# Com Sci 146: Homework #5
# 704797256

**Jingyue Shen**

## Problem 1

### Part A

We only model the word frequency in each class, but ignore the relation between current word and the words before and after it.

### Part B

$log^{Pr(D_i,y_i)} = log^{Pr(D_i|y_i)} + log^{Pr(y_i)} = log^{\frac{n!}{a_i!b_i!c_i!}\alpha_0^{a_i(1-y_i)}\alpha_1^{a_iy_i}\beta_0^{b_i(1-y_i)}\beta_1^{b_iy_i}\gamma_0^{c_i(1-y_i)}\gamma_1^{c_iy_i}} + log^{\theta^{y_i}(1-\theta)^{1-y_i}}$

### Part C

$\frac{\partial log^{\prod_{i=1}^m P(D_i,y_i)}}{\partial\alpha_1} = 0$

$=> \frac{\partial\sum_{i=1}^m a_iy_ilog^{\alpha_1}+c_iy_ilog^{1-\alpha_1-\beta_1}}{\partial\alpha_1} = 0$

$=> \frac{\sum_{i=1}^m a_iy_i}{\alpha_1} - \frac{\sum_{i=1}^m c_iy_i}{1-\alpha_1-\beta_1} = 0$

$=> \alpha_1(\sum a_iy_i + \sum c_iy_i) = (1-\beta_1)\sum a_iy_i$ (1)

Take partial derivative with respect to $\beta_1$, we can get
$\beta_1(\sum b_iy_i + \sum c_iy_i) = (1-\alpha_1)\sum b_iy_i$ (2)

Solve (1) and (2) together, we can get $\alpha_1 = \frac{\sum a_iy_i}{\sum a_iy_i+\sum b_iy_i+\sum c_iy_i}$, $\beta_1 = \frac{\sum b_iy_i}{\sum a_iy_i+\sum b_iy_i+\sum c_iy_i}$
Follow the same procedure, we can get $\gamma_1 = \frac{\sum c_iy_i}{\sum a_iy_i+\sum b_iy_i+\sum c_iy_i}$, and $\alpha_0 = \frac{\sum a_i(1-y_i)}{\sum a_i(1-y_i)+\sum b_i(1-y_i)+\sum c_i(1-y_i)}$
, $\beta_0 = \frac{\sum b_i(1-y_i)}{\sum a_i(1-y_i)+\sum b_i(1-y_i)+\sum c_i(1-y_i)}$, $\gamma_0 = \frac{\sum c_i(1-y_i)}{\sum a_i(1-y_i)+\sum b_i(1-y_i)+\sum c_i(1-y_i)}$

## Problem 2

### Part A

The missing transition probabilities are $P(q_{t+1} = 1|q_t = 2) = 1 - q_{11} = 0$ and $P(q_{t+1} = 2|q_t = 2) = 1 - q_{12} = 0$, and the two missing probabilities are $e_1(B) = P(O_t = B|q_t = 1) = 1 - e_1(A) = 0.01$ and $e_2(A) = P(O_t = A|q_t = 2) = 1 - e_2(B) = 0.49$

### Part B

$P(generateA) = \pi_1 * e_1(A) + \pi_2 * e_2(A) = 0.49 \times 0.99 + 0.51 \times 0.49 = 0.735$ $P(genearteB) = \pi_1 * e_1(B) + \pi_2 * e_2(B) = 0.49 \times 0.01 + 0.51 \times 0.51 = 0.265$
So A is the most symbnol appear in the first position of the sequence.

### Part C

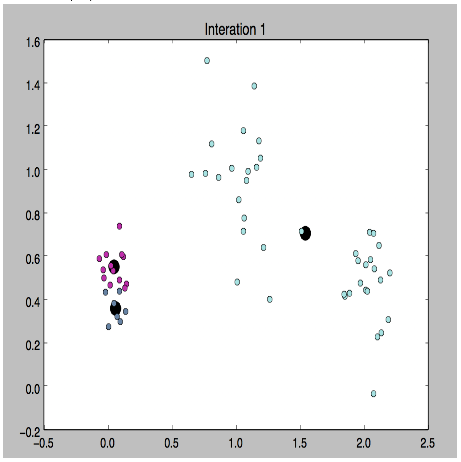
From the above we know that the most probable first symbol is A. Since $q_{11} = q_{12} = 1$, no matter the first state is 1 or 2, the second and third states must be 1. since $e_1(A) = 0.99$, so the most probable symbol in second and third positions are As. So the sequence of three output symbols that has the highest probability being generated is A A A.

