Zakaria ECHCHAIR
Zakaria.ECHCHAIR@polytechnique.edu

Lab session # 3
ALTEGRAD 2022

11/22/22

1 Question 1

Embedding layer : $n_{embd} = n_{hid} * n_{tokens} = 512 * 32000$

Positional encoding : $n_{Pos} = n_{hid} * n_{sentlength} = 512 * 1$

Self-attention heads : $n_{att} = 4 * (n_{hid} * n_{hid}) = 4 * (512 * 512)$

Feed forward : $n_{FFN} = n_{hid} * n_{hid} = 512 * 512$

Encoder : $n_{encoder} = [4 * n_{att+2*n_{FFN}}] = 4 * (512 * 512) + 2 * (512 * 512)$

Total numbers of parameters :

$n_{embd} + n_{Pos} + 4 * n_{encoder} + = 512 * 32000 + 512 * 1 + 4 * [4 * (512 * 512) + 2 * (512 * 512)] = 20675968$

2 Question 2

- We can see that the tuning of Hugging Face has more accuracy than that of Fairseq.
- Hagging face gives an accuracy of 80, while fairseq 69.7.
- Hugging face use the raw data in json form without making any transformation on the data. Unlike Fairseq, you do need to perform tokenization and binarization.
- Hugging face is faster than Fairseq.
- Depending on the usage situation. Fairseq is adaptable enough for customisation if we are researchers, however Hugging Face would be preferable if we are working on a real application and contemplating deployment.

References

https://huggingface.co/spaces/sriramelango/Social_Classification_Public/blob/main/fairseq/examples/translation/README.md

https://fairseq.readthedocs.io/en/latest/command_line_tools.html

https://factored.ai/transformer-based-language-models/#:~:text=This%20implies%20that%20the%20parameter,million%20parameters%20for%20RoBERTa%20large.

https://towardsdatascience.com/choose-the-right-transformer-framework-for-you-b7c51737d45