Problem 2

Elias Per...
1016828
...

$$p(\tau, z, W / X) = \prod_{d=1}^{D} q(w_d) \prod_{n=1}^{N} q(z_n) \prod_{d=1}^{D} q(\tau_d)$$

$$p(\tau, z, W, X) = p(X | \tau, z, W) \cdot \underbrace{p(\tau, z, W)}_{= p(\tau) \cdot p(z) \cdot p(W)}$$

$$= p(X | \tau, z, W) \cdot p(\tau) \cdot p(z) \cdot p(W)$$

$$\Rightarrow \log(p(\tau, z, W, X)) = \log(p(X | \tau, z, W)) + \log(p(\tau)) + \log(p(z)) + \log(p(W))$$

$$\log p(X | \tau, z, W) = \sum_{n=1}^{N} \log \mathcal{N}_D(x_n | W z_n, \mathrm{diag}(\tau)^{-1})$$

$$\log \mathcal{N}_D(x_n | W z_n, \mathrm{diag}(\tau)^{-1}) = -\frac{D}{2} \log 2\pi + \frac{1}{2} \log(\det(\mathrm{diag}(\tau))) - \frac{1}{2} (x_n - W z_n)^T \mathrm{diag}(\tau)(x_n - W z_n)$$

$$\Rightarrow \log p(X | \tau, z, W) = -\frac{ND}{2} \log 2\pi + \frac{N}{2} \log(\det(\mathrm{diag}(\tau))) - \frac{1}{2} \sum_{n=1}^{N} (x_n - W z_n)^T \mathrm{diag}(\tau)(x_n - W z_n)$$

Elias Pe[illegible]
1016827
[illegible]

$$\log p(\tau) = \sum_{d=1}^{D} \log \mathrm{Gamma}(\tau_d \mid a, b)$$

$$= \sum_{d=1}^{D} a \log b - \log \Gamma(a) + (a-1) \log \tau_d - b\tau_d$$

$$= D\left(a \log b - \log \Gamma(a)\right) + \sum_{d=1}^{D} (a-1) \log \tau_d - b\tau_d$$

$$\log p(z) = \sum_{n=1}^{N} \log \mathcal{N}_K(z_n \mid 0, I)$$

$$= -\frac{NK}{2} \log 2\pi + \underbrace{\frac{N}{2} \log(\det(I))}_{= \frac{N}{2} \cdot \log(1) = 0}$$

$$- \frac{1}{2} \sum_{n=1}^{N} (z_n - 0)^T \cdot I \, (z_n - 0)$$

$$= -\frac{NK}{2} \log 2\pi - \frac{1}{2} \sum_{n=1}^{N} z_n^T I z_n$$

$$\log p(W) = \sum_{d=1}^{D} \log \mathcal{N}_K(w_d \mid 0, \alpha I) \qquad I \in \mathbb{R}^{K \times K}$$

$$= -\frac{DK}{2} \log(2\pi) + \frac{D}{2} \log(\underbrace{\det(\alpha I)}_{= \alpha^K})$$

$$- \frac{1}{2} \sum_{d=1}^{D} w_d^T (\alpha I)^{-1} w_d$$

$$= -\frac{DK}{2} \log 2\pi + \frac{DK}{2} \log \alpha$$

$$- \frac{1}{2\alpha} \sum_{d=1}^{D} w_d^T w_d$$

Elias P.
10.16829
cis Ros

$$\Rightarrow \log p(\tau, Z, W, X)$$

$$= -\frac{ND}{2}\log 2\pi + \frac{N}{2}\log\det(\mathrm{diag}(\tau)) - \frac{1}{2}\sum_{n=1}^{N}(x_n - Wz_n)^T \mathrm{diag}(\tau)(x_n - Wz_n)$$

$$+ D(a\log b - \log\Gamma(a)) + \sum_{d=1}^{D}(a-1)\log\tau_d - b\tau_d$$

$$-\frac{NK}{2}\log 2\pi - \frac{1}{2}\sum_{n=1}^{N} z_n^T z_n$$

$$-\frac{NK}{2}\log 2\pi + \frac{NK}{2}\log\alpha - \frac{1}{2\alpha}\sum_{d=1}^{D} w_d^T w_d$$

