Data Mining

Clustering II – Agglomerative Hierarchical Algorithm (Part A)

Dr. Jason T.L. Wang, Professor

Department of Computer Science

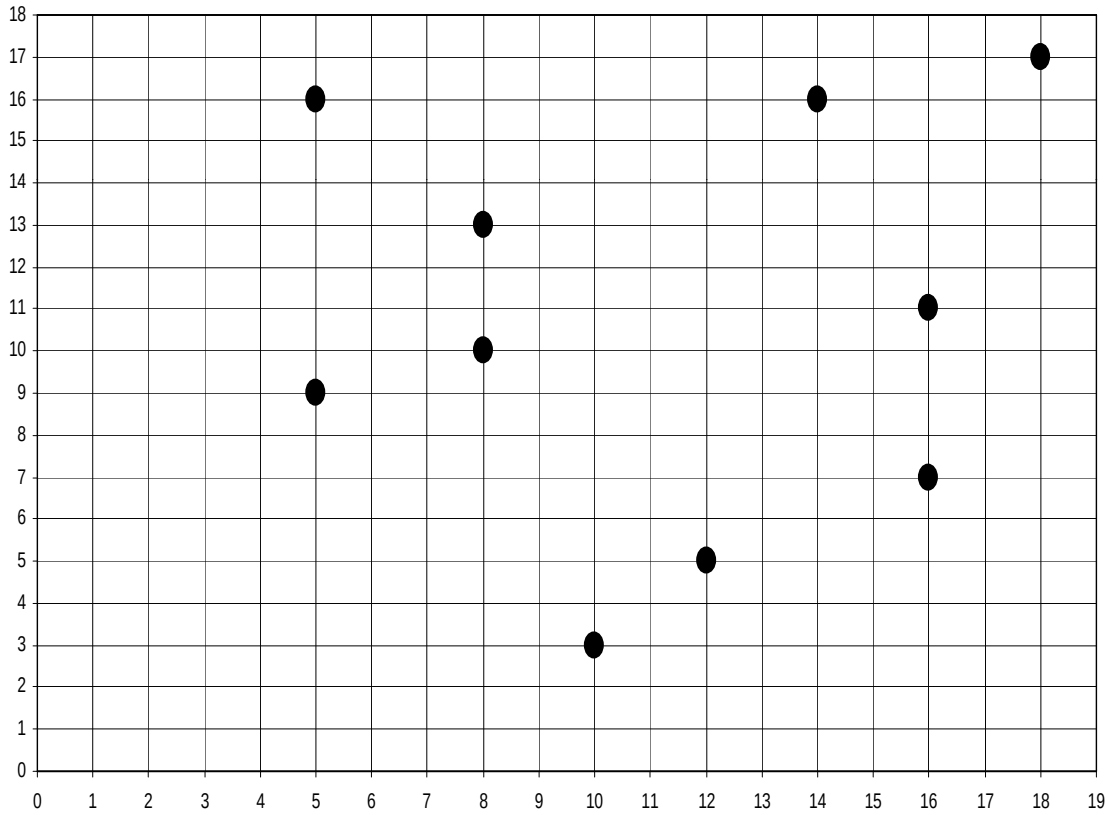
New Jersey Institute of Technology

/

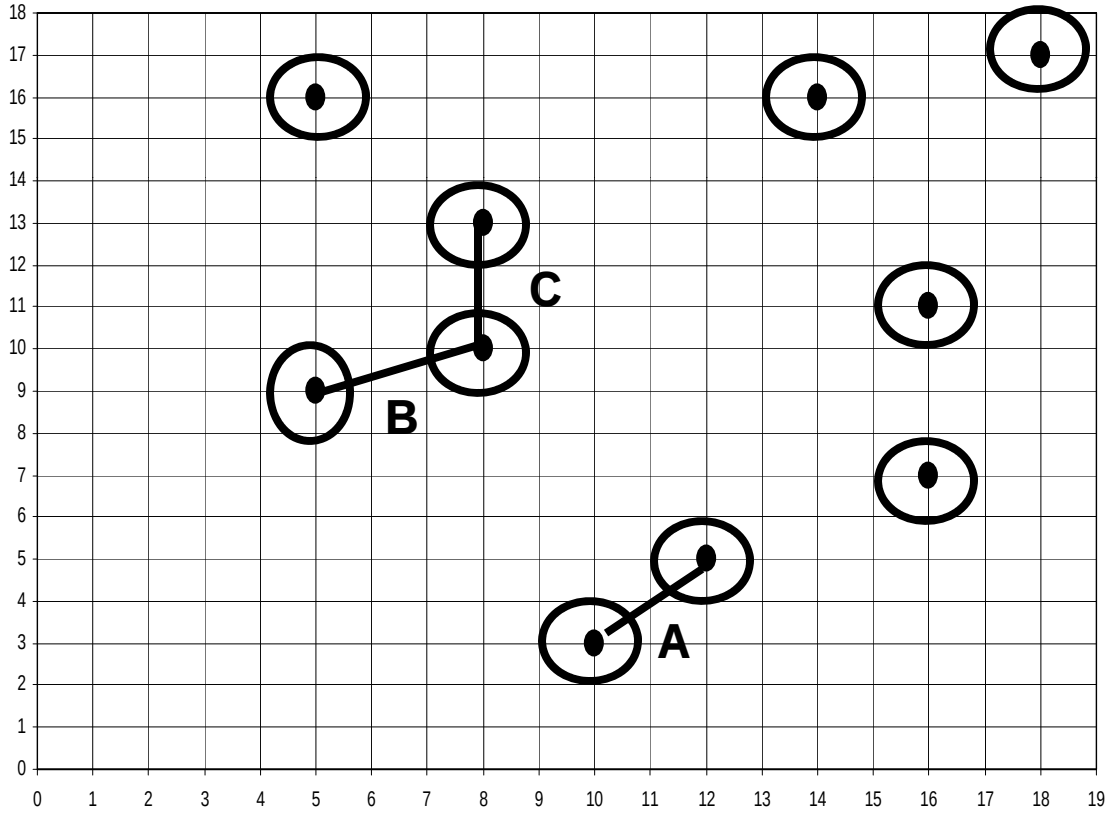
The Agglomerative Algorithm (a hierarchical clustering method)

- Initially, each object is a cluster.
- Merge two clusters that are closest. The distance between two clusters C1 and C2 is defined as the minimum distance between an object O1 in C1 and an object O2 in C2.
- Repeat the above step until a user-specified condition is met (e.g. k clusters are obtained).

