Exercises for the course
**Deep Learning 2**
Summer Semester 2024

Machine Learning Group
Faculty IV – Electrical Engineering and Computer Science
Technische Universität Berlin

# Exercise Sheet 1

### Exercise 1: Contrastive Loss (20 P)

Given the SimCLR loss from the lecture for all views $i, j$ from the same samples in a minibatch ($MB$).

$$\mathcal{L} = -\frac{1}{N} \sum_{i,j \in MB} \log \frac{\exp\left(\operatorname{sim}\left(\mathbf{z}_i, \mathbf{z}_j\right)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp\left(\operatorname{sim}\left(\mathbf{z}_i, \mathbf{z}_k\right)/\tau\right)} \tag{1}$$

with $\operatorname{sim}(u, v) = \frac{u^T v}{\|u\|\|v\|}$ being the cosine similarity $\tau$ a scalar and $N$ the number of samples.

a) Rewrite the loss explicitly into the following form:

$$\tau \mathcal{L} = \mathcal{L}_a + \mathcal{L}_d$$

with $\mathcal{L}_a = -\frac{1}{N} \sum_{i,j \in MB} \operatorname{sim}(\mathbf{z}_i, \mathbf{z}_j)$.

What is the purpose of $\mathcal{L}_a$ and $\mathcal{L}_d$ in the loss?

b) How does the parameter $\tau$ influences the distance between representations?

### Exercise 2: Lecture Questions (20 P)

a) What is a pretext task? Give four examples for pretext tasks.

b) What is a representation collapse and how is it prevented in SimCLR?

c) Given an image/text model with image encoder $f$ and text encoder $g$ which both produce a representation $z \in \mathbb{R}^d$, we want to perform zero-shot classification. Given text labels $t_1, \ldots, t_k$ that describe $k$ classes and an image $x$, how do you compute the predicted class $c$?

d) Name two other applications for representations from a pretext task other than using them for a classification downstream task.

### Exercise 3: Programming (60 P)

Download the programming files on ISIS and follow the instructions.