K-Means Clustering (KMC)

- Type: Clustering
- Purpose: Grouping data into clusters based on feature similarity.
- How it Works:
    - KMC initializes k centroids randomly.
    - Each data point is assigned to the nearest centroid, forming k clusters.
    - Centroids are recalculated as the mean of all points in the cluster.
    - Steps 2 and 3 are repeated until convergence (centroids no longer move significantly).
- Use Cases: Market segmentation, image compression, anomaly detection.
- Advantages:
    - Simple and easy to implement.
    - Scales well to large datasets.
    - Efficient in terms of computational cost.
- Disadvantages:
    - Requires specifying the number of clusters k beforehand.
    - Sensitive to initial placement of centroids.
    - Can struggle with clusters of varying sizes and densities.

K-Nearest Neighbors (KNN)

- Type: Classification (also used for regression)
- Purpose: Classifying a data point based on the majority class of its k nearest neighbors.
- How it Works:
    - Choose the number of neighbors k.
    - For a new data point, find the k closest points in the training dataset (using a distance metric like Euclidean distance).
    - Assign the new data point to the majority class among its k nearest neighbors.
- Use Cases: Pattern recognition, recommendation systems, intrusion detection.
- Advantages:

    - Simple and intuitive.
    - No training phase (instance-based learning).
    - Can adapt to changes in the training data.
- Disadvantages:
    - Computationally expensive during prediction (especially for large datasets).
    - Performance can be degraded by irrelevant or redundant features.
    - Choice of k and distance metric can significantly impact performance.

Support Vector Machine (SVM)

- Type: Classification (also used for regression)
- Purpose: Finding a hyperplane that best separates different classes in the feature space.
- How it Works:
    - For linear SVM, it finds the hyperplane that maximizes the margin (distance between the hyperplane and the nearest data point of any class).
    - For non-linear SVM, it uses kernel functions (e.g., polynomial, radial basis function) to transform the feature space and find a hyperplane in the transformed space.
- Use Cases: Text classification, image classification, bioinformatics.
- Advantages:
    - Effective in high-dimensional spaces.
    - Robust to overfitting, especially in high-dimensional space.
    - Can be used for both linear and non-linear classification.
- Disadvantages:
    - Can be computationally intensive.
    - Performance depends on the choice of kernel and regularization parameters.
    - Less interpretable compared to other algorithms like decision trees.

- Algorithm Type:
    - KMC: Clustering
    - KNN: Classification/Regression
    - SVM: Classification/Regression
- Usage:
    - KMC: Unsupervised learning for clustering.
    - KNN: Supervised learning for classification/regression.
    - SVM: Supervised learning for classification/regression.