Advanced probabilistic methods - Sketch

Christian Segercrantz 481056

March 29, 2022

$$p(\psi, \mathbf{Z}, \mathbf{W}, \mathbf{X}) = p(\mathbf{X} \mid \psi, \mathbf{Z}, \mathbf{W}) p(\psi) p(\mathbf{Z}) p(\mathbf{W}) \tag{1}$$

$$= \mathcal{N}_D(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n, \mathrm{diag}(\psi)^{-1}) \prod_{d=1}^{D} q^*(\mathbf{w}_d) \prod_{n=1}^{N} q^*(\mathbf{z}_n) \prod_{d=1}^{D} q^*(\psi_d) \tag{2}$$

$$= \mathcal{N}_D(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n, \mathrm{diag}(\psi)^{-1}) \mathcal{N}_K(\mathbf{w}_d \mid \mathbf{0}, \alpha\mathbf{I}) \mathcal{N}_K(\mathbf{z}_n \mid \mathbf{0}, \mathbf{I}) \mathrm{Gamma}(\psi_d \mid a, b) \tag{3}$$

$$\mathbb{E}_{\psi,w}\left[\log q^*(\psi, \mathbf{Z}, \mathbf{W}, \mathbf{X})\right] = \mathbb{E}_{\psi,w}[\log\left(\mathcal{N}_D(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n, \mathrm{diag}(\psi)^{-1})\right) + \log\left(\mathcal{N}_K(\mathbf{w}_d \mid \mathbf{0}, \alpha\mathbf{I})\right) \tag{4}$$

$$+ \log\left(\mathcal{N}_K(\mathbf{z}_n \mid \mathbf{0}, \mathbf{I})\right) + \log\left(\mathrm{Gamma}(\psi_d \mid a, b)\right)] \tag{5}$$

$$\log\left(\mathcal{N}_D(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n, \mathrm{diag}(\psi)^{-1})\right) \propto -\frac{1}{2}(\mathbf{x_n} - \mathbf{W}\mathbf{z}_n)^\top \mathrm{diag}(\psi)(\mathbf{x_n} - \mathbf{W}\mathbf{z}_n) \tag{6}$$

$$\log\left(\mathcal{N}_K(\mathbf{z}_n \mid \mathbf{0}, \mathbf{I})\right) \propto -\frac{1}{2}\mathbf{z_n}^\top \mathbf{z_n} \tag{7}$$

$$\mathbb{E}_{\psi,w}\left[\log q^*(\psi, \mathbf{Z}, \mathbf{W}, \mathbf{X})\right] \propto \mathbb{E}_{\psi,w}\left[-\frac{1}{2}(\mathbf{x_n} - \mathbf{W}\mathbf{z}_n)^\top \mathrm{diag}(\psi)(\mathbf{x_n} - \mathbf{W}\mathbf{z}_n) - \frac{1}{2}\mathbf{z_n}^\top \mathbf{z_n}\right] \tag{8}$$

$$= \mathbb{E}_{\psi,w}[-\frac{1}{2}(\mathbf{z_n}^\top \mathbf{z_n} + \mathbf{x_n}^\top \mathrm{diag}(\psi)\mathbf{x_n} - \mathbf{x_n}^\top \mathrm{diag}(\psi)\mathbf{W}\mathbf{z}_n \tag{9}$$

$$- (\mathbf{W}\mathbf{z}_n)^\top \mathrm{diag}(\psi)\mathbf{x_n} + (\mathbf{W}\mathbf{z}_n)^\top \mathrm{diag}(\psi)\mathbf{W}\mathbf{z}_n)] \tag{10}$$

$$\propto \mathbb{E}_{\psi,w}\left[-\frac{1}{2}(\mathbf{z_n}^\top \mathbf{z_n} - 2\mathbf{x_n}^\top \mathrm{diag}(\psi)\mathbf{W}\mathbf{z}_n + \mathbf{z}_n^\top \mathbf{W}^\top \mathrm{diag}(\psi)\mathbf{W}\mathbf{z}_n)\right] \tag{11}$$

$$= \mathbb{E}_{\psi,w}\left[-\frac{1}{2}\mathbf{z_n}^\top (\underbrace{\mathbf{I} + \mathbf{W}^\top \mathrm{diag}(\psi)\mathbf{W}}_{A})\mathbf{z_n} + \underbrace{\mathbf{x_n}^\top \mathrm{diag}(\psi)\mathbf{W}}_{b^\top}\mathbf{z}_n\right], \quad |\text{Completing the square} \tag{12}$$

$$= \mathbb{E}_{\psi,w}\left[\frac{1}{2}(\mathbf{z_n} - A^{-1}b)^\top A(\mathbf{z_n} - A^{-1}b) - \frac{1}{2}b^\top A^{-1}b\right] \tag{13}$$

$$= \frac{1}{2}(\mathbf{z_n} - \mathbb{E}_{\psi,w}[A^{-1}b])^\top \mathbb{E}_{\psi,w}[A](\mathbf{z_n} - \mathbb{E}_{\psi,w}[A^{-1}b]) - \frac{1}{2}\mathbb{E}_{\psi,w}[b^\top A^{-1}b] \tag{14}$$

