# Exercise 4

**Deadline: 09.06.2021, 4:00 pm**

In this exercise we will implement a variational autoencoder (VAE) and a conditional variational autoencoder (CVAE). We train both type of models on the MNIST dataset, visualize them and use a CVAE as a generative classifier on the MNIST dataset. We provide a Jupyter notebook `cvae.ipynb` on Moodle which contains code stubs for all tasks. You will also find additional explanations to the tasks in there. Please fill in the missing code at the marked places.

## Regulations

Please use the Jupyter notebook `cvae.ipynb` to fill in your solution and export it into `cvae.html`. Zip both files into a single archive.

Additionally, please set your Anzeigename/display name and Name in Uebungsgruppen/name in tutorials in MAMPF to your real name, which should be identical to your name in `muesli` and make sure you join the submission of your team via code before or during the submission. Check out `https://mampf.blog/handing-in-homework-assignments` for instructions.

## 1 VAE and CVAE (15 points)

Implement the CVAE class and the negated ELBO-loss in the Jupyter notebook. Detailed instructions can be found in the notebook. Check whether your implementation works by comparing the reconstruction of the VAE/CVAE with its input. Furthermore, sample latent vectors $z$ and use your decoder in order to generate samples from the data distribution. Evaluate experimentally how the number of latent dimensions influences the quality of the reconstruction and generated samples.

## 2 Visualisation VAE (10 points)

Visualization of the latent space is easiest when it has just two dimensions. Use this possibility to create two types of plots:

- Visualize the relationship between latent space positions (e.g. placed on a grid) and the corresponding decoder outputs, i.e. the generated images (see Fig. 1 below).
- Plot embeddings of different training instances in the latent space. Make sure to visualize their labels as well.

Design your plots such that they convey interesting information about the latent space structure. Give an interpretation of the visualizations with regard to this structure (2 points). Can you tell from the plot in which cases the VAE might fail to generate outputs that look like real data? Give a short explanation of possible failure cases and plot a few examples (2 points).

Ullrich Köthe
ullrich.koethe@iwr.uni-heidelberg.de

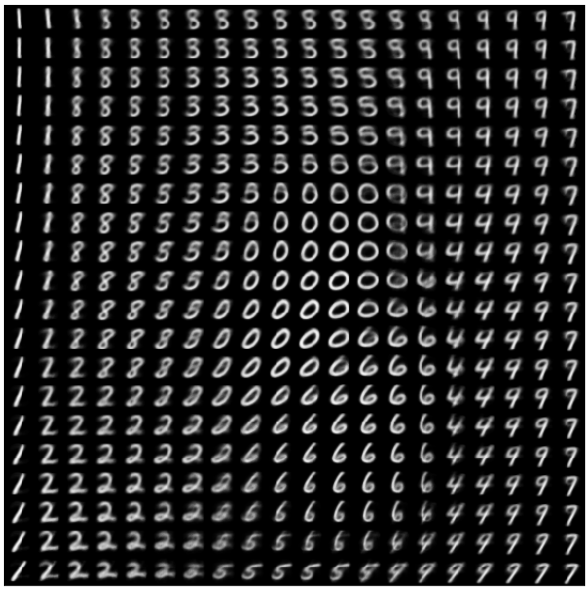

Abbildung 1: Example of how latent space can be illustrated. Output of VAE decoder for different locations of the latent space.

## 3 Visualization CVAE (5 points)

This task is similar to task 2, but we now use the CVAE with a 2 dimensional latent space. Visualize the relationship between latent code positions and corresponding decoder outputs (see Fig. 2 below) for three different conditions (=digits). Make sure your plot conveys interesting information about the latent space. What latent space structure do you observe (e.g. encoded styles)?

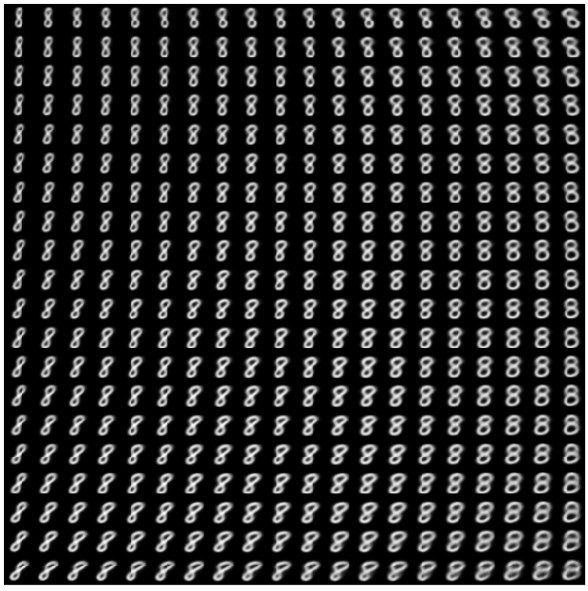

Abbildung 2: Example of how latent space can be illustrated. Output of CVAE decoder, conditioned on label 8, for different locations of the latent space.

Furthermore, plot embeddings of different training instances in the latent space for three conditions/digits. What can you say about the difference between the CVAE and VAE encodings?

# 4 Classification (10 points)

Train a CVAE on all classes as a generative model (or reuse the CVAE already learned above). Use this model as a generative classifier on MNIST.