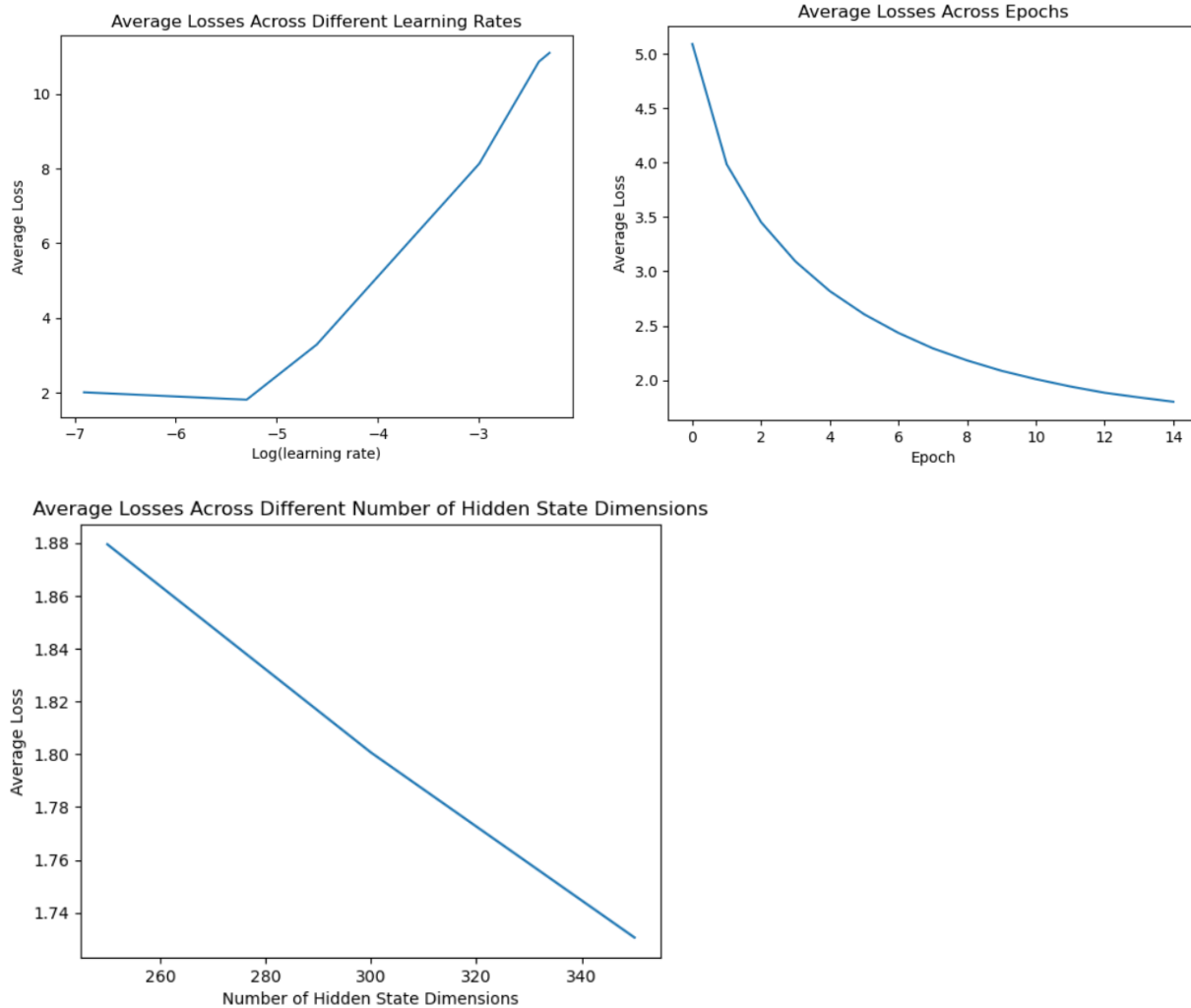# Assignment 2 Report

Name: Hongzhao Tan

MacID: Tanh10

Student #: 400136957

Task 1 RNN-based Text Generation:

After defining the architecture of RNN model with Python class and nn.Module from PyTorch, training a basic RNN model with text data preprocessed and batched (batch size = 128, sequence length = 5) from the Reuters corpus, through 15 epochs, and with several different values of learning rates [0.001, 0.005, 0.01, 0.05, 0.09, 0.1] and dimensions of hidden states [250, 300, 350], it has been found that generally the loss decreases as the number of epochs and hidden dimensions increase, while when the learning rate reached 0.01, the loss became higher than that with lower learning rate and the same number of hidden dimensions and epochs. Furthermore, when the learning rate reaches 0.05 or higher, the model will not be able to learn from training and the loss stays high across all epochs and the numbers of hidden dimensions.

