AML ASSIGNMENT 1

Name: Akash Nair Eruvattu

KNN Classifier Results - Summary Report

This report builds on the previous K-Nearest Neighbors (KNN) work by introducing a new artificial dataset with adjusted class centers and parameters. The dataset was created using the make_blobs function, and the KNN model was trained and evaluated on both the training and testing sets. The following sections present the main results, performance measures, and visual interpretations from this analysis.

Dataset Overview:

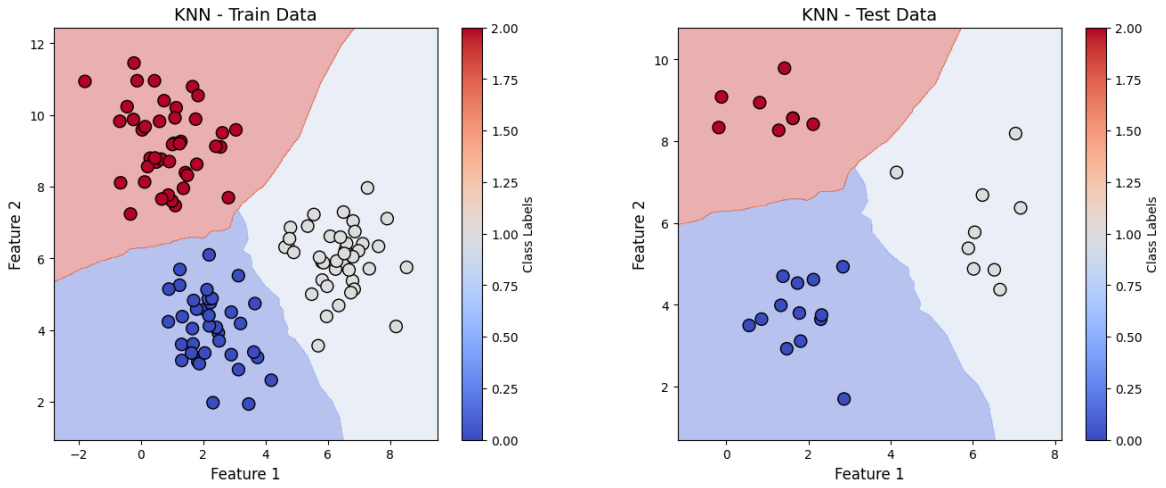
The dataset contains 150 samples grouped around three cluster centers: [2, 4], [6, 6], and [1, 9]. Each sample is described by two features, which makes it straightforward to plot and analyze the decision boundaries. The data was divided into 80% for training and 20% for testing.

KNN Classifier Parameters:

The K-Nearest Neighbors model was configured to use five neighbors (k=5) with the Euclidean distance measure. It was trained on the training portion of the dataset, and predictions were then generated for the test set.

Accuracy Results:

The accuracy score for the test data was calculated as follows: Accuracy on Test Data: 1.00 This indicates that the KNN model correctly classified 100% of the test samples, demonstrating strong performance on the modified dataset.



Explanation of the Output Plot:

The generated plots illustrate the decision boundaries formed by the KNN classifier on both the training and test sets. Each shaded region corresponds to the class label the model assigns within that area, while the data points are colored according to their true classes.

The training plot (left) shows that the classifier was able to separate the classes effectively, resulting in high accuracy.

The test plot (right) also demonstrates strong performance, though a few points near the boundary lines reveal the model's difficulty in handling overlapping regions. The shading across the plots visually represents how KNN establishes boundaries between classes, highlighting its ability to classify samples based on their closest neighbors.

Conclusion:

This analysis demonstrates that the KNN model is highly reliable when applied to artificial datasets with clearly defined cluster centers. The classifier delivered strong accuracy on both the training and testing sets, while the decision boundary visualizations provided meaningful insight into how the model classifies different regions of the feature space. Overall, the findings highlight KNN's effectiveness for classification problems where class boundaries are well-separated.