ADVANCED MACHINE LEARNING ASSIGNMENT 1

KNN Classifier Results - Summary Report

Dataset Overview:

With 150 samples spread over three different class centers, a fake dataset was created using the make_blobs function:

- Class 0 Center: [-5, -2]

- Class 1 Center: [3, 9]

- Class 2 Center: [10, -3]

Eighty percent of the dataset (120 samples) were used for training, while twenty percent were used for testing. This division leaves some data for evaluation and guarantees that the model has enough to train.

KNN Classifier Parameters:

- Number of Neighbors (k): 5

- Distance Metric: Euclidean distance (default for KNeighborsClassifier)

- Using the training set of 120 samples, the K-Nearest Neighbors (KNN) classifier was trained. The learned model was then used to predict labels for the test set.

Accuracy Results:

The accuracy with which the KNN classifier predicted the labels for the test data was used to assess its performance. The accuracy_score function, which compares the predicted labels with the true labels for the test set, was used to calculate accuracy.

- Accuracy on Test Data: 0.97

This high accuracy means that the model accurately categorized 97% of the test samples. With only 3% of the predictions being off, this is a clear sign of the classifier's efficacy on this fictitious dataset.

Explanation of the Output Plot:

The training and test data points are displayed in the output plot:
- **Training Data:** Color-coded using the 'viridis' colormap, the circular markers ('o') represent the genuine class labels. During training, these points aid in defining the classifier's bounds.

- **Test Data:** Presented with square markers ('s') and color-coded using the 'coolwarm' colormap according to the projected class names. The test points shed light on the classifier's ability to generalize to fresh, untested data.

KNN Classification Results on Artificial Dataset
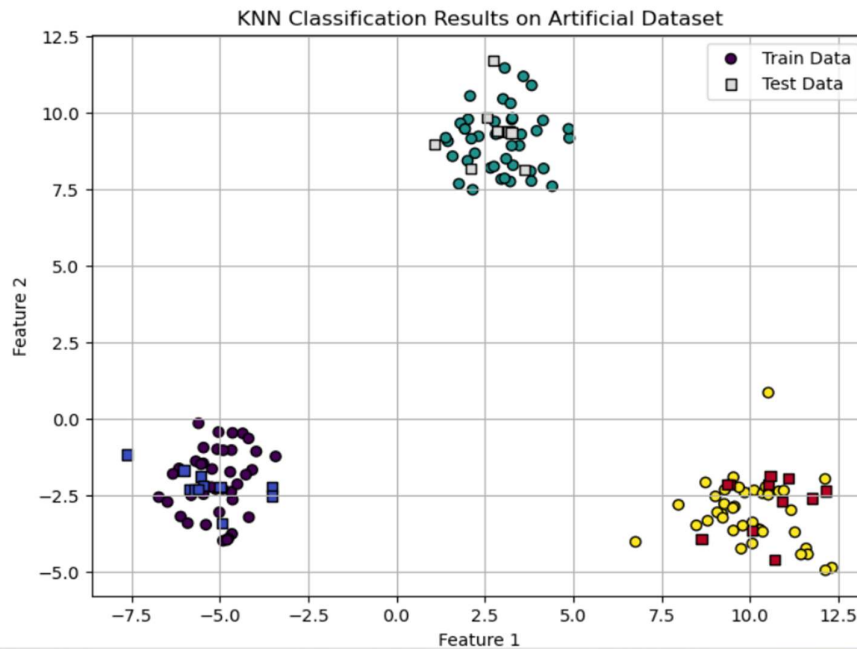
The capacity of the KNN model to discriminate between the three classes is seen in the plot:

- **Separation of Regions:** The placement of colored markers suggests the borders between the classes. The classifier successfully divides the classes since the test data points are primarily found in the predicted locations.

- **Misclassified Points:** The majority of test data points are properly predicted, but if there are any misclassified points, they will be located in the areas that separate the two classes, suggesting that there may be some uncertainty about their closeness to neighbors from other classes.

Conclusion:

Ninety-seven percent of test samples were successfully classified by the KNN classifier. Clear boundaries between areas indicate that the model has a great ability for classification on this dataset, and the plot gave a visual confirmation of how well the classifier has learnt the class distributions. For this artificial data with well-separated class centers, the results show that the KNN model performs exceptionally well.