Amartejas Manjunath

1001742606
axm2606

Machine Learning

Project 3

Functions Used:

1. **Splitting():** This function prepares the dataset and splits the data into training and test data
2. **Kmeans():** This function performs the k-means clustering.
   First it finds the Euclidean between the random points chosen and the points in the train data.
   It finds the minimum distance and adds it to it's particular cluster.
   Then we find the new centroid depending on the average of the points present in the cluster.
   This process in repeated until the new centroid is equal to the old centroid.
   Now for the testing part, we take values from the test set and find the Euclidean distance for them from the centroid we have got from the testing part. With this result we compute the accuracy
3. **Main():** This function takes the number of clusters as an input from the user.
   Reads the iris file and calls the splitting() function
   It also finds the initial random centroids and calls the kmeans() function.
   This process takes place 5 times(iterations)

Accuracy:

K = 3

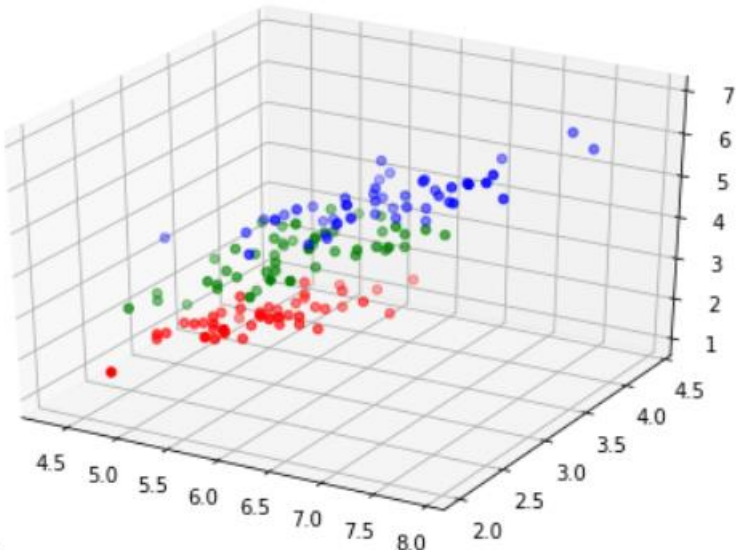
```
['6.4', '2.8', '5.6', '2.2', 'virginica'] = virginica
['6.9', '3.1', '5.4', '2.1', 'virginica'] = virginica
['6.5', '3.0', '5.8', '2.2', 'virginica'] = virginica
['6.7', '3.3', '5.7', '2.1', 'virginica'] = virginica
['4.9', '3.0', '1.4', '0.2', 'setosa'] = setosa
accuracy = 93.33333333333333
Iteration= 5
['5.4', '3.4', '1.5', '0.4', 'setosa'] = setosa
['5.9', '3.0', '5.1', '1.8', 'virginica'] = virginica
['5.1', '3.8', '1.6', '0.2', 'setosa'] = setosa
['6.4', '3.1', '5.5', '1.8', 'virginica'] = virginica
['5.5', '2.5', '4.0', '1.3', 'versicolor'] = versicolor
['5.4', '3.0', '4.5', '1.5', 'versicolor'] = versicolor
['6.8', '2.8', '4.8', '1.4', 'versicolor'] = versicolor
['5.2', '4.1', '1.5', '0.1', 'setosa'] = setosa
['4.6', '3.4', '1.4', '0.3', 'setosa'] = setosa
['6.7', '2.5', '5.8', '1.8', 'virginica'] = virginica
['5.7', '2.5', '5.0', '2.0', 'virginica'] = virginica
['5.7', '2.8', '4.1', '1.3', 'versicolor'] = versicolor
['5.2', '3.5', '1.5', '0.2', 'setosa'] = setosa
['6.2', '2.2', '4.5', '1.5', 'versicolor'] = versicolor
['7.9', '3.8', '6.4', '2.0', 'virginica'] = virginica
['6.6', '2.9', '4.6', '1.3', 'versicolor'] = versicolor
['7.0', '3.2', '4.7', '1.4', 'versicolor'] = virginica
['6.7', '3.1', '4.7', '1.5', 'versicolor'] = versicolor
['7.7', '2.6', '6.9', '2.3', 'virginica'] = virginica
['4.3', '3.0', '1.1', '0.1', 'setosa'] = setosa
['6.4', '2.8', '5.6', '2.2', 'virginica'] = virginica
['5.4', '3.4', '1.7', '0.2', 'setosa'] = setosa
['4.6', '3.1', '1.5', '0.2', 'setosa'] = setosa
['5.5', '2.3', '4.0', '1.3', 'versicolor'] = versicolor
['6.4', '2.8', '5.6', '2.1', 'virginica'] = virginica
['6.3', '2.5', '5.0', '1.9', 'virginica'] = virginica
['7.7', '3.8', '6.7', '2.2', 'virginica'] = virginica
['6.0', '2.7', '5.1', '1.6', 'versicolor'] = versicolor
['4.9', '3.0', '1.4', '0.2', 'setosa'] = setosa
['5.0', '3.4', '1.5', '0.2', 'setosa'] = setosa
accuracy = 96.66666666666667
AVG Accuracy 97.33333333333334
```

```
Enter the Number of clusters3
Iteration= 1
['4.4', '3.0', '1.3', '0.2', 'setosa'] = setosa
['6.7', '3.3', '5.7', '2.1', 'virginica'] = virginica
['5.5', '3.5', '1.3', '0.2', 'setosa'] = setosa
['6.4', '3.2', '5.3', '2.3', 'virginica'] = virginica
['6.7', '3.0', '5.2', '2.3', 'virginica'] = virginica
['5.8', '2.7', '5.1', '1.9', 'virginica'] = virginica
['5.4', '3.4', '1.7', '0.2', 'setosa'] = setosa
['4.3', '3.0', '1.1', '0.1', 'setosa'] = setosa
['6.3', '3.4', '5.6', '2.4', 'virginica'] = virginica
['6.6', '3.0', '4.4', '1.4', 'versicolor'] = versicolor
['6.7', '3.1', '4.7', '1.5', 'versicolor'] = versicolor
['4.8', '3.0', '1.4', '0.1', 'setosa'] = setosa
['5.1', '3.7', '1.5', '0.4', 'setosa'] = setosa
['6.1', '2.8', '4.7', '1.2', 'versicolor'] = versicolor
['4.7', '3.2', '1.3', '0.2', 'setosa'] = setosa
['4.8', '3.0', '1.4', '0.3', 'setosa'] = setosa
['6.4', '3.2', '4.5', '1.5', 'versicolor'] = versicolor
['6.4', '2.7', '5.3', '1.9', 'virginica'] = virginica
['5.1', '2.5', '3.0', '1.1', 'versicolor'] = versicolor
['4.4', '2.9', '1.4', '0.2', 'setosa'] = setosa
['6.3', '3.3', '6.0', '2.5', 'virginica'] = virginica
['6.7', '3.0', '5.0', '1.7', 'versicolor'] = virginica
['5.4', '3.9', '1.3', '0.4', 'setosa'] = setosa
['6.3', '2.5', '4.9', '1.5', 'versicolor'] = versicolor
['6.0', '3.0', '4.8', '1.8', 'virginica'] = virginica
['7.9', '3.8', '6.4', '2.0', 'virginica'] = virginica
['6.1', '2.9', '4.7', '1.4', 'versicolor'] = versicolor
['5.0', '3.5', '1.6', '0.6', 'setosa'] = setosa
['6.1', '2.8', '4.0', '1.3', 'versicolor'] = versicolor
['5.0', '2.3', '3.3', '1.0', 'versicolor'] = versicolor
accuracy = 96.66666666666667
```

