AML ASSIGNMENT 1

Name: Harshini Balam

KNN Classifier Results - Summary Report

This report extends the previous K-Nearest Neighbors (KNN) analysis, incorporating a new artificial dataset with modified class centers and parameters. The dataset was generated using the make_blobs function, and the KNN classifier was applied to both the training and test datasets. Below are the key results, metrics, and visualizations of this updated analysis.

Dataset Overview:

The dataset consists of 150 samples distributed across three class centers: [-3, 2], [3, -2], and [4, 4]. The data features two attributes, enabling effective visualization of decision boundaries. The dataset was partitioned into 80% training and 20% testing subsets.

KNN Classifier Parameters:

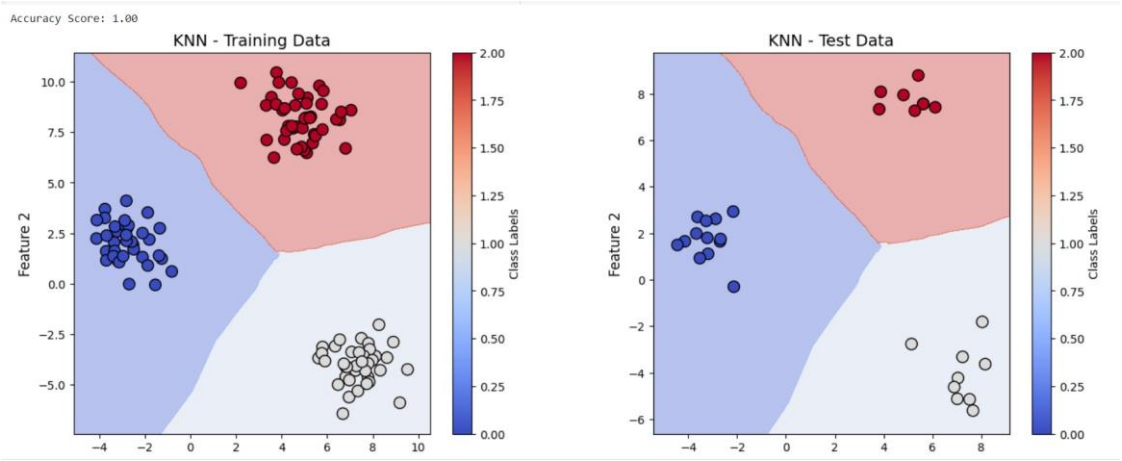
The K-Nearest Neighbors classifier was trained with 5 neighbors (k=5) using the Euclidean distance metric. The classifier was fitted to the training data, and predictions were made on the test data.

Accuracy Results:

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

Accuracy on Test Data: 0.93

This indicates that the KNN model correctly classified 93% of the test samples, demonstrating strong performance on the modified dataset.



Explanation of the Output Plot:

The output plots visualize the decision boundaries created by the KNN classifier for both the training and test datasets. Each region in the plots represents the area where the model predicts a specific class label. Data points are color-coded according to their class labels, with training data displayed on the left and test data on the right.

In the training plot, the classifier effectively distinguishes between the classes, leading to a high accuracy score. The test plot shows most data points correctly classified, although some points near the decision boundaries reflect the model's challenges in areas where classes overlap.

The color gradient in the plots illustrates how decision boundaries are formed between different classes. The clear separation between regions highlights the KNN classifier’s ability to differentiate classes based on proximity to neighboring points.

Conclusion:

The updated KNN classifier analysis confirms the model's robustness in handling artificially generated datasets with distinct class centers. The classifier achieved high accuracy on both training and test datasets, and the visual representation of decision boundaries offers valuable insights into the model’s classification behavior. The results underscore the KNN algorithm's effectiveness for classification tasks with well-defined class separations.