

# REPORT

## TASK-2

### Answer1:

The code has been run for  $k = 2, 3, 5, 7, 10, 15, 30, 45$ . In all these cases, the SSE didn't increase with iterations. Although, the rate of decrease isn't uniform.

The decrease in SSE wrt iterations is to be expected as the algorithm assigns better clustering (in terms of distance) every iteration and as mean points are chosen for the clustering during every iteration.

### Answer 2:

For mouse.csv, it is expected that the clusters would be the face and two ears. The algorithm has been tested for different seed. The algorithm does a fairly good job clustering the face and ears. We obtain a linear boundary between clusters ( as we are assigning to the closest centroid). Because the ears have a lesser number of points than the face, the ear cluster also contains a significant portion of the face points.

For 3lines.csv, it is expected that each line forms a cluster. But the algorithm does a clustering in which the extremes of the lines form a cluster and the central portion of the lines form another cluster. This is because, the initialization, for all seeds, happens to be at the extremes of the line points. Henceforth, the algorithm finds the local optimal clustering in which clusters are assigned comprising of the extreme and the central portions of the lines. The more appropriate clustering occurs if the initialization assigns points at the center of each line. Since, this is more unlikely, the algorithm doesn't produce the manually drawn clustering.

