EM Algorithm for Finite Mixture Regression

HW 5 of STAT 5361 Statistical Computing

Biju Wang[*]

10/09/2018

1 E- and M-Steps Derivations

1.1 E-Step Derivation

$$
\begin{aligned}
Q(\Psi|\Psi^{(k)}) &= \sum_{Z}\Big[p(Z|\mathbf{y}, X, \Psi^{(k)}) \log p(\mathbf{y}, Z|X, \Psi)\Big] && (1)\\
&= \sum_{Z}\left[p(Z|\mathbf{y}, X, \Psi^{(k)}) \log \prod_{i=1}^{n} p(y_i, \mathbf{z}_i|\mathbf{x}_i, \Psi)\right] && (2)\\
&= \sum_{i=1}^{n}\sum_{Z}\Big[p(Z|\mathbf{y}, X, \Psi^{(k)}) \log p(y_i, \mathbf{z}_i|\mathbf{x}_i, \Psi)\Big] && (3)\\
&= \sum_{i=1}^{n}\sum_{\mathbf{z}_i}\Big[p(\mathbf{z}_i|\mathbf{y}, X, \Psi^{(k)}) \log p(y_i, \mathbf{z}_i|\mathbf{x}_i, \Psi)\Big] && (4)\\
&= \sum_{i=1}^{n}\sum_{\mathbf{z}_i}\Big[p(\mathbf{z}_i|y_i, \mathbf{x}_i, \Psi^{(k)}) \log p(y_i, \mathbf{z}_i|\mathbf{x}_i, \Psi)\Big] && (5)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\Big[p(\mathbf{z}_i = (0, \cdots, 1, \cdots, 0)'|y_i, \mathbf{x}_i, \Psi^{(k)}) \log p(y_i, \mathbf{z}_i = (0, \cdots, 1, \cdots, 0)'|\mathbf{x}_i, \Psi)\Big] && (6)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\Big[p(z_{ij} = 1|y_i, \mathbf{x}_i, \Psi^{(k)}) \log p(y_i, \mathbf{z}_i = (0, \cdots, 1, \cdots, 0)'|\mathbf{x}_i, \Psi)\Big] && (7)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\Big[E(z_{ij}|y_i, \mathbf{x}_i, \Psi^{(k)})\{\log \pi_j + \log \varphi(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_j, 0, \sigma^2)\}\Big] && (8)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)}\{\log \pi_j + \log \varphi(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_j, 0, \sigma^2)\} && (9)\\
& && (10)
\end{aligned}
$$

where

$$\mathbf{y} = (y_1, \cdots, y_n)'$$

$$
Z = \begin{pmatrix}\mathbf{z}_1'\\ \vdots \\ \mathbf{z}_n'\end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1m}\\ \vdots & \vdots & \vdots & \vdots\\ z_{n1} & z_{n2} & \cdots & z_{nm}\end{pmatrix} \quad X = \begin{pmatrix}\mathbf{x}_1'\\ \vdots \\ \mathbf{x}_n'\end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p}\\ \vdots & \vdots & \vdots & \vdots\\ x_{n1} & x_{n2} & \cdots & x_{np}\end{pmatrix}
$$

$$
p_{ij}^{(k)} = E(z_{ij}|y_i, \mathbf{x}_i, \Psi^{(k)}) = \frac{\pi_j^{(k)}\varphi(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_j^{(k)}, 0, \sigma^{2^{(k)}})}{\sum_{j=1}^{m}\pi_j^{(k)}\varphi(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_j^{(k)}, 0, \sigma^{2^{(k)}})}
$$

The elaboration of the above steps are

[*]bijuwang@uconn.edu

- Step1→Step2: Use independence among $(y_i, \mathbf{z}_i)$
- Step3→Step4: Marginal density of $\mathbf{z}_i$
- Step4→Step5: Use the fact $\mathbf{z}_i \perp (y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n)|y_i$, we can get rid of $(y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n)$
- Step6→Step7: Easy to see conditional joint density is equal to condition marginal density

1.2 M-Step Derivation

Since we have

$$
\begin{aligned}
Q(\Psi|\Psi^{(k)}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \{\log \pi_j + \log \varphi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j, 0, \sigma^2)\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \log \pi_j - \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \log \sqrt{2\pi}\sigma - \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{2\sigma^2} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \log \frac{\pi_j}{\sqrt{2\pi}} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \log \sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{\sigma^2} \\
&= I_1 - \frac{1}{2} I_2 - \frac{1}{2} I_3
\end{aligned}
$$

From the above, we can see only I_3 contains $\boldsymbol{\beta}_j$ and

$$
I_3 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{\sigma^2} = \sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij}^{(k)} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{\sigma^2}
$$

To minimize I_3, we only need to fix j and optimize with regard to $\boldsymbol{\beta}_j$. We can directly use the formula from generazied least square method and obtain

$$
\boldsymbol{\beta}_j^{(k+1)} = (X'V^{-1}X)^{-1}X'V^{-1}\mathbf{y} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T p_{ij}^{(k)}\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i p_{ij}^{(k)} y_i\right) \quad j = 1, \cdots, m
$$

where

$$
V^{-1} = diag(p_{1j}^{(k)}, \cdots, p_{nj}^{(k)})
$$

Only I_2 and I_3 contains σ^2, since

$$
I_2 + I_3 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \log \sigma^2 + \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{\sigma^2}
$$

We minimize it with regard to σ^2 given $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^{(k+1)}$ and obtain

$$
\sigma^{2^{(k+1)}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k+1)})^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k+1)})^2}{n}
$$

Only I_1 contains π_j and

$$
I_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} \log \frac{\pi_j}{\sqrt{2\pi}} = -\frac{1}{2}\log(2\pi) \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k)} + \sum_{j=1}^{m} \left(\sum_{i=1}^{n} p_{ij}^{(k)}\right) \log \pi_j
$$

In order to maximize I_1 under constraint $\pi_1 + \cdots \pi_m = 1$. We use Lagrange multiplier method

$$
L(\pi_1, \cdots, \pi_m) = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} p_{ij}^{(k)}\right) \log \pi_j - \lambda \left(\sum_{j=1}^{m} \pi_j - 1\right)
$$

with λ a Lagrange multiplier. We can obtain

$$\pi_j^{(k+1)} = \frac{\sum_{i=1}^n p_{ij}^{(k)}}{\sum_{j=1}^m \sum_{i=1}^n p_{ij}^{(k)}} = \frac{\sum_{i=1}^n p_{ij}^{(k)}}{n} \quad j = 1, \cdots, m$$

2 A Function to Implement EM Algorithm