Pset 5

1a Information that we lose using aforementioned model is the order of words and semantic meaning of words.

b Let's assume $Pr(Y_i = 1) = \eta$ and $Pr(Y_i = 0) = 1-\eta$

Thus, $Pr(D_i, y_i) = Pr(Y = y_i)\, Pr(D_i \mid Y = y_i)$

$$= \left(Pr(Y=1)\, Pr(D_i \mid Y=1)\right)^{y_i} \left(Pr(Y=0)\, Pr(D_i \mid Y=0)\right)^{1-y_i} \quad \dots ①$$

Using naive Bayes model:

$$Pr(D_i, y_i) = \left(\eta \frac{n!}{a_i!\, b_i!\, c_i!} \alpha_1^{a_i} \beta_1^{b_i} \gamma_1^{c_i}\right)^{y_i} \left((1-\eta) \frac{n!}{a_i!\, b_i!\, c_i!} \alpha_0^{a_i} \beta_0^{b_i} \gamma_0^{c_i}\right)^{1-y_i} \quad \dots ②$$

log likelihood of $D_i$ =

$$N_i = \log Pr(D_i, y_i) = y_i \left[\log \eta + \log\left(\frac{n!}{a_i!\, b_i!\, c_i!}\right) + a_i \log \alpha_1 + b_i \log \beta_1 + c_i \log \gamma_1\right] +$$
$$(1-y_i)\left[\log(1-\eta) + \log\left(\frac{n!}{a_i!\, b_i!\, c_i!}\right) + a_i \log \alpha_0 + b_i \log \beta_0 + c_i \log \gamma_0\right] \quad \dots ③$$

c To get the value of $\gamma_1$, we will sub $\gamma_1 = 1 - \alpha_1 - \beta_1$ then derive $\alpha_1, \beta_1$.

$$\frac{\partial N}{\partial \alpha_1} = \sum y_i \left(\frac{a_i}{\alpha_1} + \frac{c_i}{\gamma_1} \frac{\partial \gamma_1}{\partial \alpha_1}\right) =$$

$$= \sum y_i \left(\frac{a_i}{\alpha_1} - \frac{c_i}{\gamma_1}\right) = \sum y_i \left(\frac{a_i \gamma_1 - c_i \alpha_1}{\alpha_1 \gamma_1}\right) = 0$$

$$\therefore \alpha_1 = \gamma_1 \frac{\sum y_i a_i}{\sum y_i c_i}$$

Same for $\beta_1$, $\beta_1 = \gamma_1 \dfrac{\sum y_i b_i}{\sum y_i c_i}$

Since we know $\gamma_1 = 1 - \alpha_1 - \beta_1$

$$\gamma_1 = 1 - \gamma_1 \left(\frac{\sum y_i a_i}{\sum y_i c_i}\right) - \gamma_1 \left(\frac{\sum y_i b_i}{\sum y_i c_i}\right)$$

$$\gamma_1 = \frac{1}{\left(\frac{\sum y_i (a_i + b_i + c_i)}{\sum y_i c_i}\right)}$$

Since, $a_i + b_i + c_i = n$,

we will know

$$\gamma_1 = \frac{\sum y_i c_i}{n \sum y_i} \qquad \alpha_1 = \frac{\sum y_i a_i}{n \sum y_i} \qquad \beta_1 = \frac{\sum y_i b_i}{n \sum y_i}$$

Using the similar method as above, we will be able to derive $\alpha_0$, $\beta_0$, and $\gamma_0$.

$$\gamma_0 = \frac{\sum (1-y_i) c_i}{n \sum (1-y_i)} \qquad \alpha_0 = \frac{\sum (1-y_i) a_i}{n \sum (1-y_i)} \qquad \beta_0 = \frac{\sum (1-y_i) b_i}{n \sum (1-y_i)}$$

2 Hidden Markov Models

a Since $\sum_i q_{ij} = 1$

$q_{11} + q_{21} = 1$

$q_{12} + q_{22} = 1$

Since we know the value of $q_{11}$ & $q_{12}$, $q_{21} = q_{22} = 0$

$\sum_b e_K(b) = 1$

$e_1(A) + e_1(B) = 1$

$e_2(A) + e_2(B) = 1$

Thus, $e_1(B) = 1 - 0.99 = 0.01$

$e_2(A) = 1 - 0.51 = 0.49$

$\therefore$ missing probabilities $q_{21} = 0$, $q_{22} = 0$, $e_1(B) = 0.01$, $e_2(A) = 0.49$

b $P(O_1 = A) = P(O_1 = A \mid q_1 = 1) + P(O_1 = A \mid q_1 = 2)$

$P(O_1 = A \mid q_1 = 1) = (0.49)(0.99) = 0.4851$

$P(O_1 = A \mid q_1 = 2) = (0.51)(0.49) = 0.2499$

$P(O_1 = A) = 0.4851 + 0.2499 = \boxed{0.735}$

$P(O_1 = B) = P(O_1 = B \mid q_1 = 1) + P(O_1 = B \mid q_1 = 2)$.

