CSE 512 - Machine Learning HW 4

Mihir Chakradeo - 111462188

Q.1 Manual calculation of one round of EM for a GMM

[I] Manual calculation of 1 round of EM for a GMM.

→ M step : GIVEN: $x = [1\ 10\ 20]$, $R = \begin{bmatrix} 1 & 0 \\ 0.3 & 0.7 \\ 0 & 1 \end{bmatrix}$

① Write down the likelihood function you are trying to optimize.

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

② After performing the M step for mixing weights π_1 & π_2 what are the new values?

Ans. 2. $\pi_c = \frac{1}{N} \sum_i r_{ic}$

$\pi_1 = \frac{1}{N} \sum_i r_{i1} = \frac{1}{3}(1+0.3+0) = \frac{1.3}{3} = \underline{0.433}$

$\pi_2 = \frac{1}{N} \sum_i r_{i2} = \frac{1}{3}(0+0.7+1) = \frac{1.7}{3} = \underline{0.567}$

③ After performing M step for means μ_1 & μ_2, what are the new values?

Ans. 3. $\mu_c = \frac{\sum_i r_{ic} x_i}{r_c}$

$\therefore \mu_1 = \frac{\sum_i r_{i1} x_i}{r_1} = \frac{(1\times1)+(0.3\times10)+(0\times20)}{1.3} = \frac{4}{1.3} = \underline{3.07}$

$\mu_2 = \frac{\sum_i r_{i2} x_i}{r_2} = \frac{(0\times1)+(0.7\times10)+(1\times20)}{1.7} = \frac{27}{1.7} = \underline{15.88}$

(4) σ_1 & σ_2 . = ? (after M steps)

<u>Ans</u>. $\sigma_c^2 = \frac{\sum_i r_{ic} x_i^2}{r_c} - \mu_c^2$

$\therefore \sigma_1^2 = \frac{(1 \times 1) + (0.3 \times 100) + (0 \times 400)}{1.3} - 9.48 = \cancel{14.42}\ 14.38$

$\sigma_1 = \sqrt{14.42} = \underline{3.79}$

$\sigma_2^2 = \frac{(0 \times 1) + (0.7 \times 100) + (1 \times 400)}{1.7} - 252.17$

$= 24.30$

$\therefore \sigma_2 = \sqrt{24.30} = \underline{4.92}$

→ <u>E step</u> :

① Write down the formula for the prob. of obs. x_i belonging to cluster c.

<u>Ans</u>. $r_{ic} = \frac{\pi_c \, p(x_i \mid \mu_c^{(t-1)})}{\sum_c \pi_c \, p(x_i \mid \mu_c^{(t-1)})}$

$= \frac{\pi_c \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{(x_i - \mu_c^{(t-1)})^2}{2\sigma_c^2}\right)}{\sum_c \left[\pi_c \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{(x_i - \mu_c^{(t-1)})^2}{2\sigma_c^2}\right)\right]}$

② After performing the E-step, what is the new value of R?

Ans.

(i) $$r_{11} = \frac{\frac{\pi_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x_1-\mu_1^{(t)})^2}{2\sigma_1^2}\right)}{\frac{\pi_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x_1-\mu_1^{(t)})^2}{2\sigma_1^2}\right) + \frac{\pi_2}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x_1-\mu_2^{(t)})^2}{2\sigma_2^2}\right)}$$

$$= \frac{\left(\frac{0.433}{3.79}\right) \exp\left(-\frac{(1-3.07)^2}{2\times 14.38}\right)}{\left(\frac{0.433}{3.79}\right) \exp\left(-\frac{(1-3.07)^2}{2\times 14.38}\right) + \left(\frac{0.567}{4.92}\right) \exp\left(-\frac{(1-15.88)^2}{2\times 24.3}\right)}$$

$r_{11} = \underline{0.987}$.

(ii) $r_{12} = 1 - r_{11} = 1 - 0.987 = \underline{0.013}$.

(iii) $$r_{21} = \frac{\frac{\pi_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x_2-\mu_1^{(t)})^2}{2\sigma_1^2}\right)}{\frac{\pi_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x_2-\mu_1^{(t)})^2}{2\sigma_1^2}\right) + \frac{\pi_2}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x_2-\mu_2^{(t)})^2}{2\sigma_2^2}\right)}$$

$$= \frac{\left(\frac{0.433}{3.79}\right) \exp\left(-\frac{(10-3.07)^2}{2\times 14.38}\right)}{\left(\frac{0.433}{3.79}\right) \exp\left(-\frac{(10-3.07)^2}{2\times 14.38}\right) + \left(\frac{0.567}{4.92}\right) \exp\left(-\frac{(10-15.88)^2}{2\times 24.3}\right)}$$

$= \underline{0.2727}$.

(iv) $r_{22} = 1 - r_{21} = 1 - 0.2727$.

$\underline{r_{22} = 0.7273}$.

