Machine Learning for Graphs and Sequential Data Exercise Sheet 02

Variational Inference

Problem 1: Consider the following latent variable model.

$$p_\theta(z) = \mathrm{Expo}(z|\theta) = \begin{cases} \theta \exp(-\theta z) & \text{if } z > 0, \\ 0 & \text{else.} \end{cases}$$

$$p(x|z) = \mathcal{N}(x|z, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - z)^2\right),$$

where $x \in \mathbb{R}$ is the observed data and $z \in \mathbb{R}_+$ is the latent variable. We have observed a single data point x and now would like to maximize the marginal log-likelihood $\log p_\theta(x) = \log\left(\int p(x|z)p_\theta(z)dz\right)$ w.r.t. the model parameters $\theta \in \mathbb{R}_+$. For this we will use variational inference.

We define the following parametric family of variational distributions

$$q_\phi(z) = \mathrm{Expo}(z|\phi) = \begin{cases} \phi \exp(-\phi z) & \text{if } z > 0 \\ 0 & \text{else;} \end{cases}$$

that is parametrized by $\phi \in \mathbb{R}_+$. We are interested in solving the optimization problem

$$\max_{\theta > 0, \phi > 0} \mathcal{L}(\theta, \phi).$$

a) Assume that θ is known and fixed. Does there exist a value of ϕ such that the ELBO is tight, i.e. $\log p_\theta(x) = \mathcal{L}(\theta, \phi)$? Justify your answer.

b) Write down the ELBO $\mathcal{L}(\theta, \phi)$ for the above probabilistic model $p_\theta(x, z)$ and the variational distribution $q_\phi(z)$ and simplify it as far as you can. Your final answer should be a closed-form expression (no integrals or expectations).

c) Compute the gradients of the ELBO $\nabla_\theta \mathcal{L}(\theta, \phi)$ and $\nabla_\phi \mathcal{L}(\theta, \phi)$.

Problem 2: You want to draw samples from an exponential distribution with rate ϕ with reparametrization. Assume that

$$q_\phi(z) = \mathrm{Expo}(z|\phi) = \begin{cases} \phi \exp(-\phi z) & \text{if } z > 0 \\ 0 & \text{else;} \end{cases}$$

where $\phi \in \mathbb{R}_+$.

a) You have access to an algorithm that produces samples ϵ from an exponential distribution with unit rate, that is

$$b(\epsilon) = \mathrm{Expo}(\epsilon|1) = \begin{cases} \exp(-\epsilon) & \text{if } \epsilon > 0 \\ 0 & \text{else.} \end{cases}$$

Write a deterministic transformation $T(\epsilon, \phi)$ that converts a sample $\epsilon \sim b(\epsilon)$ into a sample from $q_\phi(z)$. Use the change of variables formula to show that $z = T(\epsilon, \phi)$ follows the desired distribution.

b) Now, you have access to an algorithm that produces samples u from a uniform distribution on $[0, 1]$, that is

$$b(u) = \begin{cases} 1 & \text{if } u \in [0, 1] \\ 0 & \text{else.} \end{cases}$$

Write a deterministic transformation $S(u, \phi)$ that converts a sample $u \sim b(u)$ into a sample from $q_\phi(z)$. Use the change of variables formula to show that $z = S(u, \phi)$ follows the desired distribution.

Problem 3: You are given two distributions $q(\boldsymbol{z})$ and $p(\boldsymbol{z})$ over some random vector $\boldsymbol{z} \in \mathbb{R}^D$. Assume that both distributions can be factorized as

$$q(\boldsymbol{z}) = \prod_{i=1}^{D} q_i(z_i) \qquad\qquad p(\boldsymbol{z}) = \prod_{i=1}^{D} p_i(z_i).$$

(This is equivalent to saying that each component z_i is independent of z_j for $j \neq i$ under the distributions q and p). Your task is to prove that in this case the following equality holds

$$\mathbb{KL}(q(\boldsymbol{z}) \| p(\boldsymbol{z})) = \sum_{i=1}^{D} \mathbb{KL}(q_i(z_i) \| p_i(z_i)).$$