DD2434 Advanced Machine Learning
Assignment 1AD

Reuben Gezang

November 2025

D-Level

Theory 1.D.1

Question 1.1.1

We start by stating the definition of the Kullback-Leibler divergence:

$$KL(q(Z)\|p(Z|X)) = \mathbb{E}_{q(\boldsymbol{Z})}[\log(\frac{q(Z)}{p(Z|X)})] = \int q(Z)\log(\frac{q(Z)}{p(Z|X)})dZ \tag{1}$$

Now we separete the fraction inside the logarithm:

$$KL(q(Z)\|p(Z|X)) = \int q(Z)\log(\frac{q(Z)p(X)}{p(X,Z)})dZ = \int q(Z)(\log(q(Z)) - \log(p(Z|X))dZ \tag{2}$$

Note that $p(Z|X) = \frac{p(X,Z)}{p(X)}$ and using this we can rewrite the equation as:

$$KL(q(Z)\|p(Z|X)) = \int q(Z)\log(q(Z))dZ - \int q(Z)\log(p(X,Z))dZ + \log(p(X))\int q(Z)dZ \tag{3}$$

Since $q(Z)$ is a probability distribution, we know that $\int q(Z)dZ = 1$. Now we can solve for $\log(p(X))$:

$$\log(p(X)) = KL(q(Z)\|p(Z|X)) - \int q(Z)\log(q(Z))dZ + \int q(Z)\log(p(X,Z))dZ \tag{4}$$

Combining the logarithm terms, we get:

$$\log(p(X)) = KL(q(Z)\|p(Z|X)) + \mathbb{E}_{q(\boldsymbol{Z})}[\frac{\log(p(X,Z))}{q(Z)}] \tag{5}$$

Now we can identify the Evidence lower bound (ELBO)

$$\mathcal{L}(q) = \mathbb{E}_{q(\boldsymbol{Z})}[\frac{\log(p(X,Z))}{q(Z)}] \tag{6}$$

and with this we have shown that:

$$\log(p(X)) = \mathcal{L}(q) + KL(q(Z)\|p(Z|X)) \tag{7}$$

concluding the proof.

Question 1.1.2

KOLLA ÖVER IGEN In this question we are to describe (in one sentence) how the choice of variational family $q(Z)$ affects

- (i) The tightness of the ELBO
- (ii) The accuracy of the posterior approximation

(i) A more expressive variational family can lead to a tighter ELBO as it can better approximate the true posterior, reducing the KL divergence term.

(ii) The choice of variational family directly impacts the accuracy of the posterior approximation, as a limited family may not capture the true posterior's complexity, leading to a less accurate approximation.

Question 1.D.2

1.1.3

For a mean field assumption and joint distribution

$$q(Z_1, Z_2, Z_3) = q_1(Z_1)q_2(Z_2)q_3(Z_3), \quad p(X, Z)$$

Let $q_1^*(Z_1)$ be the q_1 that maximizes the ELBO. We want to show that q_1^* satisfies

$$\log q_1^*(Z_1) = \mathbb{E}_{-Z_1}[\log p(X, Z)]$$

We can start by inspecting the ELBO:

$$\mathcal{L}(q) = \mathcal{L}(q) = \mathbb{E}_{q(\boldsymbol{Z})}[\frac{\log(p(X, Z))}{q(Z)}] = \mathbb{E}_q[\log p(X, Z)] - \mathbb{E}_q[\log q(Z)]$$

and using the mean field assumption we can rewrite this as:

$$\mathcal{L}(q) = \mathbb{E}_{q(Z)}[\log p(X, Z)] - \mathbb{E}_{q(Z)}[\log(q_1(Z_1)q_2(Z_2)q_3(Z_3))]$$

and by separating the logarithm we get:

$$\mathbb{E}_{q(Z)}[\log(q_1(Z_1)q_2(Z_2)q_3(Z_3))] = \mathbb{E}_{q_1(Z_1)}[\log(q_1(Z_1))] + \mathbb{E}_{q_2(Z_2)}[\log(q_2(Z_2))] + \mathbb{E}_{q_3(Z_3)}[\log(q_3(Z_3))]$$

Since we are maximizing w.r.t $q_1(Z_1)$ we can ignore the terms that do not depend on it. Thus we can rewrite the ELBO as:

$$\mathcal{L}(q) = \mathbb{E}_{q(Z)}[\log p(X, Z)] - \mathbb{E}_{q_1(Z_1)}[\log(q_1(Z_1))] + C$$

where C is a constant w.r.t $q_1(Z_1)$ (and can thus be ignored). Now we can rewrite the expectation over $q(Z)$ as:

$$\mathbb{E}_{q(Z)}[\log p(X, Z)] = \mathbb{E}_{q_1(Z_1)}[\mathbb{E}_{q_2(Z_2)q_3(Z_3)}[\log p(X, Z)]]$$

meaning that we can rewrite the ELBO as:

$$\mathcal{L}(q) = \mathbb{E}_{q_1(Z_1)}[\mathbb{E}_{q_2(Z_2)q_3(Z_3)}[\log p(X, Z)]] - \mathbb{E}_{q_1(Z_1)}[\log(q_1(Z_1))] + C$$

Using the fact that $\int_{Z_1} q(Z_1)dZ_1 = 1$ we will now optimize the ELBO w.r.t $q_1(Z_1)$ and with a lagrange multiplier λ.

$$\frac{\partial}{\partial q_1(Z_1)} \left(\mathbb{E}_{q_1(Z_1)}[\mathbb{E}_{q_2(Z_2)q_3(Z_3)}[\log p(X, Z)]] - \mathbb{E}_{q_1(Z_1)}[\log(q_1(Z_1))] + \lambda(\int_{Z_1} q(Z_1)dZ_1 - 1) \right) = 0 \tag{8}$$

giving that

$$\mathbb{E}_{q_2(Z_2)q_3(Z_3)}[\log p(X,Z)] - \log(q_1(Z_1)) - 1 + \lambda = 0 \rightarrow \log(q_1^*(Z_1)) = \mathbb{E}_{q_2(Z_2)q_3(Z_3)}[\log p(X,Z)] + \lambda - 1 \tag{9}$$

where $\lambda - 1$ is a additive constant that can be ignored when normalizing $q_1^*(Z_1)$. Thus we have shown that:

$$\log q_1^*(Z_1) = \mathbb{E}_{-Z_1}[\log p(X,Z)] \tag{10}$$

as required.

Practice/Implementation - D level (I have chosen 1.D.3)

1.2.4

The log likelihood of the data is:

A Additional Details

Add derivations, extra figures, or ablation studies here.