

A short overview

Yonghyun Kwon

4/24/2020

Problem formulation

Let a random variable $X = (X_1, \dots, X_n)$ be hidden class variable in a domain $\{1, \dots, K\}$ and $Y = (Y_1, \dots, Y_n)$ be observable random variable that depends on class variable X . When dealing with finite mixture model, we assumed that (X_i, Y_i) 's are independent.

$$p(y, x) = \prod p(y_i, x_i)$$
$$p(y_i | x_i = l) = N_{y_i}(\mu_l, \Sigma_l)$$

Now we consider the case when (X_i, Y_i) are no longer independent. That is,

$$y | X \sim N(\tilde{X}\beta, \Sigma) \tag{1}$$

$$X_i \stackrel{iid}{\sim} Multinomial(p_1, \dots, p_k) \tag{2}$$

where \tilde{X} is a matrix of indicator functions $I_{ik} = I\{X_i = k\}$

$$\tilde{X} = \begin{pmatrix} I_{11} & I_{12} & \cdots & I_{1K} \\ I_{21} & I_{22} & \cdots & I_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ I_{n1} & I_{n2} & \cdots & I_{nK} \end{pmatrix}$$

and $\beta = (\mu_1, \dots, \mu_k)^T$.

For illustrative purpose, assume the simplest case when Σ is known. So the parameter we estimate is $\theta = (\beta, p_1, \dots, p_k)$. From (1) the complete likelihood is

$$\begin{aligned} \log f(x, y; \theta) &= \log f(y | x; \theta) + \log f(x; \theta) \\ &= -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (y - \tilde{X}\beta)^T \Sigma^{-1} (y - \tilde{X}\beta) + \\ &\quad \log p_1 \sum_{i=1}^n I(X_i = 1) + \cdots + \log p_K \sum_{i=1}^n I(X_i = K) \end{aligned}$$

To find the Q function in E-step, we need the conditional expectation of

$$\begin{aligned}
E_{\theta^*}[I(X_i = k) | Y] &= P_{\theta^*}(X_i = k | Y) \\
&= \sum_{k_1, \dots, k_n} P_{\theta^*}(X_1 = k_1, X_2 = k_2, \dots, X_i = k, \dots, X_n = k_n | Y)
\end{aligned}$$

and

$$E_{\theta^*}[\tilde{X}^T \Sigma^{-1} \tilde{X} | Y]$$

where

$$\begin{aligned}
(E_{\theta^*}[\tilde{X}^T \Sigma^{-1} \tilde{X} | Y])_{kl} &= E_{\theta^*} \left[\sum_{i,j} I_{ik} (\Sigma^{-1})_{ij} I_{jl} \right] \\
&= \sum_{i,j} (\Sigma^{-1})_{ij} P_{\theta^*}(X_i = k, X_j = l | Y) \\
&= \sum_{i,j} (\Sigma^{-1})_{ij} \sum_{k_1, \dots, k_n} P_{\theta^*}(X_1 = k_1, \dots, X_i = k, \dots, X_j = l, \dots, X_n = k_n | Y)
\end{aligned}$$

Since we assumed that Σ is known, we only have to update β and p_k 's in M-step.

$$\beta = E_{\theta^*}[\tilde{X}^T \Sigma^{-1} \tilde{X} | Y]^{-1} Y^T \Sigma^{-1} E_{\theta^*}[\tilde{X} | Y]$$

$$p_k = \frac{\sum_i P_{\theta^*}(X_i = k | Y)}{\sum_{i,l} P_{\theta^*}(X_i = l | Y)}$$

In most cases, Σ is not known.

Questions

It seems that the computation of MLE is already complex enough even when Σ is known. When Σ is unconstrained positive definite matrix, one can parametrize $\Theta = \Sigma^{-1}$ and find the mle of Θ but it still requires a lot of computations. For structured covariance, AR(1) for example, the estimation is even more complex. I, therefore, am thinking of considering the case when observed random variable y 's are conditionally independent so that a lot of elements in Σ^{-1} are zero or the covariance matrix is blockwise diagonal. (When the covariance matrix is blockwise diagonal, it can be used to apply for the case when we have replications as we discussed in class) This simplification would reduce the parameter space and therefore increase the possibility of global maximum of the likelihood.

Another problem is that we need K^n memory to compute the conditional expectations: $E_{\theta^*}[\tilde{X} | Y]$ and $E_{\theta^*}[\tilde{X}^T \Sigma^{-1} \tilde{X} | Y]$. One may consider a nice approximation for this to save memory and computation time.

One may choose other distributions for prior $X_i \stackrel{iid}{\sim} Multinomial(p_1, \dots, p_k)$ if desirable.