Nonzero Mean ASH

Chaoxing Dai

2014-Sep-03

Likelihood

By assuming a unimodal distribution, we have the hierarchical model being the following:

$$\beta_j \sim g(.) \sim \sum_{k=1}^{K} \pi_k f_k(.) \sim \sum_{k=1}^{K} \pi_k N(\cdot; \mu, \sigma_k^2)$$

$$\widehat{\beta}_j \mid \beta_j, s_j^2 \sim N(\cdot; \beta_j, s_j^2)$$

Then we would have the likelihood for π and μ being (Here we assume iid or exchangable data)

$$L(\pi, \mu) = \prod_{j=1}^{n} p(D_j \mid \pi, \mu) = \prod_{j=1}^{n} \int p(\widehat{\beta}_j \mid s_j, \pi, \mu, \beta_j) p(\beta_j \mid \pi, \mu) d\beta_j$$

$$= \prod_{j=1}^{n} \int \frac{1}{\sqrt{2\pi s_j^2}} exp\left\{-\frac{1}{2s_j^2}\left(\widehat{\beta}_j - \beta_j\right)^2\right\} \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left\{-\frac{1}{2\sigma_k^2}\left(\beta_j - \mu\right)^2\right\} d\beta_j$$

$$= \prod_{j=1}^{n} \sum_{k=1}^{K} \pi_k \int \frac{1}{\sqrt{2\pi s_j^2 \times 2\pi\sigma_k^2}} exp\left\{-\frac{1}{2s_j^2}\left(\widehat{\beta}_j - \beta_j\right)^2 - \frac{1}{2\sigma_k^2}\left(\beta_j - \mu\right)^2\right\} d\beta_j$$

$$= \prod_{j=1}^{n} \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\left(s_j^2 + \sigma_k^2\right)}} exp\left[-\frac{\left(\widehat{\beta}_j - \mu\right)^2}{2\left(s_j^2 + \sigma_k^2\right)}\right]$$

Where the term after π_k is the density of $N(\cdot; \mu, s_j^2 + \sigma_k^2)$ at $\widehat{\beta}_j$, i.e.

$$L(\pi, \mu) = \prod_{j=1}^{n} \sum_{k=1}^{K} \pi_k N(\widehat{\beta}_j; \mu, s_j^2 + \sigma_k^2)$$

EM Algorithm

For convenience we denote $x_j \equiv \widehat{\beta}_j$, and introduce a latent variable z_j specifying the mixture component that x_j belongs to. We then have

$$p(x_j, z_j \mid \mu, \pi) = \prod_{k=1}^{K} [\pi_k N(x_j; \mu, s_j^2 + \sigma_k^2)]^{I_{z_j} = k}$$

$$log\, p(x_j, z_j \mid \mu, \pi) = \sum_{k=1}^{K} [I_{z_j} = k][log(\pi_k) - \frac{1}{2} log(2\pi) - \frac{1}{2} log(s_j^2 + \sigma_k^2) - \frac{\left(x_j - \mu\right)^2}{2\left(s_j^2 + \sigma_k^2\right)}]$$

$$log\, p(X, Z \mid \mu, \pi) = \sum_{j=1}^{n} \sum_{k=1}^{K} [I_{z_j} = k][log(\pi_k) - \frac{1}{2} log(2\pi) - \frac{1}{2} log(s_j^2 + \sigma_k^2) - \frac{(x_j - \mu)^2}{2\left(s_j^2 + \sigma_k^2\right)}]$$

Then we will have the E-step, given the estimate of π and mu, we calculate the responsibiity of j-th data point to the k-th mixture component.

$$\omega_{jk} = p(z_j = k \mid x_j, \pi, \mu) = \frac{\pi_k N(x_j; \mu, s_j^2 + \sigma_k^2)}{\sum_{l=1}^{K} \pi_l N(x_j; \mu, s_j^2 + \sigma_l^2)}$$

For the M-step, we take ω_{jk} in the previous step, and maximize the expected loglikelihood over the latent variable Z.

$$Q = E_{Z|X,\mu,\pi}[log\, p(X, Z \mid \mu, \pi)] = \sum_{j=1}^{n} \sum_{k=1}^{K} [\omega_{jk}][log(\pi_k) - \frac{1}{2} log(2\pi) - \frac{1}{2} log(s_j^2 + \sigma_k^2) - \frac{(x_j - \mu)^2}{2\left(s_j^2 + \sigma_k^2\right)}]$$

We take partial derivatives of Q with respect to π and μ and set them to be 0 to get the next estimates of our parameters, subject to the constraint that $\sum_{k=1}^{K} \pi_k = 1$. Then we have

$$\widehat{\mu} = \frac{\sum_{j=1}^{n} \sum_{k=1}^{K} \frac{\omega_{jk} x_j}{s_j^2 + \sigma_k^2}}{\sum_{j=1}^{n} \sum_{k=1}^{K} \frac{\omega_{jk}}{s_j^2 + \sigma_k^2}}$$

$$\widehat{\pi_k} = \frac{1}{n} \sum_{j=1}^{n} \omega_{jk}$$

Then we go back to the E-step and repeat with the new estimate, until we get a convergence.