Problem Sections:

Loading the Data:

We have to load the given data either from the url(http://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data) or from the text file.

iris.data.txt

Separating the data into X and Y matrices:

Once we have the data, we can separate that data into x matrix which includes the parameters of x and Y matrix which contains the target values.

Applying K-means:

Once we have X and Y matrix, we need to apply k-means on X_matrix. For that, we need to randomly generate centroids based on the K_value. Then, we need to calculate the distance between each data point and each centroid by using the following formulae

$$\boldsymbol{distance} = \sqrt{(\boldsymbol{x1} - \boldsymbol{c1})^2 + (\boldsymbol{x2} - \boldsymbol{c2})^2 + \cdots}.$$

Once you find the distance you need to assign each point to its nearest centroid. After that, move the centroid to the center of all the points under it. And repeat this whole process for 300 times.

Calculate the error and accuracy:

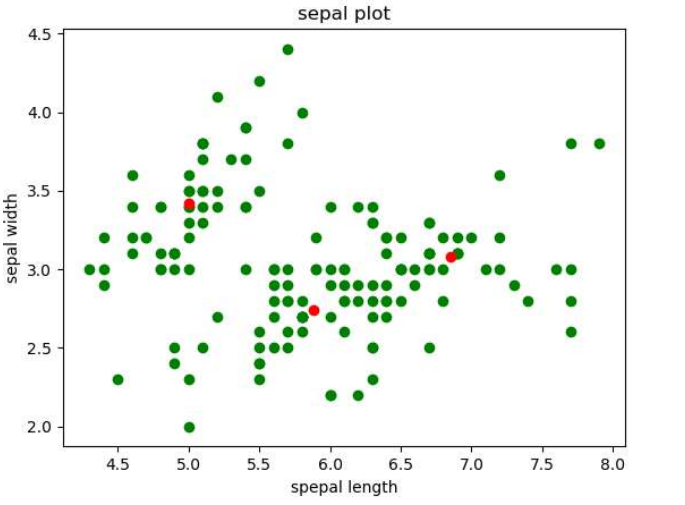
Once you have the predicted values you can calculate the error by using the formulae

Accuracy can be calculated as 100 – error.

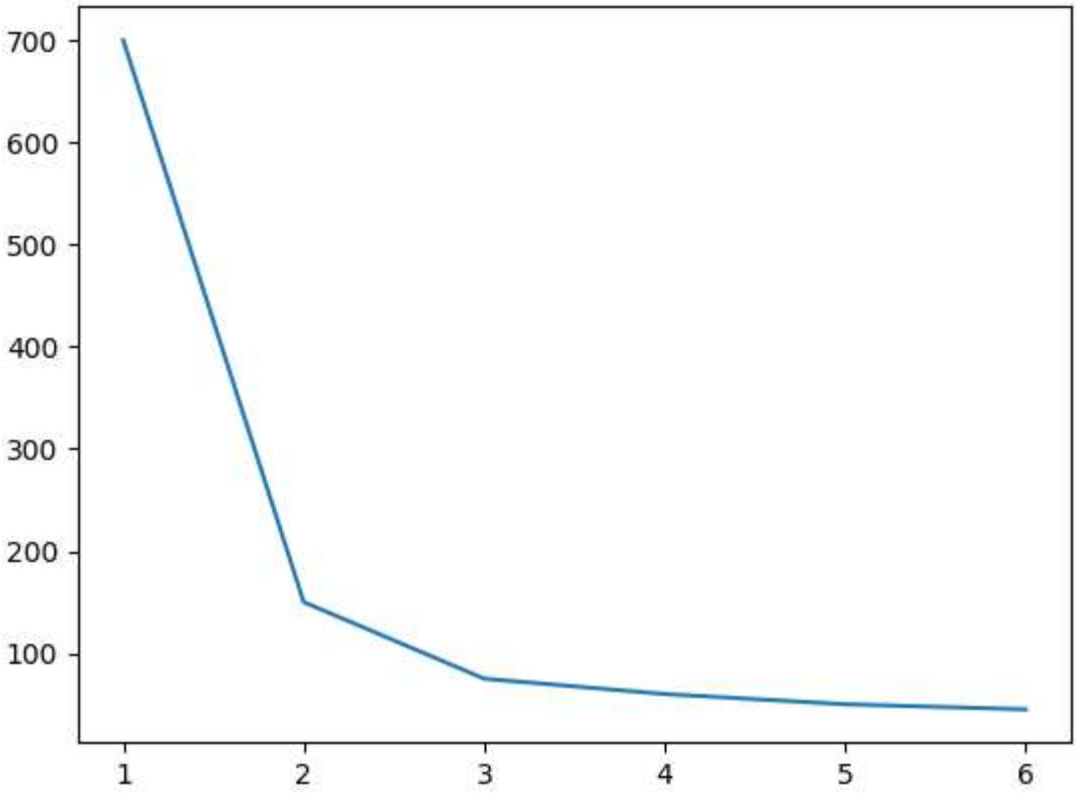
The average accuracy of the program is 86%

Results:

Here is the output of the program when used k =3 which is plotted only using two features.



K-value:



By applying elbow method the K-value is chosen as 3 from the above figure.