GOURAB BHATTACHARYYA – 17004888

ML – CSE512 – HW4

1)

M-step

1

The log-likelihood function we are trying to optimize is :-

$$\boxed{Q(\theta, \theta^{t-1}) = \sum_i \sum_k r_{ic} \log(\pi_c) + \sum_i \sum_k r_{ic} \log(P(x_i|\theta_c))}$$

(Answer)

Above taken from Murphy

2 We know that:

$$\pi_c = \frac{1}{N} \sum_i r_{ic} = \frac{r_c}{N}$$

Also given that:

$$R = \begin{pmatrix} 1 & 0 \\ 0.3 & 0.7 \\ 0 & 1 \end{pmatrix}$$

(P.T.O)

$$\pi_1 = \frac{1}{3}(1+0.3+0)$$

$$\Rightarrow \pi_1 = \frac{1}{3} \times 1.3$$

$$\Rightarrow \boxed{\pi_1 = 0.4333}$$

(Answer)

and,

$$\pi_2 = \frac{1}{3}(0+0.7+1)$$

$$\Rightarrow \pi_2 = \frac{1.7}{3}$$

$$\Rightarrow \boxed{\pi_2 = 0.5666}$$

(Answer)

(P.T.O)

(3)

we have,

$$\mu_c = \frac{\sum_i r_{ic} x_i}{r_c}$$

Plugging in the values we get :-

$$\mu_1 = \frac{(1\times1) + (0.3\times10) + (0\times20)}{1+0.3+0}$$

$$\Rightarrow \mu_1 = \frac{1+3+0}{1.3}$$

$$\Rightarrow \mu_1 = \frac{4}{1.3}$$

$$\Rightarrow \boxed{\mu_1 = 3.0769}$$ (Answer)

and,

$$\mu_2 = \frac{(0\times1) + (0.7\times10) + (1\times20)}{0+0.7+1}$$

$$\Rightarrow \mu_2 = \frac{0+7+20}{1.7}$$

$$\Rightarrow \mu_2 = \frac{27}{1.7}$$

$$\Rightarrow \boxed{\mu_2 = 15.8823}$$ (Answer)

(P.T.O)

4

we have,

$$\Sigma_c = \frac{\sum_i r_{ic}(x_i - \mu_c)(x_i - \mu_c)^T}{r_c}$$

and $$\sigma_c = \sqrt{\Sigma_c}$$

Plugging in the values :-

$$\Sigma_1 = \frac{1\cdot(1-3.0769)^2 + 0.3(10-3.0769)^2 + 0(20-3.0769)^2}{1+0.3+0}$$

$$\Rightarrow \Sigma_1 = \frac{4.314 + 14.379 + 0}{1.3}$$

$$\Rightarrow \Sigma_1 = \frac{18.693}{1.3}$$

$$\Rightarrow \underline{\Sigma_1 = 14.379}$$

$$\sigma_1 = \sqrt{\Sigma_1}$$

$$\Rightarrow \sigma_1 = \sqrt{14.379}$$

$$\Rightarrow \boxed{\sigma_1 = 3.792}$$ (Answer)

(P.T.O)

and,

$$\Sigma_2 = \frac{0(1-15.8823)^2 + 0.7(10-15.8823)^2 + 1.(20-15.8823)^2}{0+0.7+1}$$

$$\Rightarrow \Sigma_2 = \frac{0.7(-5.8823)^2 + 1.(4.118)^2}{1.7}$$

$$\Rightarrow \Sigma_2 = \frac{24.219 + 16.958}{1.7}$$

$$\Rightarrow \Sigma_2 = \frac{41.177}{1.7}$$

$$\Rightarrow \Sigma_2 = 24.222$$

finally,

$$\sigma_2 = \sqrt{\Sigma_2}$$

$$\Rightarrow \sigma_2 = \sqrt{24.222}$$

$$\Rightarrow \boxed{\sigma_2 = 4.922}$$ (Answer)

(P.T.O)

E-step

1. The probability of observation x_i belonging to cluster c:

$$\boxed{r_{ic} = \frac{\pi_k P(x_i \mid \theta_k^{(t-1)})}{\sum_{c'} \pi_{c'} P(x_i \mid \theta_{c'}^{(t-1)})}}$$

(Answer)

where the probability follows a Gaussian distribution; i.e;

$$P(x_i \mid \theta_c^{(t-1)}) = \frac{1}{\sigma_c\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu_c}{\sigma_c}\right)^2\right)$$

above is taken from Murphy.

