This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset LaTeX solutions.

1.b (ii)

The medium size model starts with higher accuracy but when the number of supports increases the two models are about the same accuracy.

1.c

11200000 * 4 bytes = 44800000 bytes = 44.8 Megabytes

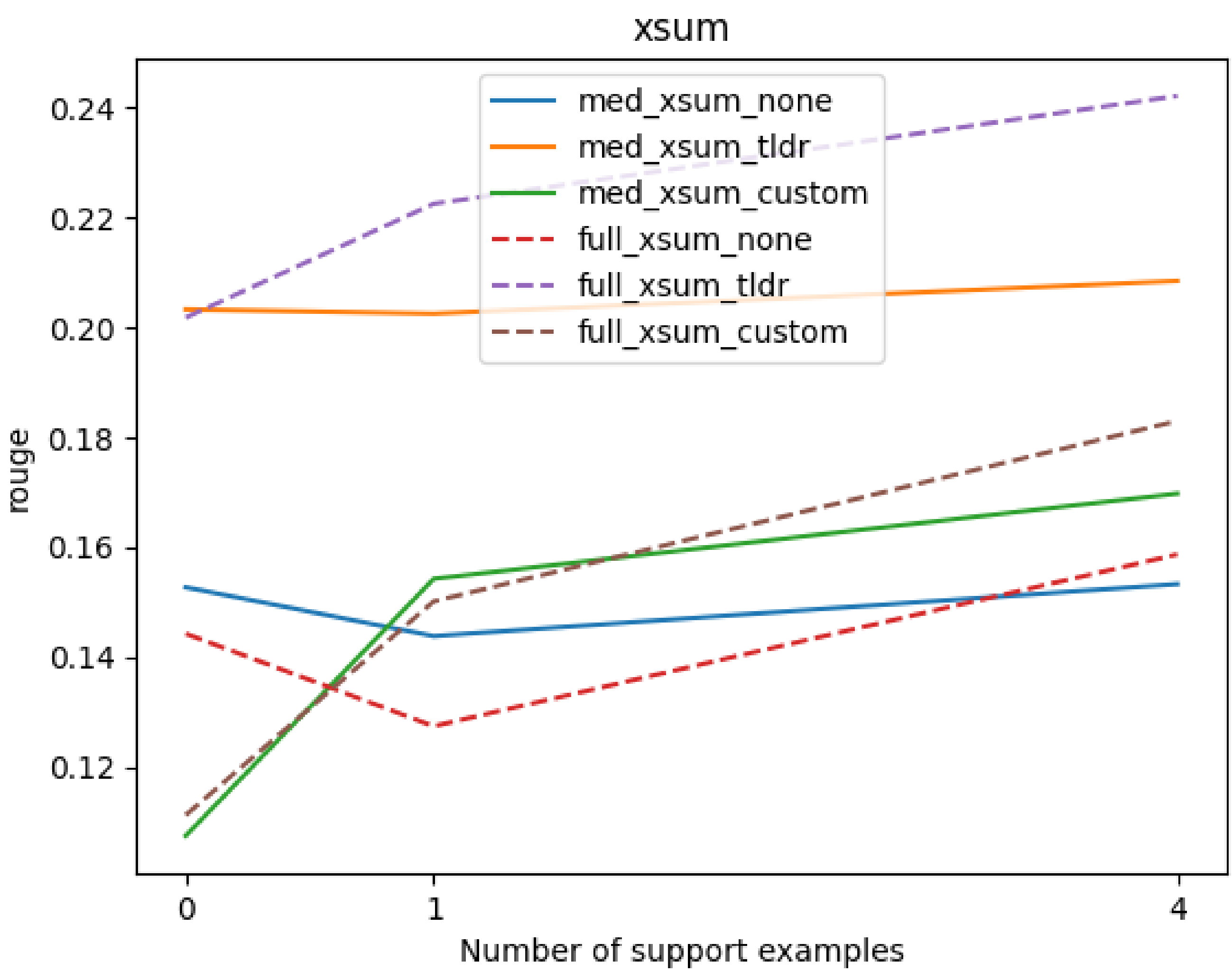
1.d

540000000000 * 4 bytes = 2.16 * 10^{12} = 2.16 Terabytes

2.b (ii)

With a small number of few shot examples the medium and full models perform similarly. As the number of shots increase the full model does slightly better but the medium model does worse.

2.c (ii)



The TL;DR: prompt does better than no prompt formatting.

My custom prompt format was to make a symbol !¿ that represents summarize with the idea that it would trigger the summarize behavior. Overall the custom prompt format performed worse than the TL;DR: format.

In the zero shot context TL;DR does better than the other two formats. The custom format does the worst and this could be because the custom symbol doesn't appear in the dataset.

In the one shot context TL;DR again does the best but the custom format does better than the no prompt format.

In the 4 shot context TL;DR stays the best and both the custom and no prompt formats do better as well.

3.b

If the full matrix is d1 x d2 then the savings is (d1 - p) * (d2 - p) parameters.

The greatest savings are when the p is rank 1 reducing the A and B matricies to vectors of size d1 x 1 and d2 x 1.

3.d (ii)

The LoRAs perform on par or better than finetuning the first and last layers.

Comparing the LoRAs and finetuning the middle layers, as the number of support examples gets larger the LoRAs perform just as well as finetuning the middle layers.

4.a

In-Context learning performs better when there are a small amount of support examples.

Fine-tuning performs better with more examples.

This show that in-context learning is limited to rare examples where getting more data is not feasible.

4.b

The value from the evals is 0.664.

In-context learning has a higher standard deviation.

4.c