When K = 5

```
Iteration= 5
['4.9', '2.4', '3.3', '1.0', 'versicolor'] = versicolor
['7.7', '3.0', '6.1', '2.3', 'virginica'] = virginica
['5.2', '3.4', '1.4', '0.2', 'setosa'] = setosa
['5.5', '2.3', '4.0', '1.3', 'versicolor'] = versicolor
['6.3', '2.3', '4.4', '1.3', 'versicolor'] = versicolor
['4.9', '3.1', '1.5', '0.1', 'setosa'] = setosa
['6.1', '3.0', '4.6', '1.4', 'versicolor'] = versicolor
['5.0', '2.0', '3.5', '1.0', 'versicolor'] = versicolor
['4.4', '3.2', '1.3', '0.2', 'setosa'] = setosa
['6.2', '2.2', '4.5', '1.5', 'versicolor'] = versicolor
['7.6', '3.0', '6.6', '2.1', 'virginica'] = virginica
['4.8', '3.1', '1.6', '0.2', 'setosa'] = setosa
['6.0', '2.7', '5.1', '1.6', 'versicolor'] = versicolor
['5.8', '2.6', '4.0', '1.2', 'versicolor'] = versicolor
['6.6', '3.0', '4.4', '1.4', 'versicolor'] = versicolor
['7.2', '3.2', '6.0', '1.8', 'virginica'] = virginica
['4.9', '3.1', '1.5', '0.1', 'setosa'] = setosa
['5.8', '2.7', '3.9', '1.2', 'versicolor'] = versicolor
['5.2', '2.7', '3.9', '1.4', 'versicolor'] = versicolor
['6.2', '2.8', '4.8', '1.8', 'virginica'] = virginica
['5.0', '3.2', '1.2', '0.2', 'setosa'] = setosa
['6.3', '2.5', '5.0', '1.9', 'virginica'] = virginica
['6.3', '3.4', '5.6', '2.4', 'virginica'] = virginica
['5.8', '2.7', '5.1', '1.9', 'virginica'] = virginica
['6.1', '2.9', '4.7', '1.4', 'versicolor'] = versicolor
['4.4', '2.9', '1.4', '0.2', 'setosa'] = setosa
['6.3', '2.7', '4.9', '1.8', 'virginica'] = virginica
['5.6', '3.0', '4.1', '1.3', 'versicolor'] = versicolor
['6.4', '2.8', '5.6', '2.1', 'virginica'] = virginica
['5.1', '3.8', '1.5', '0.3', 'setosa'] = setosa
accuracy = 100.0
AVG Accuracy 78.0
```

Scatter plot



The 3 different colors represent the 3 different species that we classify.