Project 1: Hyperparameter Tuning

This is a single-person project. It is fine to discuss your process with other students but make your own decisions and write your own code and report.

You are an ML engineer at Newcrosoft, a newly founded AI startup. Your ML researcher Andreas Carfurt has given you a notebook with the training code. It finetunes a model (distilbert-base-uncased) on paraphrase detection (MRPC). You are asked to optimize its accuracy on the validation set with careful hyperparameter tuning. Use Google Colab. Track your experiments with a tool of your choice (figure out yourself how you can use it from a Google Colab notebook).

Deliverables

At the end you should submit a report detailing your experiments for the tasks below. Also describe how you tracked your experiments, including at least one screenshot from your experiment tracking tool. The report must be at most 2 A4 pages, in PDF format. Choose a reasonable font and size. At the end, include a reflection of what went well and what you would improve next time. No code submission.

Tasks

Week 1: Manually define a list of max. 10 hyperparameters that you want to investigate in week 1. You cannot change the model or the number of epochs (=3), but you can change the optimizer, its parameters, learning rate & schedule, weight decay, or effective batch size (see this Trainer flag). Don't use the learning rate finder. With at most 20 manual training runs, select the 3 hyperparameters that you deem most promising to tune in detail in week 2.

Week 2: With at most 20 training runs, **manually** find the best setting for these 3 hyperparameters. Do not use automatic hyperparameter optimization.

Week 3: [to be announced]

Bonus task: Describe in detail the best setting you found (all hyperparameter settings, the number of training steps, everything needed to reach the same result), its score, and a plot of how your different hyperparameter settings compare on the metric.

Relevant reading

https://github.com/google-research/tuning_playbook

Grading

The following criteria will determine the grade:

- Structured and organized approach
- Clear description of experiments and conclusions
- Reflection on what went well and what you would improve next time
- Experiment tracking included
- Report adheres to specifications

Hints

- Connect your Colab notebook to a GPU (top right) for faster training and evaluation.

- If you're unsure about something in the project description, think about the possible options, pick the one that seems most appropriate to you, and mention it in an "Assumptions" section in your report.