Machine Learning - HW4

Faizaan Charania - 111463646

Q1 GMM

M Step

i) Output of E step is

$$R = \begin{pmatrix} 1 & 0 \\ 0.3 & 0.7 \\ 0 & 1 \end{pmatrix} \qquad x = [1, 10, 20]$$

where R_{ic} is the probability of observation x_i belonging to cluster c

1) ~~Maximising~~ Optimising the likelihood function is the same as optimising the log-likelihood function.

The log-likelihood function we are trying to optimise are

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

2) We know,

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

$$\therefore \pi_1 = \frac{1}{3}(1+0.3+0) = \frac{1.3}{3} = \frac{13}{30}$$

$$\pi_2 = \frac{1}{3}(0+0.7+1) = \frac{1.7}{3} = \frac{17}{30}$$

3)

$$\mu_c = \frac{\sum_{i=1}^{N} r_{ic} \cdot x_i}{r_c}$$

$$\therefore \mu_1 = \frac{1.1 + 0.3\times10 + 0\times20}{1+0.3+0} = \frac{4}{1.3} = \frac{40}{13}$$

$$\mu_2 = \frac{0.1 + 0.7\times10 + 1\times20}{0+0.7+1} = \frac{27}{1.7} = \frac{270}{17}$$

4) std. dev $\sigma_c = \sqrt{\Sigma_c}$

we know,

$$\Sigma_c = \frac{\sum_{i=1}^{N} r_{ic}(x_i - \mu_c)(x_i - \mu_c)^T}{r_c}$$

$$\therefore \Sigma_1 = \frac{1.\left(1-\frac{40}{13}\right)^2 + 0.3\left(10-\frac{40}{13}\right)^2 + 0.\left(20-\frac{40}{13}\right)^2}{1+0.3+0}$$

$$= \cancel{\frac{\frac{37^2}{9} + \frac{30}{9}\ 10}{1.3}} = \frac{2430}{169}$$

$$\therefore \sigma_1 = \sqrt{\Sigma_1} = 3.79$$

$$\Sigma_2 = \frac{0\cdot\left(1-\frac{270}{17}\right)^2 + 0\cdot 7\left(10-\frac{270}{17}\right)^2 + 1\cdot\left(20-\frac{270}{17}\right)^2}{0+0\cdot 7+1}$$

$$= \frac{7000}{289}$$

$$\therefore \sigma_2 = \sqrt{\Sigma_2} = \sqrt{\frac{7000}{289}} = 4.922$$

E-step

1) The probability that x_i belongs to cluster c is r_{ic}

$$r_{ic} = \frac{\pi_c\, p(x_i \mid \theta_c^{(t-1)})}{\Sigma_c\, \pi_c\, p(x_i \mid \theta_c^{(t-1)})}$$

where
$$p(x_i \mid \theta_c^{(t-1)}) = \frac{1}{\sigma_c\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x_i-\mu_c}{\sigma_c}\right)^2}$$

2) $$p(x_1 \mid \theta_1) = \frac{1}{3.79\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{1-\frac{40}{13}}{3.79}\right)^2} = 0.091$$

$$p(x_1 \mid \theta_2) = \frac{1}{4.92\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{1-\frac{270}{17}}{4.92}\right)^2} = 8.385\times 10^{-4}$$

$$p(x_2 | \theta_1) = \frac{1}{3.79\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{10 - \frac{40}{13}}{3.79}\right)^2} = 0.0199$$

$$p(x_2 | \theta_2) = \frac{1}{4.92\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{10 - \frac{270}{17}}{4.92}\right)^2} = 0.0397$$

$$p(x_3 | \theta_1) = \frac{1}{3.79\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{20 - \frac{40}{13}}{3.79}\right)} = 4.979 \times 10^{-6}$$

$$p(x_3 | \theta_2) = \frac{1}{4.92\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{20 - \frac{270}{17}}{4.92}\right)} = 0.0571$$

$$r_{11} = \frac{\frac{13}{30} \times 0.091}{\frac{13}{30} \times 0.091 + \frac{17}{30} \times (8.385 \times 10^{-4})} = 0.988$$

$$r_{12} = \frac{\frac{17}{30} \times (8.385 \times 10^{-4})}{\frac{13}{30} \times 0.091 + \frac{17}{30} \times (8.385 \times 10^{-4})} = 0.012$$

$$r_{21} = \frac{\frac{13}{30} \times (0.0199)}{\frac{13}{30} \times 0.0199 + \frac{17}{30} \times 0.0397} = 0.277$$

$$r_{22} = \frac{\frac{17}{30} \times (0.0397)}{\frac{13}{30} \times (0.0199) + \frac{17}{30} \times 0.0397} = 0.723$$

$$r_{31} = \frac{\frac{13}{30} \times (4.979 \times 10^{-6})}{\frac{13}{30} \times (4.979 \times 10^{-6}) + \frac{17}{30}(0.0571)} = 6.667 \times 10^{-5}$$

$$r_{32} = \frac{\frac{17}{30} \times (0.0571)}{\frac{13}{30}(4.979 \times 10^{-6}) + \frac{17}{30}(0.0571)} = 0.9999333$$

$$\therefore \quad R = \begin{pmatrix} 0.988 & 0.012 \\ 0.277 & 0.723 \\ 6.667 \times 10^{-5} & 0.999933 \end{pmatrix}$$

