DEEP LEARNING (SS2022)
SEMINAR 7

Assignment 1 (Bernoulli VAE)**.** Let us consider a VAE with binary valued latent variables $z \in \mathcal{Z} = \{0,1\}^n$. Training such VAEs by maximising the ELBO criterion requires computation of the gradient of the data term w.r.t. encoder parameters φ

$$\nabla_\varphi \mathbb{E}_{q_\varphi(z \mid x)} \log p_\theta(x \mid z) = \nabla_\varphi \sum_{z \in \mathcal{Z}} q_\varphi(z \mid x) \log p_\theta(x \mid z). \tag{1}$$

Here we can not apply the re-parametrisation trick as in the case of Gaussian latent variables.

a) We can explicitly sum over $z \in \mathcal{Z}$ if the dimension of the latent space is small. This is however not possible for high dimensional latent spaces.

b) (Score function, log-trick) Prove the following equality

$$\nabla_\varphi \sum_{z \in \mathcal{Z}} q_\varphi(z \mid x) \log p_\theta(x \mid z) = \sum_{z \in \mathcal{Z}} q_\varphi(z \mid x) \log p_\theta(x \mid z) \nabla_\varphi \log q_\varphi(z \mid x)$$

Conclude that the following procedure implements an unbiased stochastic estimator of the required gradient (1):

Sample $z \sim q_\varphi(z \mid x)$ and compute $\log p_\theta(x \mid z) \nabla_\varphi \log q_\varphi(z \mid x)$

Assignment 2 (VAE: Sticked Landing)**.** When learning VAEs by maximising the ELBO criterion, we must compute the gradient of the KL-divergence term, i.e.

$$\nabla_\varphi \sum_{z \in \mathcal{Z}} q_\varphi(z \mid x) \big[\log q_\varphi(z \mid x) - \log p(z)\big],$$

where $p(z)$ denotes the prior distribution on the latent space.

a) Usually, this KL-divergence can be computed in closed form. In case of a Bernoulli VAE discussed in the previous assignment, this amounts to compute the KL-divergence for pairs of Bernoulli distributions. Give a formula it.

b) Let us now consider the gradient of the first term in the formula above.

$$\nabla_\varphi \sum_{z \in \mathcal{Z}} q_\varphi(z \mid x) \log q_\varphi(z \mid x) = \sum_{z \in \mathcal{Z}} \log q_\varphi(z \mid x) \nabla_\varphi q_\varphi(z \mid x) + \sum_{z \in \mathcal{Z}} q_\varphi(z \mid x) \nabla_\varphi \log q_\varphi(z \mid x)$$

Prove that the second sum is always zero.

Assignment 3 (Score function)**.** The score function approach discussed in the first assignment provides an unbiased gradient estimator. However, its key drawback is its high variance. Let us study this on a simple example. Consider the derivative

$$\frac{d}{d\beta}\mathbb{E}_\beta[z] = \frac{d}{d\beta}\sum_z q_\beta(z) z,$$

where $q_\beta(z) = \beta^z(1-\beta)^{(1-z)}$ is a Bernoulli distribution for a binary variable $z = 0, 1$.

a) Show that this derivative equals 1.

b) Let us now consider the score function approach for this derivative.

$$\frac{d}{d\beta}\mathbb{E}_\beta[z] = \sum_z q_\beta(z) z \frac{d}{d\beta}\log q_\beta(z)$$

Show that $z\frac{d}{d\beta}\log q_\beta(z) = \frac{z}{\beta}$. Compute the variance of estimating this random variable on an i.i.d. sample $\{z_i \mid i = 1, \ldots, m\}$ generated from q_β. How is it depending on β?