# Exercise Sheet Deep Learning

# Part2: Generative Models
# Summer 22

This sheet includes a theoretical part and a practical assignment on the second part of the lecture Deep Learning (2_GAN). Both parts give 20 points maximum each. Please hand in solutions as a pdf in groups of at most three persons via LernraumPlus.

*name1*:

*name2*:

*name3*:

---

**PARTI – THEORY:** For the following, you might answer only YES/NO (or abstain), or you can add short arguments (at most two lines per question). If you are not sure, it is better to abstain.

1. The following generative models rely on the specific objective, ...

**yes** **~~no~~** Variational autoencoders aim for an maximization of the likelihood of data generated by the discriminator

**yes** **~~no~~** GANs minimize the discrimination loss and generation probability of observed data

**yes** **~~no~~** Flow models minimize the data log likelihood

**~~yes~~** **no** Diffusion models directly optimize the data likelihood

2. For variational autoencoders, the following is true:

**~~yes~~** **no** The encoder part maps a data points to a distribution in the latent space, represented by mean and variance.

**~~yes~~** **no** The cost function of a VAE is the reconstruction error enhanced by a regularization part based on KL divergence.

**~~yes~~** **no** The reparametrization trick of VAEs enables backpropagation training, since then sampling is restricted to a neuron without adjustable parameters.

**~~yes~~** **no** VAEs can be used together with convolutions or resnet architectures

3. The following holds for GANs and its variants:

☒ **yes** ☐ **no** Wasserstein GAN directly optimizes the difference of distributions, using a variation the JS distance of distributions.

☒ **yes** ☐ **no** Distance measures for distributions are smooth, such that WGAN is very robust.

☒ **yes** ☐ **no** WGAN CT and WGAN aim for a Lipschitz continuous discriminator.

☒ **yes** ☐ **no** Conditional GANs use specific labels for conditioning the latent space output

4. The following tasks can be addressed with generative models:

☒ **yes** ☐ **no** Cycle GAN for domain adaptation.

☒ **yes** ☐ **no** Image manipulation by arithmetics in GANs latent space

☐ **yes** ☒ **no** BigGAN provides a text generation module

☒ **yes** ☐ **no** Diffusion models are part of text to image geneators

5. The following is true:

☐ **yes** ☒ **no** Flow-based methods can be used with U-net architectures.

☐ **yes** ☒ **no** In the equilibrium of GANs, the discriminator reaches 100% accuracy.

☒ **yes** ☐ **no** Diffusion maps train a mapping which maps inputs to Gaussian noise

☒ **yes** ☐ **no** InfoGAN imposed structural elements on the latent space

**PARTII – PRACTICE:** You can use code and models which are publicly available, please clearly reference such sources. It might be a good idea to start with the examples given in the practical part of the lecture (available at `https://jgoepfert.pages.ub.uni-bielefeld.de/talk-deep-learning` ). Please give a link to your code, and please describe the experiments and results of your approach in a pdf which is well structured (e.g. modeling/training parameters/training/results/interpretation, use itemize, keywords are fine) and enables reproducability as well as easy access to your main results. Please use at most one page for both practical parts together including graphs and images.

1. Take the Fashion MNIST data set and train a variational autoencoder (VAE) on these data. Provide some insight in how good the data are represented, e.g. by reporting the reconstruction error and displaying typical results of generated images.

2. Investigate the latent space of the VAE: display how data are represented by means of a nonlinear dimensionality reduction (such as UMAP or tSNE) and investigate whether there are clusters corresponding to the classes. Investigate how generated data look like if one moves within the latent space from one point to another one from a different class on a straight line.

## Sheet 2

**Modeling**

Again we built a very simple model, the variational autoencoder (VAE). The input data is sampled from a parametrized distribution, the encoder and decoder are trained jointly such that the output minimizes the reconstruction error (Kullback–Leibler divergence). As optimizer we again used Adam. The corresponding dataset is the Fashion-MNIST.

To further investigate how the data is represented we used UMAP in a machine learning environment and show with a clustering approach to learn how the data is represented and recognized.

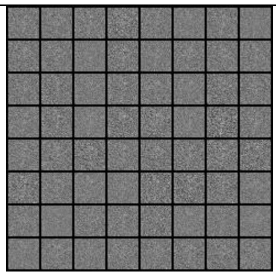
**Training Parameters**

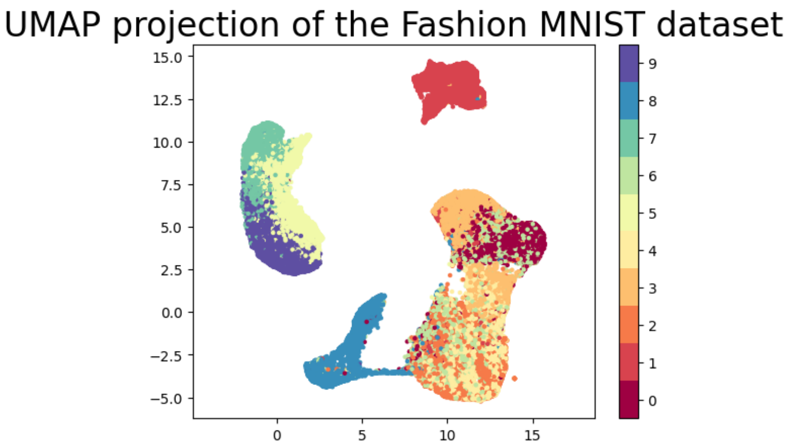
We used a learning rate of 0.001 and 10 epochs and a weight decay of 0.00001.

**Training**

The training took about 8 minutes with a batch size of 64. After training the VAE we trained the UMAP for the dimension reduction investigation.

**Results**

| **Original picture** | **Starting with noise** | **After first epoch** | **Result** |
|---|---|---|---|
|  |  |  |  |




**Interpretation**

As we can see we were able to reconstruct the picture quite good after 10 epochs. After the first epoch we have troubles to recognize the pictures, but the result is readable. We can also see how the dimension reduction works and how the data is represented and can be expressed. We can also see clear clusters representing different clothes in the latent space, some more close to each other and some clearly different from others in terms of the representation. To further improve the result, more epochs should be processed.

Code: https://github.com/itzeck/deep_learning