experimentation.

Assignment 3:

What is SVM?

Support Vector Machines (SVM) are a supervised learning algorithm used for classification, regression, and outlier detection. It is a powerful and versatile algorithm that can be used to solve a wide variety of problems.

Data Preprocessing: Import the necessary libraries, load the dataset, and perform any necessary data preprocessing steps, such as data cleaning, scaling, and normalization.

Data Splitting: Divide the dataset into training and testing sets. The training set is used to train the SVM model, while the testing set is used to evaluate the model's performance.

Model Training: Create an SVM instance and train the model on the training set. This involves selecting the appropriate kernel function, hyperparameters, and fitting the model to the training data.

Model Evaluation: Evaluate the trained SVM model on the testing set. Calculate metrics such as accuracy, precision, recall, and F1-score to assess the model's performance.

Model Tuning: Optimize the SVM model's hyperparameters using techniques like grid search or random search to improve its performance.

K-Means :

Theory of K-Means Clustering

The k-means algorithm works by iteratively assigning data points to the nearest cluster centroid, the mean value of all data points within a cluster. This process is repeated until the clusters converge to a stable state where no further reassignments occur.

Steps in K-Means Clustering

Initialize Centroids: Randomly select k data points as initial cluster centroids.

Assignment Step: Assign each data point to the cluster with the nearest centroid.

Update Centroids: Calculate the new centroid for each cluster by averaging the values of all data points assigned to that cluster.

Repeat: Repeat steps 2 and 3 until the centroids converge or a maximum number of iterations is reached.

Choosing the Value of K

The value of k, the number of clusters, is a critical parameter in k-means clustering. Choosing an appropriate value of k depends on the characteristics of the data and the desired outcome. There are various methods to estimate the optimal value of k, such as the elbow method or silhouette analysis.