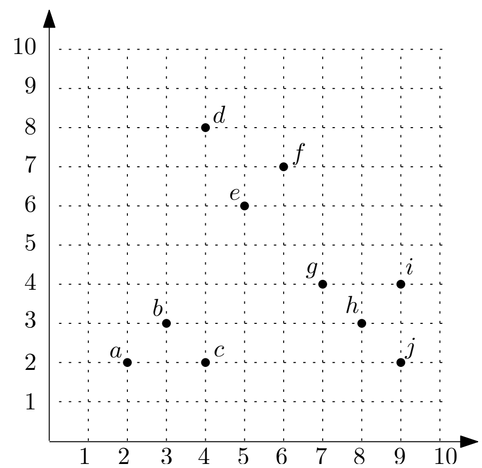
CMSC5724: Exercise List 8

Problem 1. Consider the execution of the k-center algorithm we discussed in the class on the following set P of points:



Suppose that $k = 3$ (i.e., we want to find 3 centers), and that the first center has been (randomly) decided to be f. Show what are the second and third centers found by the algorithm. The distance metric is Euclidean distance.

Problem 2. Let P be the set of points in Problem 1. What is the geometric center of the set $\{c, e, g\}$?

Problem 3. Let P be the set of points in Problem 1. Apply the k-means algorithm on P with $k = 3$ under Euclidean distance. Assume that the algorithm selects a set $S = \{c, g, h\}$ as the initial centroids. Recall that (i) the algorithm updates S iteratively, and (ii) the cost of S is defined to be $\phi(S) = \sum_{p \in P} (d_S(p))^2$ where $d_S(p) = \min_{q \in S} dist(p, q)$.

- Give the content of S after each iteration until the algorithm terminates.
- Show the value of $\phi(S)$ after every iteration.

Problem 4. The goal of this problem is for you to understand why it suffices to consider a finite number of possible solutions to the k-means problem (recall that this was needed to argue that the algorithm terminates).

Consider the k-means problem defined in the lecture notes with $k = 2$. Suppose that we have a set P of n points in $\mathbb{R}^2$ (for simplicity, we assume that the dimensionality is 2). The goal is to find centroid points c_1, c_2 in $\mathbb{R}^2$ to minimize $\sum_{p \in P} (d(p))^2$, where $d(p) = \min_{i=1}^{2} dist(p, c_i)$, with $dist$ representing Euclidean distance. Design an algorithm to solve this problem in $O(2^n \cdot n)$ time.