

Kevin Thomas

CSE 6361-001

1001544593

## PROJECT -3 REPORT

### Structure of the Code

Packages imported are,

pandas , numpy ,random, pyplot

1) The data is loaded from the external link given . After downloading the data before splitting we are changing the value for the string into integers. egs: 'Iris-setosa'  $\rightarrow$  1 using replace function.

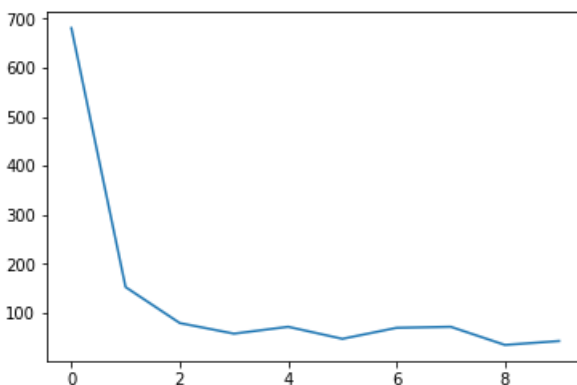
2) Then we shuffle the data so that we can obtain more accuracy while doing clustering

3)We find the centroids for the given data and classifying each points into each classes based on their distance from the centroid

4)Thus we implement the K means Clustering.

### Result:

The N value I have chosen here is 3. The reason I have chosen 3 is as it is efficient and its proven by the Elbow Method which was also given in the code Elbow.py .



### ELBOW METHOD

To run the code , it can be ran on any python IDE.