# Assignment 10
## Introduction to Machine Learning
Prof. B. Ravindran

1. Which among the following is/are some of the assumptions made by the k-means algorithm (assuming Euclidean distance measure)?

   (a) Clusters are spherical in shape

   (b) Clusters are of similar sizes

   (c) Data points in one cluster are well separated from data points of other clusters

   (d) There is no wide variation in density among the data points

2. Consider the similarity matrix given below.

	P1	P2	P3	P4	P5	P6
P1	1.00	0.70	0.65	0.40	0.20	0.05
P2	0.70	1.00	0.95	0.70	0.50	0.35
P3	0.65	0.95	1.00	0.75	0.55	0.40
P4	0.40	0.70	0.75	1.00	0.80	0.65
P5	0.20	0.50	0.55	0.80	1.00	0.85
P6	0.05	0.35	0.40	0.65	0.85	1.00

   Show the hierarchy of clustering created by the single-link clustering algorithm.

P1 P6 P5 P2 P4 P3

(a)

P1 P2 P3 P4 P5 P6

(b)

P1 P6 P4 P2 P5 P3

(c)

P1 P2 P3 P4 P5 P6

(d)
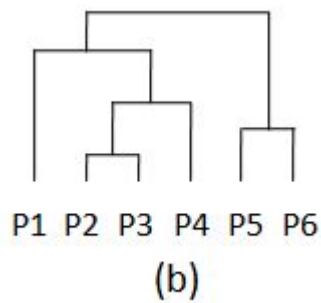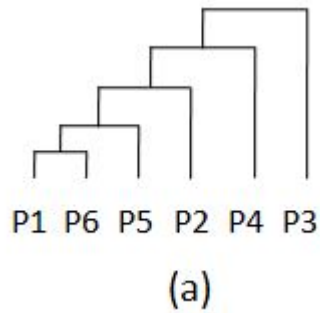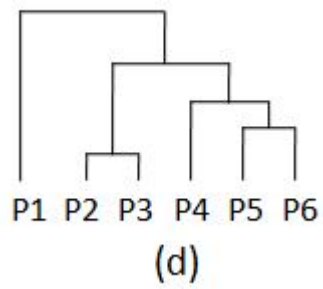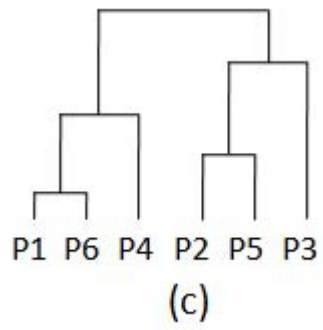
3. For the above similarity matrix, show the hierarchy of clustering obtained on performing complete-link clustering.

P1 P6 P5 P2 P4 P3

(a)

P1 P2 P3 P4 P5 P6

(b)

P1 P6 P4 P2 P5 P3

(c)


P1 P2 P3 P4 P5 P6

(d)
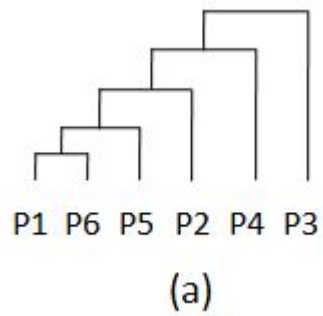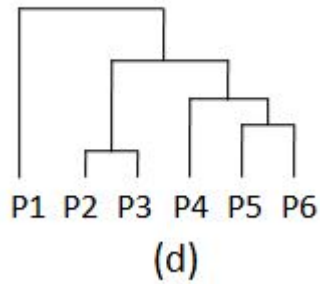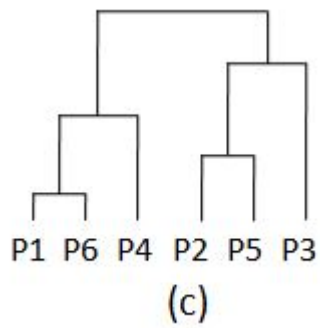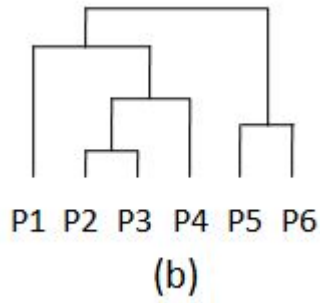
4. Considering single-link and complete-link hierarchical clustering, is it possible for a point to be closer to points in other clusters than to points in its own cluster? If so, in which approach will this tend to be observed?
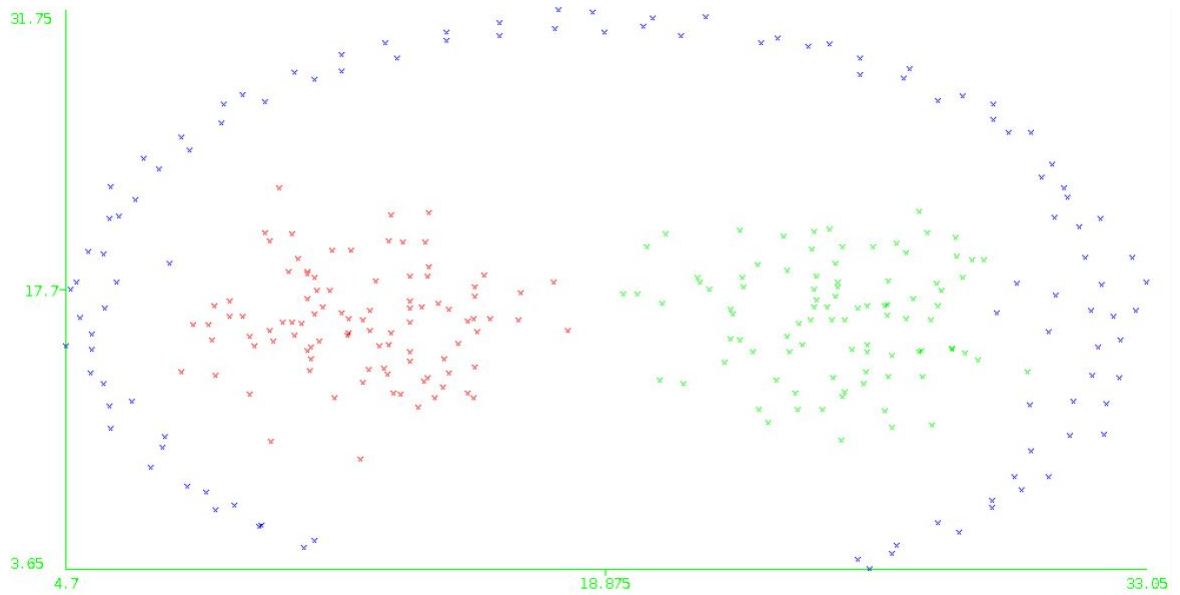
   (a) No

