

Tyler Klina

k-Means Clustering
Lab Assignment

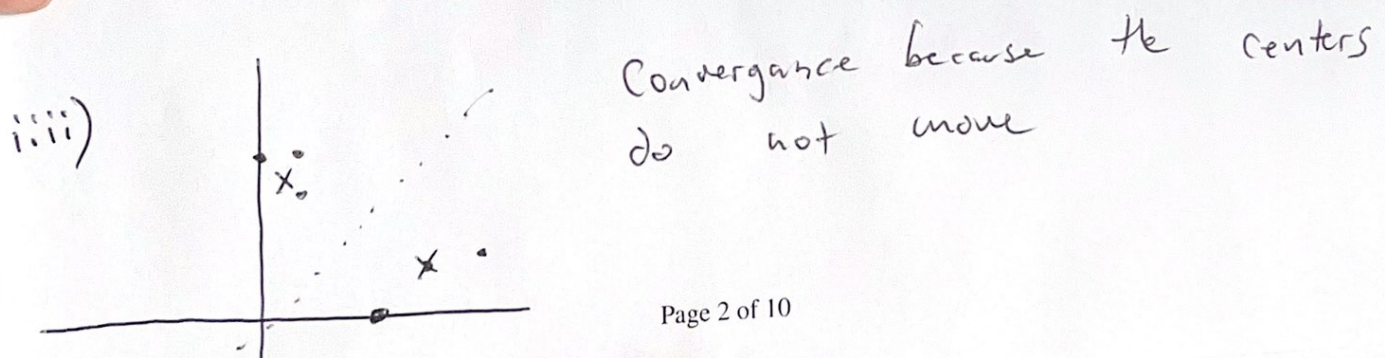
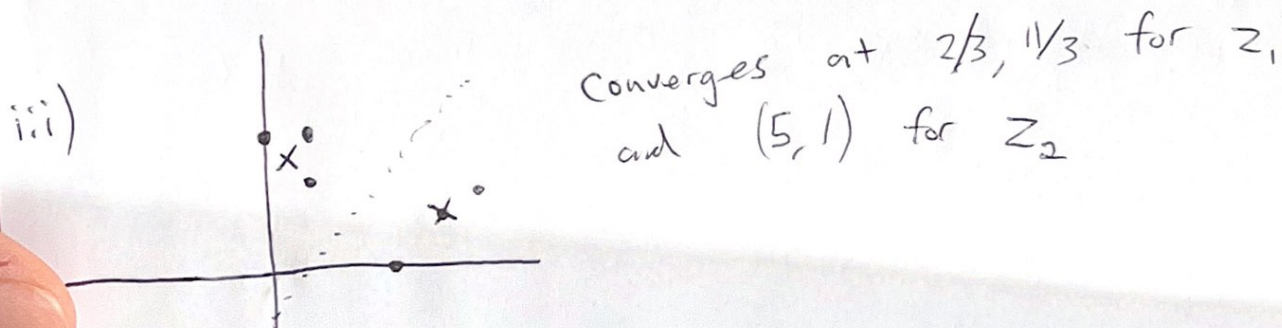
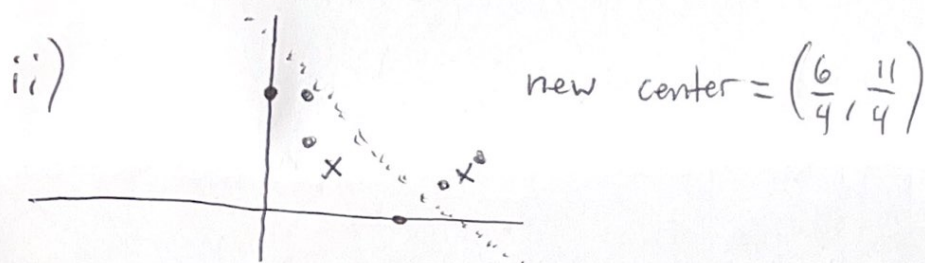
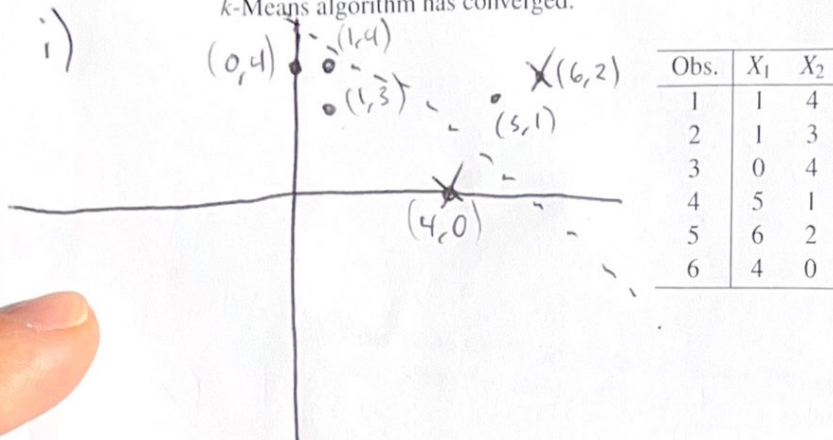
- (10) 1. Provide three example situations (maybe ones relevant to your major) that seem analogous to our "pizza-stores" model. For each situation, what are the "pizza customers" and what are the "pizza stores?"

a) Amazon drivers dispatching packages need to coordinate to drive the shortest distances from the last package they delivered.