(P.T.O)

② Using the equation shown in the previous step and plugging in the previously calculated values we get:-

below table with probability calculation:

i	c	x_i	σ_c	μ_c	$1/(\sigma_c \sqrt{2\pi})$	$\exp((-1/2) ((x_i - \mu_c) / \sigma_c)^2)$	$p(x_i \mid \theta_c)$	$\frac{1}{\sigma_c\sqrt{2\pi}} \exp -\frac{1}{2}(\frac{x_i-\mu_c}{\sigma_c})^2$
1	1	1	3.792	3.0769	0.105233	0.860715925	$p(x_1 \mid \theta_1)$	0.090575698
1	2	1	4.922	15.8823	0.0810734	0.010345900	$p(x_1 \mid \theta_2)$	0.000838778
2	1	10	3.792	3.0769	0.105233	0.188885891	$p(x_2 \mid \theta_1)$	0.019877024
2	2	10	4.922	15.8823	0.0810734	0.489614796	$p(x_2 \mid \theta_2)$	0.039694753
3	1	20	3.792	3.0769	0.105233	0.000047324	$p(x_3 \mid \theta_1)$	0.000004980
3	2	20	4.922	15.8823	0.0810734	0.704728894	$p(x_3 \mid \theta_2)$	0.057134792

below table calculates probability of observations:-

i	c	π_c	r_{ic}	$\frac{\pi_k p(x_i \mid \theta_k^{(t-1)})}{\sum_{c'} \pi_{c'} p(x_i \mid \theta_{c'}^{(t-1)})}$
1	1	0.4333	r_{11}	0.988
1	2	0.5666	r_{12}	0.012
2	1	0.4333	r_{21}	0.277
2	2	0.5666	r_{22}	0.723
3	1	0.4333	r_{31}	0
3	2	0.5666	r_{32}	1

so,

$$R_{new} = \begin{pmatrix} 0.988 & 0.012 \\ 0.277 & 0.723 \\ 0.000 & 1.000 \end{pmatrix}$$ (Answer)

2)

1) Given that,

$$\tilde{C} = \frac{1}{n} \tilde{X} \tilde{X}^T$$

$$\Rightarrow \tilde{C} = \frac{1}{n} \left((I - v_1 v_1^T) X \left((I - v_1 v_1^T) X \right)^T \right)$$ [by replacing $\tilde{X} = (I - v_1 v_1^T) X$]

Now by using the fact $(AB)^T = B^T A^T$ and $(I - v_1 v_1^T)$ is symmetric :

$$\tilde{C} = \frac{1}{n} (I - v_1 v_1^T) X X^T (I - v_1 v_1^T)$$

$$\Rightarrow \tilde{C} = \frac{1}{n} \left(X X^T - v_1 v_1^T X X^T - X X^T v_1 v_1^T + v_1 v_1^T X X^T v_1 v_1^T \right) \quad \text{——— (1)}$$

It is also given that

$$X X^T v_1 = n \lambda_1 v_1$$

$$\Rightarrow (X X^T v_1)^T = (n \lambda_1 v_1)^T$$

$$\Rightarrow v_1^T X X^T = n \lambda_1 v_1^T \quad \text{——— (2)}$$

Using value from equation — (2) and putting in equation — (1) we get :-

(09)

$$\tilde{C} = \frac{1}{n}\left(XX^T - v_1 n\lambda_1 v_1^T - n\lambda_1 v_1 v_1^T + v_1 n\lambda_1 \underline{v_1^T v_1} v_1^T\right)$$

$$\Rightarrow \tilde{C} = \frac{1}{n}\left(XX^T - v_1 n\lambda_1 v_1^T - n\lambda_1 v_1 v_1^T + v_1 n\lambda_1 \cdot 1 \cdot v_1^T\right)$$