# Problem 3

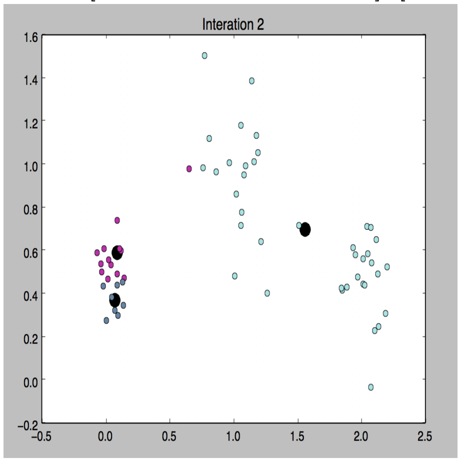
## Part A

It is a bad idea since if k is not fixed, $J(c, \mu, k)$ can reach zero by having n centroids whose positions are the n datapoints. Under this case, k =n, $\mu_i = x^{(i)}$,$c^{(i)} = i$

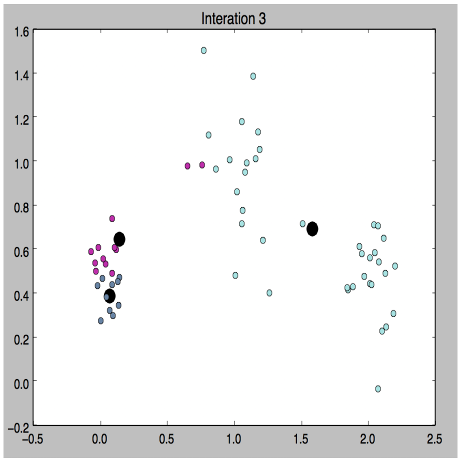
Part (d)



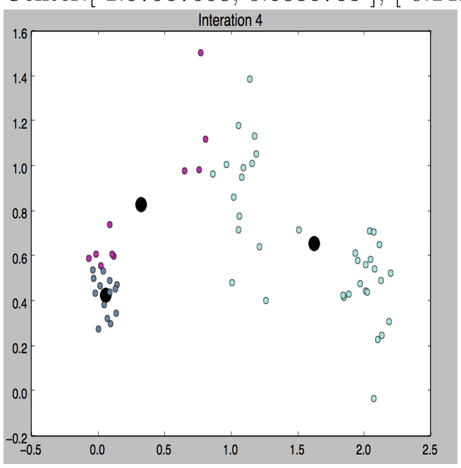

Center:[ 1.53574549, 0.70317209], [ 0.04416798,0.548694 ], [ 0.05848729, 0.35555227]



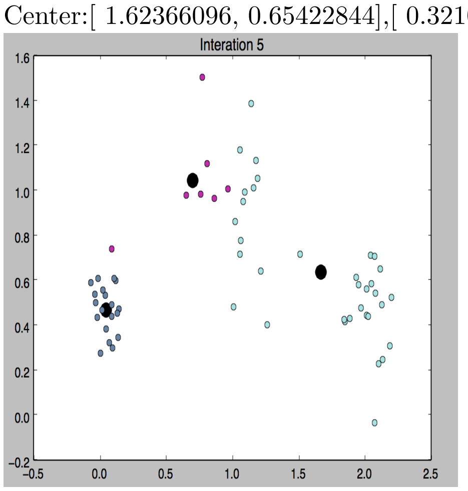

Center: [ 1.55846957, 0.69620803], [ 0.08399041, 0.58882485], [ 0.06765327, 0.36774192]



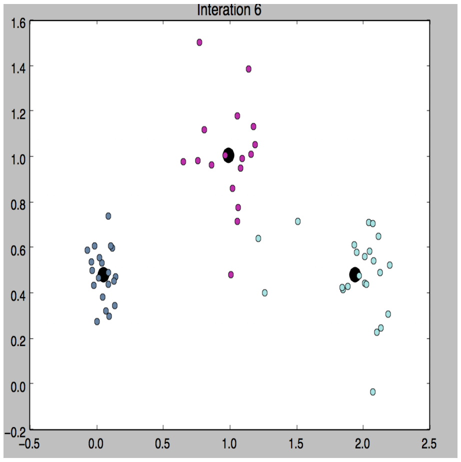

Center:[ 1.57957558, 0.6886763 ], [ 0.14104785,0.64172345],[ 0.06969682, 0.38783907]



