Assignment 10

1 KL divergence and the ELBO

Let $\boldsymbol{x}$ be observed data and $\boldsymbol{z}$ be latent variables. We wish to compute the marginal likelihood $p(\boldsymbol{x}) = \sum_z p(\boldsymbol{x}, \boldsymbol{z})$, but the integral is intractable. Define the unnormalized target $\tilde{p}(\boldsymbol{z}) = p(\boldsymbol{x}, \boldsymbol{z})$. In variational inference, we introduce a tractable variational distribution $q(\boldsymbol{z})$ to approximate the posterior $p(\boldsymbol{z} \mid \boldsymbol{x}) = \tilde{p}(\boldsymbol{z})/p(\boldsymbol{x})$.

Throughout this problem, $\mathrm{KL}(q\|p)$ denotes the KL divergence $\sum_z q(\boldsymbol{z}) \log \frac{q(\boldsymbol{z})}{p(\boldsymbol{z})}$.

1. Prove that $\mathrm{KL}(q\|p) \geq 0$ for any distributions q and p, with equality if and only if $q = p$. You might want to use Jensen's inequality: For a concave function f (such as $\log()$) and nonnegative weights $w(x)$ such that $\sum_x w(x) = 1$, $\sum_x w(x) f(x) \leq f\left(\sum_x w(x) x\right)$.

2. For the Evidence Lower Bound (ELBO),

$$\mathcal{L}(q) \;=\; \mathbb{E}_{q(\boldsymbol{z})}[\log \tilde{p}(\boldsymbol{z}) - \log q(\boldsymbol{z})]\,,$$

   prove the following decomposition:

$$\log p(\boldsymbol{x}) \;=\; \mathcal{L}(q) \;+\; \mathrm{KL}(q(\boldsymbol{z}) \,\|\, p(\boldsymbol{z} \mid \boldsymbol{x}))\,.$$

   Explain why this implies $\mathcal{L}(q) \leq \log p(\boldsymbol{x})$, with equality if and only if $q(\boldsymbol{z}) = p(\boldsymbol{z} \mid \boldsymbol{x})$.