***Assignment 3***

***Nikhil Lahoti***

***012448514***

***Answer 1:***

1. ***K Means Algorithm:***
2. Initially k random centroids are selected
3. For each point in the dataset, the point is assigned to the cluster of the nearest centroid
4. For all the clusters formed, the new centroids are recomputed in the same cluster
5. If the centroids changes, go to step 2 and perform the same step with the new centroids
6. Stop
7. For the input data provided, following test cases that should be considered
8. When a point is equidistant from the two clusters, it is assigned to the cluster which is first in the list.
9. When the selected points are on the same sides i.e. are very near to one another (i.e. they should ideally be in the same clusters in the result)
10. When the selected points are on the different sides i.e. are far from one another (i.e. they should ideally be in the different clusters in the result)

a) The algorithm does not work well with clusters of varied sizes and different densities.

b) It does not consider globular structure and breaks the clusters. For e.g. in the moon blob dataset, it breaks the global clusters.

***Please refer the file kMeans.py for the code***