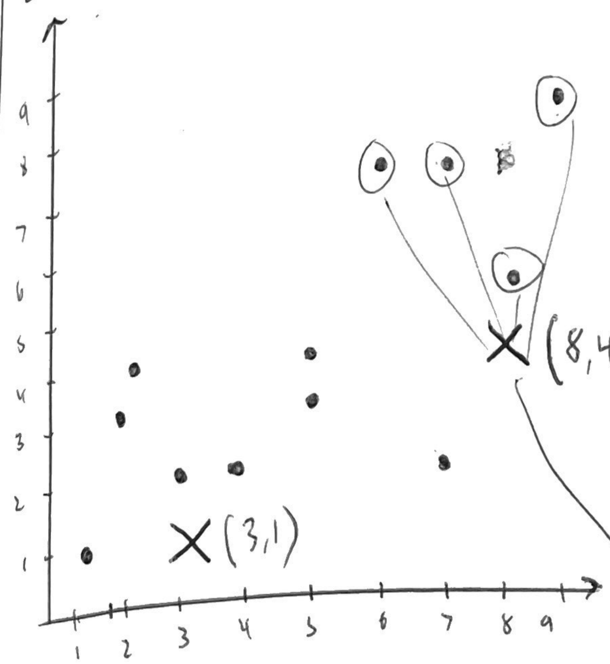
HOMEWORK #6 | Max Sokolich | efC

d) compute 4 iterations of K means

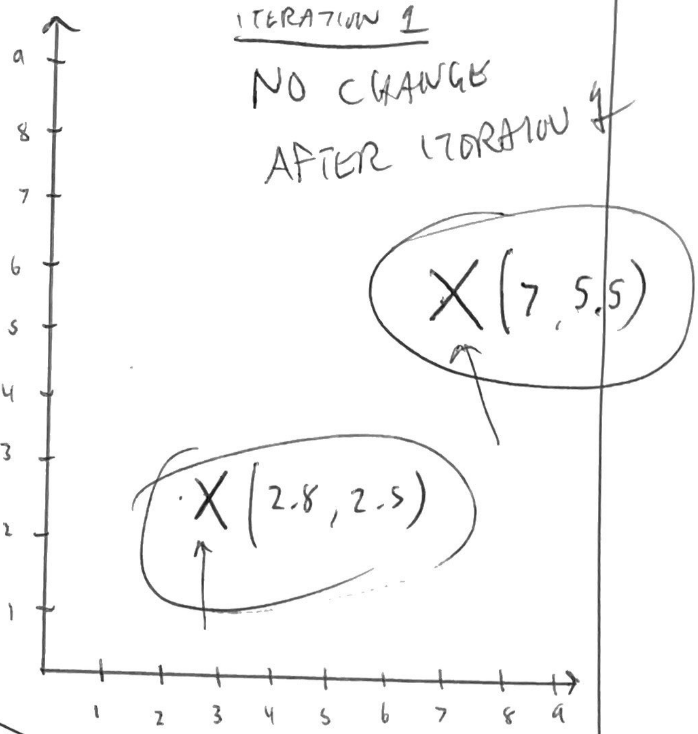
QUESTION # 1

ITERATION 0



ITERATION 1

NO CHANGE AFTER ITERATION 4



GIVEN

$K = 2$

$n = 24$

$K1 = (3,1)$

$K2 = (8,4)$

ITERATIONS

$$= \sqrt{(X_0 - X_c)^2 + (Y_0 - Y_c)^2}$$

where X_0 = 1st value of the data set

X_c = value of first centroid

Y_0 = 1st value of y dataset

Y_c = value of 2nd centroid

MIN

Iteration 1

Min $= \sqrt{(9-3)^2 + (8-1)^2} = 9.21$ | $\sqrt{(9-8)^2 + (8-4)^2} = 4.12$

Iteration 2

Min $= \sqrt{(8-3)^2 + (5-1)^2} = 6.4$ | $\sqrt{(8-8)^2 + (5-4)^2} = 1.0$

Iteration 3

Min $= \sqrt{(7-3)^2 + (7-1)^2} = 7.2$ | $\sqrt{(7-8)^2 + (7-4)^2} = 3.16$

Iteration 4

Min $= \sqrt{(7-3)^2 + (2-1)^2} = 4.1$ | $\sqrt{(7-8)^2 + (2-4)^2} = 2.23$

QUESTION # 2

Prove that K-means algorithm converges finitely

First we want to choose N centroids from K clusters which is at most K^n solutions which is <u>finite</u>.

~~Also~~ Then, the cost function $J(c_1 \dots c_m, \mu_1 \dots \mu_K)$ will always be smaller after every iteration. That is, the result will be smaller or equal to the previous result

So, 2 cases :

1) $J(c_i, \mu_K) = J(c_{i+1}, \mu_{K+1}) \rightarrow$ terminate

2) $J(c_i, \mu_K) > J(c_{i+1}, \mu_{K+1}) \longrightarrow$ reach <u>at most</u> K^n times

In conclusion K-means always converges in finite # of steps

QUESTION # 3

advantages	disadvantages
- simple to implement	- being dependent on initial values
- scales to large datasets	- clustering data of varying sizes and densities
- garantees convergence	- clustering outliers
- warm start position of centroids	- Scaling with number of dimensions
- easily adapts to new examples	

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \sum \|x_i - \mu\|_1$$

Important to note the higher you chose p in Lp (norms) the more important the largest feature distance becomes

L_1 ⟶ all ~~weights~~ distances receive the same weight and the combination of absolute value differences is Linear.

P.S. There is litterally a PhD thesis that answers this question:

Intelligent k-means clustering in L_2 and L_1 versions: experimentation and Application

Ming-Tso Chiang

QUESTION 4

Given

L_2 norm → $\|x_n - \mu_k\|$

softmax responsibility

$$r_{n,k} = \frac{\exp(-\beta\|x_n - \mu_k\|)}{\sum_{i=1}^{K} \exp(-\beta\|x_n - \mu_i\|)}$$

a) determin expression and step

↳ theres not much diffent except we use softmax to determine / label a particular $\vec{X}_i$ with its cluster

b) β is just a tuning parameter that determines how quickly it converges

c)

advantages	disadvantages
- assigns each data point multiple clusters	- computationally way more expensive especially for large data sets
- can handle non linear decision boundaries	- may require more hyperparameter tuning
- can be used for unsupervised or supervised	

P.S. also a paper that answers this question by Sibylle HESS, Wouter Duivesteijn, Decebal Mocanu