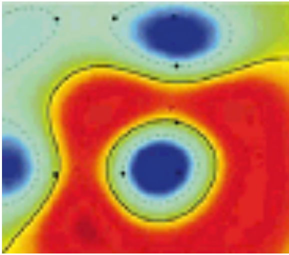
Technische Universität Berlin



Solution
Exercise 3

Generative Models - part 1

Deriving the Evidence Lower Bound

In this exercise, we want to derive the Evidence Lower Bound (ELBO) for training variational autoencoders.

Deriving the Evidence Lower Bound - Exercise A

Show:

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

Deriving the Evidence Lower Bound - Solution A

Definition KL Divergence

$$D_{KL}[Q(z|X) \| P(z|X)] = \int_z Q(z|X) \log \frac{Q(z|X)}{P(z|X)} dz \tag{2}$$

Deriving the Evidence Lower Bound - Solution A

Definition Expected Value

$$\begin{aligned} D_{KL}[Q(z|X)\|P(z|X)] &= \int_z Q(z|X)\log\frac{Q(z|X)}{P(z|X)}dz \\ &= \mathbb{E}_{z\sim Q(z|X)}\left[\log\frac{Q(z|X)}{P(z|X)}\right] \end{aligned} \tag{3}$$

Deriving the Evidence Lower Bound - Solution A

Logartihm Rule

$$
\begin{aligned}
D_{KL}[Q(z|X) \| P(z|X)] &= \int_z Q(z|X) \log \frac{Q(z|X)}{P(z|X)} dz \\
&= \mathbb{E}_{z\sim Q(z|X)} \left[\log \frac{Q(z|X)}{P(z|X)} \right] \\
&= \mathbb{E}_{z\sim Q(z|X)} [\log Q(z|X) - \log P(z|X)]
\end{aligned} \tag{4}
$$

Deriving the Evidence Lower Bound - Solution A

Bayes Theorem

$$
\begin{aligned}
D_{KL}[Q(z|X)\|P(z|X)] &= \int_z Q(z|X)\log\frac{Q(z|X)}{P(z|X)}dz \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log\frac{Q(z|X)}{P(z|X)}\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log P(z|X)\right] \\
&= \mathbb{E}_{z\sim Q(z|X)}\left[\log Q(z|X) - \log\frac{P(X|z)P(z)}{P(X)}\right]
\end{aligned} \tag{5}
$$

Deriving the Evidence Lower Bound - Solution A

Logarithm Rule

$$
\begin{aligned}
D_{KL}[Q(z|X) \| P(z|X)] &= \int_z Q(z|X) \log \frac{Q(z|X)}{P(z|X)} dz \\
&= \mathbb{E}_{z\sim Q(z|X)} \left[\log \frac{Q(z|X)}{P(z|X)} \right] \\
&= \mathbb{E}_{z\sim Q(z|X)} [\log Q(z|X) - \log P(z|X)] \\
&= \mathbb{E}_{z\sim Q(z|X)} \left[\log Q(z|X) - \log \frac{P(X|z)P(z)}{P(X)} \right] \\
&= \mathbb{E}_{z\sim Q(z|X)} [\log Q(z|X) - (\log P(X|z) + \log P(z) - \log P(X))
\end{aligned}
\tag{6}
$$

Deriving the Evidence Lower Bound - Solution A

Removing inner bracket

$$
\begin{aligned}
D_{KL}[Q(z|X) \| P(z|X)] &= \int_z Q(z|X) \log \frac{Q(z|X)}{P(z|X)} dz \\
&= \mathbb{E}_{z \sim Q(z|X)} \left[\log \frac{Q(z|X)}{P(z|X)} \right] \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - \log P(z|X)] \\
&= \mathbb{E}_{z \sim Q(z|X)} \left[\log Q(z|X) - \log \frac{P(X|z)P(z)}{P(X)} \right] \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - (\log P(X|z) + \log P(z) - \log P(X)) \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - \log P(X|z) - \log P(z) + \log P(X)]
\end{aligned}
\tag{7}
$$

Deriving the Evidence Lower Bound - Solution A

Moving constant out of expected value

$$
\begin{aligned}
D_{KL}[Q(z|X) \| P(z|X)] &= \int_z Q(z|X) \log \frac{Q(z|X)}{P(z|X)} dz \\
&= \mathbb{E}_{z \sim Q(z|X)} \left[\log \frac{Q(z|X)}{P(z|X)} \right] \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - \log P(z|X)] \\
&= \mathbb{E}_{z \sim Q(z|X)} \left[\log Q(z|X) - \log \frac{P(X|z)P(z)}{P(X)} \right] \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - (\log P(X|z) + \log P(z) - \log P(X)) \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - \log P(X|z) - \log P(z) + \log P(X)] \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - \log P(X|z) - \log P(z)] + \log P(X)
\end{aligned}
\tag{8}
$$

Deriving the Evidence Lower Bound - Exercise B

Show:

$$\log P(X) - D_{KL}[Q(z|X) \| P(z|X)] = \mathbb{E}_{z \sim Q(z|X)} [\log P(X|z)] - D_{KL}[Q(z|X) \| P(z)] \tag{9}$$

