Appendix C

Spike and slab prior KL

We can evaluate the KL divergence between the approximating distribution of section §6.6.5 and the spike and slab prior analytically:

$$\mathrm{KL}(q(\boldsymbol{\omega})||p(\boldsymbol{\omega})) = \sum_{ik} \mathrm{KL}(q(\mathbf{w}_{ik})||p(\mathbf{w}_{ik}))$$

with

$$\begin{aligned}
\mathrm{KL}(q(\mathbf{w}_{ik})||p(\mathbf{w}_{ik})) &= \int q(\mathbf{w}_{ik}) \log \frac{q(\mathbf{w}_{ik})}{p(\mathbf{w}_{ik})} \mathrm{d}\mathbf{w}_{ik} \\
&= \int q(\mathbf{w}_{ik}) \log \frac{\mathcal{N}(\mathbf{w}_{ik}; \mathbf{m}_{ik}, \sigma_{ik}^2 I) p(\mathbf{w}_{ik})/Z_q}{p(\mathbf{w}_{ik})} \mathrm{d}\mathbf{w}_{ik} \\
&= \int q(\mathbf{w}_{ik}) \log \mathcal{N}(\mathbf{w}_{ik}; \mathbf{m}_{ik}, \sigma_{ik}^2 I) \mathrm{d}\mathbf{w}_{ik} - \log Z_q \\
&= \int \left(\frac{\alpha}{\alpha+1} \delta_{\mathbf{0}} + \frac{1}{\alpha+1} \mathcal{N}\left(\mathbf{w}_{ik}; \frac{\mathbf{m}_{ik}}{1 + l^2 \sigma_{ik}^2}, \frac{\sigma_{ik}^2}{1 + l^2 \sigma_{ik}^2} I \right) \right) \\
&\qquad \cdot \log \mathcal{N}(\mathbf{w}_{ik}; \mathbf{m}_{ik}, \sigma_{ik}^2 I) \mathrm{d}\mathbf{w}_{ik} - \log Z_q \\
&= \frac{\alpha}{\alpha+1} \log \mathcal{N}(\mathbf{0}; \mathbf{m}_{ik}, \sigma_{ik}^2 I) + \frac{1}{\alpha+1} \int \mathcal{N}\left(\mathbf{w}_{ik}; \frac{\mathbf{m}_{ik}}{1 + l^2 \sigma_{ik}^2}, \frac{\sigma_{ik}^2}{1 + l^2 \sigma_{ik}^2} I \right) \\
&\qquad \cdot \log \mathcal{N}(\mathbf{w}_{ik}; \mathbf{m}_{ik}, \sigma_{ik}^2 I) \mathrm{d}\mathbf{w}_{ik} - \log Z_q
\end{aligned}$$

Note that the KL is properly defined since for every measurable set $q(\cdot)$ has mass on, $p(\cdot)$ has mass on as well (including the singleton set $\{0\}$!).

We evaluate the last integral as follows:

$$\int \mathcal{N}\left(\mathbf{w}_{ik}; \frac{\mathbf{m}_{ik}}{1 + l^2 \sigma_{ik}^2}, \frac{\sigma_{ik}^2}{1 + l^2 \sigma_{ik}^2} I \right) \log \mathcal{N}(\mathbf{w}_{ik}; \mathbf{m}_{ik}, \sigma_{ik}^2 I) \mathrm{d}\mathbf{w}_{ik}$$

$$= \int \mathcal{N}\left(\mathbf{w}_{ik}; \frac{\mathbf{m}_{ik}}{1 + l^2\sigma_{ik}^2}, \frac{\sigma_{ik}^2}{1 + l^2\sigma_{ik}^2} I\right)\left(- \frac{1}{2\sigma_{ik}^2}(\mathbf{w}_{ik}^T\mathbf{w}_{ik} - 2\mathbf{w}_{ik}^T\mathbf{m}_{ik} + \mathbf{m}_{ik}^T\mathbf{m}_{ik}) - \frac{K}{2}\log(2\pi\sigma_{ik}^2)\right)\mathrm{d}\mathbf{w}_{ik}$$

$$= -\frac{l^4\sigma_{ik}^2}{(1 + l^2\sigma_{ik}^2)^2}\frac{\mathbf{m}_{ik}^T\mathbf{m}_{ik}}{2} - \frac{K}{2(1 + l^2\sigma_{ik}^2)} - \frac{K}{2}\log(2\pi\sigma_{ik}^2)$$

leading to an analytical solution. Note that Z_q depends on the variational parameters, and thus was not omitted. This derivation results in an L_2 like regularisation, depending on the magnitude of $\mathbf{m}$, but through α as well.