Tutorial 1: Summarisation, Evaluation and PEFT

Shay Cohen, Lexi Birch, Alessandro Suglia

Semester 2, 2025-26

Instructions

- Attempt all exercises before the tutorial session.

- Discuss your reasoning, not just the final answer.

- Collaboration is encouraged, but think of your own solutions first.

Exercise 1. You are given the following three summaries:

> **Reference summary**: The film was long and interesting.
> **Summary from System A**: The film was long and boring.
> **Summary from System B**: Interesting long movie.

1. Calculate ROUGE-1 and ROUGE-2 for the summaries from System A and System B with respect to the gold summary. Base yourself on the formula given in class.

2. Which one provides a higher score for ROUGE-1 and ROUGE-2? Is that what you would expect?

3. What are some of the issues causing any disparity you see? What is a way to evaluate the systems more accurately?

Exercise 2. In class, we learned about low-rank composition for parameter-efficient finetuning, with a random dense matrix P of shape $D \times d$, with D being the dimension of the original model and d being lower dimension. With such low-rank composition, we update parameters of the form $\theta_0 + P\theta'$, where θ_0 is the base model and θ' is a set of parameters that are tuned. Note that P is fixed and selected once.

1. Joseph decided he wants to make the model even more efficient, and enforce a sparsity constraint on the P matrix. Explain what are some of the issues that could arise with that.

2. What could you recommend to a user of low-rank composition to further improve performance in a way that is additive in complexity, and even parallelisable? (a "meta" algorithm.)