$P(O_1 = B \mid q_1 = 1) = (0.49)(0.01) = 0.0049$

$P(O_1 = B \mid q_1 = 2) = (0.51)(0.51) = 0.2601$

$P(O_1 = B) = 0.0049 + 0.2601 = \boxed{0.265}$

Since we know that $P(O_1 = A) > P(O_1 = B)$, then we know that A appears more often.

c We know that we will always in state 1 because
$q_{11} = 1$ and $q_{12} = 1$
So no matter we are in symbol 2 or 3, we will always be in state 1.

Thus, $\left.\begin{array}{l} P(A) = 0.99 \\ P(B) = 0.01 \end{array}\right\}$ for $2^{nd}$ and $3^{rd}$ symbol.

Since we know that $1^{st}$ symbol will always A, from part b.
Now, we will also be able to see that $2^{nd}$ & $3^{rd}$ symbols are AA

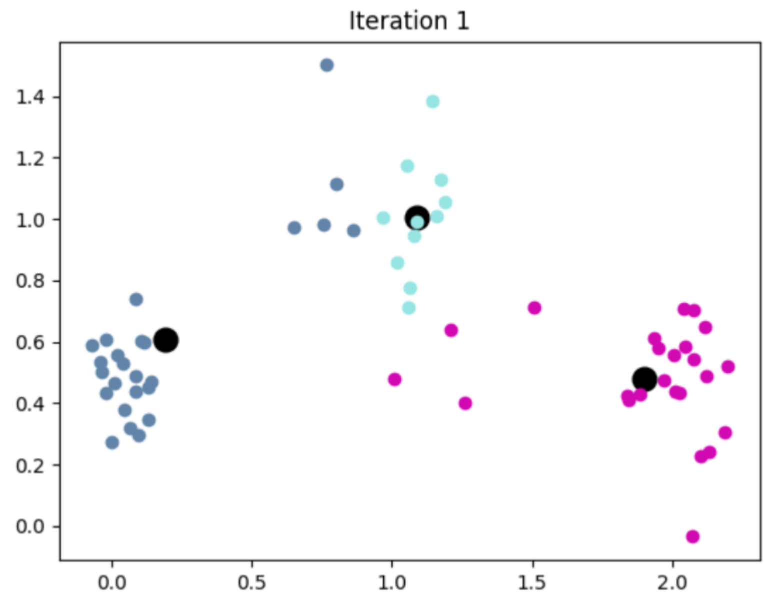
Thus, output with highest probability is AAA

3. Facial Recognition by using K-Means and K-Medoids

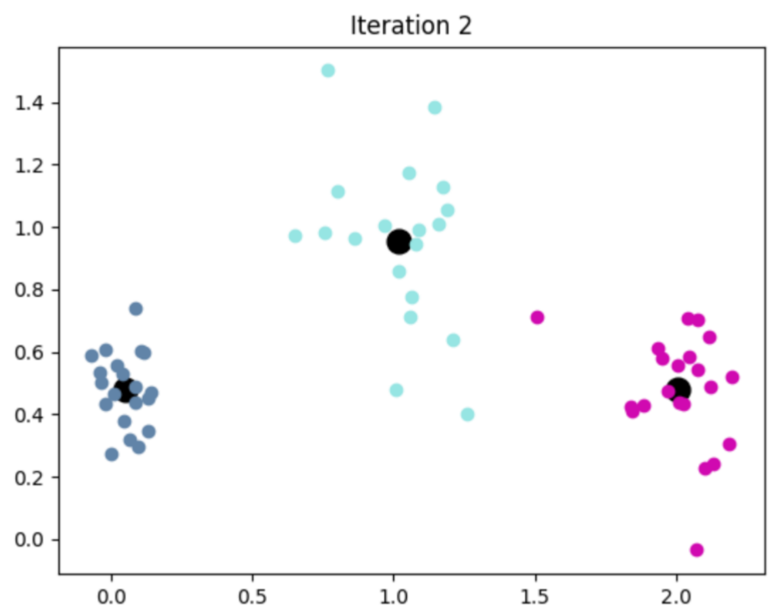
a. The minimum objective value is 0. We can get this result if we set k = n, where it means that every object has its own cluster within n different cluster. The value of $C_i = i$ and $M_j=x^{(i)}$. This is a bad idea because clustering is basically grouping some data by its similarity. However, having an individual cluster for each given data is not going to provide an ideal clustering method.

