Problem 1:

a). We lose the order of words and semantic meaning of words.

b). Assume $P(Y_i=1)=\eta$ and $P(Y_i=0)=1-\eta$

$$P(D_i, y_i) = P(Y=y_i)\,P(D_i \mid Y=y_i)$$
$$= \left(P(Y=1)\,P(D_i \mid Y=1)\right)^{y_i}\left(P(Y=0)\,P(D_i \mid Y=0)\right)^{1-y_i} \ldots$$

By using naive Bayes model:

$$P(D_i, y_i) = \left(\eta \frac{n!}{a_i!\,b_i!\,c_i!}\,\alpha_1^{a_i}\beta_1^{b_i}r_1^{c_i}\right)^{y_i}\left((1-\eta)\frac{n}{a_i!\,b_i!\,c_i!}\,\alpha_0^{a_i}\beta_0^{b_i}r_0^{c_i}\right)^{1-y_i}$$

The log likelihood of D_i =

$$N_i = \log P(D_i, y_i) = y_i\left[\log\eta + \log\left(\frac{n!}{a_i!\,b_i!\,c_i!}\right) + a_i\log\alpha_1 + b_i\log\beta_1 + c_i\log r_1\right] + (1-y_i)\left[\log(1-\eta) + \log\left(\frac{n!}{a_i!\,b_i!\,c_i!}\right) + a_i\log\alpha_0 + b_i\log\beta_0 + c_i\log\alpha_0\right]$$

c) Supeitute $r_1 = 1-\alpha_1-\beta_1$

$$\frac{\partial N}{\partial \alpha_1} = \sum y_i\left(\frac{a_i}{\alpha_1} + \frac{c_i}{r_1}\cdot\frac{\partial r_1}{\partial \alpha_1}\right)$$

$$= \sum y_i\left(\frac{a_i}{\alpha_1} - \frac{c_i}{r_1}\right) = \sum y_i\left(\frac{a_i r_1 - c_i\alpha_1}{\alpha_1 r_1}\right) = 0$$

$$\alpha_1 = r_1\frac{\sum y_i a_i}{\sum y_i c_i}, \quad \beta_1 = r_1\frac{\sum y_i b_i}{\sum y_i c_i}$$

Since $r_1 = 1-\alpha_1-\beta_1$

$$= 1 - r_1\frac{\sum y_i a_i}{\sum y_i c_i} - r_1\frac{\sum y_i b_i}{\sum y_i c_i}$$

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

Also, since $a_i + b_i + c_i = n$,

we will know

$$\gamma_1 = \frac{\Sigma y_i c_i}{n \Sigma y_i}, \quad \alpha_1 = \frac{\Sigma y_i a_i}{n \Sigma y_i}, \quad \beta_1 = \frac{\Sigma y_i b_i}{n \Sigma y_i}$$

Then we can derive $\alpha_0, \beta_0, \gamma_0$ by using the same method:

$$\gamma_0 = \frac{\Sigma(1-y_i)c_i}{n\Sigma(1-y_i)}, \quad \alpha_0 = \frac{\Sigma(1-y_i)a_i}{n\Sigma(1-y_i)}, \quad \beta_0 = \frac{\Sigma(1-y_i)b_i}{n\Sigma(1-y_i)}$$

2. Hidden Markov Models
   a. $q_{11} + q_{21} = 1$
      $q_{12} + q_{22} = 1$
      From the question, we know that the value of $q_{11} = q_{12} = 1$, then we know that the value of $q_{21} = 0$ and $q_{22} = 0$

      $e_1(A) + e_1(B) = 1$
      $e_2(A) + e_2(B) = 1$
      From the question, we know that the value of $e_1(A) = 0.99$. Thus, $e_1(B) = 1-0.99 = 0.01$. Since $e_2(A) = 0.51$, then $e_2(B) = 1-0.51 = 0.49$

      Thus, the missing probabilities are **$q_{21} = 0$, $q_{22} = 0$, $e_1(B) = 0.01$, $e_2(B) = 0.49$**

   b. $P(O_1 = A) = P(O_1 = A \mid q_1 = 1) + P(O_1 = A \mid q_1 = 2)$
      $= (0.49 * 0.99) + (0.51 * 0.49)$
      $= 0.4851 + 0.2499 =$ **0.735**