Deriving the Evidence Lower Bound - Solution B

Substitute $D_{KL}[Q(z|X) \| P(z|X)]$ according to exercise A.

$$\begin{aligned} &\log P(X) - D_{KL}[Q(z|X) \| P(z|X)] \\ = &\log P(X) - (\mathbb{E}_{z \sim Q(z|X)}[\log Q(z|X) - \log P(X|z) - \log P(z)] + \log P(X)) \end{aligned} \tag{10}$$

Deriving the Evidence Lower Bound - Solution B

$\log P(X)$ terms cancel each other out, scalar multiplication rule for expected values

$$
\begin{aligned}
&\log P(X) - D_{KL}[Q(z|X) \| P(z|X)] \\
&= \log P(X) - (\mathbb{E}_{z \sim Q(z|X)}[\log Q(z|X) - \log P(X|z) - \log P(z)] + \log P(X)) \\
&= \mathbb{E}_{z \sim Q(z|X)}[\log P(X|z) - (\log Q(z|X) - \log P(z))]
\end{aligned} \tag{11}
$$

Deriving the Evidence Lower Bound - Solution B

Rules for addition in expected values.

$$
\begin{aligned}
&\log P(X) - D_{KL}[Q(z|X) \| P(z|X)] \\
= \log P(X) - (\mathbb{E}_{z\sim Q(z|X)}&[\log Q(z|X) - \log P(X|z) - \log P(z)] + \log P(X)) \\
&= \mathbb{E}_{z\sim Q(z|X)}[\log P(X|z) - (\log Q(z|X) - \log P(z))] \\
&= \mathbb{E}_{z\sim Q(z|X)}[\log P(X|z)] - \mathbb{E}_{z\sim Q(z|X)}[\log Q(z|X) - \log P(z)]
\end{aligned}
\tag{12}
$$

Deriving the Evidence Lower Bound - Solution B

Definition expected value, definition KL Divergence

$$
\begin{aligned}
&\log P(X) - D_{KL}[Q(z|X) \| P(z|X)] \\
&= \log P(X) - (\mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - \log P(X|z) - \log P(z)] + \log P(X)) \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log P(X|z) - (\log Q(z|X) - \log P(z))] \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log P(X|z)] - \mathbb{E}_{z \sim Q(z|X)} [\log Q(z|X) - \log P(z)] \\
&= \mathbb{E}_{z \sim Q(z|X)} [\log P(X|z)] - D_{KL}[Q(z|X) \| P(z)]
\end{aligned} \tag{13}
$$

Proof Counterfactual stays on Data Manifold

In this exercise we want to show, that the counterfactual produced by the diffeomorphic counterfactuals procedure actually still lies on the data manifold.

Hint: Use Taylor Decomposition, the Chain Rule and singular value decomposition

$$g\left(z^{(i+1)}\right) = g\left(z^{(i)}\right) + \lambda U \Sigma^2 U^T \frac{\partial f_t}{\partial x} + \mathcal{O}\left(\lambda^2\right)$$

Proof Counterfactual stays on Data Manifold - Solution A

$$
\begin{aligned}
g\left(z^{(i+1)}\right) &= g(z^{(i)} + \lambda \frac{\partial (f \circ g)_t}{\partial z}(z^{(i)})) \\
&= g^{\left(z^{(i)}\right)} + \frac{\partial g}{\partial z}(z - z^{(i)}) + \mathcal{O}\left(|z - z^{(i)}|^2\right) \\
&= g^{\left(z^{(i)}\right)} + \frac{\partial g}{\partial z}(z^{(i)} + \lambda \frac{\partial (f \circ g)_t}{\partial z} - z^{(i)}) + \mathcal{O}\left(|z^{(i)} + \lambda \frac{\partial (f \circ g)_t}{\partial z} - z^{(i)}|^2\right) \\
&= g^{\left(z^{(i)}\right)} + \lambda \frac{\partial g}{\partial z} \frac{\partial (f \circ g)_t}{\partial z} + \mathcal{O}\left(\lambda^2\right) \\
&= g^{\left(z^{(i)}\right)} + \lambda \frac{\partial g}{\partial z} \frac{\partial f_t}{\partial g} \frac{\partial g}{\partial z} + \mathcal{O}\left(\lambda^2\right) \\
&= g^{\left(z^{(i)}\right)} + \lambda \frac{\partial g}{\partial z} \frac{\partial g}{\partial z}^T \frac{\partial f_t}{\partial x} + \mathcal{O}\left(\lambda^2\right) \\
&= g\left(z^{(i)}\right) + \lambda U \Sigma V V^T \Sigma^T U^T \frac{\partial f_t}{\partial x} + \mathcal{O}\left(\lambda^2\right) \\
&= g\left(z^{(i)}\right) + \lambda U \Sigma^2 U^T \frac{\partial f_t}{\partial x} + \mathcal{O}\left(\lambda^2\right)
\end{aligned}
$$