$$\propto \frac{1}{2}(\mathbf{z_n} - \mathbb{E}_{\psi,w}[A^{-1}b])^\top \mathbb{E}_{\psi,w}[A](\mathbf{z_n} - \mathbb{E}_{\psi,w}[A^{-1}b]) \tag{15}$$

$$A^{-1} = \left(\mathbf{I} + \underbrace{\mathbf{W}^\top}_{\mathbf{K}\times\mathbf{D}} \underbrace{\mathrm{diag}(\psi)}_{\mathbf{D}\times\mathbf{D}} \underbrace{\mathbf{W}}_{\mathbf{D}\times\mathbf{K}}\right)^{-1} = \left(\mathbf{I} + \sum_{\mathbf{d=1}}^{\mathbf{D}} \psi_{\mathbf{d}} \mathbf{w_d} \mathbf{w_d}^\top\right)^{-1} \tag{16}$$

$$b = \mathbf{W}^\top \mathrm{diag}(\psi)\mathbf{x}_n \tag{17}$$

$$\mathbb{E}_{\psi,\mathbf{w}}[\mu_n] = \mathbb{E}_{\psi,\mathbf{w}}[A^{-1}b] = \mathbb{E}_{\psi,\mathbf{w}}[A^{-1}]\mathbb{E}_{\psi,\mathbf{w}}[b] \tag{18}$$

$$\mathbb{E}_{\psi,\mathbf{w}}[A^{-1}] = \mathbb{E}_{\psi,\mathbf{w}}\left[\mathbf{I} + \sum_{\mathbf{d=1}}^{\mathbf{D}} \psi_{\mathbf{d}} \mathbf{w_d} \mathbf{w_d}^\top\right]^{-1} \tag{19}$$

$$= \left(\mathbf{I} + \sum_{\mathbf{d=1}}^{\mathbf{D}} \mathbb{E}_\psi[\psi_{\mathbf{d}}]\mathbb{E}_{\mathbf{w}}[\mathbf{w_d}\mathbf{w_d}^\top]\right)^{-1} = \left(\mathbf{I} + \sum_{\mathbf{d=1}}^{\mathbf{D}} \langle\psi_{\mathbf{d}}\rangle \langle\mathbf{w_d}\mathbf{w_d}^\top\rangle\right)^{-1} \tag{20}$$

$$\mathbb{E}_{\psi,\mathbf{w}}[b] = \mathbb{E}_{\psi,\mathbf{w}}\left[\mathbf{W}^\top \mathrm{diag}(\psi)\mathbf{x}\right] \tag{21}$$

$$= \mathbb{E}_{\mathbf{w}}[\mathbf{W}^\top]\mathbb{E}_\psi[\mathrm{diag}(\psi)]\mathbf{x} \tag{22}$$

$$= \langle\mathbf{W}^\top\rangle\, \mathrm{diag}\,\langle\psi\rangle\, \mathbf{x} \tag{23}$$

$$\mathbb{E}_{\psi,w}\left[\log q^*(\psi, \mathbf{Z}, \mathbf{W}, \mathbf{X})\right] \propto \frac{1}{2}(\mathbf{z_n} - \underbrace{\mathbb{E}_{\psi,w}[A^{-1}]\mathbb{E}_{\psi,w}[b]}_{\mu_n})^\top \underbrace{\mathbb{E}_{\psi,w}[A^{-1}]^{-1}}_{\mathbf{K}_n}(\mathbf{z_n} - \underbrace{\mathbb{E}_{\psi,w}[A^{-1}]\mathbb{E}_{\psi,w}[b]}_{\mu_n}) \tag{24}$$

$$\mathbb{E}_{\psi,w}\left[\log q^*(\psi, \mathbf{Z}, \mathbf{W}, \mathbf{X})\right] \propto \mathcal{N}(\mu_n, \mathbf{K}_n) \tag{25}$$