(v) $$r_{31} = \frac{\frac{\pi_1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x_3-\mu_1^{t+1})^2}{2\sigma_1^2}\right)}{\frac{\pi_1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x_3-\mu_1^{t+1})^2}{2\sigma_1^2}\right) + \frac{\pi_2}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(x_3-\mu_2^{t+1})^2}{2\sigma_2^2}\right)}$$

$$= \frac{\left(\frac{0.433}{0.79}\right)\exp\left(-\frac{(20-3.07)^2}{2\times 14.38}\right)}{\left(\frac{0.433}{0.79}\right)\exp\left(-\frac{(20-3.07)^2}{2\times 14.38}\right) + \left(\frac{0.567}{4.92}\right)\exp\left(-\frac{(20-15.88)^2}{2\times 24.3}\right)}$$

$$\underline{r_{31} = 6.6\times 10^{-5}}$$

(vi) $r_{32} = 1 - r_{31} = 1 - 6.6\times 10^{-5}$

$$\underline{r_{32} = 0.999934}$$

$$\therefore R = \begin{bmatrix} 0.987 & 0.013 \\ 0.2727 & 0.7273 \\ 6.6\times 10^{-5} & 0.999934 \end{bmatrix}$$

Q.2 PCA via Successive Deflation

(Q.2) PCA via Successive Deflation

[1] ST: $\tilde{C} = \frac{1}{n}\tilde{X}\tilde{X}^T$

covariance of $\tilde{C}$ is given by: $\tilde{C} = \frac{1}{n}XX^T - \lambda_1 v_1 v_1^T$.

Ans. Given: $\tilde{X} = (I - v_1v_1^T)X$.

$$\therefore \tilde{C} = \frac{1}{n}(I - v_1v_1^T)X[(I - v_1v_1^T)X]^T$$

$$= \frac{1}{n}(I - v_1v_1^T)XX^T(I - v_1v_1^T) \qquad \left\{ \begin{array}{c}(I - v_1v_1^T)^T = (I - v_1v_1^T) \\ \text{as symmetric}\end{array} \right\}$$

$$= \frac{1}{n}(I - v_1v_1^T)XX^T - \frac{1}{n}(I - v_1v_1^T)XX^Tv_1v_1^T$$

$$= \frac{1}{n}XX^T - \frac{1}{n}v_1v_1^TXX^T - \frac{1}{n}XX^Tv_1v_1^T + \frac{1}{n}v_1v_1^TXX^Tv_1v_1^T$$

given that: $XX^Tv_1 = n\lambda_1v_1$.

$$\therefore \tilde{C} = \frac{1}{n}XX^T - \frac{1}{n}v_1v_1^TXX^T - \lambda v_1v_1^T + v_1v_1^T\lambda_1v_1v_1^T.$$

$$= \frac{1}{n}XX^T - \lambda_1v_1v_1^T - \frac{1}{n}v_1v_1^TXX^T + v_1v_1^T\lambda_1v_1v_1^T.$$

Now, we know $XX^Tv_1 = n\lambda_1v_1$

i.e $v_1^TXX^T = n\lambda_1v_1^T$ {taking transpose}

$$\therefore \tilde{C} = \frac{1}{n}XX^T - \lambda_1v_1v_1^T - \frac{1}{\cancel{n}}v_1\cancel{n}\lambda_1v_1^T + v_1v_1^T\lambda_1v_1v_1^T$$

$$= \frac{1}{n}XX^T - \lambda_1v_1v_1^T - \cancel{\lambda_1v_1v_1^T} + \cancel{\lambda_1v_1v_1^T} \qquad \{v_1^Tv_1 = 1\}$$

$$\boxed{\tilde{C} = \frac{1}{n}XX^T - \lambda_1v_1v_1^T.}$$

[2] ST: for $j \neq 1$, if v_j is principal eigenvector ...

Ans. Given: $Cv_j = \lambda_j v_j$

To show that: $\tilde{C} v_j = \lambda_j v_j$.

From part 1, $\tilde{C} = \frac{1}{n} XX^T - \lambda_1 v_1 v_1^T$

post multiply by v_j: $\tilde{C} v_j = \frac{1}{n} XX^T v_j - \lambda_1 v_1 v_1^T v_j$

- Now, it is given that $j \neq 1$, and also that

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

Let's consider $v_1^T v_j$. $\{i=1,\ i \neq j,\ \therefore v_i^T v_j = 0\}$

this term $= 0$.

$$\therefore \tilde{C} v_j = \frac{1}{n} XX^T v_j - 0. = \frac{1}{n} XX^T v_j.$$

But, $\frac{1}{n} XX^T v_j = C$.

$$\therefore \tilde{C} v_j = C v_j.$$

But, given that $Cv_j = \lambda_j v_j$.

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

ie v_j is also the principal eigenvector of $\tilde{C}$ with same eigenvalue λ_j.

[3] Let u be the first principal eigenvector of $\tilde{C}$. Explain why $u = v_2$. (you may assume u is unit norm).

<u>Ans</u>. Let us see the proof for part 2 again.

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

if $j=1$. {first eigenvector}

$$\rightarrow \quad \tilde{C} v_1 = \cancel{XX} \, C v_1 - \lambda_1 v_1 \underline{v_1^T v_1}$$

$$= \lambda_1 v_1 - \lambda_1 v_1 \qquad \{v_1^T v_1 = 1\}$$

$$\tilde{C} v_1 = \bar{0}.$$

This is not a eigenvector (v_1).

$\therefore$ The first eigenvector start from $j=2$. (v_2).

We know that the first principal eigenvector has the largest eigenvalue.

$\therefore$ From above and from part 2, we can say that v_2 is the first principle eigenvector of $\tilde{C}$.

i.e $\boxed{u = v_2}$.

[4] Pseudocode

```
Algorithm (C, k):
    list = [ ]
    for i : 1 to k do
        (λi, vi) = f(C)
        list.append(vi)
        C = C − λi vi
    end for
    return list
end
```