Advanced probabilistic methods - Sketch

Christian Segercrantz 481056

March 30, 2022

$$
\begin{aligned}
-\log p(\mathcal{D} \mid \mathbf{W}) &= -\log \prod_{i=0}^{1} \mathcal{N}(y_i \mid w_0 + w_1 x_i, \sigma_l^2) && (1) \\
&= \sum_{i=0}^{N} -\log \mathcal{N}(y_i \mid w_0 + w_1 x_i, \sigma_l^2) && (2) \\
-\log \mathcal{N}(y_i \mid w_0 + w_1 x_i, \sigma_l^2) &= -\log \frac{1}{\sigma_l \sqrt{2\pi}} + \frac{1}{2\sigma_l^2}(y_i - w_0 - w_1 x_i)^2 && (3) \\
&= \log \sigma_l \sqrt{2\pi} + \frac{1}{2\sigma_l^2}(y_i - w_0 - w_1 x_i)^2 && (4) \\
-\log p(\mathcal{D} \mid \mathbf{W}) &= \sum_{i=0}^{N} \log \sigma_l \sqrt{2\pi} + \frac{1}{2\sigma_l^2} \sum_{i=0}^{N} (y_i - w_0 - w_1 x_i)^2 && (5) \\
&= N \log \sigma_l \sqrt{2\pi} + \frac{1}{2\sigma_l^2} \sum_{i=0}^{N} (y_i - w_0 - w_1 x_i)^2, \qquad \log \sigma_l \sqrt{2\pi} \approx 0.001 && (6) \\
&\approx \frac{1}{2\sigma_l^2} \sum_{i=0}^{N} (y_i - w_0 - w_1 x_i)^2 && (7)
\end{aligned}
$$

Since the prior $p(\mathbf{w})$ is a MVN with diagonal covariance and that the mean-field approximation of the posterior $q(\mathbf{w})$ is product of gaussians we can write the KL-divergence as a sum of the separate parts

$$\mathrm{KL}\left[q(\mathbf{w})|p(\mathbf{w})\right]) = \mathrm{KL}_0\left[q(w_0)|p(w_0)\right]) + \mathrm{KL}_1\left[q(w_1)|p(w_1)\right]) \tag{8}$$

$$\mathrm{KL}_i[q(w_i) \mid p(w_i)] = \int -q(w_i)\log\frac{p(w_i)}{q(w_i)}dw_i \tag{9}$$

$$= \underbrace{\int q(w_i)\log q(w_i)dw_i}_{\text{entropy}=\frac{1}{2}\log(2\pi\sigma_i^2)+\frac{1}{2}} - \int q(w_i)\log p(w_i)dw_i \tag{10}$$

$$q(w_i) = \mathcal{N}(w_i|\mu_i,\sigma_i^2) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp(-\frac{1}{2\sigma_i^2}(w_i-\mu_i)^2), \qquad \sigma_i \geq 0 \tag{11}$$

$$p(w_i) = \mathcal{N}(w_i|0,\alpha) = \frac{1}{\alpha\sqrt{2\pi}}\exp(-\frac{1}{2\alpha^2}w_i^2), \qquad \alpha \geq 0 \tag{12}$$

$$\log p(w_i) = -\log(\alpha\sqrt{2\pi}) - \frac{1}{2\alpha^2}w_i^2 \tag{13}$$

$$\mathrm{KL}_i[q(w_i) \mid p(w_i)] = \frac{1}{2}\log(2\pi\sigma_i^2) + \frac{1}{2} - \int q(w_i)\log p(w_i)dw_i \tag{14}$$

$$\mathrm{KL}_i[q(w_i) \mid p(w_i)] = \frac{1}{2}\log(2\pi\sigma_i^2) + \frac{1}{2} - \int \mathcal{N}(w_i|\mu_i,\sigma_i^2)\left(-\log(\alpha\sqrt{2\pi}) - \frac{1}{2\alpha^2}w_i^2\right)dw_i \tag{15}$$

$$= \frac{1}{2}\log(2\pi\sigma_i^2) + \frac{1}{2} + \log(\alpha\sqrt{2\pi})\underbrace{\int \mathcal{N}(w_i|\mu_i,\sigma_i^2)dw_i}_{=1} + \frac{1}{2\alpha^2}\int \mathcal{N}(w_i|\mu_i,\sigma_i^2)w_i^2dw_i \tag{16}$$

$$= \frac{1}{2}\log(2\pi\sigma_i^2) + \frac{1}{2} + \log(\alpha\sqrt{2\pi}) + \frac{1}{2\alpha^2}\underbrace{\int \frac{1}{\sigma_i\sqrt{2\pi}}\exp(-\frac{1}{2\sigma_i^2}(w_i-\mu_i)^2)w_i^2dw_i}_{\mu^2+\sigma_i^2} \tag{17}$$

$$= \frac{1}{2}\log(2\pi\sigma_i^2) + \frac{1}{2} + \log(\alpha\sqrt{2\pi}) + \frac{\mu^2+\sigma_i^2}{2\alpha^2} \tag{18}$$

$$\tag{19}$$