Quiz 1

K Means

1. Points A (7,50), B(23,34), C(12,12)

Similarity $d_{AC} = \sqrt{(7-12)^2 + (50-12)^2} = \sqrt{5^2 + 38^2} = 38.33$

$d_{BC} = \sqrt{(23-12)^2 + (34-12)^2} = \sqrt{11^2 + 22^2} = 24.6$

Based on euclidean distance as Points B,C less distance
→ B point closer to C

2. Centroid of 5 observations

$$\text{Centroid } (C_x, C_y, C_z) = \left(\frac{12+31+17+19+13}{5}, \frac{23+31+15+27+11}{5}, \frac{45+31+25+45+27}{5}\right)$$

$$= (18.4, 21.4, 34.6)$$

3. $d_{AD} = \sqrt{1^2+1^2} = \sqrt{2}$

$d_{BD} = \sqrt{3^2+3^2} = 3\sqrt{2}$

$d_{CD} = \sqrt{5^2+0^2} = 5$

Point (1,2) belongs to cluster A(2,3)

4. Chebyshev distance

$$d_C(P,Q) = \max(|P_i - Q_i|)$$

$$= \max(2, 4, 1, 1)$$

Chebyshev distance $= 4$

5. It is not convergence

1. Convergence occurs when the centroid points don't change or vary for several iterations.

2. The given clusters can further be changed and the points in cluster also change. So it is not in convergence state.

3. The vertical cluster centroid is much near to end point of horizontal cluster than its orginal cluster.

4. So, convergence still did not occur.

7.

Convergence Criterion:

→ No re-assignments of data points of clusters.

That means centroids position doesnot change.

→ Minimum decrease of SSE (sum of squared errors)

$$SSE = \sum_{j=1}^{K} \sum_{x \in C_j} d(x, m_j)^2$$

C_j is j^{th} cluster,

m_j is centroid of cluster C_j,

$d(x, m_j)$ is euclidean distance.

8. True

Since each run of k-means is independent multiple runs can find Different local optima and this can help in choosing the global optimum value.

9. From given 3 images

a) Intra cluster distance

Clustering in images 1,2 same intra cluster distance
And comparatively image 3 has more intra cluster distance than 1,2.

b) Inter cluster distance.

Clustering in images 1,2,3 can be seen to say that image 1,3 has more inter cluster distance than image 2.

10.

Silhouette coefficient	clusters
close to 1	correctly clustered
Equal to 0	Overlapping clusters
close to -1	Wrongly clustered

11. Data standardization:

$$x_{scaled} = \frac{x - mean}{SD}$$

ID	Height	Weight
1	−0.5	−2
2	1	−1

12. a) Assign points 1 (1,4) , 2 (1,3) as centroids

cluster 1 1,3,

cluster 2 2,4,5,6

centroids 1 = (0.5, 4)

Centroid 2 = (4, 1.5)

b) Reassign points to cluster

cluster 1 = 1,2,3

cluster 2 = 4,5,6

Centroid 1 = $(2/3, 11/3)$

Centroid 2 = $(5, 1)$

c) Re computing clusters

Cluster 1 = 1,2,3

Cluster 2 = 4,5,6

Same cluster centroids