Exercise sheet 5

The Exercises are meant to be done in groups of three, **please enroll your group in OLAT**. Even if you are doing the exercies on your own, please enroll yourself in an enmpty group in OLAT. Without enrollment, you wont be able to submit.
Submission of your solutions is done via OLAT. Please add the names of all group members to your solution.

Excercise 5.1: Kullback-Leibler Divergence (5 Points)

Given functions $p : \mathbb{R}^k \to \mathbb{R}$ and $q : \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}$, let the function $D : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$D(x) := \int_{\mathbb{R}^k} q(z,x) \ln \frac{q(z,x)}{p(z)}\, dz, \forall x \in \mathbb{R}^n. \tag{5.1}$$

Assume that p and q are given by the formulas

$$\begin{aligned} q(z,x) &:= \frac{\det(\Sigma(x))^{-\frac{1}{2}}}{(2\pi)^{\frac{k}{2}}} \exp\left(-\frac{1}{2}\big(z-\mu(x)\big)^\top \Sigma(x)^{-1}\big(z-\mu(x)\big)\right), \quad \forall x \in \mathbb{R}^n, z \in \mathbb{R}^k \\ p(z) &:= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\|z\|^2}{2}\right), \quad \forall z \in \mathbb{R}^k \end{aligned} \tag{5.2}$$

where, $\mu(x) : \mathbb{R}^n \to \mathbb{R}^k$ and $\Sigma(x) : \mathbb{R}^n \to \mathbb{R}^{k\times k}$, $\Sigma(x) > 0$ for all $x \in \mathbb{R}^n$.

(a) Calculate the term $\ln \frac{q(z,x)}{p(z)}$ using (5.2)

(b) Calculate the term $q(z,x) \ln \frac{q(z,x)}{p(z)}$ using (5.2)

(c) Substitute the result obtained from (b) in (5.1) and prove that

$$D(x) = \frac{1}{2}\Big(\mathrm{tr}(\Sigma(x)) + \|\mu(x)\|^2 - k - \ln\left[\det(\Sigma(x))\right]\Big).$$

Hints:

(i) Use the fact that p and q are (Gaussian) probability density functions with respect to z, i.e.

$$\begin{aligned} 1 &= \int_{\mathbb{R}^k} p(z)\ dz, \text{ and } 1 = \int_{\mathbb{R}^k} q(z,x)\ dz, \forall x \in \mathbb{R}^n \\ 0 &= \int_{\mathbb{R}^k} z\ p(z)\ dz, \text{ and } I = \int_{\mathbb{R}^k} zz^\top\ p(z)\ dz,\ I \in \mathbb{R}^{k\times k} \\ \mu(x) &= \int_{\mathbb{R}^k} z\ q(z,x)\ dz, \text{ and } \Sigma(x) = \int_{\mathbb{R}^k} (z-\mu)(z-\mu)^\top\ q(z,x)\ dz,\ \forall x \in \mathbb{R}^n \\ \mathrm{tr}(I) &= \int_{\mathbb{R}^k} (z-\mu(x))^\top \Sigma^{-1}(x)(z-\mu(x))\ q(z,x)\ dz,\ \forall x \in \mathbb{R}^n. \end{aligned}$$

Excercise 5.2: ADAM optimizer (8 Points)

Consider the ADAM optimizer with the following update rules at step t:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \quad m_0 = 0$$
$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2, \quad v_0 = 0$$
$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t},$$
$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon},$$

where $g_t = \nabla f_t(\theta_t)$ is the gradient of the stochastic loss function f_t, and $\alpha, \beta_1, \beta_2, \epsilon$ are hyperparameters. Assume that,

(i) the feasible set $\Theta \subseteq \mathbb{R}^d$ is convex and bounded with diameter D,

(ii) for all $t \geq 0$, the loss function $f_t : \Theta \to \mathbb{R}^+$ is convex with respect to Θ

(iii) $\exists\, \theta^* \in \Theta$ such that $\mathbb{E}[f_t(\theta^*)] = 0$ for all $t \geq 0$

(iv) the stochastic gradients g_t satisfy $\mathbb{E}[\|g_t\|^2] \leq G^2$

(v) $\mathbb{E}[\|\nabla f_{t_1}(\theta_{t_1}) - \nabla f_{t_2}(\theta_{t_2})\|] \leq \frac{G}{\sqrt{t_1}} |t_1 - t_2|$

Define a scalar function $R : \mathbb{R}^+ \to \mathbb{R}^+$, called the regret function, as $R(T) = \sum_{t=1}^{T} [f_t(\theta_t) - f_t(\theta^*)]$. For the ADAM update-rule prove that

1. the regret function has sublinear growth in expectation, i.e. it satisfies the following inequality:
$$\mathbb{E}[R(T)] \leq C\sqrt{T}, \quad C > 0.$$
2. $\mathbb{E}[f_T(\bar{\theta}_T)] \to 0$ as $T \to \infty$, where $\bar{\theta}_T := \frac{1}{T} \sum_{t=1}^{T} \theta_t, \quad T > 1.$

Excercise 5.3: Project 4 (5 Points)

The objective of this assignment is to get familiarized with implementing considerably larger network and also to get introduced to probabilistic modelling using networks. The project files are presented as python notebook files in Material/Projects/Proj4 folder. It contains the script file: celeb_face_gen.ipynb. This file implements a variational autoencoder (VAE) to new faces. The VAE uses the following design architecture Let $\phi : \mathbb{R}^n \to \mathbb{R}^k \times \mathbb{R}^{k \times k}$ denote the encoder network, let $\psi : \mathbb{R}^k \to \mathbb{R}^n$ denote the decoder network, then the VAE network is given as $\mathcal{V} : \mathbb{R}^n \to \mathbb{R}^n$, $\mathcal{V}(x) = \psi \circ \phi(x)$, where

$$\mu, \Sigma = \phi(X)$$
$$Z = \mu + \Sigma\epsilon, \quad \epsilon \sim \mathcal{N}(0,1), \text{a standard normally distributed random variable}$$
$$\hat{X} = \psi(Z)$$

Task: Implement the encoder (Task1) and decoder (Task2) networks using convolutional layers.

Submission guidelines:

1. Please submit the solved file (with outputs for each code-block) along with the theory exercise solutions. Please name your submission file as celeb_face_gen_solved_group_id.ipynb, where id is your two digit group number. For example, if you belong to Group01 please name the file as: celeb_face_gen_solved_group_01.ipynb
2. Please ensure that the submitted ipynb files are runnable and does not produce any errors. The project shall be considered for marking if and only if there are zero errors in the script file.