1. **Problem and Data**

Classification of Iris plants dataset.

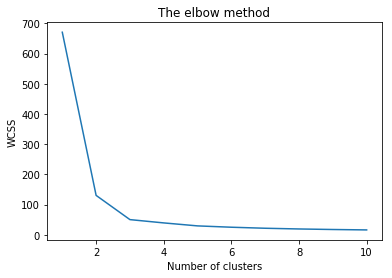
Number of Instances: 150 (50 in each of three classes) Number of Attributes: 4 Attribute Information: a) sepal length in cm b) sepal width in cm c) petal length in cm d) petal width in cm

Class Information: Iris-Setosa (-1) Iris-Versicolour (0) Iris-Virginica (1)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Summary Statistic: | | MIN | | MAX | | Mean | SD | | CORRELATION | |
| SEPAL LENGTH | 4.3 | | 7.9 | | 5.84 | | | 0.83 | | 0.7826 |
| SEPAL WIDTH | 2.0 | | 4.4 | | 3.05 | | | 0.43 | | -0.4194 |
| PETAL LENGTH | 1.0 | | 6.9 | | 3.76 | | | 1.76 | | 0.9490 |

1. **Evaluating the value of ‘K’**

We will implement 'The elbow method' on the Iris dataset. The elbow method allows us to pick the optimum amount of clusters for classification. This is the graph that we obtain from the elbow method. The optimum clusters is where the elbow occurs. This is when the within cluster sum of squares (WCSS) doesn't decrease significantly with every iteration. Hence we assign k = 3 in the code.



1. **Implementation and Results:**Here, I have split the data randomly in the following parts: 80% Clustering and 20% for evaluating. For K-means, we Initially set the centroid as 0 for all classes and features. We then calculate the Euclidian distance for the features and find the distance to the closest centroid and assign it. We iterate the data until the centroids don’t change. Then we assign the testing data set to the centroids and cross verify the results by comparing the results to the actual labels. We do this process 5 times and report the average accuracy.  
   The code gives output for the testing data in the format:  
   [feature1, feature2, feature3, feature4, actual\_label] = predicted\_label
2. **Sample Output of the Code**