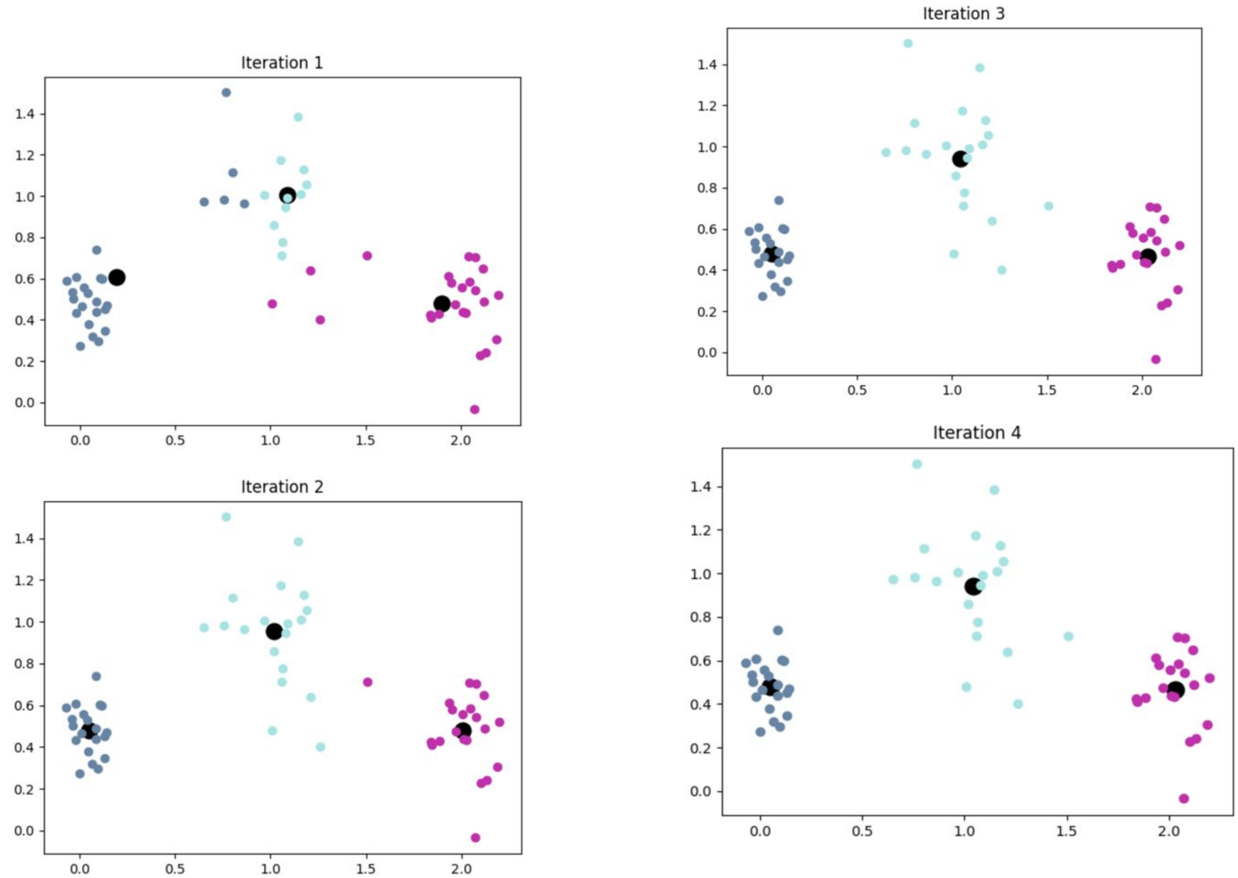
      $P(O_1 = B) = P(O_1 = B \mid q_1 = 1) + P(O_1 = B \mid q_1 = 2)$
      $= (0.49 * 0.01) + (0.51 * 0.51)$
      $= 0.0049 + 0.2601 =$ **0.265**

      By the calculation, we know that **A appears more often** since $P(O_1 = A) > P(O_1 = B)$.

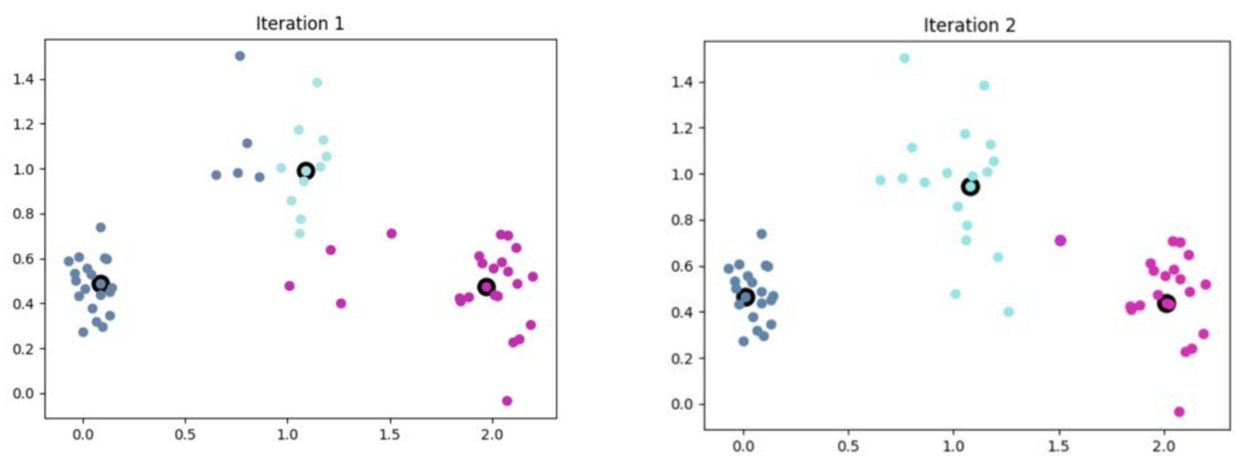
   c. Output with highest probability is AAA. The reason is the following:
      - From part B, we know that the 1st symbol will be A.
      - Second and 3rd symbol will always be in the 1st state. We also know that probability of it being symbol A is 0.99, and symbol B is 0.01. Thus, it is clear that for second and third symbol, it will also be A.

