http://ml3.leuphana.de/lectures/summer24/DL
Machine Learning Group, Leuphana University of Lüneburg

Soham Majumder (soham.majumder@leuphana.de)

Exercise 6

Due on: Thursday, 13.06.2024

Task 15 Kullback–Leibler divergence

The Kullback–Leibler divergence (or short KL divergence) is a measure of distance between two probability distributions. Let $p(x)$ and $q(x)$ be two probability distributions on the same probability space $\mathcal{X}$. Then, the Kullback–Leibler divergence is given by

$$D_{\mathrm{KL}}\left(q \| p\right) = \int_{\mathcal{X}} q(x) \log\left(\frac{q(x)}{p(x)}\right) \mathrm{d}x.$$

(i) Using the inequality $\log(t) \leq t - 1$ for $t > 0$, show that the KL divergence is non-negative.

Hint: Start with $-D_{\mathrm{KL}}\left(q \| p\right)$.

(ii) In mathematics, a metric or distance function measures the distance between two objects. The definition is as follows. Let $\mathcal{X}$ be a set. A function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a metric on $\mathcal{X}$ if it satisfies for any three objects x, y, and z of $\mathcal{X}$ the following three properties:

(a) $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$

(b) $d(x, y) = d(y, x)$

(c) $d(x, y) \leq d(x, z) + d(z, y)$

Is the Kullback–Leibler divergence a metric?

(iii) We often need the KL divergence between two Gaussians. Let $\mathcal{N}_i$ be a Gaussian with mean $\mu_i \in \mathbb{R}^d$ and covariance matrix $\Sigma_i \in \mathbb{R}^{d \times d}$. Then, the KL divergence is given by

$$D_{\mathrm{KL}}\left(\mathcal{N}_0 \| \mathcal{N}_1\right) = \frac{1}{2}\left(\operatorname{tr}\left(\Sigma_1^{-1}\Sigma_0\right) + \left(\mu_1 - \mu_0\right)^\top \Sigma_1^{-1}\left(\mu_1 - \mu_0\right) - d + \ln\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right)\right).$$

Assume that $\mathcal{N}_0$ has an arbitrary mean $\mu_0 \in \mathbb{R}^d$ and a diagonal covariance matrix $\Sigma_0 = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_d^2\right)$ and that $\mathcal{N}_1$ is a standard Gaussian, *i.e.*, $\mathcal{N}_1$ has a mean of zero and the covariance matrix is the identity.

Show that the KL divergence between $\mathcal{N}_0$ and $\mathcal{N}_1$ is given by

$$D_{\mathrm{KL}}\left(\mathcal{N}_0 \| \mathcal{N}_1\right) = \frac{1}{2}\left(\sum_{j=1}^{d} \sigma_j^2 + \mu_j^2 - 1 - \ln \sigma_j^2\right).$$

Task 16 Autoencoder

(i) Define encoder and decoder networks for the MNIST data set. Use the MSE loss and think about proper activation functions for the last layers of both networks.

Hint: A ReLU activation in the last layer of the decoder is a bad choice if the input data is within $[-1, 1]^d$.

(ii) Experiment with some latent dimensionalities, but make sure to also try a two-dimensional latent space.

(iii) Train the autoencoder and show the two-dimensional latent space. Color the data points according to their label. You can use `matplotlib.pyplot.scatter` for that.

(iv) Pick two points and linearly interpolate them in latent space. Show the decoded interpolation path. Use at least five intermediate points.