Advanced probabilistic methods - Sketch

Christian Segercrantz 481056

March 24, 2022

$$
\begin{aligned}
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}) && (1)\\
=& \mathcal{N}_D(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n, \operatorname{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) && (2)\\
=& \mathcal{N}_D(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n, \operatorname{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}) \operatorname{Gamma}(\psi_d \mid a, b) && (3)\\
\mathbb{E}_{\psi,w}\left[\log p(\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, \operatorname{diag}(\psi)^{-1})\right) + \log\left(\mathcal{N}_K(\mathbf{w}_d \mid \mathbf{0}, \alpha\mathbf{I})\right) && (4)\\
&+ \log\left(\mathcal{N}_K(\mathbf{z}_n \mid \mathbf{0}, \mathbf{I})\right) + \log\left(\operatorname{Gamma}(\psi_d \mid a, b)\right)] && (5)\\
\log\left(\mathcal{N}_D(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n, \operatorname{diag}(\psi)^{-1})\right) \propto& -\frac{1}{2}(\mathbf{x_n} - \mathbf{W}\mathbf{z}_n)^\top \operatorname{diag}(\psi)(\mathbf{x_n} - \mathbf{W}\mathbf{z}_n) && (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} && (7)\\
\mathbb{E}_{\psi,w}\left[\log p(\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 \operatorname{diag}(\psi)(\mathbf{x_n} - \mathbf{W}\mathbf{z}_n) - \frac{1}{2}\mathbf{z_n}^\top \mathbf{z_n}\right] && (8)\\
=& \mathbb{E}_{\psi,w}[-\frac{1}{2}(\mathbf{z_n}^\top \mathbf{z_n} + \mathbf{x_n}^\top \operatorname{diag}(\psi)\mathbf{x_n} - \mathbf{x_n}^\top \operatorname{diag}(\psi)\mathbf{W}\mathbf{z}_n && (9)\\
&- (\mathbf{W}\mathbf{z}_n)^\top \operatorname{diag}(\psi)\mathbf{x_n} + (\mathbf{W}\mathbf{z}_n)^\top \operatorname{diag}(\psi)\mathbf{W}\mathbf{z}_n)] && (10)\\
\propto& \mathbb{E}_{\psi,w}\left[-\frac{1}{2}(\mathbf{z_n}^\top \mathbf{z_n} - 2\mathbf{x_n}^\top \operatorname{diag}(\psi)\mathbf{W}\mathbf{z}_n + \mathbf{z}_n^\top \mathbf{W}^\top \operatorname{diag}(\psi)\mathbf{W}\mathbf{z}_n)\right] && (11)\\
=& \mathbb{E}_{\psi,w}\left[-\frac{1}{2}\mathbf{z_n}^\top(\underbrace{\mathbf{I} + \mathbf{W}^\top \operatorname{diag}(\psi)\mathbf{W}}_{A})\mathbf{z_n} + \underbrace{\mathbf{x_n}^\top \operatorname{diag}(\psi)\mathbf{W}}_{b^\top}\mathbf{z}_n\right], \quad |\text{Completing the square} && (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] && (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] && (14)\\
\mathbf{K}_n =& A^{-1}, \qquad A^{-1} = \left(\mathbf{I} + \underbrace{\mathbf{W}^\top}_{\mathbf{K\times D}} \underbrace{\operatorname{diag}(\psi)}_{\mathbf{D\times D}} \underbrace{\mathbf{W}}_{\mathbf{D\times K}}\right)^{-1} = \left(\mathbf{I} + \sum_{\mathbf{d=1}}^{\mathbf{D}} \psi_{\mathbf{d}} \mathbf{w_d} \mathbf{w_d}^\top\right)^{-1} && (15)\\
\mu_n =& A^{-1}b, \qquad b = \mathbf{W}^\top \operatorname{diag}(\psi)\mathbf{x}_n && (16)
\end{aligned}
$$