Agglomerative Clustering



Agglomerative Clustering

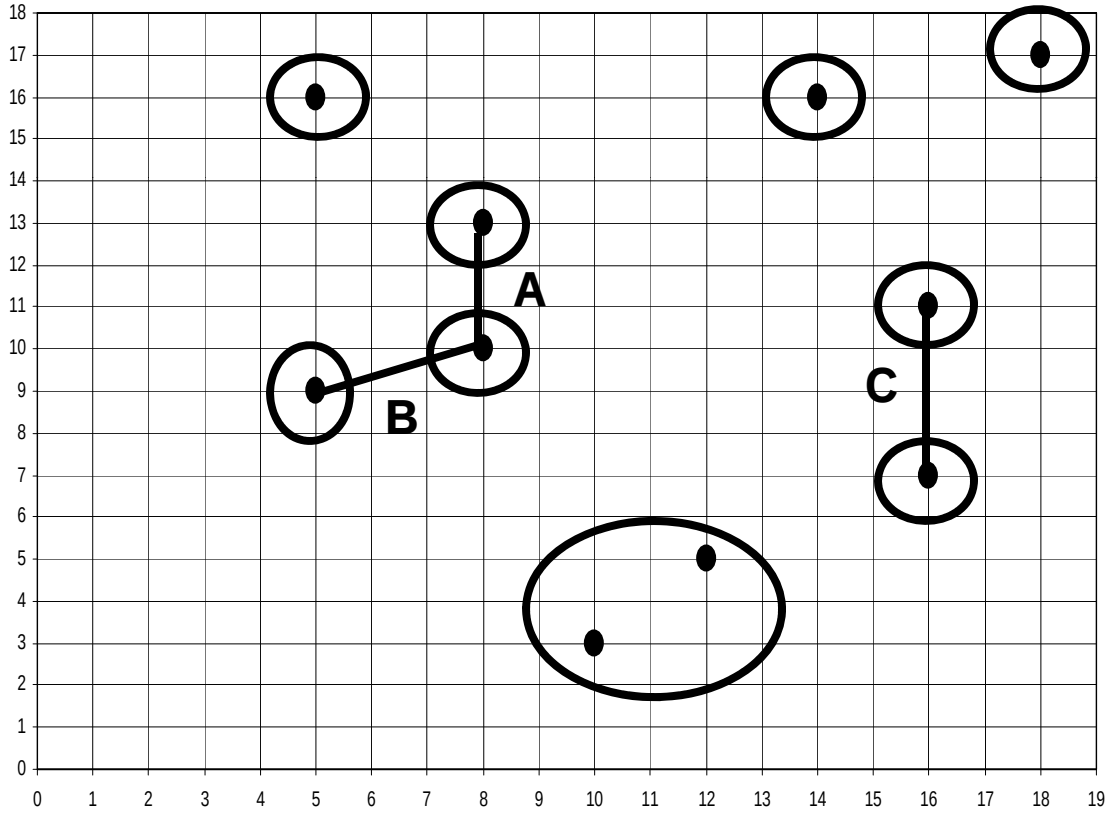


$$A = \sqrt{2^2 + 2^2} = 2.83$$

$$B = \sqrt{3^2 + 1^2} = 3.16$$

$$C = 3$$

Agglomerative Clustering

