Below is an **English translation** of the **SimCLR** exercises, including both **computation** and **application** questions. These questions cover the InfoNCE loss formula, numerical calculations, and practical scenarios for SimCLR-based self-supervised learning.

I. Computation Questions

Question 1: Batch Similarity Matrix Calculation

Suppose we have **batch size = 3**, and each image is augmented twice, so there are **(2N = 6)** total embeddings ({z_1, z_2, z_3, z_4, z_5, z_6}).

- The positive sample pairs are ((z_1, z_2)), ((z_3, z_4)), ((z_5, z_6)).
- All other ((z_i, z_j)) with (i \neq j) are considered negative pairs.

We have a **similarity matrix** (S), where (S_{ij} = \mathrm{sim}(z_i, z_j)) (cosine similarity), as follows:

[
S =
\begin{pmatrix}

- & 0.95 & 0.10 & 0.05 & 0.02 & 0.00 \
  0.95 & - & 0.12 & 0.07 & 0.03 & 0.01 \
  0.10 & 0.12 & - & 0.90 & 0.06 & 0.04 \
  0.05 & 0.07 & 0.90 & - & 0.08 & 0.02 \
  0.02 & 0.03 & 0.06 & 0.08 & - & 0.93 \
  0.00 & 0.01 & 0.04 & 0.02 & 0.93 & -
  \end{pmatrix}
  ]

where "-" indicates self-similarity (not needed). Suppose the temperature (\tau = 0.1).

1. **Compute** the InfoNCE loss for the positive pair ((z_1, z_2)), i.e.
   [
   L*{1,2} = -\log \frac{\exp(\mathrm{sim}(z_1, z_2) / \tau)}{\sum*{k=1}^{6} \mathbf{1}_{k \neq 1} \exp(\mathrm{sim}(z_1, z_k)/\tau)}.
   ]
   - Write out the numerator and denominator numerically, and find the resulting (-\log(\dots)) value.
2. Similarly, **compute** the loss (L_{3,4}) for the positive pair ((z_3, z_4)).
3. **Explain** how to combine these losses to get the total batch loss (L) if we have 3 positive pairs in total.

> **Hint**: Each positive pair ((i,j)) yields an individual loss (L_{i,j}), and typically we average them over the batch.

Question 2: Effect of the Temperature Parameter

1. If we change the temperature (\tau) from 0.1 to 0.5 in the above example, how would the exponential terms in the denominator behave? What intuitive effect would this have on the final loss values?

2. In practice, **raising the temperature** often makes the similarity differences “less pronounced.” Briefly explain why this happens using the formula.

Question 3: Impact of a Small Batch Size

1. Suppose we can only use **batch size = 2** (i.e., (N=1)). Then we effectively have 1 positive pair and **0 negative samples**.
    - In the InfoNCE formula, how would the denominator appear in this scenario? Could we still compute the loss properly?
    - What practical problem arises in training?
2. Summarize why SimCLR typically needs a large batch size (e.g., 256~2048) to achieve good performance.

II. Application Questions

Question 4: Applying SimCLR on a Small Dataset

You have a small dataset (only 10,000 images) and plan to use **SimCLR** for self-supervised pretraining, followed by a linear probe for classification.

1. How would you design **data augmentation** to ensure contrastive learning works effectively? Give at least 3 augmentation methods and explain their significance.
2. If GPU memory is limited and you cannot use a large batch size, what strategies could you employ to approximate “plentiful negative samples”? (e.g., memory bank, queue, or MoCo-like approaches)

Question 5: Visualizing and Interpreting SimCLR Representations

After training SimCLR, you want to **visualize** the learned features to confirm that it distinguishes different images appropriately:

1. How can you inspect the **distance** in the representation space between two different augmentations of the same image?
2. If you notice that some embeddings of different images are extremely close, what might cause this?
3. If you apply **t-SNE** or **UMAP** on the learned embeddings, what phenomenon would you look for to verify that SimCLR indeed clusters similar images closer?

Question 6: Comparing SimCLR to Supervised Learning

1. Suppose you train a **supervised ResNet** and a **SimCLR ResNet (self-supervised)** on ImageNet, then fine-tune them on a downstream task (e.g., classification or detection) with only a small fraction of labeled data.
    - How might their performance differ?
    - If the labeled subset is extremely small (e.g., 1% of ImageNet), which method has an advantage, and why?

2. Can the representations learned by SimCLR be used **directly** for classification without fine-tuning? If so, how? Would the accuracy approach that of a fully supervised model?

Summary

- These **computation questions** and **application questions** address the **InfoNCE loss formula**, the role of the **temperature** parameter, **batch size** considerations, **small dataset scenarios**, and how to **visualize** or **evaluate** the learned representations.
- By working through these problems, you can deepen your understanding of **SimCLR**'s mathematical foundation and practical deployment, including why large batch sizes and strong data augmentation are critical for success.