<u>part 2</u>

only those terms which depend on z denoted by:

$$\log q(z) = -\frac{1}{2}\sum_{n=1}^{N}(x_n - Wz_n)^T \mathrm{diag}(\tau)(x_n - Wz_n)$$

$$-\frac{1}{2}\sum_{n=1}^{N} z_n^T z_n$$

$$= \sum_{n=1}^{N} -\frac{1}{2}(x_n - Wz_n)^T \mathrm{diag}(\tau)(x_n - Wz_n) - \frac{1}{2}z_n^T z_n$$

lets look at the summands:

$$-\frac{1}{2}(x_n - Wz_n)^T \mathrm{diag}(\tau)(x_n - Wz_n) - \frac{1}{2}z_n^T z_n$$

$$= \underbrace{-\frac{1}{2}x_n^T \mathrm{diag}(\tau)x_n}_{\text{discard since does not depend on } z} + x_n^T \mathrm{diag}(\tau)Wz_n - \frac{1}{2}z_n^T W^T \mathrm{diag}(\tau)Wz_n$$

$$-\frac{1}{2}z_n^T z_n$$

Elias Pel
1016828
[illegible]

$$= -\frac{1}{2} z_n^T \underbrace{(W^T \mathrm{diag}(\tau) W + I)}_{A} z_n + \underbrace{x_n^T \mathrm{diag}(\tau) W}_{= b^T} z_n$$

$$= -\frac{1}{2} z_n^T A z_n + b^T z_n$$

apply the completing square technique since indeed $A^T = A$

$$\Rightarrow -\frac{1}{2} z_n^T A z_n + b^T z_n = -\frac{1}{2}(z_n - A^{-1}b)^T A (z_n - A^{-1}b) \underbrace{- \frac{1}{2} b^T A^{-1} b}$$

does not depend on $z$
$\Rightarrow$ discard

$$\Rightarrow \log q(z) = \sum_{n=1}^{N} -\frac{1}{2}(z_n - A^{-1}b)^T A (z_n - A^{-1}b)$$

where $A = W^T \mathrm{diag}(\tau) W + I$, $b = W^T \mathrm{diag}(\tau) x_n$

now lets look at

$$\log q(z_n) = -\frac{1}{2}(z_n - A^{-1}b)^T A (z_n - A^{-1}b)$$

we notice that this is the form of a normal distribution. The covariance matrix is $A^{-1}$ and the mean is $A^{-1}b$.

$$\Rightarrow q(z_n) \sim \mathcal{N}_K(A^{-1}b, A^{-1})$$

but since now we need to replace $\psi$ and $W$ with their expectations since we assumed that they were given

(sorry for the shitty notation)

$$\Rightarrow \quad A = \mathbb{E}_{w,\psi}\left[W^T \operatorname{diag}(\psi) W + I\right], \quad b = \mathbb{E}_{w,\psi}\left[W^T \operatorname{diag}(\psi) x_n\right]$$

$$A = \mathbb{E}_{w,\psi}\left[W^T \operatorname{diag}(\psi) W\right] + I$$

using the hint

$$= \mathbb{E}_{w,\psi}\left[\sum_{d=1}^{D} \psi_d w_d w_d^T\right] + I \qquad \text{linearity}$$

$$= \sum_{d=1}^{D} \mathbb{E}_{w,\psi}\left[\psi_d w_d w_d^T\right] + I \qquad \text{independence assumption between } \psi \text{ and } W$$

$$= I + \sum_{d=1}^{D} \underbrace{\mathbb{E}_{\psi}[\psi_d]}_{\langle \psi_d \rangle} \underbrace{\mathbb{E}_{w}[w_d w_d^T]}_{\langle w_d w_d^T \rangle}$$

$$= I + \sum_{d=1}^{D} \langle \psi_d \rangle \cdot \langle w_d w_d^T \rangle$$

$$\Rightarrow \quad A^{-1} = \left(I + \sum_{d=1}^{D} \langle \psi_d \rangle \cdot \langle w_d w_d^T \rangle\right)^{-1} \qquad \text{covariance matrix}$$

Elias Pelo
1016828
CCS P...

now the mean similarly:

independent "constant"

$$b = \mathbb{E}_{W,\tau}\left[W^T \operatorname{diag}(\tau)\, x_n\right] = \mathbb{E}_{W,\tau}\left[W^T \operatorname{diag}(\tau)\right] x_n$$

$$= \underbrace{\mathbb{E}_W\left[W^T\right]}_{\langle W^T \rangle} \cdot \underbrace{\mathbb{E}_{\tau}\left[\operatorname{diag}(\tau)\right]}_{\operatorname{diag}(\langle \tau \rangle)} x_n$$

$$= \langle W^T \rangle \operatorname{diag}(\langle \tau \rangle)\, x_n \Rightarrow A^{-1} b = A^{-1} \cdot \langle W^T \rangle \operatorname{diag}(\langle \tau \rangle)\, x_n$$

via independence assumption

thus we have shown that the update factor is equal to:

$$q(z_n) = \mathcal{N}_K(\mu_n, K_n)$$

where $K_n = A^{-1} = \left(I + \sum_{d=1}^{D} \langle \tau_d \rangle \langle w_d w_d^T \rangle\right)^{-1}$

$$\mu_n = A^{-1} b = K_n \langle W^T \rangle \operatorname{diag}(\langle \tau \rangle)\, x_n$$

$\square$

Elias Pelo
1016828
CIS [illegible]