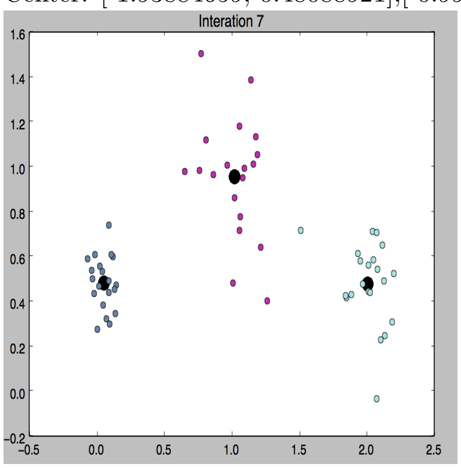

Center:[ 1.62366096, 0.65422844],[ 0.32109265, 0.82675015],[ 0.05362096, 0.42350331]



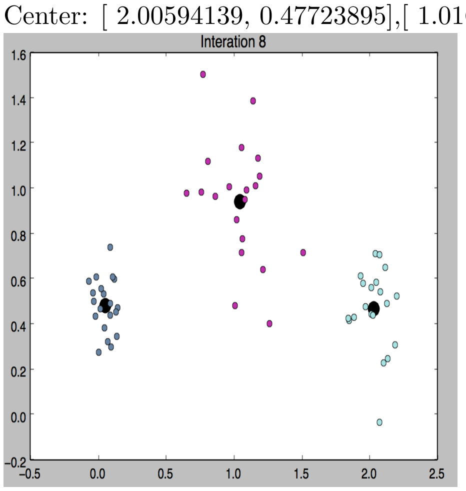

Center:[ 1.66538394, 0.63481783], [ 0.69870477, 1.04031048], [ 0.04733828, 0.46751536]



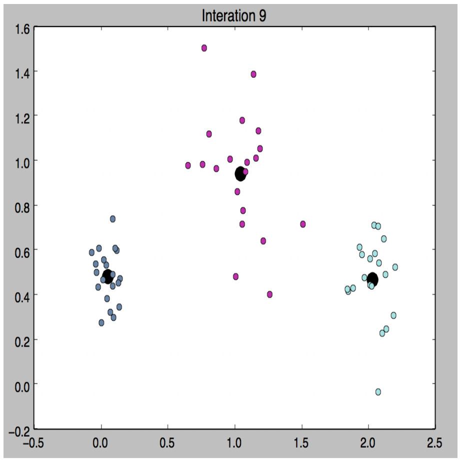

Center: [ 1.93884659, 0.48088921],[ 0.99037342, 1.00390775]),[ 0.04917974, 0.4810944 ]




Center: [ 2.00594139, 0.47723895],[ 1.01605529, 0.95288767], [ 0.04917974, 0.4810944 ]



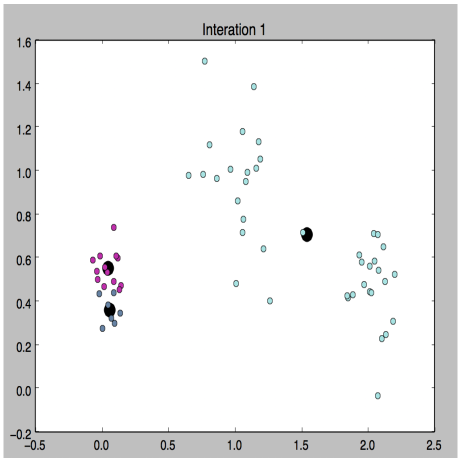

Center:[ 2.03085592, 0.46538378], [ 1.04063507, 0.9409604 ],[ 0.04917974, 0.4810944 ]



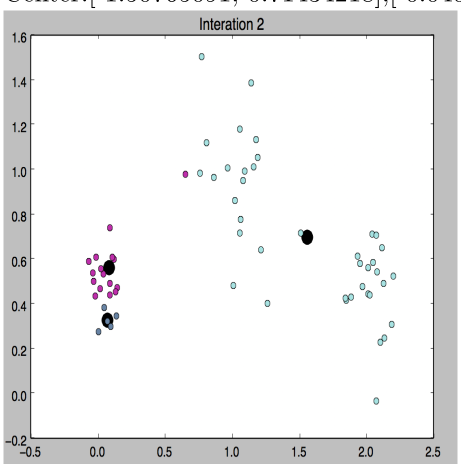

Center:[ 2.03085592, 0.46538378], [ 1.04063507,0.9409604 ],[ 0.04917974, 0.4810944 ]

So k-mean converges in 9 iterations in this case.

