

KNN Classifier Report

Name: Bitla Hari Krishna

Dataset Overview

I created a dataset with 150 points using the `make_blobs` function. The dataset had three centers located at $[-3, 2]$, $[7, -4]$, and $[5, 8]$. The dataset was then divided into 80 percent training and 20 percent testing. To visualize the split, training data points are shown in blue while testing data points are shown in red. This helped us clearly understand how the data was divided before applying the model.

KNN Classifier Setup

I used the K-Nearest Neighbors (KNN) method to classify the dataset. In the first model I set k equal to 5 neighbors, which is the default value. In the second model I changed k to 3 neighbors to see how the performance of the model would change. Both models were trained with the training data and then tested with the testing data.

Accuracy Results

The KNN model with k equal to 5 gave 100 percent accuracy on the test data. The model with k equal to 3 also gave 100 percent accuracy. This shows that in this dataset, the choice of k did not change the performance since all points were classified correctly.

Visual Results

The Train-Test Split plot should show how the dataset was divided into training and testing points. The Decision Boundary plot for k equal to 5 should show neat regions where the model predicted each class correctly. The Decision Boundary plot for k equal to 3 also should show clear boundaries and give the same result. These visualizations helped us understand how the KNN model makes predictions.

Explanation of the Plots

The Train-Test Split plot explained how the dataset was divided for training and testing purposes. The Decision Boundary plots show how the KNN model divides the space into regions for predicting class labels. Points that lie near the decision boundaries show that sometimes the model can be uncertain, but in this case the dataset was very well separated, so the accuracy reached 100 percent.

Conclusion

The KNN model performed perfectly on this dataset, reaching 100 percent accuracy for both $k=5$ and $k=3$. Changing the value of k did not affect the results since the dataset was simple and clearly separated. The visual plots gave us a clear picture of how the model separates the classes. From this experiment I can say that KNN is a simple and effective method for classifying data into groups.