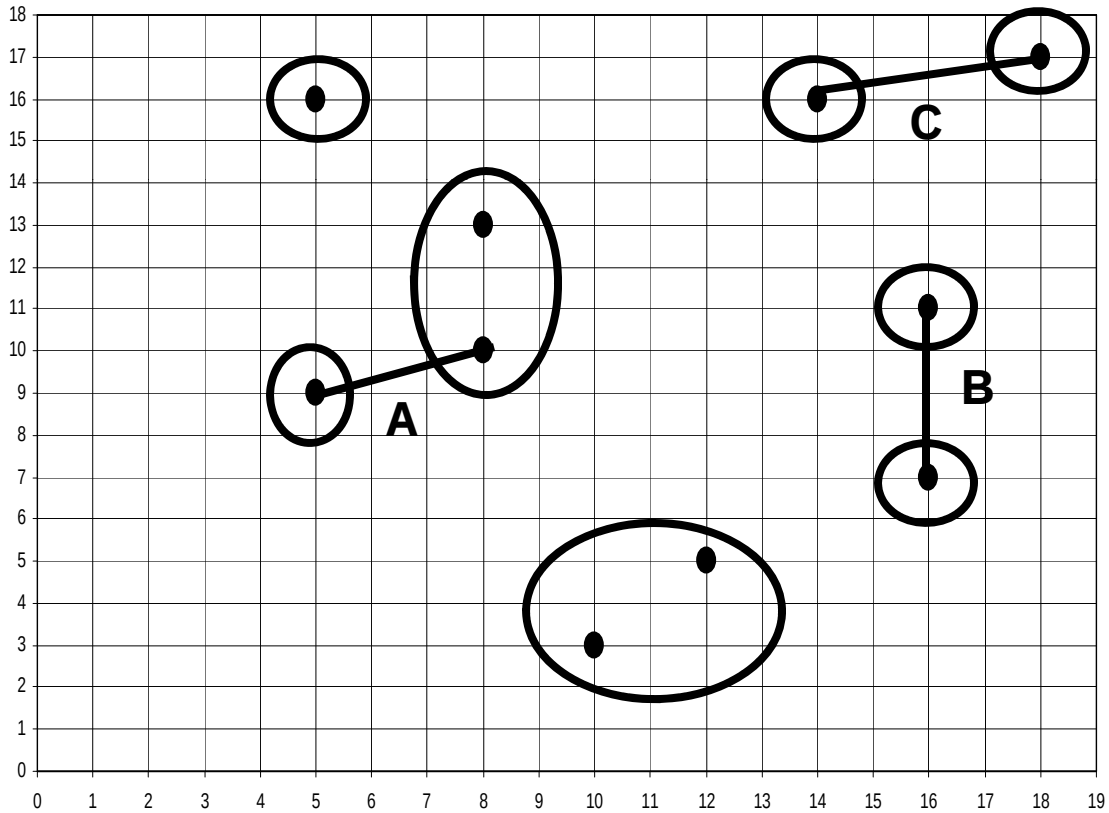


$$A = 3$$

$$B = \sqrt{3^2 + 1^2} = 3.16$$

$$C = 4$$

Agglomerative Clustering

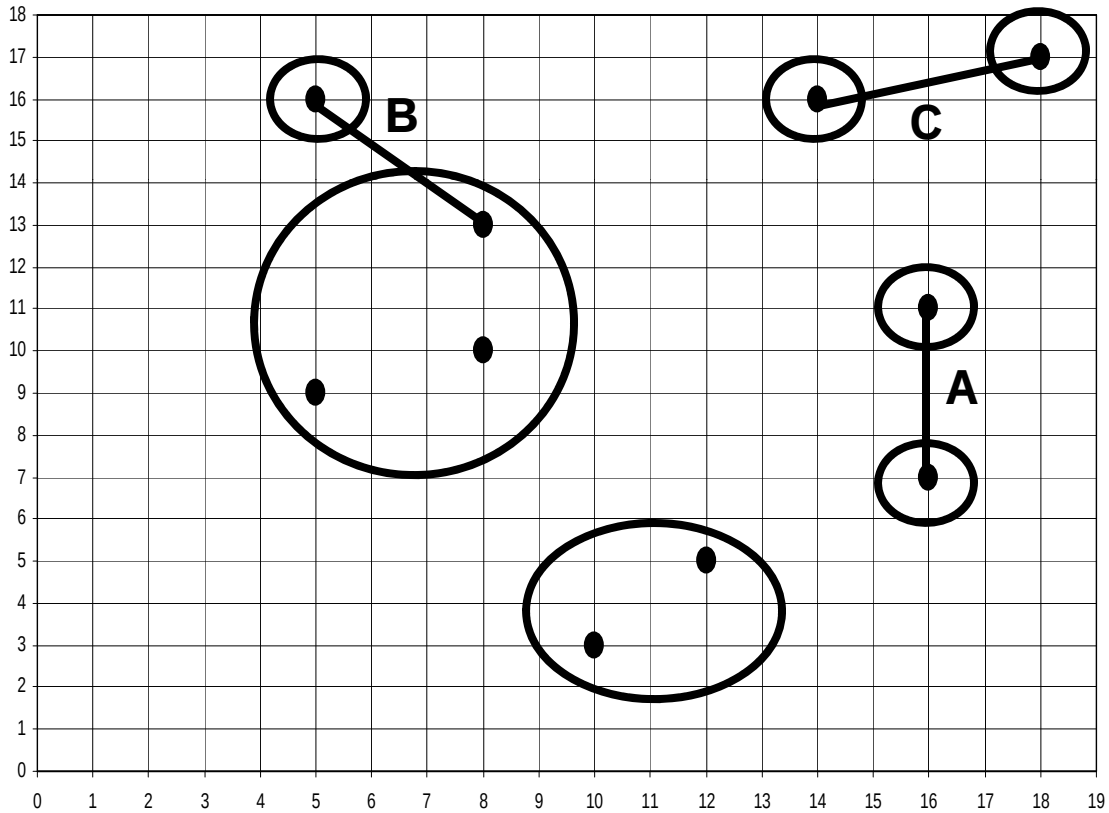


$$A = \sqrt{3^2 + 1^2} = 3.16$$

$$B = 4$$

