Finite Mixture Regression

5361 Homework 5

Qinxiao Shi *

13 October 2018

1 Verify The Validity of E- And M-steps

Suppose the density of y_i (conditional on x_i, $i = 1, ..., n$), is given by

$$f(y_i|x_i, \Psi) = \sum_{j=1}^{m} \pi_j \varphi(y_i; x_i^T \beta_j, \sigma^2)$$

whose complete log-likelihood is

$$\ell_c^n(\Psi) = \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij} \log[\pi_j \varphi(y_i - x_i^T \beta_j; 0, \sigma^2)]$$

E-step:

$$\begin{aligned}
Q(\Psi|\Psi^{(k)}) &= \mathbb{E}[\ln L(\Psi|(\mathbf{x}, \mathbf{y}, \mathbf{z})|\mathbf{x}, \mathbf{y}, \Psi^{(k)}] \\
&= \sum_{\mathbf{z}} p(\mathbf{z}|(\mathbf{x}, \mathbf{y}), \Psi^{(k)}) \ln p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \Psi^{(k)}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \left\{ \left[\sum_{\mathbf{z}} z_{ij} p(z_{ij}|(x_i, y_j), \Psi^{(k)}) \right] \left[\log \pi_j + \log(\varphi(y_i - x_i^T \beta_j; 0, \sigma^2)) \right] \right\} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \left\{ E(z_{ij}; y_i, x_i, \Psi^{(k)}) \left[\log \pi_j + \log(\varphi(y_i - x_i^T \beta_j; 0, \sigma^2)) \right] \right\} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \left\{ p_{ij}^{(k+1)} \left[\log \pi_j + \log(\varphi(y_i - x_i^T \beta_j; 0, \sigma^2)) \right] \right\}
\end{aligned}$$

By conditon, $z_{ij} = 1$ if ith observation is from jth component, and 0 otherwise,

$$\begin{aligned}
p_{ij}^{(k+1)} &= E(z_{ij}; y_i, x_i, \Psi^{(k)}) \\
&= p(z_{ij} = 1|y_i, x_i, \Psi^{(k)}) \\
&= \frac{p(y_i, x_i, z_{ij} = 1, \Psi^{(k)})}{p(y_i, x_i, \Psi^{(k)})} \\
&= \frac{\pi_j^k \varphi(y_i; x_i^T \beta_j^{(k)}, \sigma^{2k})}{\sum_{j=1}^{m} \pi_j^k \varphi(y_i; x_i^T \beta_j^{(k)}, \sigma^{2k})}
\end{aligned}$$

*qinxiao.shi@uconn.edu

M-step:

$$\sum f(y_i|x_i,\Psi) = \sum_{j=1}^{m} \pi_j \times 1 = \sum_{j=1}^{m} \pi_j = 1$$

(1) Let

$$\begin{aligned}
\frac{\partial Q(\Psi|\Psi^{(k)})}{\partial \pi_j} &= \frac{\partial}{\partial \pi_j}(\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log \pi_j) \\
&= \frac{\partial}{\partial \pi_j}\left\{\sum_{i=1}^{n}\sum_{j=1}^{m-1} p_{ij}^{(k+1)}\log \pi_j + \sum_{i=1}^{n} p_{im}^{(k+1)}\log(1-\pi_1-...-\pi_{m-1})\right\} \\
&= \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\pi_j} - \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\pi_m} \\
&= 0
\end{aligned}$$

Then

$$\begin{aligned}
&\Rightarrow \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\pi_j} = \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\pi_m} \\
&\Rightarrow \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\sum_{j=1}^{m}\pi_j} = \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\pi_m} \\
&\Rightarrow \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij}^{(k+1)}}{1} = \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\pi_m} = \sum_{i=1}^{n} 1 = n \\
&\Rightarrow \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{\pi_j} = n \\
&\Rightarrow \pi_j^{(k+1)} = \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)}}{n}
\end{aligned}$$

