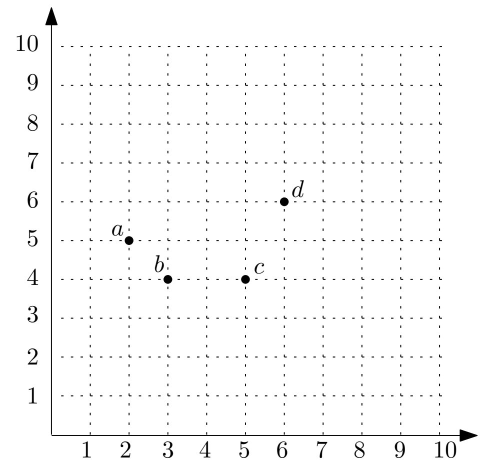
# CMSC5724: Exercise List 9

Answer Problems 1-2 based on the following dataset:



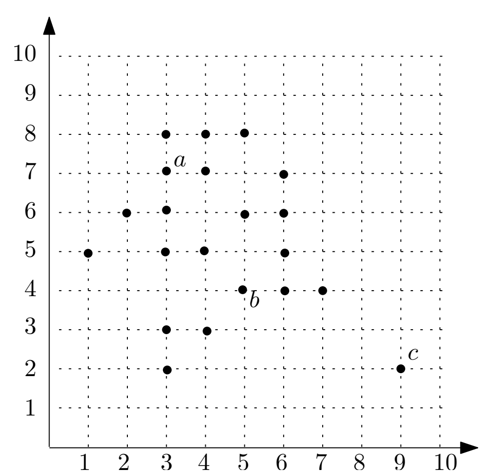
**Problem 1.** Recall that, in discussing hierarchical clustering, we introduced 3 distance metrics on two sets of points: *min*, *max*, and *mean*. Let $S_1 = \{a, c\}$ and $S_2 = \{b, d\}$. What is the distance between $S_1$ and $S_2$ under those three metrics, respectively (assuming that the distance of two points is calculated by Euclidean distance)?

**Problem 2.** Show the dendrogram returned by the Agglomerative algorithm under the min and max metrics, respectively.

**Problem 3.** Suppose that we use $d_{min}$ to define the similarity of two clusters $C_1, C_2$. Give an algorithm to compute the dendrogram on $n$ points in $O(n^2 \log n)$ time.

**Problem 4.** Suppose that we use $d_{mean}$ to define the similarity of two clusters $C_1, C_2$. As discussed in the lecture, $d_{mean}(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{(p_1,p_2) \in C_1 \times C_2} dist(p_1, p_2)$. Give an algorithm to compute the dendrogram on $n$ points in $O(n^2 \log n)$ time.

**Problem 5.** Consider the set $P$ of points below:



Set $\epsilon = 1$ and *minpts* $= 3$. Show the clusters output by DBSCAN, assuming that the distance metric is Euclidean distance.

**Problem 6.** Given a pair of parameters $\epsilon$ and *minpts*, describe an algorithm to compute the DBSCAN clusters in $O(n^2)$ time, assuming that the distance metric is Euclidean distance, and that the dimensionality of the data space is a constant.