Will Lavanakul (February 9, 2023)

CS288 HW1 Writeup
Will Lavanakul

1 Descriptions of LSTM Modifications

1.1 Ensembling

The main method I used for improving perplexity is ensembling. The process of ensembling trains a set of LSTMs initialized at different weights. Each model is trained independently. When choosing the probability for the next word in a sequence, we take the mean of the logits output by all the models, and then softmax over that mean. We then choose the word with the highest probability after this softmax. Specifically, I train 6 LSTM models and average their output logits before softmax.

2 Motivation of modification

2.1 Ensembling

The motivation is that using only one model is not robust enough. For some sequences, it will not have a good enough understanding to predict well. By using a 'majority vote' of multiple models, we can make the entire system more robust. Where one model might not predict well, others might predict better, giving more weight to the prediction of the better-performing models. This should decrease perplexity.

3 Effectiveness of modifications/results

Since we are training the models individually with the same training method as before, we do not see in improvement in just training loss. But since we implement an ensemble procedure during validation, we can see that the perplexity improves better over training and reaches a better perplexity after the same number of epochs.
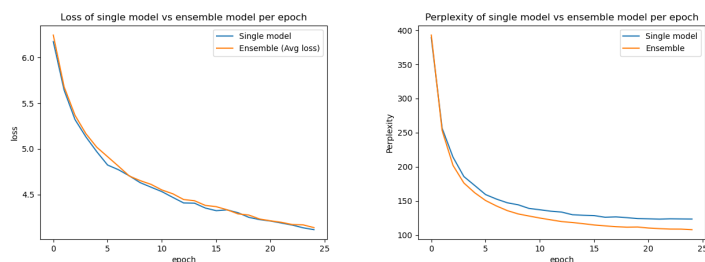
4 Summary of results

With the ensemble implementation, we see a better perplexity score as well as longer improvements of perplexity. We see that only one model eventually reaches a perplexity and does not improve. The ensemble model continues to improve and the past 25 epochs will probably improve more.