$$
\begin{aligned}
\mathbb{E}_{\psi,\mathbf{z}}\left[\log q^*(\psi,\mathbf{Z},\mathbf{W},\mathbf{X})\right] =& \mathbb{E}_{\psi,\mathbf{z}}[p(\mathbf{W})p(\mathbf{X} \mid \mathbf{Z},\mathbf{W},\mathbf{\Psi})p(\mathbf{Z})p(\mathbf{\Psi})] && (26)\\
\propto& \mathbb{E}_{\psi,\mathbf{z}}[p(\mathbf{W})p(\mathbf{X} \mid \mathbf{Z},\mathbf{W},\mathbf{\Psi})] && (27)\\
\propto& \mathbb{E}_{\psi,\mathbf{z}}[\sum_{n=1}^{N}\log\left(\mathcal{N}(\mathbf{x}_{nd} \mid \mathbf{w}_d^\top\mathbf{z}_n,\psi_d^{-1})\right) + \log\left(\mathcal{N}_K(\mathbf{w}_d \mid \mathbf{0},\alpha\mathbf{I})\right)] && (28)\\
\log\left(\mathcal{N}(\mathbf{x}_{nd} \mid \mathbf{w}_d^\top\mathbf{z}_n,\psi_d^{-1})\right) \propto& -\frac{1}{2}(\mathbf{x}_{nd}-\mathbf{w}_d^\top\mathbf{z}_n)^\top\psi_{\mathbf{d}}(\mathbf{x}_{nd}-\mathbf{w}_d^\top\mathbf{z}_n) && (29)\\
\log\left(\mathcal{N}_K(\mathbf{z}_n \mid \mathbf{0},\mathbf{I})\right) \propto& -\frac{1}{2}\mathbf{w_d}^\top(\alpha^{-1}\mathbf{I})\mathbf{w_d} && (30)\\
\mathbb{E}_{\psi,\mathbf{z}}\left[\log q^*(\psi,\mathbf{Z},\mathbf{W},\mathbf{X})\right] \propto& \mathbb{E}_{\psi,\mathbf{z}}\left[\sum_{n=1}^{N}\left(-\frac{1}{2}(\mathbf{x}_{nd}-\mathbf{w}_d^\top\mathbf{z}_n)^\top\psi_{\mathbf{d}}(\mathbf{x}_{nd}-\mathbf{w}_d^\top\mathbf{z}_n)\right)-\frac{1}{2}\mathbf{w_d}^\top(\alpha^{-1}\mathbf{I})\mathbf{w_d}\right] && (31)\\
\propto& \mathbb{E}_{\psi,\mathbf{z}}\left[\sum_{n=1}^{N}\left(-\frac{\psi_d}{2}(\mathbf{x}_{nd}^2-2\mathbf{x}_{nd}\mathbf{w}_d^\top\mathbf{z}_n+\mathbf{w}_d^\top\mathbf{z}_n\mathbf{z}_n^\top\mathbf{w}_d)\right)-\frac{1}{2}\mathbf{w_d}^\top(\alpha^{-1}\mathbf{I})\mathbf{w_d}\right] && (32)\\
\propto& \mathbb{E}_{\psi,\mathbf{z}}\left[\sum_{n=1}^{N}\left(\psi_d\mathbf{x}_{nd}\mathbf{w}_d^\top\mathbf{z}_n-\frac{\psi_d}{2}\mathbf{w}_d^\top\mathbf{z}_n\mathbf{z}_n^\top\mathbf{w}_d\right)-\frac{1}{2}\mathbf{w_d}^\top(\alpha^{-1}\mathbf{I})\mathbf{w_d}\right] && (33)\\
\propto& \mathbb{E}_{\psi,\mathbf{z}}\left[\psi_d\mathbf{w}_d^\top\sum_{n=1}^{N}\mathbf{x}_{nd}\mathbf{z}_n-\frac{\psi_d}{2}\mathbf{w}_d^\top\sum_{n=1}^{N}\left(\mathbf{z}_n\mathbf{z}_n^\top\right)\mathbf{w}_d-\frac{1}{2}\mathbf{w_d}^\top(\alpha^{-1}\mathbf{I})\mathbf{w_d}\right] && (34)\\
\propto& \mathbb{E}_{\psi,\mathbf{z}}\left[\psi_d\mathbf{w}_d^\top\sum_{n=1}^{N}\mathbf{x}_{nd}\mathbf{z}_n-\frac{1}{2}\mathbf{w}_d^\top\left(\psi_d\sum_{n=1}^{N}\left(\mathbf{z}_n\mathbf{z}_n^\top\right)+\alpha^{-1}\mathbf{I}\right)\mathbf{w_d}\right] && (35)\\
\propto& \mathbb{E}_{\psi}[\psi_d]\sum_{n=1}^{N}(\mathbf{x}_{nd}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n^\top])\mathbf{w}_d-\frac{1}{2}\mathbf{w}_d^\top\left(\mathbb{E}_{\psi}[\psi_d]\sum_{n=1}^{N}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n\mathbf{z}_n^\top]+\alpha^{-1}\mathbf{I}\right)\mathbf{w_d} && (36)\\
\propto& \underbrace{\langle\psi_d\rangle\sum_{n=1}^{N}(\mathbf{x}_{nd}\left\langle\mathbf{z}_n^\top\right\rangle)}_{\mathbf{b}^\top}\mathbf{w}_d-\frac{1}{2}\mathbf{w}_d^\top\underbrace{\left(\langle\psi_d\rangle\sum_{n=1}^{N}\left\langle\mathbf{z}_n\mathbf{z}_n^\top\right\rangle+\alpha^{-1}\mathbf{I}\right)}_{\mathbf{A}}\mathbf{w_d} && (37)\\
& \text{Complete the square} && (38)\\
\propto& -\frac{1}{2}\left(\mathbf{w_d}-\mathbf{A}^{-1}\mathbf{b}\right)^\top\mathbf{A}\left(\mathbf{w_d}-\mathbf{A}^{-1}\mathbf{b}\right) && (39)\\
\text{So we can conclude that} & && (40)\\
\propto& \mathcal{N}_K(\mathbf{w_d} \mid \mathbf{A}^{-1}\mathbf{b},\mathbf{A}^{-1}) && (41)\\
& && (42)
\end{aligned}
$$