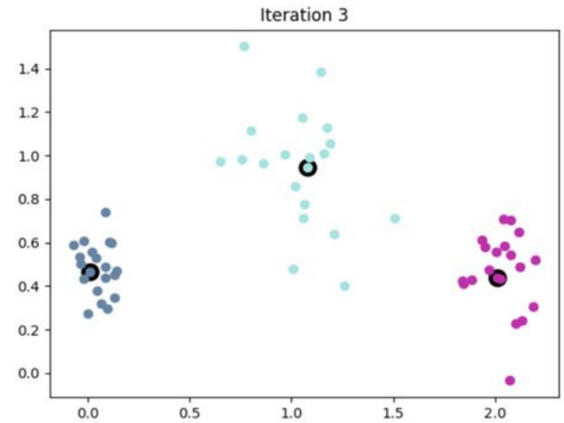
3. Facial Recognition by using K-means and K-medoids.
   a. If we set k = n, we will get the minimum value 0 because it means that every object has its own cluster in n different cluster. In addition, value of $C_i = i$ and $M_j = x^{(i)}$. This implementation is not a good idea because the goal of clustering is we want to group some objects by its similarity. However, since every object has its own cluster, then the clustering method does not really work.

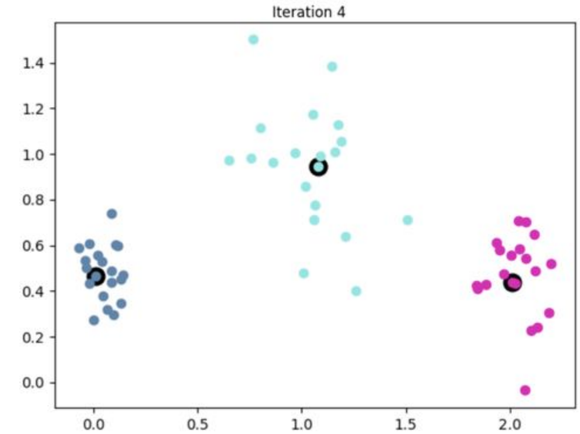
d. K-means plot



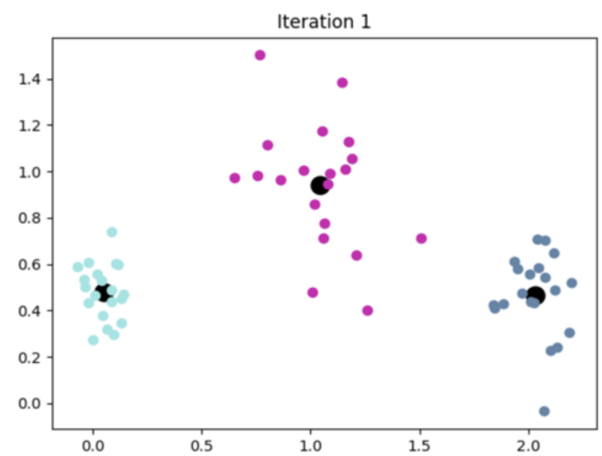
E. k-medoids

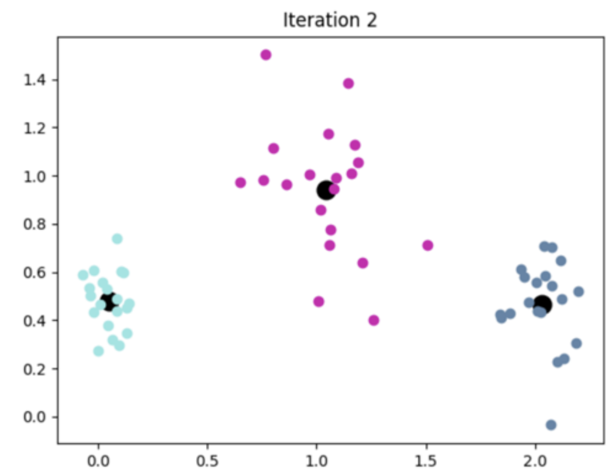






F. cheat_init() - k-means





cheat_init() - k-medoids