Exercises for the course
Deep Learning 2
Summer Semester 2023

Machine Learning Group
Faculty IV – Electrical Engineering and Computer Science
Technische Universität Berlin

Exercise Sheet 3

Exercise 1: Deriving the Evidence Lower Bound (25 + 25 P)

In this exercise we want to derive the evidence lower bound mentioned in the lecture.

(a) Therefore you first show:

$$D_{KL}[Q(z|X)\|P(z|X)] = \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log P(X|z) - \log P(z)\right] + \log P(X) \tag{1}$$

where D_{KL} is the KL divergence.

$$\begin{aligned}
D_{KL}[Q(z|X)\|P(z|X)] &= \int_z Q(z|X)\log\frac{Q(z|X)}{P(z|X)}dz \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log\frac{Q(z|X)}{P(z|X)}\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log P(z|X)\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log\frac{P(X|z)P(z)}{P(X)}\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - (\log P(X|z) + \log P(z) - \log P(X))\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log P(X|z) - \log P(z) + \log P(X)\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log P(X|z) - \log P(z)\right] + \log P(X)
\end{aligned}$$

(b) Show:

$$\log P(X) - D_{KL}[Q(z|X)\|P(z|X)] = \mathbb{E}_{z\sim Q(z|X)}\left[\log P(X|z)\right] - D_{KL}[Q(z|X)\|P(z)] \tag{2}$$

Now you can choose suitable tractable functions $P(X|z)$, $Q(z|X)$ (e.g. as neural networks) and $P(z)$ (e.g. as isotropic gaussian) and one can train the variational autoencoder by maximizing the ELBO.

$$\begin{aligned}
\log P(X) - D_{KL}[Q(z|X)\|P(z|X)] &= \log P(X) - (\mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log P(X|z) - \log P(z)\right] + \log P(X)) \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log P(X|z) - (\log Q(z|X) - \log P(z))\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log P(X|z)\right] - \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log P(z)\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log P(X|z)\right] - D_{KL}[Q(z|X)\|P(z)]
\end{aligned}$$

Exercise 2: Programming (50 P)

Download the programming files on ISIS and follow the instructions.
Fill in the gaps in the GPT implementation.