(2) Let

$$\frac{\partial Q(\Psi|\Psi^{(k)})}{\partial \beta_j} = -\frac{1}{2} \times 2\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} x_i \frac{y_i - x_i^T\beta_j}{\sigma^2} = 0$$

$$\begin{aligned}
&\Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} x_i \frac{y_i - x_i^T\beta_j}{\sigma^2} = \sum_{i=1}^{n} p_{ij}^{(k+1)} x_i \frac{y_i - x_i^T\beta_j}{\sigma^2} = 0 \\
&\Rightarrow \sum_{i=1}^{n} p_{ij}^{(k+1)} x_i x_i^T \beta_j = \sum_{i=1}^{n} p_{ij}^{(k+1)} x_i y_i \\
&\Rightarrow \beta_j^{(k+1)} = (\sum_{i=1}^{n} p_{ij}^{(k+1)} x_i x_i^T)^{-1}(\sum_{i=1}^{n} p_{ij}^{(k+1)} x_i y_i)
\end{aligned}$$

(3) Let

$$\frac{\partial Q(\Psi|\Psi^{(k)})}{\partial} = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}[\frac{(y_i - x_i^T\beta_j)^2}{2\sigma^4} - \frac{1}{2\sigma^2}]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}[(y_i - x_i^T\beta_j)^2 - \sigma^2] = 0$$
$$\Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}(y_i - x_i^T\beta_j)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}\sigma^2 = n\sigma^2$$

$$\Rightarrow \sigma^{2(k+1)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}(y_i - x_i^T\beta_j^{k+1})^2}{n}$$

2 EM Algorithm Function Code

```
regmix_em <- function(y, xmat, pi.init, beta.init, sigma.init, control) {
  xmat <- as.matrix(xmat)
  P <- matrix(0, nrow = nrow(xmat), ncol = length(pi.init))
  beta <- matrix(0, nrow = ncol(xmat), ncol = length(pi.init))
  conv <- 1

  ###pij^(k+1)
  for (i in 1:control$maxit) {
    for (j in 1:ncol(xmat)) {
      P[j, ] <- pi.init * dnorm(y[j] - xmat[j, ] %*% beta.init, mean = 0, sd = sigma.init) / su
    }


    ###pi^(k+1)
    p_i <- colMeans(P)

    ###beta^(k+1)
    for (j in 1:length(pi.init)){
      beta[ ,j] <- solve(t(xmat) %*% diag(P[, j]) %*% xmat) %*% t(xmat) %*% diag(P[, j]) %*% y
    }
      ###sigma^2(k+1)
      sigma <- sqrt(sum(P * (y %*% t(rep(1, length(pi.init))) - xmat %*% beta.init)^2)/n)
      if (sum(abs(pi.init-p_i))+sum(abs(beta.init-beta))+abs(sigma.init-sigma) < control$tol)
        break
  }
  return(list(p_i, beta, sigma, conv))
}
```

3 Generation Data and Estimating

```
regmix_sim <- function(n, pi, beta, sigma) {
    K <- ncol(beta)
    p <- NROW(beta)
    xmat <- matrix(rnorm(n * p), n, p) # normal covaraites
    error <- matrix(rnorm(n * K, sd = sigma), n, K)
    ymat <- xmat %*% beta + error # n by K matrix
    ind <- t(rmultinom(n, size = 1, prob = pi))
    y <- rowSums(ymat * ind)
    data.frame(y, xmat)
}

n <- 400
pi <- c(.3, .4, .3)
bet <- matrix(c( 1,  1,  1,
                -1, -1, -1), 2, 3)
sig <- 1
set.seed(1205)
dat <- regmix_sim(n, pi, bet, sig)

pi.init <- pi/pi/length(pi)
beta.init <- bet*0
sigma.init <- sig/sig
control = list(maxit = 500, tol = 1e-5)

es <- regmix_em(y = dat[,1], xmat = dat[,-1], pi.init, beta.init, sigma.init, control)
```

So the estimator of $\pi_j^{(k+1)}$ is

```
es[[1]]
```

```
## [1] 0.001666667 0.001666667 0.001666667
```

The estimator of $\beta_j^{(k+1)}$ is

```
es[[2]]
```

```
##           [,1]      [,2]      [,3]
## [1,] -3.207512 -3.207512 -3.207512
## [2,]  1.867485  1.867485  1.867485
```

The estimator of $\sigma^{2(k+1)}$ is

```r
es[[3]]
```

```
## [1] 0.06094433
```