Average Losses Across Different Learning Rates

Average Losses Across Epochs

Average Losses Across Different Number of Hidden State Dimensions

As a result, the optimal hyper-parameters that yield the lowest value of loss was when the number of epochs, the number of hidden dimensions and the learning rate equal to 15, 350 and 0.005 respectively. A basic RNN model and a RNN model with LSTM layers have then been defined and tranined with the optimal combination of hyper-parameters and also some other parameters including the number of word embeddings' dimensions (equals to 200), and the threshold for clipping gradient (equals to 5). The two trained RNN models were then used to generate text of 100 words in length with the word 'england' as the input word. The generated texts were as shown below:

100-word text generated using basic RNN:

```
england
said the central bank intervened to buy dollar
usx corp chairman david roderick said.
" i think we're near a daily basis," hamley said.
the dollar has nearly halved its value against major currencies, but u. s.
government securities market to arrange two billion dlrs at 10- 3/4 pct, and the
u. s. dlrs from july 1986, the bank added.
the dollar has hovered around 40 mln stg in band one at 9- 3/4 pct, the national bank
```

100-word text generated using RNN with LSTM layers:

```
england
would force the u. s. and japan will not want it to
become uncompetitive because of a lower house budget committee in parliament on indonesia 2 in the year ending march 1988, leading to a
meeting of
the group of five(g- 5) industrial nations reaffirmed their willingness to
cooperate on securities repurchase pacts
the bank of england has recently
used and a news conference yesterday's drop as
as the dollar remains in a
41- day repurchase pacts
the fed is quite adequate as
```

Because the Reuters corpus mainly consists of financial news in English language, the generated texts were also in English words and some words that commonly take place in financial news such as "bank" and "dollars" have shown up in both generated texts. Both generated texts were correct on syntax overall, while do not really make sense on semantic. The text generated by RNN with LSTM layers has shown a long-term memory while the text from the basic RNN has not. In the text generated by basic RNN, the word 'england' or some words related to it such as 'u. k.' has not shown up other than the first input word 'england'. However, in the text generated by RNN model with LSTM layers, a series of words "the bank of england" has appeared at a place that was tens of words away from the first input word 'england', indicating the information of the input word has been carried through the long-term memory and in turn affected the result of text generation.

Task 2 Seq2Seq NMT with Optional Attention:

