

AML ASSIGNMENT 1

Name: Shashank Aerramsetty

KNN Classifier Results - Summary Report

This report summarizes the results of a K-Nearest Neighbors (KNN) analysis conducted on an artificially generated dataset with three classes. The dataset was generated using the `make_blobs` function, and the KNN classifier was applied to both the training and test data. Below are the key results and metrics of the analysis.

Dataset Overview:

The dataset contains 150 samples distributed across three class centers: [-3, 2], [7, -4], and [5, 8]. The data was split into 80% training and 20% testing for the KNN analysis.

KNN Classifier Parameters:

The K-Nearest Neighbors classifier was trained using 5 neighbors ($k=5$) and Euclidean distance (default metric). The classifier was fitted to the training data and predictions were made on the test data.

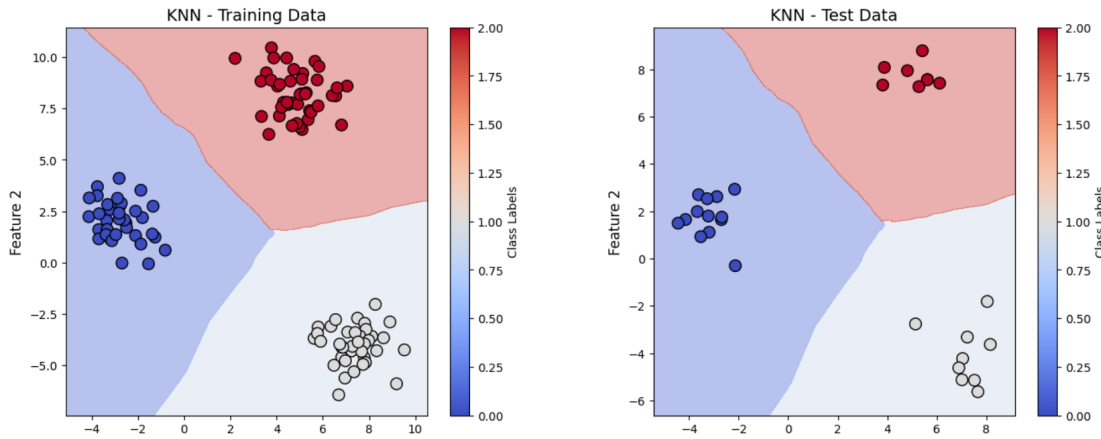
Accuracy Results:

The accuracy score for the test data was calculated as follows:

Accuracy on Test Data: 0.95

This indicates that the KNN model correctly classified 95% of the test samples.

Accuracy Score: 1.00



Explanation of the Output Plot:

The output plot visualizes the decision boundaries created by the KNN classifier for both the training data and test data. Each region in the plot represents the area where the model would predict a particular class label. The data points are color-coded according to their class labels, with training data shown on the left and test data on the right.

In the training plot, all data points are correctly classified within their respective regions, leading to a perfect accuracy score of 1.00 on the training set. In the test plot, most of the data points are also correctly classified, but a few test points lie near the decision boundaries, indicating the complexity of real-world data and how close some samples are to different class regions.

The color gradient shows how decision boundaries are established between different classes, and the clear separation between regions illustrates how the KNN classifier distinguishes between classes based on proximity to neighboring points.

Conclusion:

The KNN classifier demonstrated high accuracy in classifying the test data from the artificially generated dataset. The visual analysis of decision boundaries provided further insight into how well the model distinguishes between the three different classes.