Tutorial & Practical 8: EM Algorithm

Question 1

Consider a mixture distribution of the form

$$p(\mathbf{y}) = \sum_{k=1}^{K} p_k p(\mathbf{y}|k)$$

where the elements of $\mathbf{y}$ could be discrete or continuous or a combination of these. Denote the mean and the covariance of $p(\mathbf{y}|k)$ by $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ respectively.

1. Show that the mean of the mixture is given by

$$\mathbb{E}[\mathbf{y}] = \sum_{k=1}^{K} p_k \boldsymbol{\mu}_k$$

2. Show that the covariance of the mixture is given by

$$\text{Cov}[\mathbf{y}] = \sum_{k=1}^{K} p_k \left\{ \boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\top \right\} - \mathbb{E}[\mathbf{y}] \mathbb{E}[\mathbf{y}]^\top$$

Question 2

Assume we wish to use the EM algorithm to maximize the posterior distribution over parameters $p(\boldsymbol{\theta}|\mathbf{Y})$ for a model containing latent variables $\mathbf{Z}$ and where $\mathbf{Y}$ is the observed data set.

1. Describe the modification of the E-step compared to the maximum likelihood case.

2. Show that the quantity to be maximized in the M-step is given by

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{i-1}\right) + \log p(\boldsymbol{\theta})$$

   where

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{i-1}\right) = \sum_{\mathbf{z}} \log p(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\theta})\, p\left(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^{i-1}\right)$$

Question 3

Consider a random sample $T_1, \cdots, T_n$ from a Weibull distribution which has the *pdf*

$$f(t|\beta) = \frac{4}{\beta} t^3 e^{-t^4/\beta}, \quad t > 0; \quad \beta > 0.$$

Suppose we have observed $T_1 = y_1, \cdots, T_m = y_m$ and $T_{m+1} > c_{m+1}, \cdots, T_n > c_n$, where m is given, $m < n$, and $y_1, \cdots, y_m$ and $c_{m+1} \cdots, c_n$ are given numerical values. This implies that $T_1, \cdots, T_m$ are completely observed whereas $T_{m+1}, \cdots, T_n$ are partially observed in that they are right-censored. We want to use an EM algorithm to find the MLE of β.

1. Find the complete-data log-likelihood function $\ell(\beta) = \ln L(\beta)$.

2. In the E-step, we calculate

$$Q(\beta, \beta^{(k)}) = E[\ln L(\beta) \mid T_1 = y_1, \cdots, T_m = y_m, T_{m+1} > c_{m+1}, \cdots, T_n > c_n; \beta^{(k)}]$$

where $\beta^{(k)}$ is the kth-step estimate of β. Show that

$$\begin{aligned} Q(\beta, \beta^{(k)}) &= \ln 4^n - n \ln \beta + \sum_{i=1}^{m} \ln y_i^3 + \sum_{i=m+1}^{n} E[\ln T_i^3 | T_i > c_i, \beta^{(k)}] \\ &\quad - \frac{1}{\beta} \sum_{i=1}^{m} y_i^4 - \frac{1}{\beta} \sum_{i=m+1}^{n} c_i^4 - \frac{(n-m)\beta^{(k)}}{\beta}. \end{aligned}$$

(*Note*: The result $E(T^4 | T > c) = c^4 + \beta$ for the Weibull random variable T considered here can be used without proof.)

3. In the M-step, we maximise $Q(\beta, \beta^{(k)})$ with respect to β to find an update $\beta^{(k+1)}$ from $\beta^{(k)}$. Show that

$$\beta^{(k+1)} = \frac{1}{n} \left[\sum_{i=1}^{m} y_i^4 + \sum_{i=m+1}^{n} c_i^4 + (n-m)\beta^{(k)} \right].$$

Question 4

Suppose $Y_1, \cdots, Y_n$ are to be independently observed from a two-component normal mixture model which has the following pdf

$$g(y_i | \theta) = \frac{p}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_1)^2\} + \frac{1-p}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_2)^2\}; \quad i = 1, \cdots, n$$

where $\theta = (p, \mu_1, \mu_2, \sigma)$ are unknown parameters with $0 \leq p \leq 1$ and $\sigma > 0$. We want to use the EM algorithm to find the MLE $\hat{\theta}$ of θ. In order to do this, we introduce an unobserved indicator variable Z_i for each $i = 1, \cdots, n$, with $P(Z_i = 1) = p$ and $P(Z_i = 0) = 1-p$. Then we know the conditional pdf of Y_i given Z_i is $(Y_i | Z_i = 1) \stackrel{d}{=} N(\mu_1, \sigma^2)$ and $(Y_i | Z_i = 0) \stackrel{d}{=} N(\mu_2, \sigma^2)$.

1. Write down the observed-data log-likelihood function $\ln L(\theta | \mathbf{y}_n)$ where $\mathbf{y}_n = (y_1, \cdots, y_n)^T$ are observations of $Y_1, \cdots, Y_n$.

2. Write down the complete-data log-likelihood function $\ln L(\theta | \mathbf{y}_n, \mathbf{Z}_n)$ where $\mathbf{Z}_n = (Z_1, \cdots, Z_n)^T$.

