Probabilistic Machine Learning

Exercise Sheet #11

Variational Inference

1. **EXAMple Question — Gibbs' inequality.** The Kullback-Leibler divergence between two probability distributions p, q has been used several times in the lectures already:

$$D_{\mathrm{KL}}(p\|q) := \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx.$$

Show that this expression is non-negative for any p, q. In Physics, this fact is known as *Gibbs' inequality.* Hint: Use Jensen's inequality, which states that for all probability densities p, any *convex* function f, and any real-valued (p-integrable) g,

$$f\mathbb{E}_p[g(x)] \leq \mathbb{E}_p[f \circ g(x)].$$

Note that log is a *concave* function.

2. **Theory Question — Free energy for Gaussians.** In the lecture, a sketch was shown to argue that an approximation q to a distribution p found by minimizing $D_{\mathrm{KL}}(p\|q)$ tends to be "too wide", while an approximation found by minimizing $D_{\mathrm{KL}}(q\|p)$ tends to be "too narrow". The following argument supports this insight using the simple case of Gaussian distributions:

   (a) Show that the Kullback-Leibler divergence between two scalar, centered Gaussian distributions is given by

$$D_{\mathrm{KL}}(\mathcal{N}(x; 0, \sigma_q^2)\|\mathcal{N}(x; 0, \sigma_p^2)) = \frac{1}{2}\left(\log\left(\frac{\sigma_p^2}{\sigma_q^2}\right) - 1 + \frac{\sigma_q^2}{\sigma_p^2}\right).$$

   (b) Consider the two-dimensional Gaussian distribution p and a spherical approximation q, each given by (with $\sigma_1 \neq \sigma_2$)

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right), \qquad q(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_q^2 \end{bmatrix}\right),$$

   Find the values of σ_q that minimize $D_{\mathrm{KL}}(p\|q)$ and $D_{\mathrm{KL}}(q\|p)$, respectively.

3. **Practical Question.** In this week's coding exercise, we will replace the Gibbs sampler introduced in last week's lectures and exercise with the variational method as developed in the lectures. Use `Exercise_11.ipynb` as your workspace.