Balaji Paiyur Mohan
1001576836

Structure of Code:

- First, the iris data file is copied to an iris csv file and headers 'SepalLength', 'SepalWidth', 'PetalLength', 'PetalWidth', and 'Species' are added to the data points for accessing the data points.
- Features of the iris are stored in a list while categories are stored in a separate list.
- Since there are only three classes, we initialize k to 3.
- We find the centroids of the cluster using the mean and standard deviation formula which includes error.
- We remove the error by normalizing the distance between each centroid and data points and assign it to the closest centroid and repeating this until the error comes to zero.
- Output displays the number of data points that are assigned to an incorrect cluster.
- Plotting the data points and centroids in the graph and saving it in an external file 'KMeans.png'

Screenshot of the output:

Number of data points incorrectly clustered:
52