Q1. Present the clustering of Example 7.2 if we use the minimum of the distances between any two points (one from each cluster) as the distance between two clusters.
(These points live in a 2-dimensional Euclidean space, and each point is named by its (x, y) coordinates.)

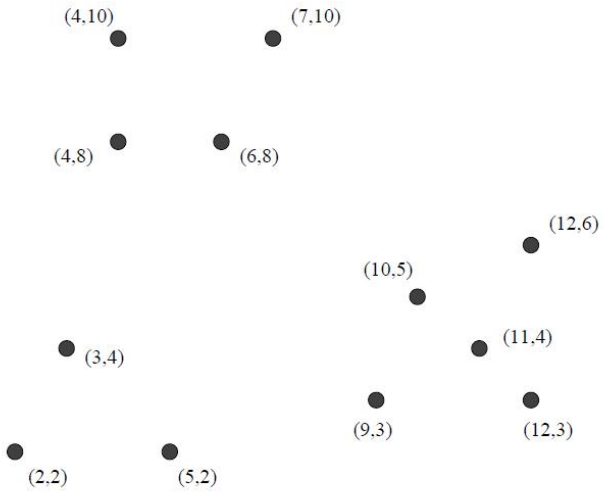

Figure 7.2: Twelve points to be clustered hierarchically

Q2.Let us consider the twelve points of Fig. 7.2, which we reproduce here as Fig. 7.8. In the worst case, our initial choice of a point is near the center, say (6,8).

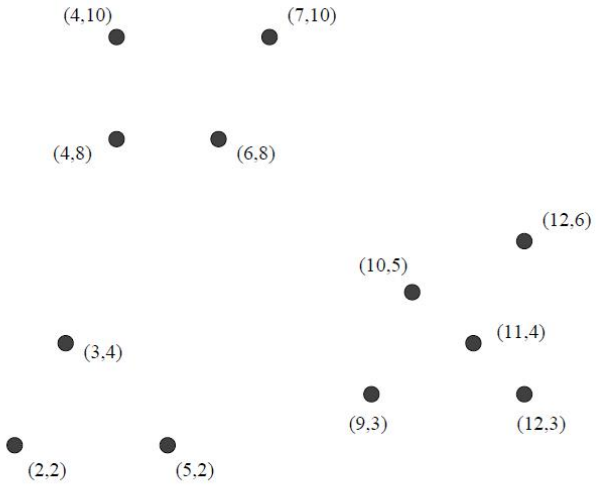

Figure 7.8: Repeat of Fig. 7.2

A. Find the other 2 starting point for the clustering by seeking the point with the largest minimum distance to the selected starting point(s).
B. Compute the representation of the cluster as in the BFR Algorithm. That is, compute N, SUM, and SUMSQ.
C. Compute the variance and standard deviation of each cluster in each of the two dimensions.

Q3.Suppose a cluster of three-dimensional points has standard deviations of 2, 3, and 5, in the three dimensions, in that order. Compute the Mahalanobis distance between the origin (0, 0, 0) and the point (1,−3, 4).

Q4.The following figure gives a matrix with six rows.

Element	S1	S2	S3	S4
0	0	1	0	1
1	0	1	0	0
2	1	0	0	1
3	0	0	1	0
4	0	0	1	1
5	1	0	0	0

A. Compute the permutations by using the following three hash functions respectively:

$$h1(x) = (2x + 1) \bmod 6$$
$$h2(x) = (3x + 2) \bmod 6$$
$$h3(x) = (5x + 2) \bmod 6$$

Indicate which of these permutations is true permutation and explain why?

B. Compute the Minhash signature for each column if we use the three permutations in A.

C. How close are the estimated Jaccard similarities for the six pairs of columns to the true Jaccard similarities?