Quick Update on NLP Transformer

Gobind Puniani
MBG 2021
05/26/21

Recap

- Previously tried integrating activation functions into AWD-LSTM model from FastAI
    - Working on Google Colab notebooks adapted from FastAI:
        - Overview of AWD-LSTM: https://colab.research.google.com/drive/1sxuVp6KifSXEvTKWvcKoFEN0ZfiCjCEY?usp=sharing
        - Building an NLP architecture from scratch: https://colab.research.google.com/drive/1wBfBAacZtBcjVMecv1yqq6hAZ9FFxDkC?usp=sharing
- Could not run Jupyter notebooks on TLALOC
- Could not run FastAI Python scripts on TLALOC

Transformer

- Transformer architecture is basis of NLP state-of-the-art
  - GPT-3 uses Transformer architecture (Generative Pre-trained Transformer)
  - Transformer supplants LSTM: https://www.youtube.com/watch?v=S27pHKBEp30
  - Human attention mechanism inspired Transformer architecture: https://thegradient.pub/attention-in-human-brain-and-its-applications-in-ml/
- Only one instance of activation function used in implementation
  - Good for testing purposes (less variation)

HuggingFace

- Previously only had access to HuggingFace GitHub profile
    - Base code for architecture was insufficient for testing
- Found HuggingFace documentation: https://huggingface.co/transformers/index.html
    - Can now build own model and run after pre-processing data

Next Steps

- Use HuggingFace documentation to build Transformer model with our activation functions
- LaTeX tables are still in progress
    - Difficulty in formatting different numbers of columns for various rows