$$C = \sqrt{4^2 + 1^2} = 4.12$$

Agglomerative Clustering

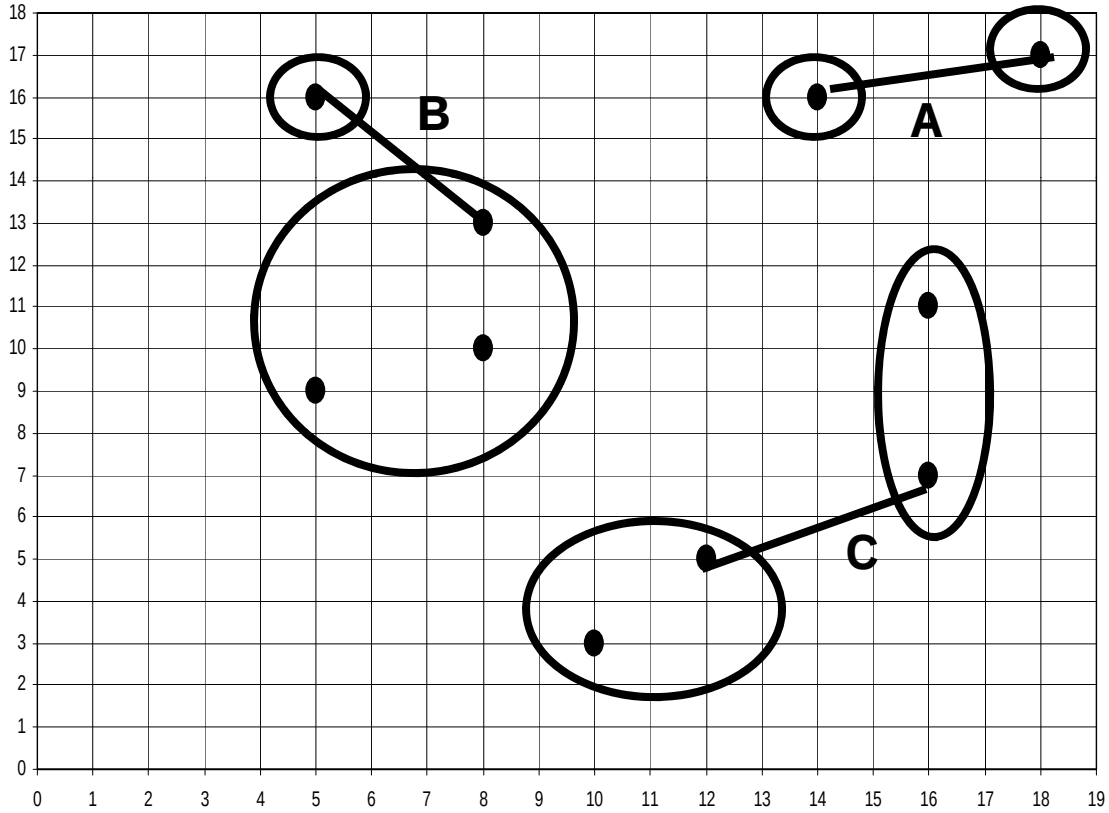


$$A = 4$$

$$B = \sqrt{3^2 + 3^2} = 4.24$$

$$C = \sqrt{4^2 + 1^2} = 4.12$$

Agglomerative Clustering

