

## 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 ( $MB$ ).

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

with  $\text{sim}(u, v) = \frac{u^T v}{\|u\| \|v\|}$  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  $\mathcal{L}_a = -\frac{1}{N} \sum_{i,j \in MB} \text{sim}(\mathbf{z}_i, \mathbf{z}_j)$ .

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

$$\mathcal{L}_d = \frac{\tau}{N} \sum_i \log \sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k) / \tau) \quad (2)$$

$\mathcal{L}_a$  encourages representations of augmented views to be consistent (alignment) while  $\mathcal{L}_d$  encourages representations (or a random subset of them) to match a prior distribution (of high entropy).  $\mathcal{L}_d$  prevents a representation collapse and helps to distribute the representations across the whole space (uniformity).

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

$\tau$  is the temperature parameter that is usually used to calibrate a softmax distribution. In contrast to regular classification, the logits are in this case determined by the cosine similarities and therefore bounded in the range  $[-1, 1]$ . Using a low temperature makes the distribution to have less entropy. Therefore, the differences in the probabilities (distances/similarities) after the softmax also become larger. Using a larger temperature leads to a more uniform distribution of the similarities/distances after the softmax.

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

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

A pretext task is an auxiliary task that has an objective that does not require explicit labels. The pretext task is usually performed on a property that is inherent in the dataset itself. Examples:

- Masking and Reconstruction
- Contrastive Learning Objective
- Color prediction (grayscale to color)
- Jigsaw Puzzle task
- Rotation prediction
- Super-resolution

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

A representation collapse means that the encoder maps all representations to a single point (e.g.  $0 \in \mathbb{R}^d$ ). SimCLR uses negative samples which are used in the contrastive loss to minimize similarity (push representations apart) in order to avoid a representation collapse.

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  $t_1, \dots, t_k$  that describe  $k$  classes and an image  $x$ , how do you compute the predicted class  $c$ ?

$$c = \operatorname{argmax}_{i=1, \dots, k} \frac{f(x)^T \cdot g(t_i)}{\|f(x)\| \|g(t_i)\|}$$

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

- Clustering
- Semantic Search/retrieval tasks
- Anomaly Detection
- Finding independent components
- Matching them to other modalities

### Exercise 3: Programming (60 P)

Download the programming files on ISIS and follow the instructions.