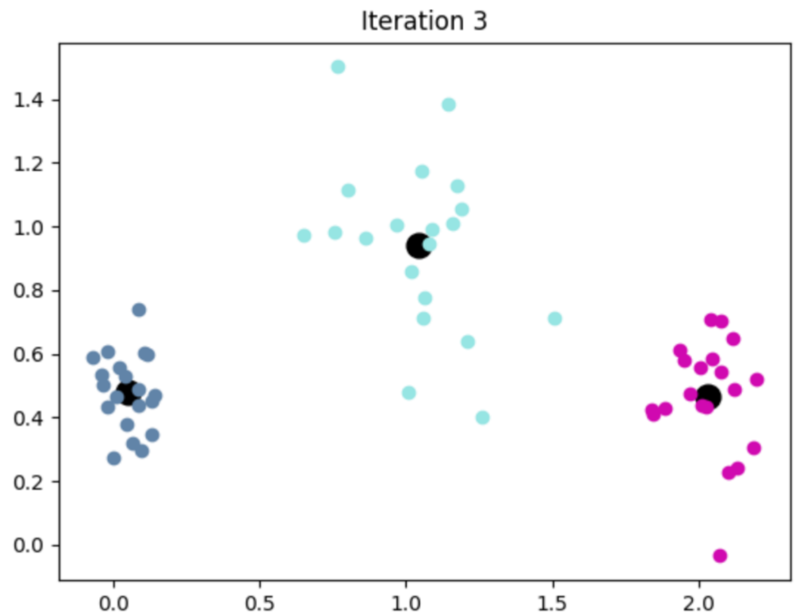
b. Implemented in the code.

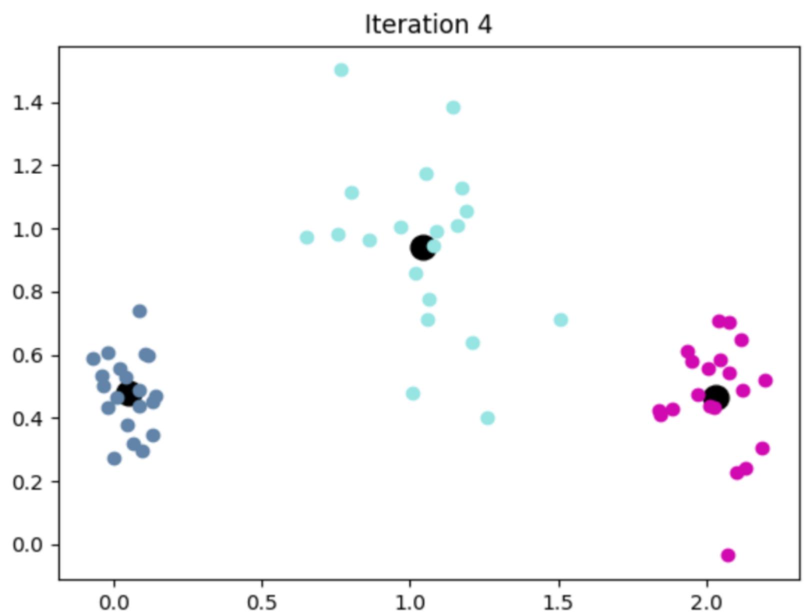
c. Implemented in the code.




d.

