Ethan Bodell

CMPS 451

Final Project Transformer Model Documentation

**Context:** Due to being unable to consistently communicate with a group due to my recent poor health I will be working on this assignment alone. This document is still a work in progress, it will document my current thoughts and plans as I work through this assignment.

**Project Work:**

* First I needed to find a dataset of English and Spanish pairs of phrases for my model. I ended up receiving one from [kaggle.com](http://kaggle.com). The dataset was within a TSV file.
* Now that I had my file I needed to create the model itself and train it. For this I used a Python3 Jupyter Notebook, I imported libraries like pytorch to build and train the model. The first step was to load this data from the dataset into English-Spanish pairs that were usable. The first block in my Jupyter notebook accomplishes this.
* Next I needed to convert the data into a DataFrame before moving on to the next step. I accomplished this with the second block in my notebook.
* In the next step I utilized Pytorch to process the previous data into token IDs so that the transformer can actually utilize them. This was accomplished in the third block in my notebook.
* The last step I was able to complete before running into several errors I was unable to troubleshoot before my time of submission was building the transformer model itself. I built a small transformer to stay within the limit of 100 neural units. This was accomplished in the fourth block of my notebook.
* Like previously mentioned when I attempt to train the model in the next steps I am repeatedly receiving errors that I have been unable to debug.