CM146, Winter 2018
Problem Set xx: yy

1 Problem 1

(a) **Solution:**
TODO

(b) **Solution:**
We know from Conditional Probability that

$$P(D_i, y_i) = P(D_i|y = 0) \cdot P(y = 0) + P(D_i|y = 1) \cdot P(y = 1)$$

$$P(y_i = 1) = \theta, P(y_i = 0) = 1 - \theta$$

$$P(D_i, y_i) = \left(\frac{n!}{a_i!b_i!c_i!}\alpha_1^{a_i}\beta_1^{b_i}\gamma_1^{c_i}\right) \cdot \theta + \left(\frac{n!}{a_i!b_i!c_i!}\alpha_0^{a_i}\beta_0^{b_i}\gamma_0^{c_i}\right) \cdot (1 - \theta)$$

$$\log P(D_i, y_i) =$$

2 Problem 2

(a) **Solution:**
The two unspecified transition probabilities are:

- $q_{22} = P(q_{t+1} = 2|q_t = 2) = 1 - P(q_{t+1} = 1|q_t = 2) = 1 - 1 = 0$
- $q_{21} = P(q_{t+1} = 2|q_t = 1) = 1 - P(q_{t+1} = 1|q_t = 1) = 1 - 1 = 0$

The two unspecified outcome probabilities are:

- $e_1(B) = P(O_t = B|q_t = 1) = 1 - P(O_t = A|q_t = 1) = 1 - 0.99 = 0.01$
- $e_2(A) = P(O_t = A|q_t = 2) = 1 - P(O_t = B|q_t = 2) = 1 - 0.51 = 0.49$

(b) **Solution:**
The probability that the first symbol is A can be calculated as follows:

$$P(first = A) = P(O_1 = A|q_1 = 1) \cdot P(q_1 = 1) + P(O_1 = A|q_1 = 2) \cdot P(q_1 = 2)$$

$$P(first = A) = (0.99)(0.49) + (0.49)(0.51) = 0.735$$

The probability that the first symbol is B can be calculated as follows:

$$P(first = B) = P(O_1 = B|q_1 = 1) \cdot P(q_1 = 1) + P(O_1 = B|q_1 = 2) \cdot P(q_1 = 2)$$

$$P(first = B) = (0.01)(0.49) + (0.51)(0.51) = 0.265$$

Hence, the more frequent output in the first position is A

(c) **Solution:**
TODO

3 Problem 3

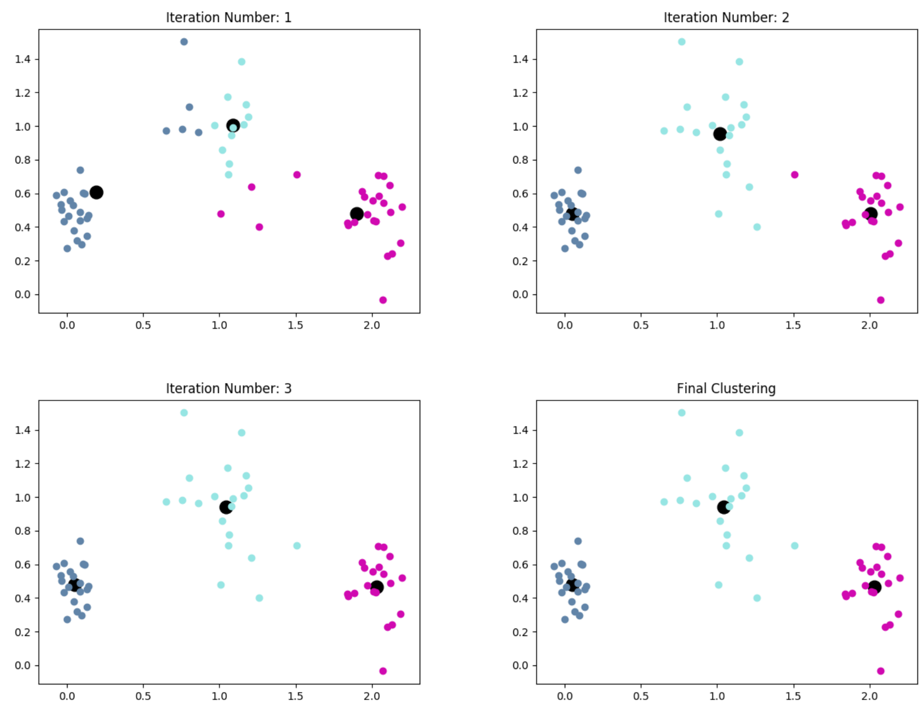
(a) The problem with minimizing over μ, c, k is that we are extremely prone to overfitting. With this kind of loss function, we will generate a model that achieves $J(\mu, c, k) = 0$ on a training set of size n, by using n cluster centers, where each cluster has the label that is assigned to it in the training data. This is an overfitted model and would perform terribly on test sets.
For a training set S with labels y,

