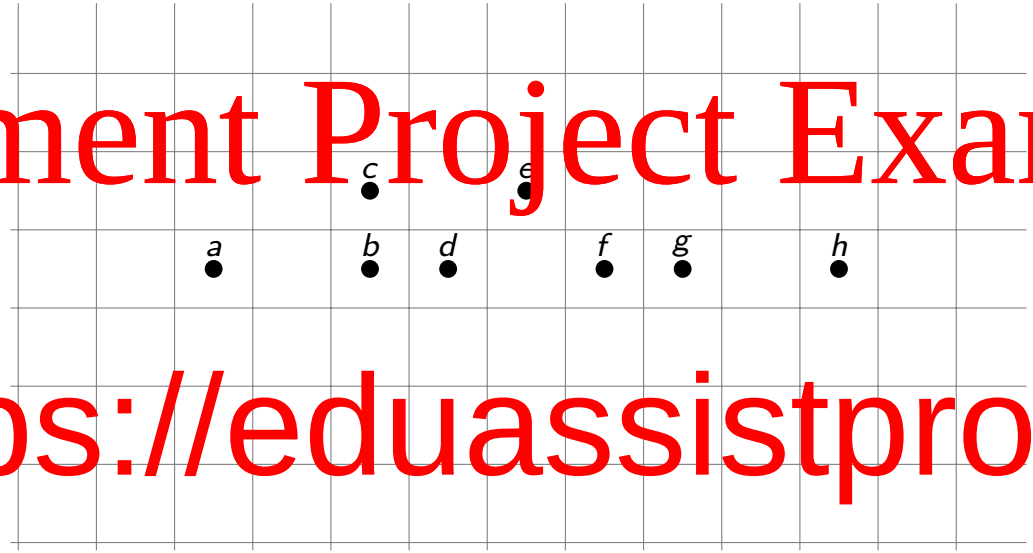
Q1 I

Consider eight tuples represented as points in the two dimensional space as follows:



Assume that (1) each point lies within the center of the grid; (2) uniform partition of the data space; and (3) each grid is a square of length 1.

We consider applying DBSCAN clustering algorithm on this dataset. Specifically, we set the minimum number of points (excluding the point in the center) within the ϵ-neighborhood (*MinPts*) to be 3, the radius of the ϵ-neighborhood (ϵ) to be 2, and we adopt the **Manhattan Distance** metric in the computation (i.e., a point p is within the ϵ-neighborhood of point o if and only if the Manhattan distance between p and o is no larger than ϵ).

Q1 II

1. What is the Manhattan distance between point *a* and *e*?
2. List all the *core objects*.
3. What is the clustering result of the DBSCAN algorithm on this dataset if points are accessed following the alphabetical order? You need to write out all the cluster (with points that belonging to the cluster) as well as the outliers (if any).

Q2 I

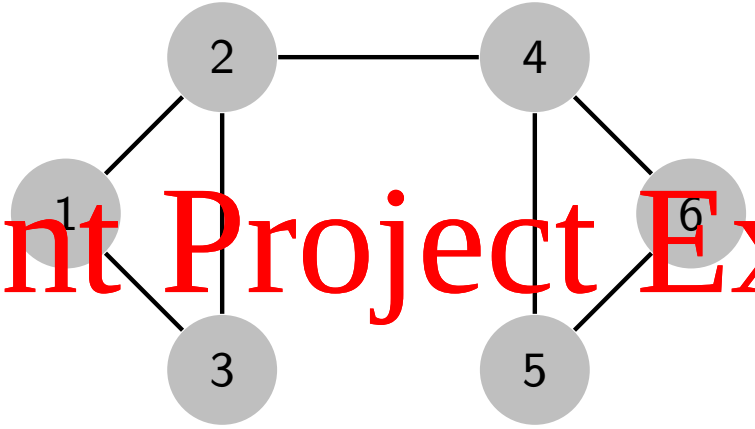
Consider the same dataset as Q1. What is the result of applying **centroid-based** hierarchical clustering algorithm on the dataset (still using the Manhattan distance)?

Q3 I

Consider the k-means clustering algorithm.

1. Construct a simple example (with at most four data points) which shows that the clustering result of k-means algorithm can be arbitrarily worse than the optimal clustering results in terms of the cost. (The cost of a clustering is measured as the sum of square distance to the cluster center)

Q3 I



Consi

1. Wr
2. Compute $\mathbf{x}^\top \mathbf{L}\mathbf{x}$ for $\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$
3. Show the major steps of embedding the graph into two di