CS 559 Machine Learning HW3

Problem 1:

1. What's the center of the first cluster (red) after one iteration?

**a) [5.171, 3.171]**

2) What's the center of the second cluster (green) after two iteration?

**a) [5.3, 4.0]**

3) What's the center of the third cluster (blue) when the clustering converges?

**a) [6.2, 3.025]**

**4)** How many iterations are required for the clusters to converge?

**a) 2**

Problem 2:

1. Write down the compact form of p(z) and p(xjz).
2. Marginal Distribution over z:

p(zk=1)=πk

p(z)=­­

The conditional distribution of x given a particular value for z is gaussian

P(x| zk=1)= ­­N(x|µk,Ʃk)

P(x|z)= zk

2) Show that the marginal distribution p(x) has the following form:

p(x) =

XK

k=1

\_kN(xj\_k;\_k)

a)Marginal distribution of x is then obtained by summing joint distribution over all possible values of z.

p(x)=

=

3) If we want to \_nd the MLE solution for parameters \_k; \_k;\_k in such model, what

algorithm should we use? Briey describe its major di\_erence compared to K-means algorithm.

1. Expectation-Maximization algorithm should be used for calculating MLE solutions for Gaussian Mixture Model.

K Means hard assigns a data point to one particular cluster on convergence.It uses L2 Norm when optimizing.

While EM soft assigns a point to clusters depending on probability.

It does not use L2 norm but uses probability.

Problem 3:

1)Draw the corresponding bayesian network.

**B3**

**B2**

**A2**

**B1**

**A1**

1. Based on the bayesian network in (1), write down the joint distribution p(A1;A2;B1;B2;B3).

=P(A1)P(A2)P(B1|A1,A2)P(B3|A1)P(B2|A2)

1. How many independent parameters are needed to fully specify the joint distribution
2. in (2).

1+1+4+2+2 =10

4) Suppose we do not have any independence assumption, write down one possible fac-

torization of p(A1;A2;B1;B2;B3), and state how many independent parameters are required to

describe joint distribution in this case.

a)

**B2**

**A2**

**B1**

**A1**

**B3**

p(A1;A2;B1;B2;B3).=P(A1|B2)P(A2|B3)P(B1|A1,A2)P(B3|A1)P(B2|A2)

Independent parameters= 2+2+4+2+2=12

Problem 4:

1)the loss is much smaller when using 3 layered network , also the designing is less complex

Model name: model\_3

Test loss: 0.176312612162696

Test accuracy: 0.9555555555555556

The model converges faster