Iteration # 1

['4.4', '2.9', '1.4', '0.2', '-1'] = -1

['5.9', '3', '4.2', '1.5', '0'] = 0

['6.7', '2.5', '5.8', '1.8', '1'] = 0

['5.1', '3.5', '1.4', '0.3', '-1'] = -1

['4.9', '2.4', '3.3', '1', '0'] = 0

['7.7', '2.6', '6.9', '2.3', '1'] = 0

['5.5', '2.4', '3.8', '1.1', '0'] = 0

['6.4', '2.9', '4.3', '1.3', '0'] = 0

['4.3', '3', '1.1', '0.1', '-1'] = -1

['4.9', '2.5', '4.5', '1.7', '1'] = 0

['6.9', '3.2', '5.7', '2.3', '1'] = 0

['5.4', '3.7', '1.5', '0.2', '-1'] = -1

['4.8', '3.4', '1.6', '0.2', '-1'] = -1

['6.4', '2.8', '5.6', '2.2', '1'] = 0

['5.7', '4.4', '1.5', '0.4', '-1'] = -1

['5', '3.5', '1.3', '0.3', '-1'] = -1

['6.4', '3.1', '5.5', '1.8', '1'] = 0

['6.4', '2.7', '5.3', '1.9', '1'] = 0

['4.8', '3', '1.4', '0.1', '-1'] = -1

['6', '2.2', '5', '1.5', '1'] = 0

['5.1', '2.5', '3', '1.1', '0'] = -1

['5.6', '3', '4.1', '1.3', '0'] = 0

['5.8', '2.7', '3.9', '1.2', '0'] = 0

['7.7', '2.8', '6.7', '2', '1'] = 0

['5', '2.3', '3.3', '1', '0'] = 0

['6.7', '3.1', '5.6', '2.4', '1'] = 0

['6.7', '3', '5', '1.7', '0'] = 0

['5.5', '3.5', '1.3', '0.2', '-1'] = -1

['6.1', '3', '4.9', '1.8', '1'] = 0

['6.3', '2.5', '4.9', '1.5', '0'] = 0

accuracy = 60.0

################################

Iteration # 2

['6.2', '2.2', '4.5', '1.5', '0'] = 0

['6.1', '2.6', '5.6', '1.4', '1'] = 1

['4.9', '3.1', '1.5', '0.1', '-1'] = -1

['5.3', '3.7', '1.5', '0.2', '-1'] = -1

['7.2', '3', '5.8', '1.6', '1'] = 1

['5.6', '2.5', '3.9', '1.1', '0'] = 0

['6.1', '2.8', '4', '1.3', '0'] = 0

['5.4', '3.9', '1.3', '0.4', '-1'] = -1

['6.8', '3.2', '5.9', '2.3', '1'] = 1

['5.6', '3', '4.5', '1.5', '0'] = 0

['5.4', '3.4', '1.5', '0.4', '-1'] = -1

['5.8', '2.8', '5.1', '2.4', '1'] = 1

['7', '3.2', '4.7', '1.4', '0'] = 1

['5.1', '3.8', '1.5', '0.3', '-1'] = -1

['5.8', '2.7', '4.1', '1', '0'] = 0

['6.8', '3', '5.5', '2.1', '1'] = 1

['5.5', '3.5', '1.3', '0.2', '-1'] = -1

['5.7', '2.6', '3.5', '1', '0'] = 0

['4.9', '3', '1.4', '0.2', '-1'] = -1

['5.5', '4.2', '1.4', '0.2', '-1'] = -1

['5.2', '3.4', '1.4', '0.2', '-1'] = -1

['5.7', '2.5', '5', '2', '1'] = 1

['5', '3.4', '1.5', '0.2', '-1'] = -1

['5.2', '4.1', '1.5', '0.1', '-1'] = -1

['6.3', '2.8', '5.1', '1.5', '1'] = 0

['6.7', '2.5', '5.8', '1.8', '1'] = 1

['7.2', '3.6', '6.1', '2.5', '1'] = 1

['6.5', '3.2', '5.1', '2', '1'] = 1

['6.3', '2.7', '4.9', '1.8', '1'] = 1

['7.1', '3', '5.9', '2.1', '1'] = 1

accuracy = 93.33333333333333

################################

Iteration # 3

['5', '3', '1.6', '0.2', '-1'] = -1

['5.1', '3.4', '1.5', '0.2', '-1'] = -1

['5.2', '3.4', '1.4', '0.2', '-1'] = -1

['4.8', '3.1', '1.6', '0.2', '-1'] = -1

['6.4', '3.2', '4.5', '1.5', '0'] = 0

['7.1', '3', '5.9', '2.1', '1'] = 1

['7.2', '3', '5.8', '1.6', '1'] = 1

['6.3', '3.4', '5.6', '2.4', '1'] = 1

['5.4', '3', '4.5', '1.5', '0'] = 0

['5.9', '3', '5.1', '1.8', '1'] = 1

['4.9', '2.4', '3.3', '1', '0'] = 0

['6.3', '2.9', '5.6', '1.8', '1'] = 1

['7.7', '2.8', '6.7', '2', '1'] = 1

['5.1', '3.7', '1.5', '0.4', '-1'] = -1

['6.7', '3.1', '4.7', '1.5', '0'] = 0

['6.7', '2.5', '5.8', '1.8', '1'] = 1

