COMP4650/COMP6490 Document Analysis
2025 Semester 2

Computing Lab 5

Q1: Use of Pre-trained Models

(This is a theory question that does not require coding.)

In class, we have seen several pre-trained models: BERT, ELMo, and GPT models. They all can provide contextual word embeddings. In addition, BERT and GPT models have often been fine-tuned for downstream tasks.

Explain the differences between (1) using contextual word embeddings as input feature representations to train a model for a downstream task (e.g., sentiment classification), and (2) fine-tuning a pre-trained BERT or GPT model for a downstream task.

Q2: Self-supervised Objectives

(This is a theory question that does not require coding.)

(a) GPT and BERT are trained on different supervision tasks. Explain the different tasks used for training these models.

(b) Explain what the special tokens `[CLS]` and `[SEP]` are used for in the BERT model.

Q3: Practical Exercise: Attention

In this exercise you will fine-tune a pre-trained transformer model using the Hugging Face transformers[1] library. The dataset is the IMDb movie review data where the task is to classify a review as either positive (if the reviewer liked the movie) or negative (if the reviewer did not like the movie). The input is the text of the review and the output is a binary label either 0 (negative) or 1 (positive).

You will need to work through the notebook `lab5-finetune_transformer.ipynb` and complete the practical exercise in it.

[1] `https://github.com/huggingface/transformers`