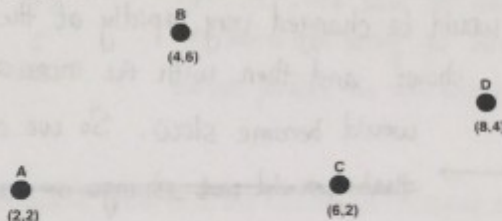


Name: Yijun LinUSC ID: 3689281438INF 553 – Spring 2016
(10 Points)

10/10

1. (3 pts) Consider **Hierarchical Clustering** of the data points at the coordinates below. Assume you know in advance that the **final number of clusters for this data set is 2**.



- a) What two points will be clustered first? Show mathematically how you would make this determination (i.e., don't just cluster points by visual inspection).

$$|AB| = \sqrt{(4-2)^2 + (6-2)^2} = 2\sqrt{5}$$

$$|AC| = \sqrt{(6-2)^2 + (2-2)^2} = 4$$

$$|AD| = \sqrt{(8-2)^2 + (4-2)^2} = 2\sqrt{10}$$

$$|BC| = \sqrt{(6-4)^2 + (6-2)^2} = 2\sqrt{5}$$

$$|BD| = \sqrt{(8-4)^2 + (6-4)^2} = 2\sqrt{5}$$

$$|CD| = \sqrt{(8-6)^2 + (4-2)^2} = 2\sqrt{2}$$

\therefore CD will be clustered first

- b) What is the centroid of the new cluster?

$$\text{new centroid is } \left(\frac{6+8}{2}, \frac{2+4}{2} \right) = (7, 3)$$

- c) What is the **final clustering**? Show mathematically how you determine this.

A (2, 2)

B (4, 6)

new centroid D_1 (7, 3)

$$AD_1 = \sqrt{(7-2)^2 + (3-2)^2} = \sqrt{26}$$

$$BD_1 = \sqrt{(7-4)^2 + (6-3)^2} = 3\sqrt{2}$$

$$AB = 2\sqrt{5}$$

\therefore BD₁ will be clustered together

so final clusters will be

A is a cluster with centroid (2, 2)

B, C, D is a cluster with centroid (6, 4)