After defining the architecture and functions of the Seq2Seq NMT model with Python class and nn.Module from PyTorch, NMT models without using attention mechanism have been created and trained with 40,000 preprocessed and batched English and French sentences pairs where French was used as the source language while English was the target. Several values of learning rate [0.001, 0.005, 0.01, 0.05, 0.1], the number of hidden state dimensions [150, 200, 250], and size of each batch [100, 150, 200] has been combined with each other and experimented by traning the NMT models through 12 epochs. The average loss (negative log likelihood) and average perplexity of each epoch have been recorded during the experiments. Similar to the result of Task 1, the average loss and perplexity reached their minimum when learning rate equaled to 0.005 then both generally increased when learning rate equalled to 0.01, and eventually stayed high when learning rate equalled to 0.05 or larger. When using 0.005 as the value of learning rate, both the average loss and perplexity have generally decreased as the number of epochs increases, while the marginal gain on minimizing these two metrics has been very little after the number of epochs reaches 7. The optimal values for batch size and the number of hidden state dimensions when learning rate equalled to 0.005 was found to be 100 and 250 respectively.
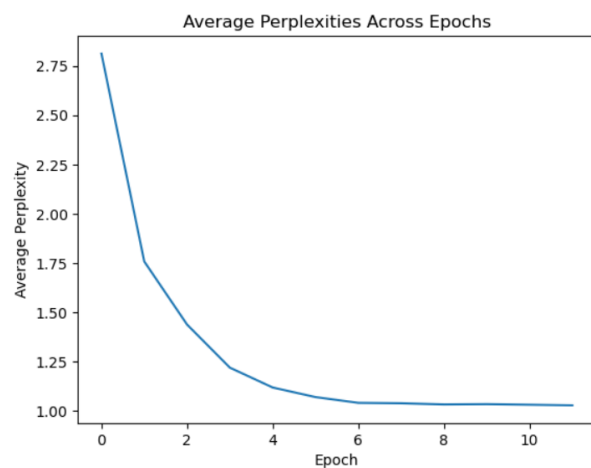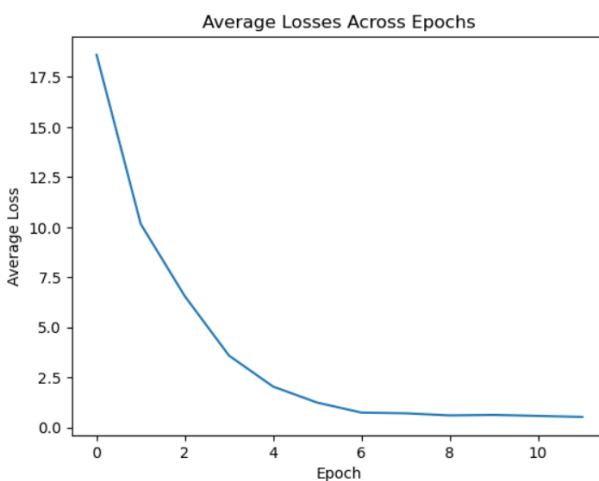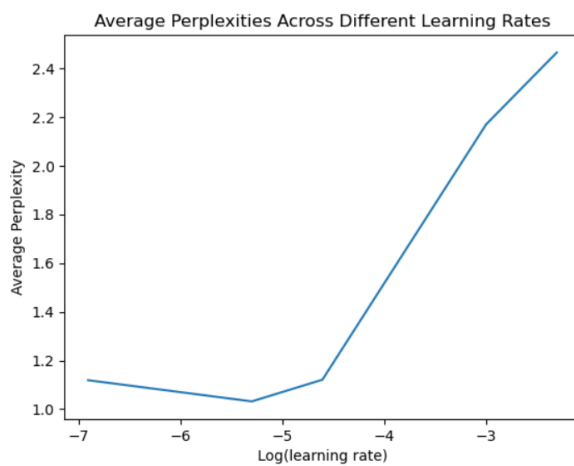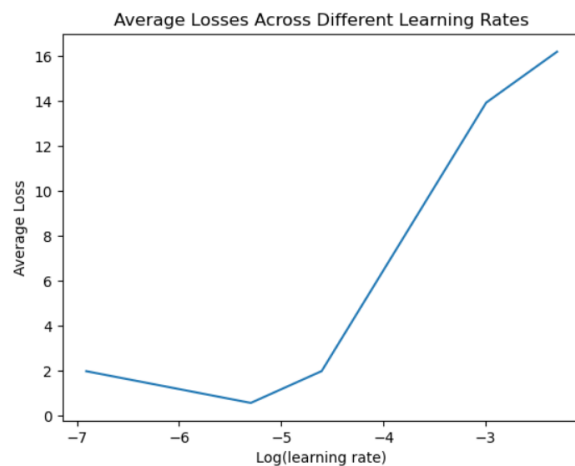
Heatmap of average loss against the number of hidden dimensions and batch size:
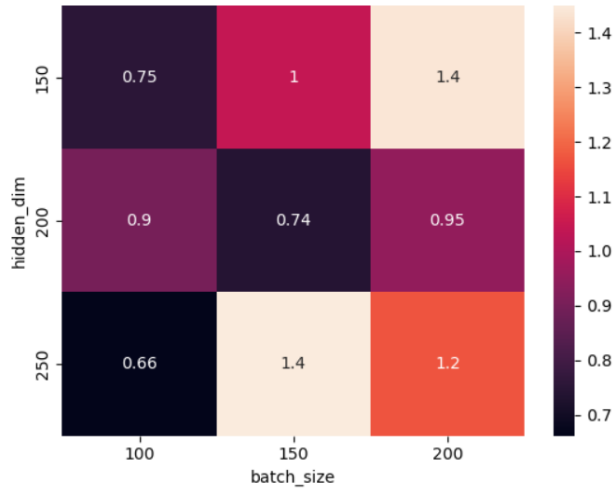
With the optimal hyper-parameters' values (i.e. learning rate = 0.005, number of epochs = 7, hidden dimension = 250, and batch size = 100), two NMT models that are with and without using the attention mechanism respectively, were trained with the same 40,000 English and French sentences pairs mentioned above. Some other parameters including threshold for clipping gradient (equalled to 5) and words embeddings' dimensions (equalled to 100) were also used for training. These two trained NMT models were then used to decode/translate a test set of 10,000 French source sentences to English with the beam search mechanism. Eventually, a corpus-level BLEU score was computed for each of the two NMT models, by comparing the hypotheses that were yielded from beam search and had the highest log probabilities, against the ground-truth English target sentences of the 10,000 French sentences. As a result, the translations from the NMT model without using attention mechanism have reached a BLEU score of 0.636, when the translations from the NMT model that has employed attention mechanism have reached a BLEU score of 0.705.