$$\mu = S$$

$$c = y$$

$$k = |S|$$

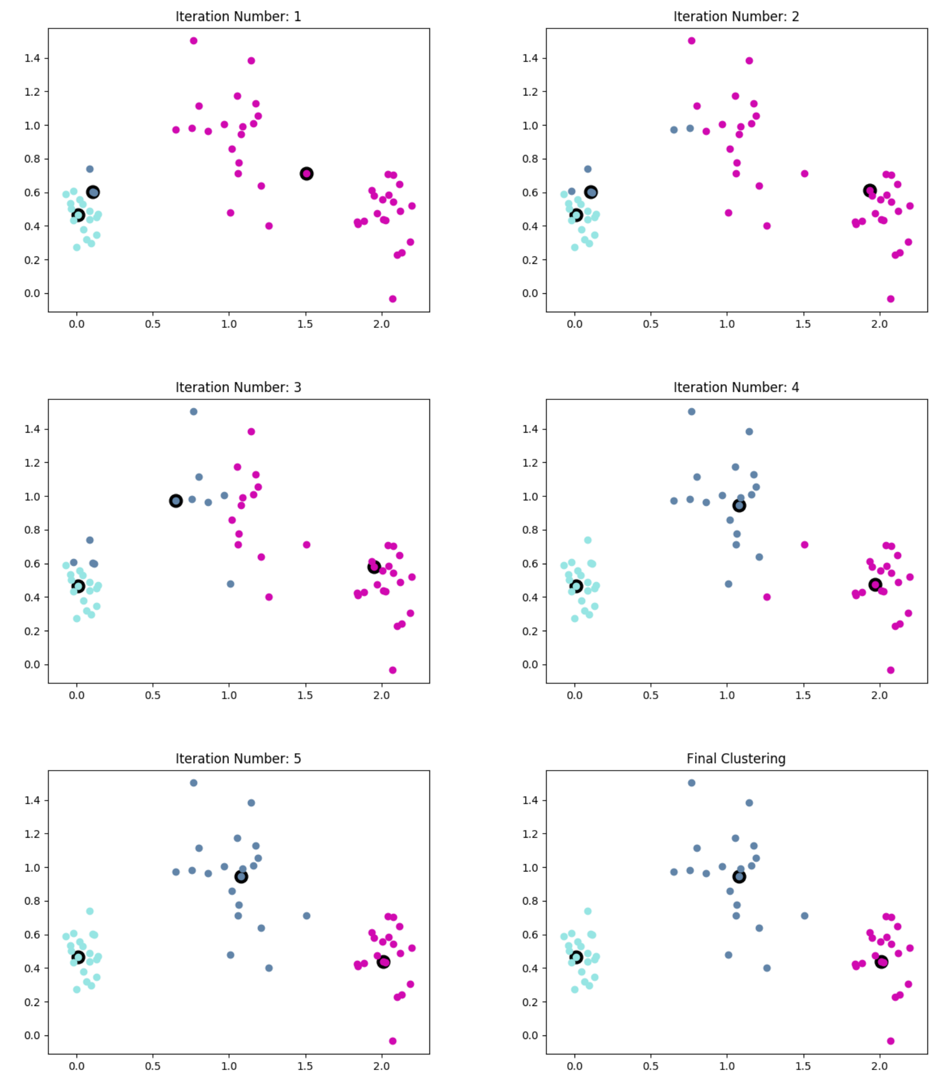
(d) The graphs for random initialization with kMeans look like the following (for 20 points, and 3 clusters):



The cluster centers for each of the above iterations is as follows

Iteration Number	Cluster Center 1	Cluster Center 2	Cluster Center 3
1	(1.061948,0.775711)	(1.210252,0.638547)	(0.649507,0.974770)
2	(1.089668,1.004243)	(1.900104,0.480904)	(0.192982,0.606416)
3	(1.016055,0.952888)	(2.005941,0.477239)	(0.049180,0.481094)
Final	(1.040635,0.940960)	(2.030856,0.465384)	(0.049180,0.481094)

(e) The graphs for random initialization with kMedoids looks like the following:



The cluster centers for each of the above iterations is as follows

Iteration Number	Cluster Center 1	Cluster Center 2	Cluster Center 3
1	(-0.033408,0.500212)	(1.157995,1.009878)	(0.084167,0.739096)
2	(0.012471,0.467721)	(1.507651,0.714342)	(0.104758,0.604594)
3	(0.012471,0.467721)	(1.932385,0.612375)	(0.104758,0.604594)
4	(0.012471,0.467721)	(1.948285,0.579243)	(0.649507,0.974770)
5	(0.012471,0.467721)	(1.969410,0.472674)	(1.076992,0.947875)
Final	(0.012471,0.467721)	(2.011976,0.440005)	(1.076992,0.947875)