K-Nearest Neighbors (KNN) Algorithm

Shah Md. Arshad Rahman Ziban

The K-Nearest Neighbors (KNN) Algorithm is a simple, non-parametric machine learning method used for both classification and regression tasks. The basic idea behind KNN is that it makes predictions based on the 'k' nearest data points to a given query point. Here's a breakdown of how KNN works:

Steps in KNN:

1. **Choose the number of neighbors (k):** Select the value of 'k', which determines how many nearest neighbors will be considered for making the prediction.
2. **Calculate distance:** For a given data point, calculate the distance between the point and all other points in the dataset. Common distance metrics include:
   - **Euclidean distance:**
   $$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$
   - **Manhattan distance:**
   $$d = |x_1 - x_2| + |y_1 - y_2|$$
   - **Minkowski distance:** (generalized distance metric)
3. **Find the nearest neighbors:** Sort the distances and pick the 'k' nearest neighbors.
4. **Predict the output:**
   - **For Classification:** The output class is determined by the majority class among the 'k' nearest neighbors.
   - **For Regression:** The output is typically the average or weighted average of the 'k' nearest neighbors' target values.