Homework 2
Lm2221

Problem 1)

$z_n \sim N(0, I) \rightarrow \mathbb{R}^K$
$W \sim N(0, \Upsilon^{-1}) \rightarrow \mathbb{R}^{d \times K}$
$x_n \sim N(W z_n, \sigma^2 I) \rightarrow \mathbb{R}^d$

First we will solve the general solution. For the E-step, we are looking for the posterior distribution:
$p(z \mid x, w)$, where Υ, σ^2 are fixed & W is our parameter.

$$p(z \mid x, w) = \frac{p(x \mid w, z)\, p(z)}{\int p(x \mid w, z)\, p(z)}$$

$$= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2\sigma^2 (x - Wz)^T (x - Wz)} \frac{1}{\sqrt{2\pi}} e^{-1/2 z^T z}}{\int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2\sigma^2 (x - Wz)^T (x - Wz)} \frac{1}{\sqrt{2\pi}} e^{-1/2 z^T z}\, dz}$$

In this case we know $z \sim N(\mu_z, \Sigma_z)$. Given a resource $x \mid z \sim N(Wz + \mu_x, \Sigma_x)$

$z \mid x \sim N(R(x - W\mu_z - \mu_x), \Sigma_z - RW\Sigma_z^T$

$R = \Sigma_z W^T (W \Sigma_z W^T + \Sigma_x)^{-1}$

$\mu_z = 0, \Sigma_z = I$
$\mu_x = 0, \Sigma_x = \sigma^2 I$

$R = W^T (WW^T + \sigma^2 I)^{-1}$
$\mu_{z \mid x, w} = Rx$
$\Sigma_{z \mid x, w} = I - RW$

$\rightarrow$

Now solve the M-step:

$$\underbrace{\ln p(x, z, w)}_{Q} = \ln \Big[\underbrace{p(x|z,w)}_{\text{solved in last part}} \underbrace{p(z)\, p(w)}_{\text{joint}}\Big]$$

$$\ln [p(x|z,w)] + \underbrace{\ln p(z)}_{\text{ignore does not depend on } w} + \ln p(w)$$

$$= -\frac{1}{2\sigma^2}\Big(\underbrace{x^T x - x^T W z - z^T W^T x + z^T W^T W z}_{(x-Wz)^T(x-Wz)}\Big) - \frac{\lambda}{2} \operatorname{tr}(W^T W)$$

$$\frac{\partial}{\partial W} = -\frac{1}{\sigma^2}(x z^T + W z z^T) - \lambda W$$

plugging back in to full equation

$$\frac{\partial Q}{\partial W} = \int p(z|x, W_{old}) \left[\frac{\partial}{\partial W} \ln p(x, z, w)\right] dz = 0$$

$$= \int p(z|x, W_{old} \left[-\frac{1}{\sigma^2} x z^T + \frac{1}{\sigma^2} W z z^T + \lambda W\right] dz = 0$$

this looks like $\int p(x) f(x) dx = E[f(x)]$

$$= -E(x z^T) + W E(z z^T) + \sigma^2 \lambda W \underbrace{\int p(z|x, W^{old}) dz}_{1}$$

$$W(E(z z^T) + \sigma^2 \lambda I) = E(x z^T)$$

$$W = E(x z^T)(E(z z^T) + \sigma^2 \lambda I)^{-1}$$

$$= x \mu_{z|x,w} \left(\Sigma_{z|x,w} + \mu_{z|x,w} \mu_{z|x,w}^T + \sigma^2 \lambda I\right)^{-1}$$

$\rightarrow$

Now converting it from the general case, (3)
when $\ln p(X, Z, W) = \ln \left[\prod_{n=1}^{N} p(x_n, z_n, W) \right]$

$$\boxed{W = \left[\sum_{i=1}^{n} x_i \mu_{z_i|x_i,W} \right] \left[\sum_{i=1}^{n} \left(\Sigma_{z_i|x_i,W} + \mu_{z_i|x_i W} \mu^T_{z_i x_i, W} \right) + \sigma^2 I \right]^{-1}}$$

Sudo Code Summary:

from section 9.4 in Bishop the EM algorithm →

1. Initialize θ_{old}, where we have $p(X, Z|\theta)$ and goal is to maximize $p(X|\theta)$

2. E-step = Calculate $E[Z]$ where $p(Z|X, \theta^{old})$

3. M-step = Calculate θ_{new} through

$$\theta^{new} = \arg\max \underbrace{Q(\theta, \theta^{old})}$$

$$Q(\theta, \theta_{old}) =$$

$$\sum_{Z} p(Z|X, \theta^{old}) \ln p(X, Z|\theta)$$

4. Do this over T interactions once the update value is smaller than ϵ → defined in problem stop.

$$\boxed{\ln P_T - \ln P_{T-1} < \epsilon}$$

□