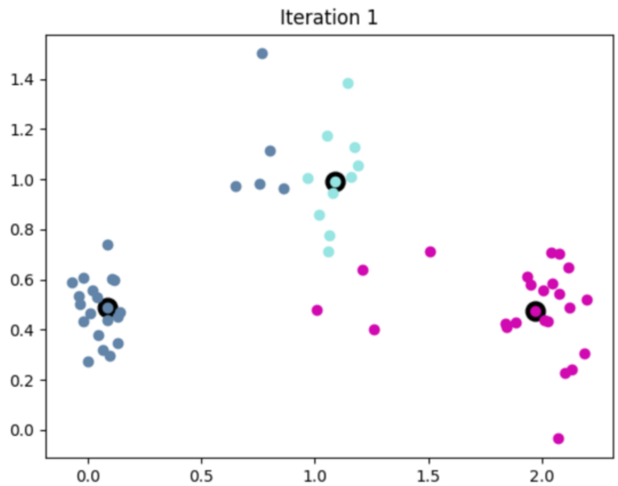


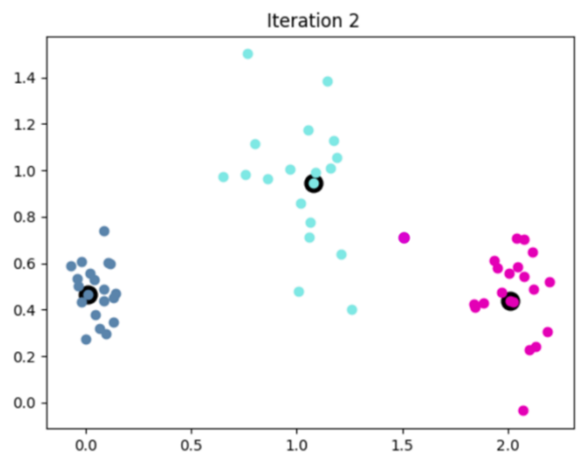


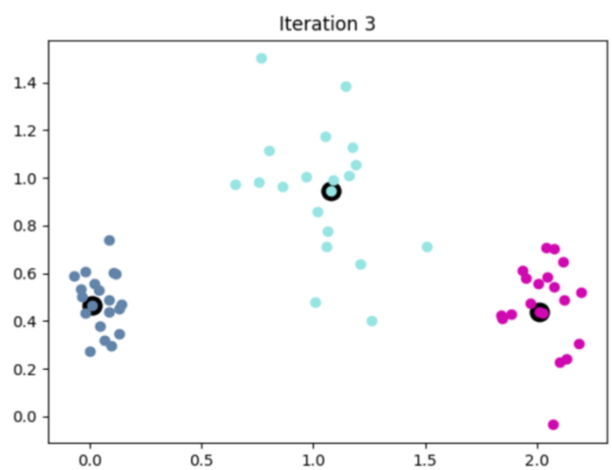



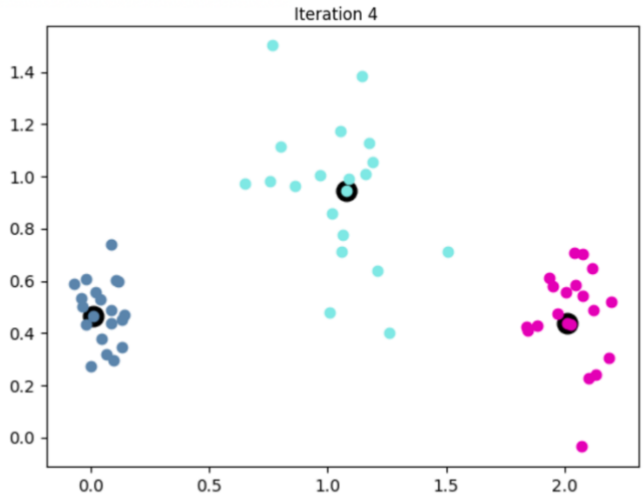

Plots above is implemented using k-means cluster assignments in iteration 1, 2, 3, and 4 with random initialization.

e.



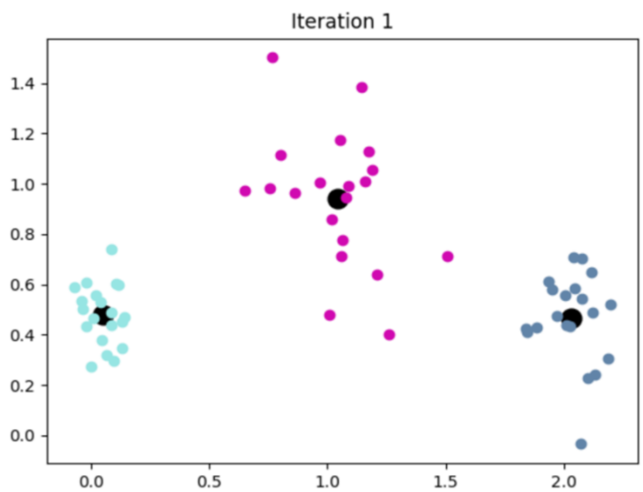



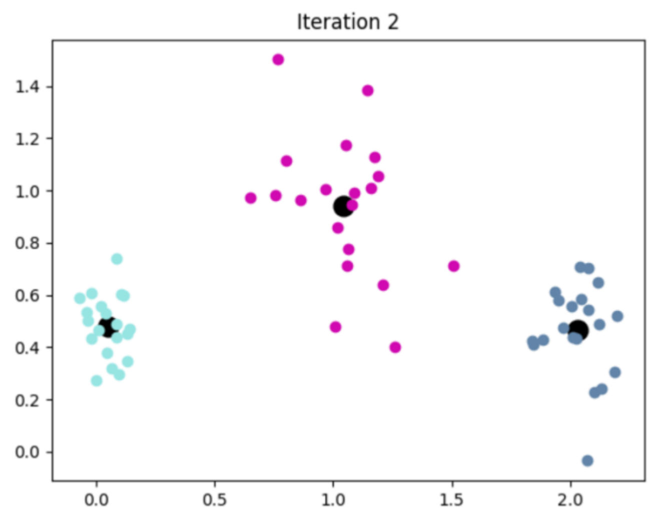






Plots above is implemented using k-medoids cluster assignments in iteration 1, 2, 3, and 4 with random initialization.

f. k-means plot using cheat_init():



k-medoids plot cheat_init():