3. Show that the conditional pdf of Z_i given $Y_i = y_i$ and θ is

$$P(Z_i = 1 | y_i, \theta) = \frac{p \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_1)^2\}}{p \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_1)^2\} + (1-p) \exp\{-\frac{1}{2\sigma^2}(y_i - \mu_2)^2\}}$$

and $P(Z_i = 0 | y_i, \theta) = 1 - P(Z_i = 1 | y_i, \theta)$. Then find $E(Z_i | y_i, \theta)$ and $\mathrm{Var}(Z_i | y_i, \theta)$.

4. Use the results in (3) to derive and simplify $Q(\theta|\theta^{(k)}) = E[\ln L(\theta|\mathbf{y}_n, \mathbf{Z}_n)|\mathbf{y}_n, \theta^{(k)}]$ that is to be used in the E-step. Here $\theta^{(k)} = (p^{(k)}, \mu_1^{(k)}, \mu_2^{(k)}, \sigma^{(k)})$ is the kth update of θ obtained from the EM algorithm.

5. Denote $z_i^{(k)} = E(Z_i|y_i, \theta^{(k)})$. By implementing the M-step show that the $(k+1)$th update of θ is

$$
\begin{aligned}
p^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n} z_i^{(k)}, \quad \mu_1^{(k+1)} = \frac{\sum_{i=1}^{n} z_i^{(k)} y_i}{\sum_{i=1}^{n} z_i^{(k)}}, \quad \mu_2^{(k+1)} = \frac{\sum_{i=1}^{n}(1-z_i^{(k)})y_i}{\sum_{i=1}^{n}(1-z_i^{(k)})} \\
\sigma^{(k+1)} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[z_i^{(k)}(y_i-\mu_1^{(k+1)})^2 + (1-z_i^{(k)})(y_i-\mu_2^{(k+1)})^2\right]} \\
z_i^{(k+1)} &= \frac{p^{(k+1)}\exp\{-\frac{(y_i-\mu_1^{(k+1)})^2}{2(\sigma^{(k+1)})^2}\}}{p^{(k+1)}\exp\{-\frac{(y_i-\mu_1^{(k+1)})^2}{2(\sigma^{(k+1)})^2}\} + (1-p^{(k+1)})\exp\{-\frac{(y_i-\mu_2^{(k+1)})^2}{2(\sigma^{(k+1)})^2}\}}
\end{aligned}
$$

6. The following R function implements the EM algorithm for the problem. Test the function using a simulated sample of $n = 400$ observations where the true value of θ is $(p = 0.25, \mu_1 = 5, \mu_2 = 10, \sigma = 1.5)$. Use different initial values, e.g. $\hat{\theta} = (0.4, \min(y), \max(y), \text{sd}(y))$; or $(0.4, \max(y), \min(y), \text{sd}(y))$; or $(0.9, \min(y), \max(y), \text{sd}(y) + 1)$; etc..

```
set.seed(1234)
y1=rnorm(n=100, mean=5, sd=1.5)
y2=rnorm(n=300, mean=10, sd=1.5)
y=c(y1,y2)

########begin tut3em.f function
tut3em.f=function(y, theta0, iter=20){
        theta=matrix(0, iter+1,4)
        n=length(y)
        z=matrix (0, n, iter+1) # matrix of conditional mean of Z_is

        loglik=rep(-1,iter+1)
        theta[1,]=theta0
        tem1=theta[1,1]*exp(-0.5*(y-theta[1,2])^2/(theta[1,4])^2)
        tem2=(1-theta[1,1])*exp(-0.5*(y-theta[1,3])^2/(theta[1,4])^2)
        loglik[1]=-n*log(sqrt(2*pi)*theta[1,4])+sum(log(tem1+tem2))
        z[,1]=tem1/(tem1+tem2)
        for(k in 1:iter){
                theta[k+1,1]=mean(z[,k]) #calculate p(k+1)
                theta[k+1,2]=sum(z[,k]*y)/sum(z[,k]) #calculate mu1(k+1)
                theta[k+1,3]=sum((1-z[,k])*y)/sum(1-z[,k]) #calculate mu2(k+1)
                theta[k+1,4]=sqrt((1/n)*(sum(z[,k]*(y-theta[k+1,2])^2)+sum((1-z[,k])*(y-theta[k+1,3])^2)))
                tem1=theta[k+1,1]*exp(-0.5*(y-theta[k+1,2])^2/(theta[k+1,4])^2)
                tem2=(1-theta[k+1,1])*exp(-0.5*(y-theta[k+1,3])^2/(theta[k+1,4])^2)
                loglik[k+1]=-n*log(sqrt(2*pi)*theta[k+1,4])+sum(log(tem1+tem2))
                z[,k+1]=tem1/(tem1+tem2)
                }
        result=list(theta=theta, z=z, loglik=loglik)
####end tut3em.f function

em1=tut3em.f(y, theta0=c(0.4, min(y), max(y), sd(y)))
em2=tut3em.f(y, theta0=c(0.4, max(y), min(y), sd(y)))
em3=tut3em.f(y, theta0=c(0.9, min(y), max(y), sd(y)+1), iter=70)
```