Bayesian Models for Machine Learning
Problem Set #2

Si Kai Lee sl3950@columbia.edu

October 21, 2016

Problem 1

We want to find $W' = \arg\max_W p(x_{1:n}, W)$. Split $\ln p(x_{1:n}, W) = \sum_{i=1}^n \ln p(x_i, W)$ and expand $\ln p(x_i, W)$ to obtain $\ln p(x_i, W) = \ln p(x_i, W, z_i) - \ln p(z_i|x_i, W)$. We also assume that the joint $p(x_i, W, z_i)$ can be represented as $p(x_i, W, z_i) = p(w)p(x_i|z_i, W)p(z_i)$

The Expectation-Maximisation (EM) equation is derived as:

$$\begin{aligned}
\ln p(x_{1:n}, W) &= \sum_{i=1}^n \ln p(x_i, W, z_i) - \sum_{i=1}^n \ln p(z_i|x_i, W) \\
&= \sum_{i=1}^n \int_{q(\phi)} q(\phi) \frac{\ln p(x_i, W, z_i)}{q(\phi)} d\phi + \sum_{i=1}^n \int_{q(\phi)} q(\phi) \frac{q(\phi)}{\ln p(z_i|x_i, W)} d\phi \\
&= \mathbb{E}_{q(\phi)}[\ln p(x_i, W, z_i)] + KL(q||p)
\end{aligned}$$

We can express $\ln p(x_i, W, z_i)$ as the following:

$$\ln p(x_i, W, z_i) = \frac{dk}{2} \ln \frac{\lambda}{2\pi} - \frac{\lambda}{2} tr(W^T W) + \frac{d}{2} \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x_i - Wz_i)^T (x_i - Wz_i) + \frac{k}{2} \ln \frac{1}{2\pi} - \frac{1}{2} z_i^T z_i$$

Taking expectations w.r.t to $q(\phi)$, we have:

$$\begin{aligned}
\mathbb{E}_{q(\phi)}[\ln p(x_i, W, z_i)] &= \frac{dk}{2} \ln \frac{\lambda}{2\pi} - \frac{\lambda}{2} tr(W^T W) + \frac{d}{2} \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2} \mathbb{E}_{q(\phi)}[x_i^T x_i - 2x_i^T W z_i + z_i^T W^T W z_i] \\
&\quad + \frac{k}{2} \ln \frac{1}{2\pi} - \frac{1}{2} \mathbb{E}_{q(\phi)}[z_i^T z_i] \\
&= \frac{dk}{2} \ln \frac{\lambda}{2\pi} - \frac{\lambda}{2} tr(W^T W) + \frac{d}{2} \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2} x_i^T x_i + \frac{1}{\sigma^2} x_i^T W \mathbb{E}_{q(\phi)}[z_i] \\
&\quad - \frac{1}{2\sigma^2} \mathbb{E}_{q(\phi)}[tr(z_i^T W^T W z_i)] - \frac{1}{2} \mathbb{E}_{q(\phi)}[tr(z_i^T z_i)] \\
&= \frac{dk}{2} \ln \frac{\lambda}{2\pi} - \frac{\lambda}{2} tr(W^T W) + \frac{d}{2} \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2} x_i^T x_i + \frac{1}{\sigma^2} x_i^T W \mathbb{E}_{q(\phi)}[z_i] \\
&\quad - \frac{1}{2\sigma^2} \mathbb{E}_{q(\phi)}[tr(z_i z_i^T W^T W)] - \frac{1}{2} \mathbb{E}_{q(\phi)}[tr(z_i z_i^T)] \\
&= \frac{dk}{2} \ln \frac{\lambda}{2\pi} - \frac{\lambda}{2} tr(W^T W) + \frac{d}{2} \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2} x_i^T x_i + \frac{1}{\sigma^2} x_i^T W \mathbb{E}_{q(\phi)}[z_i] \\
&\quad - \frac{1}{2\sigma^2} tr(\mathbb{E}_{q(\phi)}[z_i z_i^T] W^T W) - \frac{1}{2} tr(\mathbb{E}_{q(\phi)}[z_i z_i^T])
\end{aligned}$$

