Name: Aman Bassi

StudentId: 1001393217

This report is pertaining to the third project that is K-means clustering on the iris data set.

DataSet:-

I am working on the irisdataset which has 4 features namely sepal length, sepal width, petal length, petal width and 3 classes of iris flower namely  :-

* Iris Setosa   
  -- Iris Versicolour   
  -- Iris Virginica

I have converted this classes to 1,2,3.

Results: -

Since its k means clustering so we can not expect same accuracy every time as it is unstable so sometimes it is coming 33, 65.33 and so on

Algorithm:-

It involves 3 steps :-

1. Initial assignments of centroid which I have generated randomly from dataset only so that my algorithm converges easily.
2. Computing distances from the centroids and assigning clusters.
3. Computing centroids new for the clusters

We keep on doing it unless the clusters or centroid gets constant.

I have taken k =3 for this dataset:

1. First of all, I already know that this dataset has 3 classes so its better to cluster the datapoints in 3 clusters for good results.
2. Secondly, data is uniformly distributed among three classes.
3. I also check through running different k values.

Setup:

You can simply run it any ide like Pycharm with 3.6 interpretor.