Brief Article

The Author

May 3, 2014

1 EM with prior on θ

On pp. 454 of PRML, we have an expression for $\ln p(\boldsymbol{\theta} \mid X)$. Let's derive it.

$$
\begin{align}
p(\boldsymbol{\theta} \mid X) &= \frac{\sum_Z p(\boldsymbol{\theta}, Z, X)}{p(X)} \tag{1} \\
&= \frac{\sum_Z p(Z, X \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(X)}, \tag{2} \\
\ln p(\boldsymbol{\theta} \mid X) &= p(X \mid \theta) + \ln p(\boldsymbol{\theta}) - \ln p(X) \tag{3} \\
&= L(q, \boldsymbol{\theta}) + \mathrm{KL}(q \| p) + \ln p(\boldsymbol{\theta}) - \ln p(X). \tag{4}
\end{align}
$$

2 9.12

Basically, the idea is to represent mean and covariance in terms of moments.

$$E(\boldsymbol{x}) = \sum_{\boldsymbol{x}} \sum_{k=1}^{K} \pi_k p(\boldsymbol{x} \mid k)\boldsymbol{x} \tag{5}$$

$$= \sum_{k=1}^{K} \pi_k \sum_{\boldsymbol{x}} p(\boldsymbol{x} \mid k)\boldsymbol{x} \tag{6}$$

$$= \sum_{k=1}^{K} \pi_k \boldsymbol{\mu}_k, \tag{7}$$

$$\mathrm{cov}(\boldsymbol{x}) = \sum_{\boldsymbol{x}} \sum_{k=1}^{K} \pi_k p(\boldsymbol{x} \mid k)\boldsymbol{x}\boldsymbol{x}^T - E(\boldsymbol{x})E(\boldsymbol{x})^T \tag{8}$$

$$= \sum_{k=1}^{K} \pi_k \sum_{\boldsymbol{x}} p(\boldsymbol{x} \mid k)\boldsymbol{x}\boldsymbol{x}^T - E(\boldsymbol{x})E(\boldsymbol{x})^T \tag{9}$$

$$= \sum_{k=1}^{K} \pi_k [\Sigma_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T] - E(\boldsymbol{x})E(\boldsymbol{x})^T \tag{10}$$

3 9.15

Taking the derivative of expected complete data log likelihood w.r.t. $\mu_{k,i}$ and set it to zero, we have

$$
\begin{aligned}
\frac{\partial L}{\partial \mu_{k,i}} &= \sum_{n=1}^{N} \gamma_{n,k} \left[\frac{x_{n,i}}{\mu_{k,i}} - \frac{1 - x_{n,i}}{1 - \mu_{k,i}} \right] && (11) \\
&= 0, && (12) \\
0 &= \sum_{n=1}^{N} \gamma_{n,k} \left[\frac{x_{n,i}}{\mu_{k,i}} - \frac{1 - x_{n,i}}{1 - \mu_{k,i}} \right] [\mu_{k,i}(1 - \mu_{k,i})] && (13) \\
&= \sum_{n=1}^{N} \gamma_{n,k} \left[x_{n,i}(1 - \mu_{k,i}) - (1 - x_{n,i})\mu_{k,i} \right] && (14) \\
&= \sum_{n=1}^{N} \gamma_{n,k} \left[x_{n,i} - \mu_{k,i} \right], && (15) \\
\mu_{k,i} &= \frac{\sum_{n=1}^{N} \gamma_{n,k} x_{n,i}}{\sum_{n=1}^{N} \gamma_{n,k}} && (16)
\end{aligned}
$$

Doing this for every i, then arrange them into a vector, we have desired result.

4 9.24

$$
\begin{aligned}
\ln p(X, Z \mid \boldsymbol{\theta}) &= \ln p(Z \mid X, \boldsymbol{\theta}) + \ln(X \mid \boldsymbol{\theta}), && (17) \\
L(q, \boldsymbol{\theta}) &= \sum_{Z} q(Z) \ln p(X, Z \mid \boldsymbol{\theta}) - \sum_{Z} q(Z) \ln q(Z) && (18) \\
&= \sum_{Z} q(Z) [\ln p(Z \mid X, \boldsymbol{\theta}) + \ln(X \mid \boldsymbol{\theta})] - \sum_{Z} q(Z) \ln q(Z) && (19) \\
&= \sum_{Z} q(Z) [\ln p(Z \mid X, \boldsymbol{\theta}) - \ln q(Z)] + \sum_{Z} q(Z) \ln(X \mid \boldsymbol{\theta}) && (20) \\
&= -\mathrm{KL}(q \| p) + \ln(X \mid \boldsymbol{\theta}). && (21)
\end{aligned}
$$