Milan Zanussi

CSCI 4350/5350

Homework 9

Due: Tue. Nov. 27, 11:00 PM

1. (3 points) What are the advantages and disadvantages of using a parametric model instead of a non-parametric model for unsupervised learning problems?

   Since parametric models make assumptions about existence and relationships between parameters, parametric learning procedures do not need to as much time to generate parameters to describe the data, and thus take less time to compute.

2. (2 points) Can unsupervised learning approaches be used to solve supervised learning problems? Explain why or why not.

   Yes. Since unsupervised approaches can use the labels from supervised training data and determine which labels correlate to the classes constructed by the unsupervised algorithm.

3. K-means Clustering - Given the following data set, answer the questions below. Note that data points are filled, black circles and current cluster centers are double, red circles.

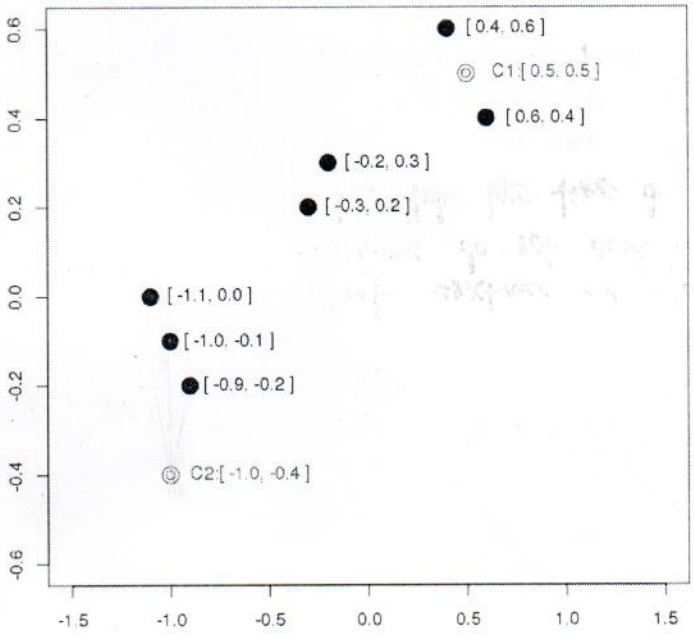
   a. (3 points) What will be the updated location of C1?

      [Work on next page]

   b. (3 points) What will be the updated location of C2?

      [Work on next page]

   c. (4 points) Will additional iterations of the algorithm be required to reach convergence?



Yes, since C_2 was updated on this iteration.

a.) (My work is done assuming the centers are not overlapping any data points in the diagram,)

	(-1.1, 0.0)	(-1.0, -0.1)	(-0.9, -0.2)	(-0.3, 0.2)	(-0.2, 0.3)	(0.4, 0.6)	(0.6, 0.4)
Distance from $C_1 = (0.5, 0.5)$	1.676	1.616	1.565	0.854	0.728	0.141	0.141
Distance from $C_2 = (-1.0, -0.4)$	1.077	0.949	0.825	0.632	0.707	1.118	1.063

a.) Average of cluster 1: $(0.5, 0.5)$

C_1's position is not updated

b.) Average of cluster 2: $(-0.7, 0.4)$

C_2's position is updated.