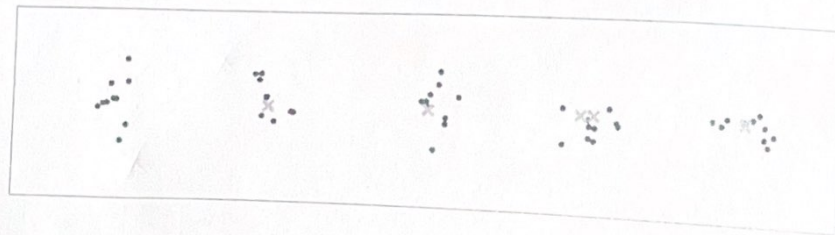
b) Police Stations automatically choose which station you are connected to based off your location to get you quickest dispatch possible

c) School districts have to have somewhat equal distribution among schools so they cluster neighborhoods based on distance and density of students

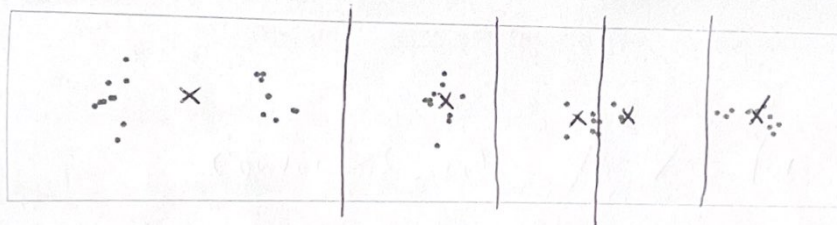
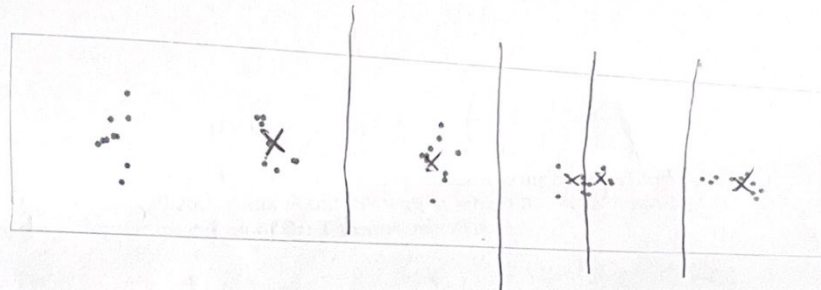
- (20) 2. Perform k -Means clustering by hand on the following data with $k = 2$ and assuming the initial cluster centers are located at $(4,0)$ and $(6,2)$. Show your work step-by-step... what are the clusters at each step, and the new cluster centers at each step? Be sure to identify the final cluster assignments after the k -Means algorithm has converged.



- (20) 3. Consider the data set shown below, with initial cluster locations denoted by the \times 's.

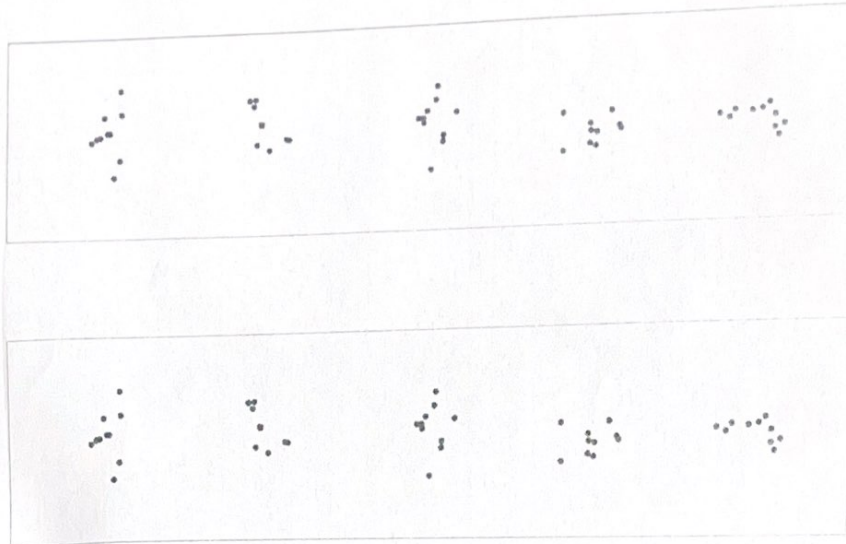


- (a) What is the result of *k*-Means clustering of this data with these initial cluster locations?
(Data are replicated below to assist you in iterating through *k*-Means.)



converges at this point





- (b) In what way are the clusters found by k -means not satisfying? Is this result a failure of the goal (k -Means is inappropriate for this dataset), or a failure of the algorithm (Lloyd's Algorithm failed)? Explain why this happened... think about the way in which the initialization of the cluster centers affects subsequent minimization of the k -means cost function.

k -means could be effective here, but a poor initialization caused the algorithm to fail because two initial guesses were very close to each other and it partitioned a single cluster into 2 and one guess covered two clusters.