[by using fact $v_1^T v_1 = 1$]

$$\Rightarrow \tilde{C} = \frac{1}{n}\left(XX^T - v_1 n\lambda_1 v_1^T + v_1 n\lambda_1 v_1^T - n\lambda_1 v_1 v_1^T\right)$$

[rearranging the terms]

$$\Rightarrow \tilde{C} = \frac{1}{n}\left(XX^T - n\lambda_1 v_1 v_1^T\right)$$

$$\Rightarrow \boxed{\tilde{C} = \frac{1}{n}XX^T - \lambda_1 v_1 v_1^T}$$

(proved)

(2)

2) we have,

$$\tilde{C} v_j = \left(\frac{1}{n}XX^T - \lambda_1 v_1 v_1^T\right) v_j$$

$$\Rightarrow \tilde{C} v_j = \frac{1}{n}(XX^T v_j) - \lambda_1 v_1 v_1^T v_j$$

$$\Rightarrow \tilde{C} v_j = \frac{1}{n}(n\lambda_j v_j) - \lambda_1 v_1 v_1^T v_j$$

[as $XX^T v_1 = n\lambda_1 v_1$
$\Rightarrow XX^T v_j = n\lambda_j v_j$]

$$\Rightarrow \tilde{C} v_j = \lambda_j v_j - \lambda_1 v_1 v_1^T v_j$$

(P.T.O)

$$\tilde{C}v_j = \lambda_j v_j - \lambda_1 v_1 v_1^T v_j$$

$$\Rightarrow \tilde{C}v_j = \lambda_j v_j - 0 \quad [\text{since } v_1^T v_j = 0 \text{ as } j \neq 1]$$

$$\Rightarrow \underline{\tilde{C}v_j = \lambda_j v_j}$$

From the above equation it's clear that for $j \neq 1$, v_j is a principal eigenvector of $\tilde{C}$ with same eigenvalue of λ_j. (proved)

If we consider $j=1$ then, we can write,

$$\tilde{C}v_1 = \lambda_1 v_1 - \lambda_1 v_1 v_1^T v_1$$

$$\Rightarrow \tilde{C}v_1 = \lambda_1 v_1 - \lambda_1 v_1 \cdot 1 \quad [\text{as } v_1^T v_1 = 1 \text{ as } j = 1]$$

$$\Rightarrow \tilde{C}v_1 = \lambda_1 v_1 - \lambda_1 v_1$$

$$\Rightarrow \underline{\tilde{C}v_1 = 0}$$

The above also proves that v_1 is an eigenvector of $\tilde{C}$ with eigenvalue of 0.

(P.T.O)

3) As given that:

$v_1, v_2, \cdots, v_k$ are the first k eigenvectors with largest eigenvalues of C, i.e; the principal basis vectors, therefore,

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k \quad \text{——— ③}$$

From – part 2 of this problem, we know that

→ for $\tilde{C}$, v_j are the principle eigenvectors with eigenvalues $(0, \lambda_2, \lambda_3, \cdots, \lambda_k)$.

Therefore, from equation – ③ above, λ_2 is the largest eigenvalues of $\tilde{C}$ (since, λ_1 is not an eigenvalue of $\tilde{C}$).

Hence, v_2 is the first principle eigenvector.

2

4) Below is the pseudocode for finding the first k principal eigenvectors of C :-

(Note:- I have used python a programming language for the below pseudocode)

findEigenVectors.py

```python
def findEigenVectors(C, k, f):
    lambda_List = []
    v_List = []

    for i in range(k):
        lambdaVal, v = f(C)
        C = C − lambdaVal ∗ v ∗ v.T

        lambda_List.append(lambdaVal)
        v_List.append(v)


    return v_List, lambda_List

```