   (b) Yes, single-link clustering

   (c) Yes, complete-link clustering

   (d) Yes, both single-link and complete-link clustering

5. A graph is said to be k-connected if there does not exist a set of k-1 vertices whose removal disconnects the graph. If we define clusters as comprising of k-connected components of the thresholded graphs, does this result in a well-defined clustering algorithm?

   (a) Yes
   (b) No

6. A set of nodes forms a p-cluster, if at least p percentage of the edges from the nodes in the set go to another node in the set. If we define clusters as comprising of p-clusters of the thresholded graphs, does this result in a well-defined clustering algorithm?

   (a) Yes
   (b) No

7. In the CURE clustering algorithm, representative points of a cluster are moved a fraction of the distance between their original location and the centroid of the cluster. Would it make more sense to move them all a fixed distance towards the centroid instead? Why or why not?

   (a) Yes, because this approach will ensure that the original cluster shape is preserved.
   (b) No, because this approach will not be as effective against outliers as the original approach.

8. Suppose while performing DBSCAN we randomly choose a point which has less than MinPts number of points in its neighbourhood. Which among the following is true for such a point?

   (a) It is treated as noise, and not considered further in the algorithm
   (b) It becomes part of its own cluster
   (c) Depending upon other points, it may later turn out to be a core point
   (d) Depending upon other points, it may be density connected to other points

9. Consider the following image showing data points belonging to three different clusters (indicated by the colours of the points). Which among the following clustering algorithms will perform well in accurately clustering the given data?
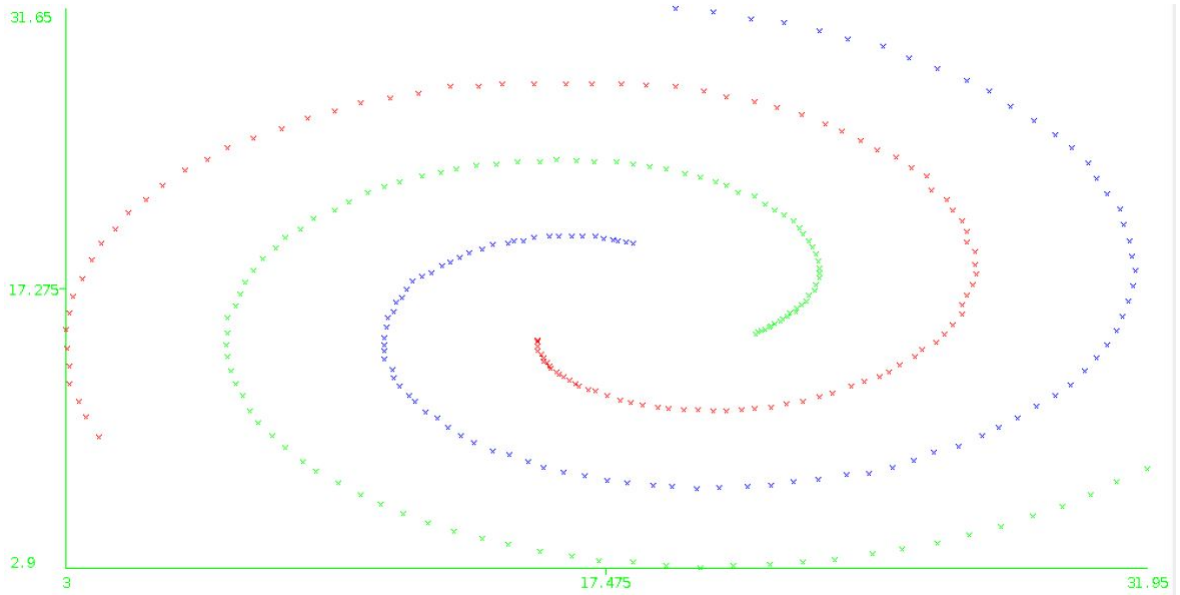
(a) K-means

(b) Single-link hierarchical

(c) Complete-link hierarchical

(d) DBSCAN

10. Consider the following image showing data points belonging to three different clusters (indicated by the colours of the points). Which among the following clustering algorithms will perform well in accurately clustering the given data?

5

(a) K-means

(b) Single-link hierarchical

(c) Complete-link hierarchical

(d) DBSCAN