The Gaussian distribution

butapro7@gmail.com

January 2015

1 M-step. Re-estimate the parameters using the current responsibilities

We shall call $\{X, Z\}$ the complete data set, and we shall refer to the actual observed data X as incomplete.

$$p(x|\theta) = \sum_z p(x, z|\theta) = \sum_z p(z|x)p(x|\theta) = E[p(x|\theta)] \tag{1}$$

1.1 Optimization with respect to the mean

$$\begin{aligned}
\frac{\partial}{\partial \mu_k} E[\log p(x|z;\theta] =& \frac{\partial}{\partial \mu_k} \sum_{t=1}^{T} \sum_{k=1}^{K} E[z_{tk}]\{\log \pi_k + \log \mathcal{N}(x_t|\theta)\} \\
=& \sum_{t=1}^{T} E[z_{tk}] \frac{\partial}{\partial \mu_k} \log \mathcal{N}(x_t|\theta) \\
=& \sum_{t=1}^{T} \gamma(z_{tk}) \frac{\frac{\partial}{\partial \mu_k} \mathcal{N}(x_t|\theta)}{\mathcal{N}(x_t|\theta)} \\
=& \sum_{t=1}^{T} \gamma(z_{tk}) \frac{\frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{\Sigma^{\frac{1}{2}}} e^{\{-\frac{1}{2}(x_t-\mu_k)^T \Sigma^{-1}(x_t-\mu_k)\}}}{\mathcal{N}(x_t|\theta)} \\
& * \frac{\partial}{\partial \mu_k} \{-\frac{1}{2}(x_t - \mu_k)^T \Sigma^{-1}(x_t - \mu_k)\} \\
=& \sum_{t=1}^{T} \gamma(z_{tk}) \frac{\mathcal{N}(x_t|\theta)}{\mathcal{N}(x_t|\theta)} \Sigma^{-1}(x_t - \mu_k) \\
=& \sum_{t=1}^{T} \gamma(z_{tk}) \Sigma^{-1}(x_t - \mu_k) \\
=& \sum_{t=1}^{T} \gamma(z_{tk})(x_t - \mu_k) = 0
\end{aligned} \tag{2}$$

So, rearranging we obtain:

$$\mu_k = \frac{1}{N_k} \sum_{t=1}^{T} \gamma(z_{tk}) x_t \tag{3}$$

Where we have defined:

$$N_k = \sum_{t=1}^{T} \gamma(z_t k) \tag{4}$$

1.2 Optimization with respect to the sigma

First use logarithm to simplify the log-likelihood function

$$\ln p(\vec{x}|\mu, \Sigma) = -\frac{ND}{2} \ln(2\pi) - \frac{N}{2} \ln|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^T \Sigma^{-1} (x_n - \mu) \tag{5}$$

and next take the derivative of the log-likelihood function of the Gaussian mixture

Next use the facts that

$$-\log|\Sigma| = \log|\Sigma^{-1}| \tag{6}$$

$$\frac{\partial}{\partial A} \log(A) = A^{-T} \tag{7}$$

and

$$a^T A a = tr(a^T A a) = tr(a a^T A) \tag{8}$$

$$\frac{\partial}{\partial A} tr(AB) = B^T \tag{9}$$

and differentiate $E[\log p(x|z;\theta]$ by Σ^{-1}

$$\begin{aligned}
\frac{\partial}{\partial \Sigma_k^{-1}} E[\log p(x|z;\theta] &= \frac{\partial}{\partial \Sigma_k^{-1}} \sum_{t=1}^{T} \sum_{k=1}^{K} E[z_{tk}] \{\log \pi_k + \log \mathcal{N}(x_t|\theta)\} \\
&= \sum_{t=1}^{T} E[z_{tk}] \frac{\partial}{\partial \Sigma_k^{-1}} \log \mathcal{N}(x_t|\theta) \\
&= \sum_{t=1}^{T} \gamma(z_{tk}) \frac{\partial}{\partial \Sigma_k^{-1}} (\frac{1}{2} \log|\Sigma_k^{-1}| - \frac{1}{2} tr[(x_t - \mu_k^{new})(x_t - \mu_k^{new})^T \Sigma_k^{-1}]) \\
&= \sum_{t=1}^{T} \gamma(z_{tk}) \frac{1}{2} \Sigma_k^T - \sum_{t=1}^{T} \gamma(z_{tk}) \frac{1}{2} (x_t - \mu_k^{new})^T (x_t - \mu_k^{new}) = 0
\end{aligned} \tag{10}$$

$$\Sigma_k^T = \frac{\sum_{t=1}^T \gamma(z_{tk})(x_t - \mu_k^{new})^T(x_t - \mu_k^{new})}{\sum_{t=1}^T \gamma(z_{tk})} \tag{11}$$

Totally

$$\Sigma_{ML} = \frac{1}{N_k}\sum_{t=1}^T \gamma(z_{tk})(x_t - \mu_k^{new})(x_t - \mu_k^{new})^T \tag{12}$$

1.3 Optimization with respect to the mixing coefficient

To find the derivative of the function with constraint $\sum_{k=1}^K \pi_k = 1$ we must use a Lagrangian multipliers

$$\begin{aligned}
&\frac{\partial}{\partial \pi_k}(E[\log p(x|z;\theta]+\lambda(\sum_{j=1}^K \pi_j - 1)) \\
&\qquad =\frac{\partial}{\partial \pi_k}\sum_{t=1}^T\sum_{k=1}^K E[z_{tk}]\{\log \pi_k + \log \mathcal{N}(x_t|\theta)\} + \frac{\partial}{\partial \pi_k}\lambda(\sum_{j=1}^K \pi_j - 1) \\
&\qquad =\sum_{t=1}^T \gamma(z_{tk})\frac{\partial}{\partial \pi_k}\{\log \pi_k + \log \mathcal{N}(x_t|\theta)\} + \frac{\partial}{\partial \pi_k}\lambda(\sum_{j=1}^K \pi_j - 1) \\
&\qquad =\sum_{t=1}^T \gamma(z_{tk})\frac{\partial}{\partial \pi_k}\{\log \pi_k\} + \frac{\partial}{\partial \pi_k}\lambda(\sum_{j=1}^K \pi_j - 1) \\
&\qquad =\sum_{t=1}^T \gamma(z_{tk})\frac{1}{\pi_k} + \lambda \\
&\qquad =\sum_{t=1}^T \frac{\mathcal{N}(x_t|\theta)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_t|\theta_j)} + \lambda = 0
\end{aligned} \tag{13}$$

So, let us obtain the first equation by multiplying expression by π_k

$$\sum_{t=1}^T \gamma(z_{tk}) + \pi_k \lambda = 0 \tag{14}$$

Sum over k make the second equation

$$\sum_{t=1}^T\sum_{k=1}^K \gamma(z_{tk}) + \sum_{k=1}^K \pi_k \lambda = 0 \tag{15}$$

Using

$$\sum_{k=1}^{K} \gamma(z_{tk}) = 1 \tag{16}$$

and

$$\sum_{k=1}^{K} \pi_k = 1 \tag{17}$$

we obtain from the second equation (17)

$$T + \lambda = 0 \tag{18}$$

$$\lambda = -T \tag{19}$$

Next, substitute λ in the first equation (16)

$$\pi_k = -\frac{\sum_{t=1}^{T} \gamma(z_{tk})}{\lambda} = \frac{N_k}{T} \tag{20}$$

Finally

$$\pi_{ML} = \frac{N_k}{T} \tag{21}$$