EECS 182	Deep Neural Networks	Homework 9
Spring 2023	Anant Sahai	

This homework is due on Friday, April 14, 2022, at 10:59PM.

1. Read a Research Paper: FaceNet

In class, you have learnt how self-supervised learning can be used to learn useful representations from large datasets without labels (e.g., learning features from ImageNet). While these features may inherently pick out some notion of "similarity" between different images in the dataset, they are not incentivized to *cluster* different data points based on any interesting similarity measure.

The paper "FaceNet: A Unified Embedding for Face Recognition and Clustering" explores how we can view task of face recognition through the lens of self-supervised (or to be more accurate, *slightly* supervised) learning.

Read the paper and answer the questions below.

(a) **What are the two neural network architectures considered by the authors?**

(b) **Briefly describe the *triplet loss* and how it differs from a typical supervised learning objective.**

(c) **What is the challenge with generating all possible triplets? Briefly describe how the authors address this challenge.**

(d) **How many parameters and floating point operations (FLOPS) do the authors use for their neural network? How does this compare to a ResNet-50?**

(e) Briefly explain what the authors mean by *semi-hard* negatives. What are harmonic embeddings?

(f) How does the performance vary with embedding dimensionality?

(g) How does the performance vary with increasing amounts of training data?

(h) Briefly share your favorite *emergent* property/result of the learned behavior with a triplet loss from the paper.

(i) **Which approach taken by the authors interested you the most? Why?** (≈ 100 words)

2. Masked Auto-Encoding

Please follow the instructions in this notebook. You will implement Vision Transformer and Masked Autoencoder in PyTorch. Once you finished with the notebook, download `hw9_submission.zip` and submit it to "Homework 9 (Code) (MAE)" in Gradescope.

3. Coding Question: Summarization (Part I)

Please follow the instructions in this notebook. You will implement a Transformer using fundamental building blocks in PyTorch. You'll apply the Transformer encoder-decoder model to a sequence-to-sequence NLP task: document summarization. Refer to the Attention is All You Need paper for details on the model architecture. Once you finished with the notebook,

- Download `submission_log.json` and submit it to "Homework 9 (Code) (Summarization)" in Gradescope.
- Answer the following questions in your submission of the written assignment:

(a) **Please submit the screenshots of the training loss and the validation loss displayed on Tensorboard.**

4. Coding Question: Visualizing Attention

Please run the cells in the Visualizing_BERT.ipynb notebook, then answer the questions below.

(a) Attention in GPT: Run part a of the notebook and generate the corresponding visualizations

   i. **What similarities and differences do you notice in the visualizations between the examples in this part?** Explore the queries, keys, and values to identify any interesting patterns associated with the attention mechanism.

   ii. **How does attention differ between the different layers of the GPT model? Do you notice that the tokens are attending to different tokens as we go through the layers of the network?**

(b) BERT pays attention: Run part b of the notebook and generate the corresponding visualizations.

   i. Look at different layers of the BERT model in the visualizations of part (b) and identify different patterns associated with the attention mechanism. Explore the queries, keys, and values to further inform your answer. **For instance, do you notice that any particular type of tokens are attended to at a given timestep?**

   ii. **Do you spot any differences between how attention works in GPT vs. BERT? Think about how the model architectures are different.**

   iii. For the example with syntactically similar but definitionally different sentences, look through the different layers of the two BERT networks associated with sentence a and sentence b, and take a look at the queries, keys, and values associated with the different tokens. **Do you notice any differences in the embeddings learned for the two sentences that are essentially identical in structure but different in meaning?**

   iv. **For the pre-training related examples, do you notice BERT's bi-directionality in play? Do you think pre-training the BERT helped it learn better representations?**

(c) BERT has multiple heads!: Run part c of the notebook and generate the corresponding visualizations.

   i. **Do you notice different features being learned throughout the different attention heads of BERT? Why do you think this might be?**

   ii. **Can you identify any of the different features that the different attention heads are focusing on?**

(d) Visualizing untrained attention weights

   i. **What differences do you notice in the attention patterns between the randomly initialized and trained BERT models?**

   ii. **What are some words or tokens that you would expect strong attention between? What might you guess about the gradients of this attention head for those words?**

(e) **Were you able to identify interesting patterns in the visualizations?** If yes, please share some examples (describe in text or paste a screenshot). If not, feel free to use this space for your frustrations.

5. Homework Process and Study Group

Citing sources and collaborators are an important part of life, including being a student!
We also want to understand what resources you find helpful and how much time homework is taking, so we can change things in the future if possible.

(a) **What sources (if any) did you use as you worked through the homework?**

(b) **If you worked with someone on this homework, who did you work with?**
List names and student ID's. (In case of homework party, you can also just describe the group.)

(c) **Roughly how many total hours did you work on this homework? Write it down here where you'll need to remember it for the self-grade form.**

Contributors:

- Dhruv Shah.

- CS182 Staff from past semesters.

- Jake Austin.

- Olivia Watkins.

- Linyuan Gong.

- Sheng Shen.

- Hao Liu.

- Allie Gu.

- Anant Sahai.

- Shivam Singhal.

- Kevin Li.

- Bryan Wu.