

# Midterm

Brian Henderson

10/23/17

## 1 Overview

Using the RNN lab entitled "Character Modeling with RNNs/LSTMs," run experiments modifying the hidden units (32, 64, 128, 256, 512), the sequence (20, 50, 75), and analyze the perplexity of the sentence outputted and any variations. The dataset used for training the neural network is a dataset from the authorship of William Shakespear.

## 2 Hidden Units Variation

### 2.1 Final Results Table

Shape	Training Loss	Time/Batch	Last Generated Sentence
(60, 50, 32)	1.832	0.011	The bound. Wenabry 'tiea-dordech swo bith's have our s
(60, 50, 64)	1.622	0.017	The otneing of his is our word King
(60, 50, 128)	1.418	0.050	Felling where than ventich more thee fortun
(60, 50, 256)	1.200	0.139	Are not home at your lands of
(60, 50, 512)	0.897	0.585	For being ignortion of his pernicial cramet

### 2.2 Analysis

With the variation of the hidden units, it is easily shown that there is a correlation in regard to the training loss, time for each epoch, and perplexity of the generated sentences. As the hidden units increase, the training loss decreases, the time to complete a batch decreases, and the sentences become more clear. As shown in the table, the perplexity of the sentence with lower hidden units (32), "*The bound. Wenabry 'tiea-dordesh swo bith's have our s,*" is very high as the RNN struggles to make comprehensive sentences with no sentence structure and disoriented words. As the hidden units increase (512), "*For being ignortion of his pernicial cramet,*" the perplexity decreases as the sentence starts to form more of a defined sentence structure, most of the words are valid, and the sentence is more readable.

The input data is fed into a layer of hidden units and then that is fed into more layers, etc., which produces the output. The reason why the sentences with the better comprehension are the experiments run with higher hidden units is because they train exponentially longer. The higher the hidden units, the longer it takes as there are more CPU computations, resulting in longer batch time, however the result is a more accurate result as shown by the lower training loss. Just with anything, the more one trains, the better the output is.

### 3 Sequence Length Variation

#### 3.1 Final Results Table

Shape	Training Loss	Time/Batch	Last Generated Sentence
(60, 25, 128)	1.331	0.031	The ruleties: thou rather oft o' mayer and primple
(60, 50, 128)	1.408	0.065	I burgh thy pin
(60, 75, 128)	1.364	0.101	With delay's did maski

#### 3.2 Analysis

The results of the sequence length show less of a trend than the previous set of experiments. With no clear trend, it is hard to concisely identify the sequence effect on the perplexity of the sentence. However, it can be leaned towards the conclusion that longer sequences have a lower perplexity as the sentences seem to flow better. It would be expected that with the higher sequence length the RNN would have more inference on determining the next character due to the amount of data prior. Since getting a naturally flowing sentence is heavily dependent on the surrounding words, a higher sequence length would account for a more fluent sequence of words. Although the training loss is better for the lower sequence length, I believe that is due to the training set, also since the sequence is longer, it has to deal with more words that the RNN may not seem necessary to the sentence.