['5.9', '3.2', '4.8', '1.8', '0'] = 1

['6.2', '2.8', '4.8', '1.8', '1'] = 1

['6.3', '2.7', '4.9', '1.8', '1'] = 1

['5', '3.3', '1.4', '0.2', '-1'] = -1

['7.9', '3.8', '6.4', '2', '1'] = 1

['5.2', '4.1', '1.5', '0.1', '-1'] = -1

['5.4', '3.7', '1.5', '0.2', '-1'] = -1

['4.6', '3.1', '1.5', '0.2', '-1'] = -1

['4.9', '3.1', '1.5', '0.1', '-1'] = -1

['6.4', '2.7', '5.3', '1.9', '1'] = 1

['4.4', '3', '1.3', '0.2', '-1'] = -1

['5', '2', '3.5', '1', '0'] = 0

['6.2', '2.2', '4.5', '1.5', '0'] = 0

['6.7', '3.1', '5.6', '2.4', '1'] = 1

accuracy = 96.66666666666667

################################

Iteration # 4

['4.8', '3.1', '1.6', '0.2', '-1'] = -1

['4.9', '3.1', '1.5', '0.1', '-1'] = -1

['5.8', '2.7', '3.9', '1.2', '0'] = 0

['7.7', '2.8', '6.7', '2', '1'] = 1

['7.3', '2.9', '6.3', '1.8', '1'] = 1

['6.7', '3.1', '4.4', '1.4', '0'] = 0

['5.1', '3.8', '1.9', '0.4', '-1'] = -1

['5.8', '2.6', '4', '1.2', '0'] = 0

['6.8', '3', '5.5', '2.1', '1'] = 1

['6.1', '3', '4.6', '1.4', '0'] = 0

['4.4', '2.9', '1.4', '0.2', '-1'] = -1

['4.9', '3', '1.4', '0.2', '-1'] = -1

['5.4', '3.4', '1.7', '0.2', '-1'] = -1

['6.9', '3.1', '5.4', '2.1', '1'] = 1

['5.7', '3.8', '1.7', '0.3', '-1'] = -1

['4.8', '3.4', '1.6', '0.2', '-1'] = -1

['5.1', '3.8', '1.5', '0.3', '-1'] = -1

['6.8', '3.2', '5.9', '2.3', '1'] = 1

['6.3', '3.3', '6', '2.5', '1'] = 1

['4.7', '3.2', '1.3', '0.2', '-1'] = -1

['5.1', '2.5', '3', '1.1', '0'] = 0

['4.6', '3.1', '1.5', '0.2', '-1'] = -1

['4.5', '2.3', '1.3', '0.3', '-1'] = -1

['5.5', '2.4', '3.8', '1.1', '0'] = 0

['5.1', '3.5', '1.4', '0.2', '-1'] = -1

['5.5', '2.4', '3.7', '1', '0'] = 0

['5.7', '4.4', '1.5', '0.4', '-1'] = -1

['6.1', '3', '4.9', '1.8', '1'] = 1

['7.7', '3.8', '6.7', '2.2', '1'] = 1

['4.4', '3.2', '1.3', '0.2', '-1'] = -1

accuracy = 100.0

################################

Iteration # 5

['5.5', '2.4', '3.8', '1.1', '0'] = 0

['6.4', '2.8', '5.6', '2.2', '1'] = 1

['6.4', '2.8', '5.6', '2.1', '1'] = 1

['5.7', '2.8', '4.5', '1.3', '0'] = 0

['5.7', '4.4', '1.5', '0.4', '-1'] = -1

['6.9', '3.2', '5.7', '2.3', '1'] = 1

['5', '2', '3.5', '1', '0'] = 0

['6.6', '3', '4.4', '1.4', '0'] = 0

['5.1', '3.5', '1.4', '0.3', '-1'] = -1

['5.5', '3.5', '1.3', '0.2', '-1'] = -1

['5.6', '3', '4.5', '1.5', '0'] = 0

['5.6', '2.8', '4.9', '2', '1'] = 0

['6.5', '3', '5.5', '1.8', '1'] = 1

['4.9', '3.1', '1.5', '0.1', '-1'] = -1

['6.4', '2.7', '5.3', '1.9', '1'] = 1

['5.3', '3.7', '1.5', '0.2', '-1'] = -1

['7.9', '3.8', '6.4', '2', '1'] = 1

['6.3', '2.5', '4.9', '1.5', '0'] = 0

['7.3', '2.9', '6.3', '1.8', '1'] = 1

['4.4', '2.9', '1.4', '0.2', '-1'] = -1

['5', '3.2', '1.2', '0.2', '-1'] = -1

['4.8', '3.1', '1.6', '0.2', '-1'] = -1

['7.1', '3', '5.9', '2.1', '1'] = 1

['5.4', '3.7', '1.5', '0.2', '-1'] = -1

['6.8', '3.2', '5.9', '2.3', '1'] = 1

['5.6', '2.9', '3.6', '1.3', '0'] = 0

['4.9', '3.1', '1.5', '0.1', '-1'] = -1

['6.6', '2.9', '4.6', '1.3', '0'] = 0

['5.6', '3', '4.1', '1.3', '0'] = 0

['5', '3', '1.6', '0.2', '-1'] = -1

accuracy = 96.66666666666667

################################

average accuracy = 89.33333333333334