1. K-Nearest Neighbors (KNN) Theory (2 hours)

Introduction to KNN:

K-Nearest Neighbors is a simple, yet powerful, instance-based learning algorithm used for both classification and regression tasks. It makes predictions based on the majority label of the nearest points in the feature space.

Key Concepts:

Instance-Based Learning: KNN does not learn an explicit model but stores training instances to make predictions.

K Value: The number of nearest neighbors considered for the prediction. A larger K value smoothens the decision boundary, but too large a K value may cause over-smoothing.

Distance Metrics: Commonly used distance measures include:

Euclidean Distance (most common for continuous variables)

Manhattan Distance (useful for grid-like data)

Minkowski Distance (a generalized version of Euclidean and Manhattan distances)

Hamming Distance (used for categorical variables)

KNN for Classification:

In KNN classification, the majority class of the K-nearest points determines the class label for the new data point.

KNN for Regression:

In KNN regression, the target value for a new instance is predicted by averaging the values of its K-nearest neighbors.

Advantages and Disadvantages:

Advantages: Simple to implement, intuitive, non-parametric, works well with small datasets.

Disadvantages: Computationally expensive with large datasets, sensitive to irrelevant features, requires feature scaling.

Applications of KNN:

Image classification, handwriting recognition, recommendation systems, anomaly detection, and more.

gsw6775