EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Computer Science Department
Chair of Cognitive Modeling

Winter semester 2019/2020

Advanced Neural Networks

Exercise sheet 04

Release: January 9, 2019 Deadline: January 29, 2020

General remarks:

- Download the file `exercisesheet04.zip` from the lecture site (ILIAS). This archive contains files (Python code and data), which are required for the exercises.
- Ideally, the exercises should be completed in teams of two students. Larger teams are not allowed.
- Add a brief documentation (pdf) to your submission. The documentation should contain protocols of your experiments, parameter choices, and discussions of the results.
- Please do not submit the pretrained weights from Exercise 2 again on ILIAS.

Exercise 1 – (Variational) Autoencoder and the KL Divergence [60 points]

(a) What is the general purpose of autoencoders (AEs) and what are their main components? Give an illustration. [8 points]

(b) Specify the general formula of the *Kullback-Leibler divergence* for discrete probability distributions. [4 points]

(c) Briefly explain the purpose of the KL divergence. [4 points]

(d) Given two discrete probability distributions Q, P. Show (by means of an example) that $KL(Q \parallel P) = KL(P \parallel Q)$ does **not** hold in general. [6 points]

(e) Given two k-dimensional normal distributions $\mathcal{N}_1(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ with $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ being k-dimensional mean vectors and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ are positive definite, non-singular $k \times k$ covariance matrices, then it holds

$$KL(\mathcal{N}_1 \parallel \mathcal{N}_2) = \frac{1}{2}\left\{ \operatorname{tr}\left(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\right) + (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) - k + \log \frac{\det \boldsymbol{\Sigma}_2}{\det \boldsymbol{\Sigma}_1} \right\}. \quad (1)$$

Show that when the reference distribution is standard normal the following simplification applies

$$KL\left(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I})\right) = \frac{1}{2}\sum_{i=1}^{k}(\sigma_i^2 + \mu_i^2 - \log(\sigma_i^2) - 1), \quad (2)$$

if $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_k^2)$ ($\boldsymbol{\Sigma}$ is a diagonal matrix). [20 points]

(f) What are the major differences of the *variational autoencoders* (VAEs) in comparison with standard autoencoders (name an essential advantage)? Relate your answer to the two sub-losses that are optimized during VAE training. [6 points]

(g) Briefly describe the *re-parametrization trick* and explain why it is necessary to train a VAE. [6 points]

(h) Name a frequently mentioned drawback of VAEs and give one example from the recent literature (reference and brief description) that addresses this issue. [6 points]

Exercise 2 – β-VAE in PyTorch [40 points]

(a) Image Generation [20 points]

The zip file of this exercise contains a pretrained β-VAE[1]. The first task is to modify the model such that images can be generated given 32-dimensional real valued vectors using only the decoder part of the VAE. What did the VAE learn? Generate at least ten clearly different example images. Hint: apply a sigmoid on the decoder's output.

(b) Latent Space Analysis [14 points]

Identify at least four general variables within the latent space that describe characteristic features of the visual appearance in the generated images. Document your findings also with examples images.

(c) Disentanglement [6 points]

Briefly explain the term *disentanglement* in the context of VAEs and Exercise 2 (b) in particular and how this is achieved.

[1] `https://openreview.net/pdf?id=Sy2fzU9gl`
`https://arxiv.org/pdf/1804.03599.pdf`