EECS 182 Deep Neural Networks
Spring 2023 Anant Sahai

Review: Fine-tuning

1. Finetuning

(a) If you pretrain using a masked autoencoder, when you finetune the autoencoder encoder for downstream tasks, **do you still mask the inputs?**

(b) Let's say you want to train an LSTM encoder/decoder model on translating from English to Spanish using a paired English/Spanish training set. You also have a much larger corpus of unpaired English sentences. **Describe one way to pretrain the LSTM encoder using the unpaired data.**

(c) For each of the following finetuning problems, **describe whether you should prefer to use (a) feature extraction (also called linear probing), (b) full finetuning, (c) hard prompting, or (d) soft prompting.**

(i) You are using a 175B parameter language model for a question-answering task. You have a dataset of 100k examples.

(ii) You are using a 90B parameter language model for a spam classification task. You have a task description but no training data.

(iii) You are using a 1B parameter conv net pretrained on ImageNet for wildlife classification task. You have 100 training examples.

(iv) You are using a 1B parameter vision transformer pretrained on ImageNet for an X-Ray fracture localization task. You have 100M training examples.

2. Prompting

(a) Typically, when you create a soft prompt for use with a GPT model, you prepend the prompt to the left of the input. **Could you also get good performance by appending it to the right**?

(b) Let's say you would like to use hard prompting with a large GPT model. You have a dataset of 10 thousand training examples for your downstream task. Would it be a good idea to include all of these examples (except for a held-out validation set) in your prompt?

3. Finetuning Pretrained NLP Models (From Discussion 11)

In this problem, we will compare finetuning strategies for three popular architectures for NLP.

(a) **BERT** - encoder-only model
(b) **T5** - encoder-decoder model
(c) **GPT** - decoder-only model

(a) For each of the three models, **state the objective used for pretraining.**

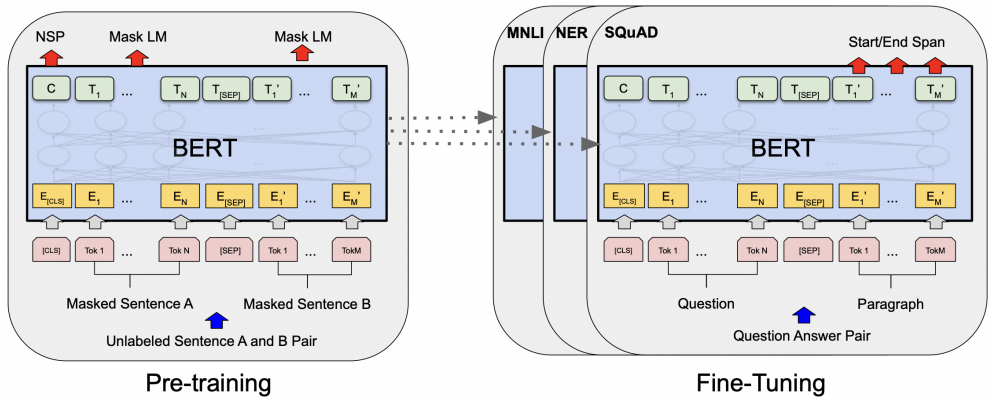

Figure 1: Overall pre-training and fine-tuning procedures for BERT.

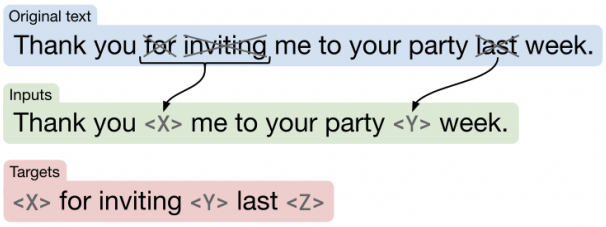

Figure 2: T5 Training procedure

(b) Consider the MNLI (Multi-Genre Natural Language Inference Corpus) task. It provides a passage and a hypothesis, and you must state whether the hypothesis is an entailment, contradition, or neutral.

> **EXAMPLE:**
> **Passage**: At the other end of Pennsylvania Avenue, people began to line up for a White House tour.
> **Hypothesis**: People formed a line at the end of Pennsylvania Avenue.
> **Classification**: entailment

(i) **With each of the 3 models, state whether it is possible to use the model for this task with no finetuning or additional parameters.** If so, state how.

(ii) With each of the 3 models, **state how you would use the model for this task if you were able to add additional parameters and/or finetune existing parameters.**

(c) **Compare and contrast the ways we use pretrained representations in BERT to the way we use pretrained autoencoder representations.**

