Homework 2 BMML E6720 ves

1) $x_1 \ldots x_N$) $x \in \mathbb{R}^d$ $x_n \sim N(Wz_n, \sigma^2 I)$ $W \in \mathbb{R}^{d \times k}$

$z_n \sim N(0, I)$

$$p(W) = (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{1}{2} \operatorname{trace}(W^T W)\right]$$

$$p(X_n | W, z_n) = (2\pi\sigma^2 I)^{-d/2} \exp\left[\frac{-1}{2\sigma^2 I}(x_n - Wz_n)^T (x_n - Wz_n)\right]$$

$$p(Z_N) = (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{1}{2} z_n^T z_N\right]$$

goal -> $W' = \underset{W}{\arg\max} \ln p(x_1 \ldots x_n, W)$

now, from class

$$\ln p(x_1 \ldots x_n, W) \geq \int q(z_n) \ln \frac{p(x_1 \ldots x_n, W, z_n)}{q(z_n)} dz_n$$

$$q'(z) = p(z | W, x) \propto \prod_{i=1}^{n} p(x_n | W, z_n) p(z_n)$$

$$q'(z) \propto \prod_{i=1}^{n} (2\pi\sigma^2 I)^{-\frac{d}{2}} \exp\left[\frac{-1}{2\sigma^2 I}(x_n - Wz_n)^T (x_n - Wz_n)\right] (2\pi)^{-\frac{k}{2}} \exp\left[-\frac{1}{2} z_N^T z_N\right]$$

constants drop out

$$\propto \exp\left[-\frac{1}{2} \sum_{n=1}^{N} \left[\frac{1}{\sigma^2} z_N^T W^T W z_n - \frac{2}{\sigma^2} x_n^T W z_n + z_n^T z_n\right]\right]$$

$$\propto \exp\left[\frac{1}{2\sigma^2} \sum_{n=1}^{N} \left[-2 x_n^T W z_n + z_n^T (\sigma^2 I + W^T W) z_n\right]\right]$$

$$\mu^* = \left(\frac{\sigma^2 I}{\sigma^2 I + W^T W}\right)^{-1} W^T x_{all}$$

$\sigma^2 z_n^T z_n + z_n^T W^T W z_n$

$z_n^T (\sigma^2 I + W^T W) z_n$

$$\Sigma^* = \left(\frac{\sigma^2 I}{\sigma^2 I + W^T W}\right)$$ ... I think that's it.

This is a $N(\mu^*, \Sigma^*)$

expectation of joint likelihood

Y(w)

$$\mathcal{L}(W) = E_q[\ln p(z,x,W)]$$

$$p(z_n, x_1 \dots x_n, W) \propto p(W) \prod_{i=1}^{N} p(x_i | z_i W)\, p(z_i)$$

$$\ln p(z_n, x_n, W) \propto \ln p(W) + \sum_{i=1}^{N} \ln p(x_n | z_n, W) + \text{some constant}$$

$$\propto E_q \ln p(W) + E_q \sum_{i=1}^{N}\left[\ln\left[(2\pi\sigma^2 I)^{-\frac{1}{2}} \exp\left[\frac{1}{2\sigma^2 I}\sum_{i=1}^{N}(x_i - Wz_i)^T(x_i - Wz_i)\right]\right]\right]$$

$$\propto \ln p(W) + E_q\left[\sum_{i=1}^{N}\ln(2\pi\sigma^2 I) + \frac{1}{2\sigma^2 I}\sum_{i=1}^{N}(x_i - Wz_i)^T(x_i - Wz_i)\right]$$

$$\propto \ln p(W) + E_q\left[\frac{1}{2\sigma^2 I}\sum_{i=1}^{N}(x_i - Wz_i)^T(x_i - Wz_i)\right]$$

$$\propto \ln p(W) + \frac{1}{2\sigma^2 I}\sum_{i=1}^{N} E_q\left[x_i^T x_i - 2x_i^T W z_i + W^T z_i^T z_i W\right]$$

$$\propto \ln p(W) + \frac{1}{2\sigma^2 I}\sum^{N}\left[E_q(x_i^T x_i) + E_q[z_n^T W^T W z_n] - 2E_q[x_i^T W z_i]\right]$$

$$\propto \ln p(W) + \frac{1}{2\sigma^2 I}\sum^{N} \underbrace{\cancel{x_i^T x_i}}_{\text{don't care}} - 2x_i^T W \mu^* + E_q[z_n^T W^T W z_n]$$

$$= E[z_i W]^T[z_i W] + \text{trace}(W^T W \,\text{cov}(z_i))$$

$$\propto -\frac{\lambda}{2}\text{trace}(W^T W) + \frac{1}{2\sigma^2}\sum(\mu^* W)^T(\mu^* W) + \text{trace}(W^T W \Sigma^*) - 2x_i^T W \mu^*$$

M Step

$$\nabla \mathcal{L}_t(W) = -\lambda W + \frac{1}{2\sigma^2}\sum\left[W(\mu^{*T}\mu^* + \Sigma^*)\right] - \frac{2}{2\sigma^2}\sum x_i^T \mu^*$$

$$= 0 = -\lambda W + \frac{1}{2\sigma^2}\sum W(\mu^{*T}\mu^* + \Sigma^*) - \frac{1}{\sigma^2}\sum x_i^T \mu^*$$

$$-\lambda W + \frac{W}{2\sigma^2}\cancel{\sum}(\mu^{*T}\mu^* + \Sigma^*) = \frac{1}{\sigma^2}\sum x_i^T \mu^*$$

$$\boxed{W_t = \frac{\sum_{i=1}^{N} x_i^T \mu^*}{-\lambda\sigma^2 + \frac{1}{2}(\mu^{*2} + \Sigma^*)}}$$

Scanned by CamScanner

EM algorithm

1) Initialize W with dimensions

2) for iteration t,

a) calculate $E_{q_t}[z]$, which is $\frac{W^T x_n}{(\sigma^2 I + W^T W)}$ E step

b) calculate $W_t = \frac{\sum_{i=1}^{N} x_i \mu^*}{-\lambda\sigma^2 + \frac{1}{2}(\mu^{*2} + \Sigma^*)}$

3) once you have this,

calculate $\ln p_t(x_1 \ldots x_N, W)$ from $\ln p(x_1 \ldots x_n, W, z_n)$ until converges → so difference is sufficiently small

Scanned by CamScanner