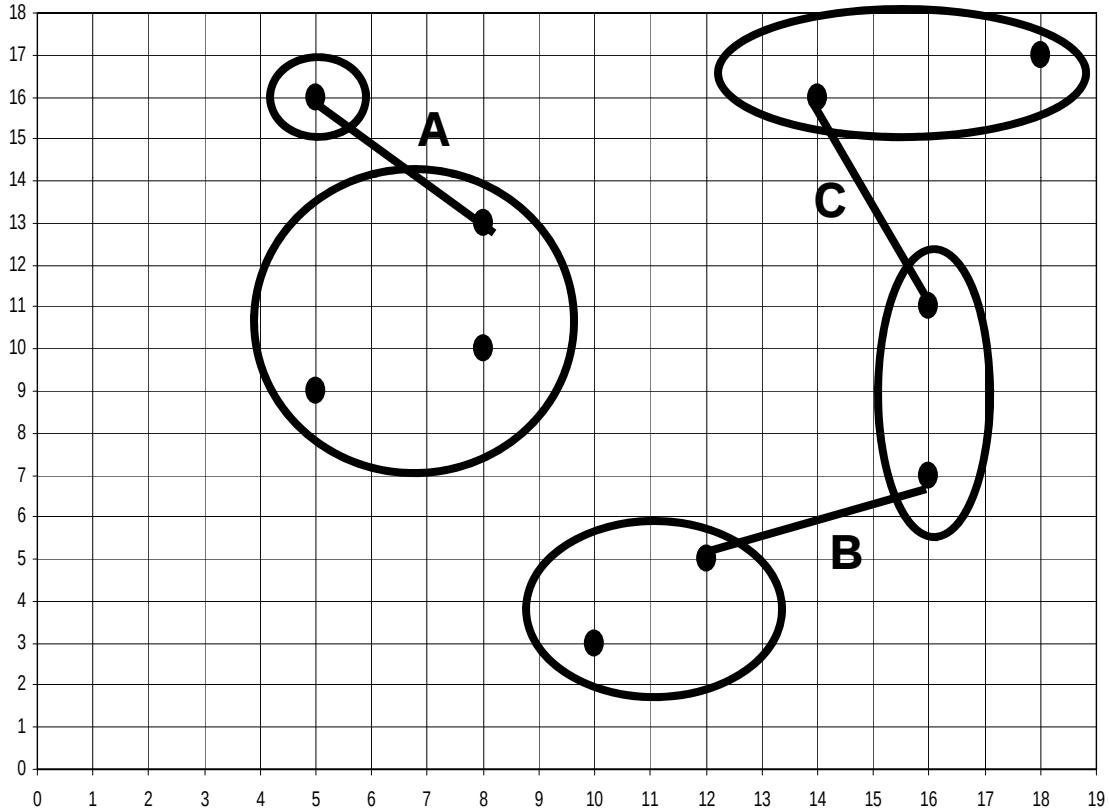


$$A = \sqrt{4^2 + 1^2} = 4.12$$

$$B = \sqrt{3^2 + 3^2} = 4.24$$

$$C = \sqrt{4^2 + 2^2} = 4.47$$

Agglomerative Clustering

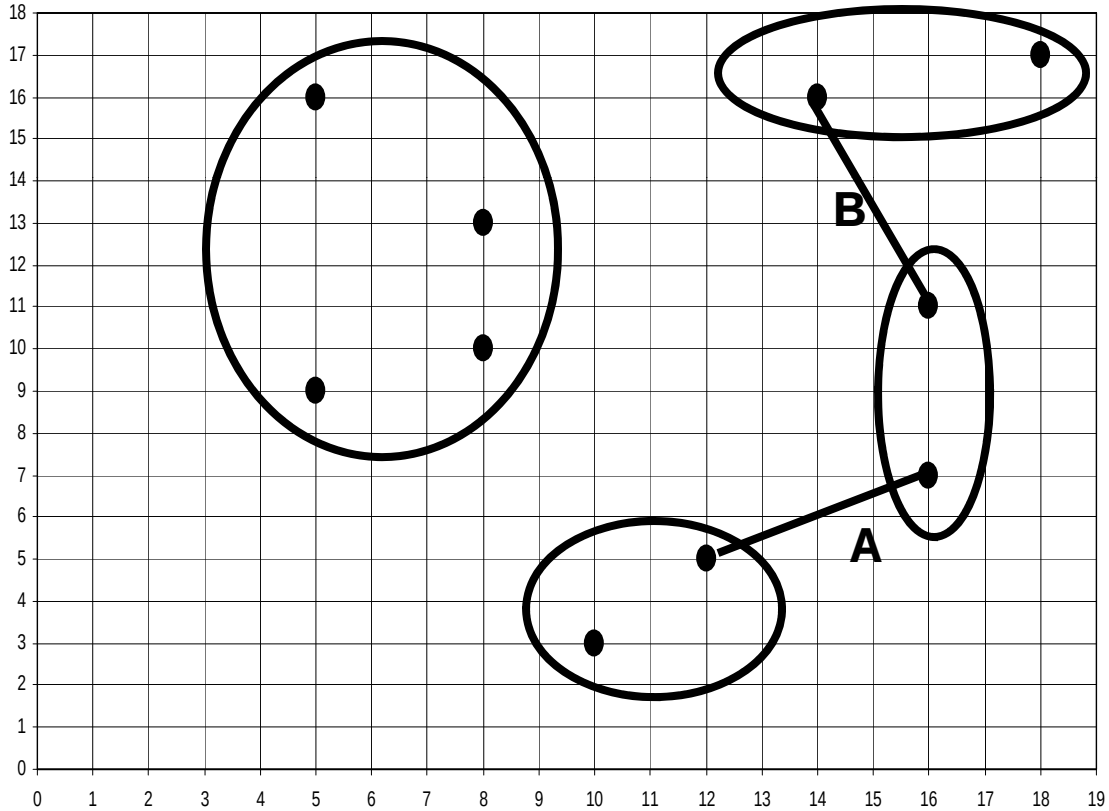