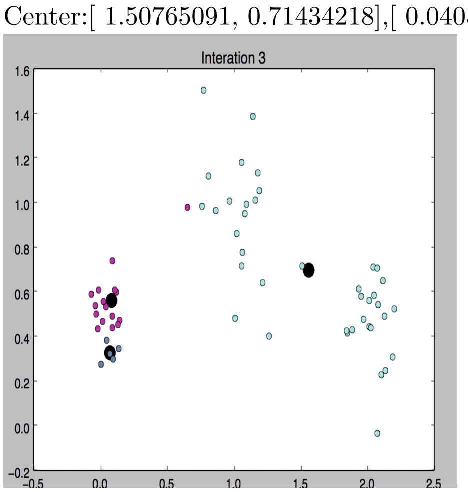
**Part E**




Center:[ 1.50765091, 0.71434218],[ 0.04054534, 0.52890919],[ 0.06755541, 0.31829728]




Center:[ 1.50765091, 0.71434218],[ 0.04054534, 0.52890919],[ 0.06755541, 0.31829728]

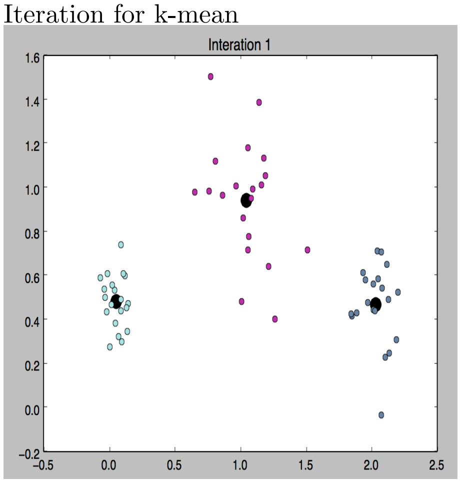



Center: [ 1.50765091, 0.71434218], [ 0.04054534, 0.52890919],[ 0.06755541, 0.31829728]

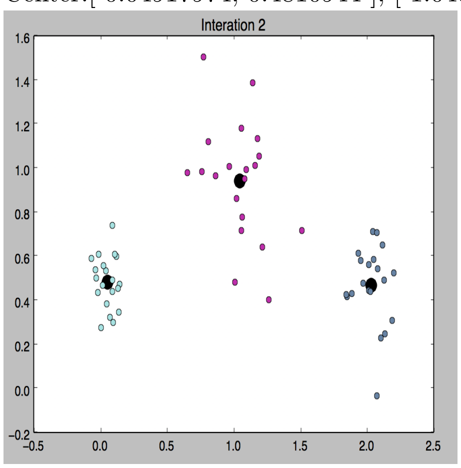
So k-medoids converges in 3 iterations , but I think it does not cluster as well as k-means in this case

**Part F**

Iteration for k-mean

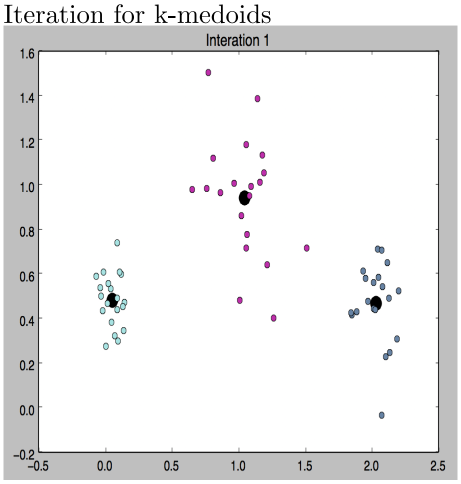



Center:[ 0.04917974, 0.4810944 ], [ 1.04063507, 0.9409604 ], [ 2.03085592, 0.46538378]

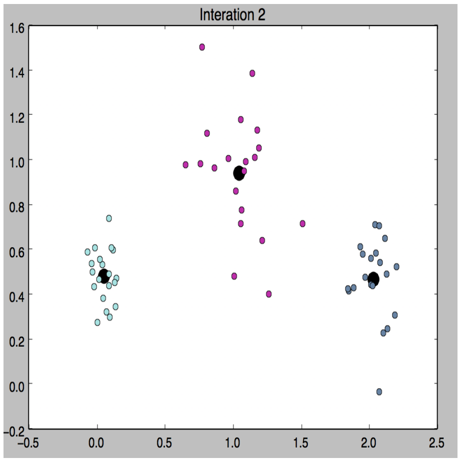



Center: [ 0.04917974, 0.4810944 ], [ 1.04063507, 0.9409604 ], [ 2.03085592, 0.46538378]

Iteration for k-medoids




Center:[ 0.0124713 , 0.46772052], [ 1.07699221, 0.94787531], [ 2.01197635, 0.44000531]




Center:[ 0.0124713 , 0.46772052], [ 1.07699221, 0.94787531], [ 2.01197635, 0.44000531]

When using cheat initialization, both k-means and k-medoids converges in 2 iterations and both generalize well.