Exercise Sheet 2

- ▷ Sheets should be submitted in teams of 3-4 students.
- ▷ Accepted format is a single PDF file. Photos of (readable) handwritten text and math are accepted inside the PDF. Please put the full names and student ID numbers of all team members on the title page of your submission.
- ▷ Submissions are made via a Sciebo dropoff link: `https://fz-juelich.sciebo.de/s/GwXodd3mA6Vh202`.
- ▷ Submissions have to follow the naming scheme *sheet01_USER1_USER2_USER3_USER4.pdf* using your university username (”Benutzername” in ILIAS/IDM). If a second version is uploaded use the ending *_v2.pdf* for the file. Only the latest version will be rated.
- ▷ At least 50 points are required to pass this sheet.

1. **Variational Autoencoder (VAE) (25 Points)**

   (a) (5 Points) Explain why a standard Autoencoder cannot directly be used to generate new samples.

   (b) (10 Points) Explain how the KL-divergence term $\mathrm{D_{KL}}(q(z|x) \,\|\, p(z))$ in the loss regularizes the latent space.

   (c) (5 Points) We assume a standard Gaussian prior $p(z)$ to compare against the learned latent encoder distribution $q(z|x)$. Why does the encoder not always predict constant $\mu(x) = 0$ and $\Sigma(x) = I$ for all inputs?

   (d) (5 Points) VAEs with a Gaussian likelihood $p(x|z)$ often produce blurry images compared to Generative Adversarial Networks (GANs). Explain why this happens.

2. **Generative Adversarial Network (GAN) (50 Points)**

   Generative Adversarial Networks are used for generative modeling, where a generator G and a discriminator D compete in a minimax game. The discriminator $D(x)$ attempts to distinguish between real samples drawn from the data distribution $p_{\text{data}}(x)$ and fake samples generated by G, which follow the distribution $p_{\text{gen}}(x)$. From lecture 5, we know that the optimal discriminator is given by:

$$D^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) \; + \; p_{\text{gen}}(x)}$$

   Using this result, we explore the properties and implications of the following GAN loss functions:

$$\mathcal{L}_D = -\mathbb{E}_{x\sim p_{\text{data}}}[\log D^*(x)] - \mathbb{E}_{x\sim p_{\text{gen}}}[\log(1 - D^*(x))] \tag{1}$$

$$\mathcal{L}_G = \mathbb{E}_{x\sim p_{\text{gen}}}[\log(1 - D^*(x))] \tag{2}$$

(a) (25 Points) Compute the value of the discriminator loss $\mathcal{L}_D$ and the generator loss $\mathcal{L}_G$ at convergence when p_{data} and $p_{\text{generator}}$ do not overlap (refers to the regions in the data space where both distributions assign non-zero probability). Explain what this means for the training process.

(b) (25 Points) Compute the value of the discriminator loss $\mathcal{L}_D$ and the generator loss $\mathcal{L}_G$ at convergence when $p_{\text{data}} = p_{\text{generator}}$ (100% overlap) over the entire training set. Explain what this means for the training process.

3. **GAN Training (25 Points)**

(a) (5 Points) Define mode collapse in the context of GANs.

(b) (10 Points) Explain why mode collapse happens during GAN training. Include in your explanation the roles of both the generator and the discriminator.

(c) (10 Points) Describe and explain a specific method that can be used to address mode collapse.