Q2 PCA via successive Deflation

We know that,

$$X = [x_i; \ldots ; x_n] \qquad (d \times n) \text{ matrix}$$

Covariance of $X \Rightarrow C = \frac{1}{n} X X^T$

$v_1 v_2 \ldots v_k$ are the first k eigenvectors of C

$$v_i^T v_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

The Deflated matrix $\quad \bar{X} = (I - v_1 v_1^T) X \quad$ —(a)

1) To prove,

$$\bar{C} = \frac{1}{n} X X^T - \lambda_1 v_1 v_1^T$$

We know that,

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

from (a),

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

$$= \frac{1}{n} (I - v_1 v_1^T) X X^T (I - v_1 v_1^T)^T \quad \left[(AB)^T = B^T A^T \right]$$

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

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

$$= \frac{1}{n} \left[(X - v_1 v_1^T X)(X^T - X^T v_1 v_1^T) \right]$$

$$= \frac{1}{n} \left[X X^T - X X^T v_1 v_1^T - v_1 v_1^T X X^T + v_1 v_1^T X X^T v_1 v_1^T \right]$$

—(b)

We know,

$$XX^T v_1 = n\lambda_1 v_1 \quad \text{— (c)}$$

Taking transpose on both sides

$$(XX^T v_1)^T = (n\lambda_1 v_1)^T$$

$$v_1^T XX^T = n\lambda_1 v_1^T \quad \text{— (d)}$$

Substitute (c) & (d) in (b), we get

$$\bar{C} = \frac{1}{n}\left[XX^T - (XX^T v_1)v_1^T - v_1(v_1^T XX^T) + v_1 v_1^T (XX^T v_1) v_1^T\right.$$

$$\therefore \bar{C} = \frac{1}{n}\left[XX^T - n\lambda_1 v_1 v_1^T - n\lambda_1 v_1 v_1^T + n\lambda_1 v_1 v_1^T v_1 v_1^T\right]$$

$$= \frac{1}{n}\left[XX^T - n\lambda_1 v_1 v_1^T - \cancel{n\lambda_1 v_1 v_1^T} + \cancel{n\lambda_1 v_1 v_1^T}\right] \quad (\because v_1^T v_1 = 1)$$

$$= \frac{1}{n}XX^T - \lambda_1 v_1 v_1^T$$

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

Hence, proved

2) Given $C v_j = \lambda_j v_j$ for $j \neq 1$ ... (e)

To prove $\bar{C} v_j = \lambda_j v_j$

$$\Rightarrow \bar{C} v_j = \left(\frac{1}{n} X X^T - \lambda_1 v_1 v_1^T \right) v_j$$

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

$$\because v_i^T v_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

$v_1^T v_j = 0 \quad [\because j \neq 1]$

$$\bar{C} v_j = \left(\frac{1}{n} X X^T \right) v_j - 0$$

$$= \frac{1}{n} X X^T v_j = \frac{1}{n} \left(X X^T v_j \right)$$

$$= \frac{1}{\cancel{n}} \times \cancel{n} \lambda_j v_j \quad \left(\because X X^T v_j = n \lambda_j v_j \right)$$

$$= \lambda_j v_j$$

$$\therefore \boxed{\bar{C} v_j = \lambda_j v_j}$$

3) Consider $\bar{C} v_1$

$$\bar{C} v_1 = \frac{1}{n}\left(X X^T - \lambda v_1 v_1^T\right) v_1$$

$$= \frac{1}{n} X X^T v_1 - \lambda v_1 v_1^T v_1$$

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

$$= \lambda_1 v_1 - \lambda_1 v_1 (1) \quad \left[\because v_1^T v_1 = 1\right]$$

$$= 0$$

$\therefore$ we have removed any component that lies in the direction of the first principal vector.

Consider $\bar{C} v_2$

$$\bar{C} v_2 = \frac{1}{n}\left(X X^T - n \lambda_1 v_1 v_1^T\right) v_2$$

$$= \frac{1}{n} X X^T v_2 - \lambda_1 v_1 \left(v_1^T v_2\right)$$

$$= \frac{1}{n}\left(X X^T v_2\right) - 0 \quad \left[\because v_1^T v_2 = 0\right]$$

$$= \frac{1}{n}\left(n \lambda_2 v_2\right)$$

$$= \lambda_2 v_2$$

$$\therefore \boxed{\bar{C} v_2 = \lambda_2 v_2}$$

$\therefore$ This implies that the first ~~eign~~ eigen vector of $\bar{C}$ is v_2

$\therefore \boxed{u = v_2}$

(4)

Consider function f that gives leading eigenvector & eigen value for C

$$[\lambda, u] = f(C)$$

After getting λ_1 & v_1, we need to update C for the next iteration.

$$\bar{C} = C - \lambda u u^T$$

$\bar{C}$ can then be used in the next iteration to get the next leading eigenvector

<u>**Pseudocode:**</u> Start

```
function [e_vecs e_vals] = my_func(f, C, k)

    e_vals = zeros(k, 1)
    e_vecs = zeros(k, size(C, 2))
    for i = 1 ... k
        [ λ , u ] = f(C)
        e_vals (i) = λ
        e_vecs(i, :) = u
        C = C - λ × u × u.transpose()
    end for

End
```