Homework #8

In this assignment, we will revisit our English/Spanish sentence classifier but using Transformers. You may use any online resources but must cite your sources and indicate clearly what portions of your code have been copied and modified from elsewhere. You may work individually or with a partner on this assignment. At the top of your notebook, please clearly indicate the names of both team members and briefly summarize who did what.

Submit your assignment as a single jupyter notebook on Sakai. To speed up training, you may want to run your jupyter notebook in Google Colab with a GPU. Note: The datasets provided below are very large, and you don't need to train on everything! In fact, as you develop your code, I would recommend using a tiny subset of data to iterate quickly, and wait until your code is debugged to start training on larger subsets of data. Since our goal in this assignment is learning, it is perfectly acceptable to only train on a small subset (~1%) of the data.

Part 1: Character-level Transformer Model (35 points)

In the first part of the assignment you will do the following:

- Prepare the data (5 points). Get two large text files: one English file (WikiText-103, 181MB) and one Spanish file (e.g. Spanish text corpus, 155MB). Convert to lowercase and remove all punctuation except "." so the data only contains alphabet characters, whitespace, and periods. Determine a set of unique characters and map all characters to integers. Split the data into train & validation sets, and split each into individual sentences. You may reuse your work from last week's assignment.
- Train 1-layer model (15 Points). Define a Transformer model containing 1 Transformer encoder layer followed by an output linear layer. Your model should classify a fixed-length sequence of characters as English or Spanish. Show your training & validation loss curves, along with your validation classification accuracy. Compare your results to the LSTM model from last week and comment.
- Experimentation (10 points). Experiment with one aspect of the model: the number of Transformer layers, encoder vs decoder layers, the size of the hidden layer, etc. Train the corresponding models, compare their performance, and provide plots to demonstrate the effect of the hyperparameter of interest.
- Intuition (5 points). Show the output of your model for several specific sentences. Pick inputs that demonstrate the behavior of the system, and try to figure out what things the model is focusing on. Explain your intuition about what the model is doing.

Part 2: Linear Probe on Pretrained Language Model (35 points)

In the second part of the assignment you will do the following:

- Extract features (15 points). Find a pretrained BERT language model and learn how to extract features from the output of the last Transformer layer. Include a code example in your notebook for how to extract features from an input sentence.
- Train model (20 points). Use a suitably chosen output from the Transformer layers as the feature representation for each sentence, train a linear probe on (a subset of) the English/Spanish training data, and test your finetuned model on the English/Spanish test data. Show your training/validation loss curves, report your validation classification accuracy, and compare the results to the model in part 1.

Part 3: Prompting a Foundation Model (20 points)

- Complete the ChatGPT Prompt Engineering for Developers Short Course.
- Design prompt (5 points). Experiment with different prompts to find a suitable prompt for the English/Spanish classification task that we are interested in. Provide examples from your prompt experiments and indicate the final prompt you settled on.
- Measure performance (15 points). Use your chosen prompt to classify (a subset of) the validation data in your English/Spanish classification task. Report your accuracy and compare your results to the models in parts 1 & 2.

An additional 10 points will be graded for the organization and clarity of your notebook. Your notebook should read like a tutorial and be understandable to others!