Maximise $\mathbb{E}_{q(\phi)} \ln p(x_i, W, z_i)$ w.r.t. to W

$$\begin{aligned}
\nabla_W \mathbb{E}_{q(\phi)} \ln p(x_i, W, z_i) &= -\lambda W + \frac{1}{\sigma^2} x_i^T \mathbb{E}_{q(\phi)}[z_i] - \frac{1}{\sigma^2} W \mathbb{E}_{q(\phi)}[z_i z_i^T] \\
0 &= -\lambda W + \frac{1}{\sigma^2} x_i^T \mathbb{E}_{q(\phi)}[z_i] - \frac{1}{\sigma^2} W \mathbb{E}_{q(\phi)}[z_i z_i^T] \\
x_i^T \mathbb{E}_{q(\phi)}[z_i] &= \sigma^2 \lambda W + W \mathbb{E}_{q(\phi)}[z_i z_i^T] \\
x_i^T \mathbb{E}_{q(\phi)}[z_i] &= W(\sigma^2 \lambda + \mathbb{E}_{q(\phi)}[z_i z_i^T])) \\
W &= (x_i^T \mathbb{E}_{q(\phi)}[z_i])(\sigma^2 \lambda + \mathbb{E}_{q(\phi)}[z_i z_i^T])^{-1}
\end{aligned}$$

The above W refers to its value according to the contribution of the i^{th} example. To obtain W, we would have to sum up all n examples.

Since we need $q(\theta)$ to match $\ln p(z_i|x_i, W)$ so we have $q(\theta)$ set as $p(z_i|x_i, W)$. The corresponding distribution is:

$$\begin{aligned}
p(z_i|x_i, W) &\propto p(z_i, x_i, W) \\
&\propto \exp(-\frac{1}{2} z_i^T z_i - \frac{1}{2\sigma^2}(x_i - W z_i)^T (x_i - W z_i) - \frac{\lambda}{2} W^T W) \\
&\propto \exp(-\frac{1}{2\sigma^2}(\sigma^2 z_i^T z_i + x_i^T x_i - 2 x_i^T W z_i + z_i^T W^T W z_i)) \\
&\propto \exp(-\frac{1}{2\sigma^2}(z_i^T \underbrace{(W^T W + \sigma^2 I)}_{M} z_i - 2 z_i^T W^T x_i)) \\
&\propto \exp(-\frac{1}{2\sigma^2}(z_i^T M z_i - 2 z_i^T M M^{-1} W^T x_i + (M^{-1} W^T x_i)^T (M^{-1} W^T x_i))) \\
&\propto \exp(-\frac{1}{2}(z_i - M^{-1} W^T x_i)^T (\sigma^{-2} M)(z_i - M^{-1} W^T x_i)) \\
&= \mathcal{N}(M^{-1} W^T x_i, \sigma^2 M^{-1})
\end{aligned}$$

Hence $\mathbb{E}_{p(z_i|x_i)}[z_i] = M^{-1} W^T x_i$ and $\mathbb{E}_{p(z_i|x_i)}[z_i z_i^T] = \sigma^2 M^{-1} - \mathbb{E}_{p(z_i|x_i)}[z_i] \mathbb{E}_{p(z_i|x_i)}[z_i^T]$.

An EM algorithm for Bayesian PCA

1. Initialise $W \sim N(0, \lambda^{-1})$ and $z_{1:n} \sim N(0, I)$.
2. For iteration $t = 1, ..., T$:
   (a) E-Step: Compute the following where $M = W^T W + \sigma^2 I$
   $\mathbb{E}_{p(z_{1:n}|p(x_{1:n})}[z_{1:n}] = \sum_{i=1}^n M^{-1} W^T x_i$
   $\mathbb{E}_{p(z_{1:n}|p(x_{1:n})}[z_{1:n} z_{1:n}^T] = \sum_{i=1}^n \sigma^2 M^{-1} - \mathbb{E}_{p(z_i|x_i)}[z_i] \mathbb{E}_{p(z_i|x_i)}[z_i^T]$
   (b) M-Step: Update W with the calculated values above with
   $W = \sum_{i=1}^n (x_i^T \mathbb{E}_{q(\phi)}[z_i])(\sigma^2 \lambda + \mathbb{E}_{q(\phi)}[z_i z_i^T])^{-1}$
   (c) Calculate $\ln p(x, W, z)$ using equation
   $\ln p(x, W, z) = \sum_{i=1}^n \frac{dk}{2} \ln \frac{\lambda}{2\pi} - \frac{\lambda}{2} tr(W^T W) + \frac{d}{2} \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x_i - W z_i)^T (x_i - W z_i) + \frac{k}{2} \ln \frac{1}{2\pi} - \frac{1}{2} z_i^T z_i$