$$A = \sqrt{3^2 + 3^2} = 4.24$$

$$B = \sqrt{4^2 + 2^2} = 4.47$$

$$C = \sqrt{2^2 + 5^2} = 5.39$$

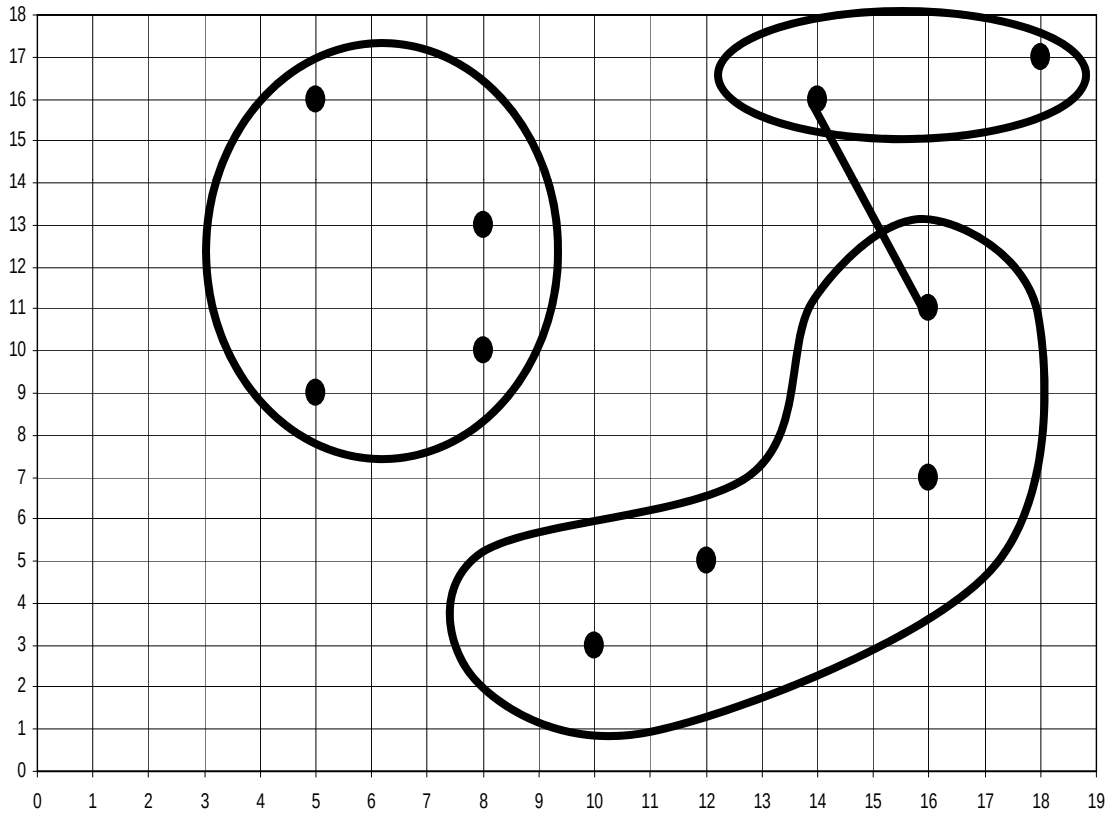
Agglomerative Clustering



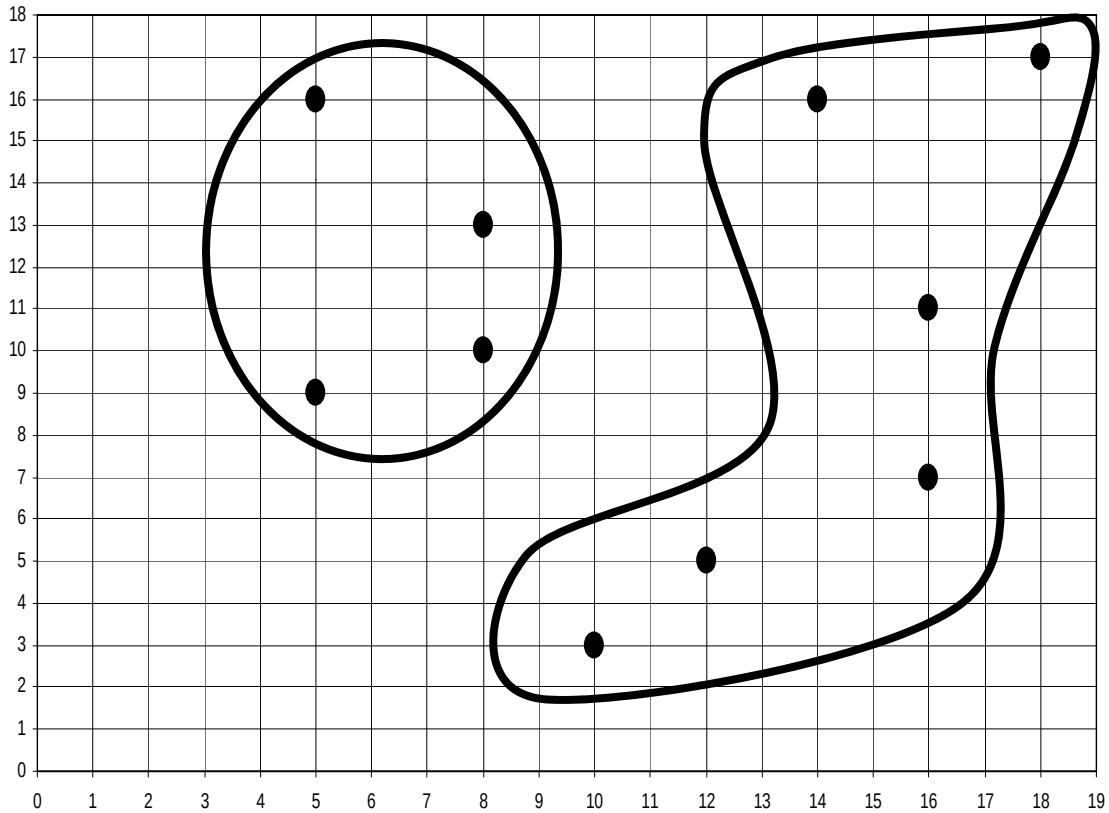
$$A = \sqrt{4^2 + 2^2} = 4.47$$

$$B = \sqrt{2^2 + 5^2} = 5.39$$

Agglomerative Clustering



Agglomerative Clustering



End of Agglomerative Hierarchical Clustering Algorithm Module (Part A)