

CSL7020 Assignment 3

K-MEANS CLUSTERING

Dataset:

Generating randomly 120 points in 3D Space (X, Y, Z) or Range in between (0,100)

Approach:

- 1) Using numpy we generated random points and calculate the mean of centroids.
- 2) Using Panda Library we make the data frame of generating points.
- 3) Using AXES3D and matplotlib we plotted data in 3D space.

Model Description:

- 1) We first generated four color specific random points (Centroids)
- 2) Calculate the distance of each point from all 4 centroids.
- 4) Closest point from centroid takes in same color or same cluster
- 4) Update the cluster points to minimize the mean of all points in same cluster
- 5) We update the centroids till there is no change in cluster.

Accuracy:

Created 4 clusters of 120 points.

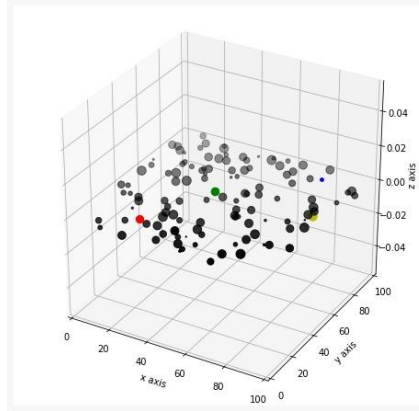
Number of points in each cluster (AT Particular one instance):

- 1) Red Cluster: 37
- 2) Green Cluster: 30
- 3) Blue Cluster: 29
- 4) Yellow cluster: 24

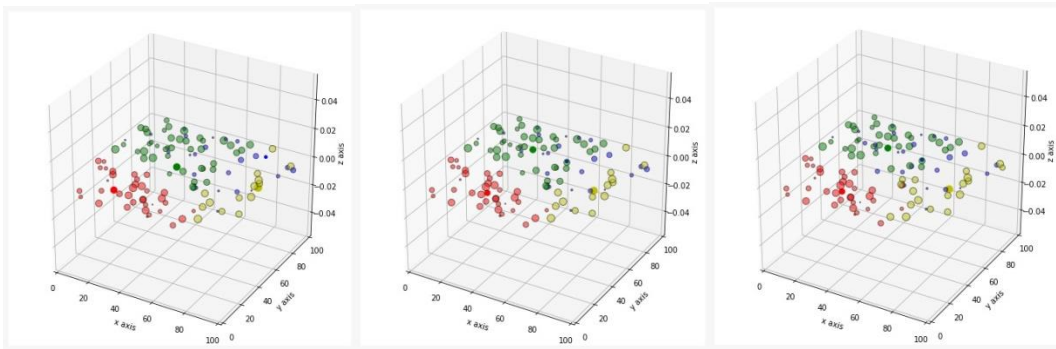
PLOTS:

K-MEANS CLUSTERING:

1) DATA SET:



2) INTERMEDIATE IMAGE:



3) FINAL IMAGE:

