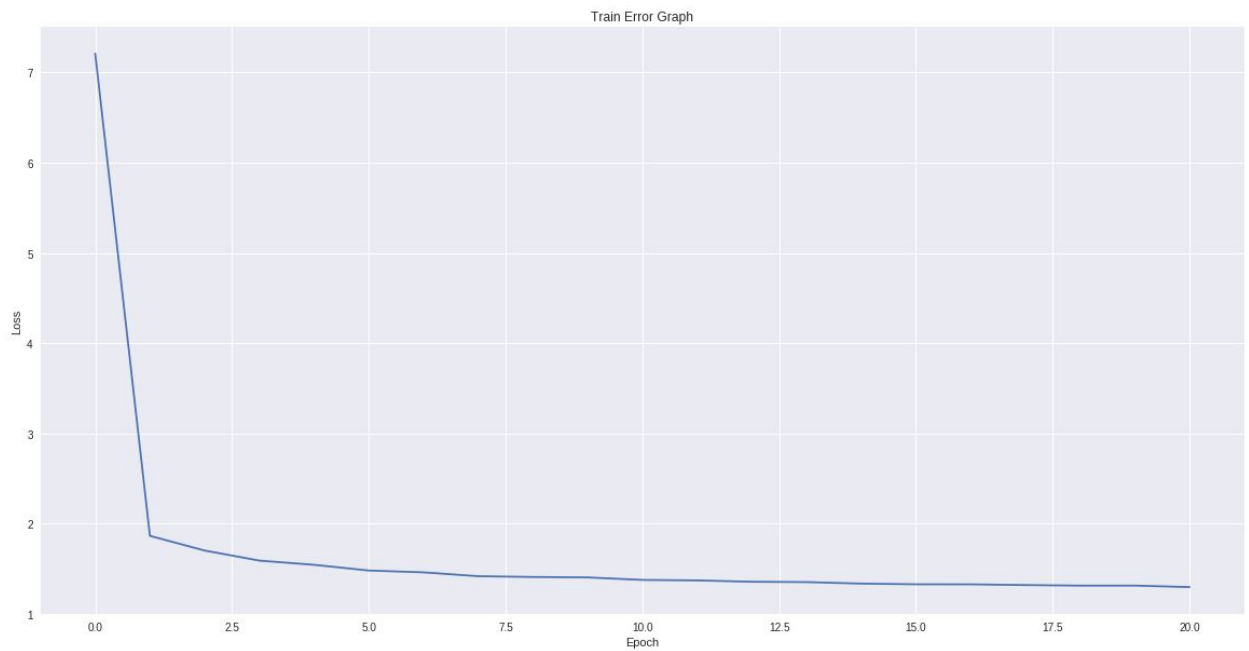


## Homework Questions

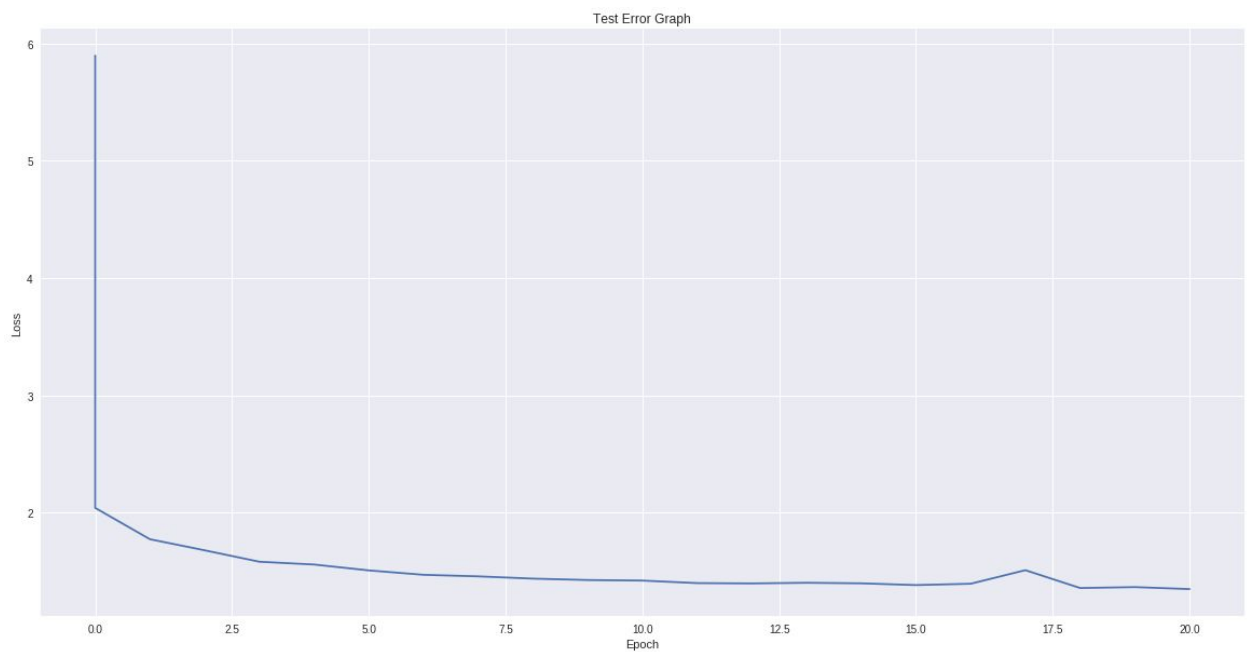
1. Modification & Hyperparameters:

- Temperature = 1; Num\_layers = 3; Beam\_width = 15; Sequence\_length = 20; Batch\_size = 256; Feature\_size = 512; Test\_batch\_size = 256; Epochs = 20; Learning\_Rate = 0.002; Weight\_Decay = 0.0005

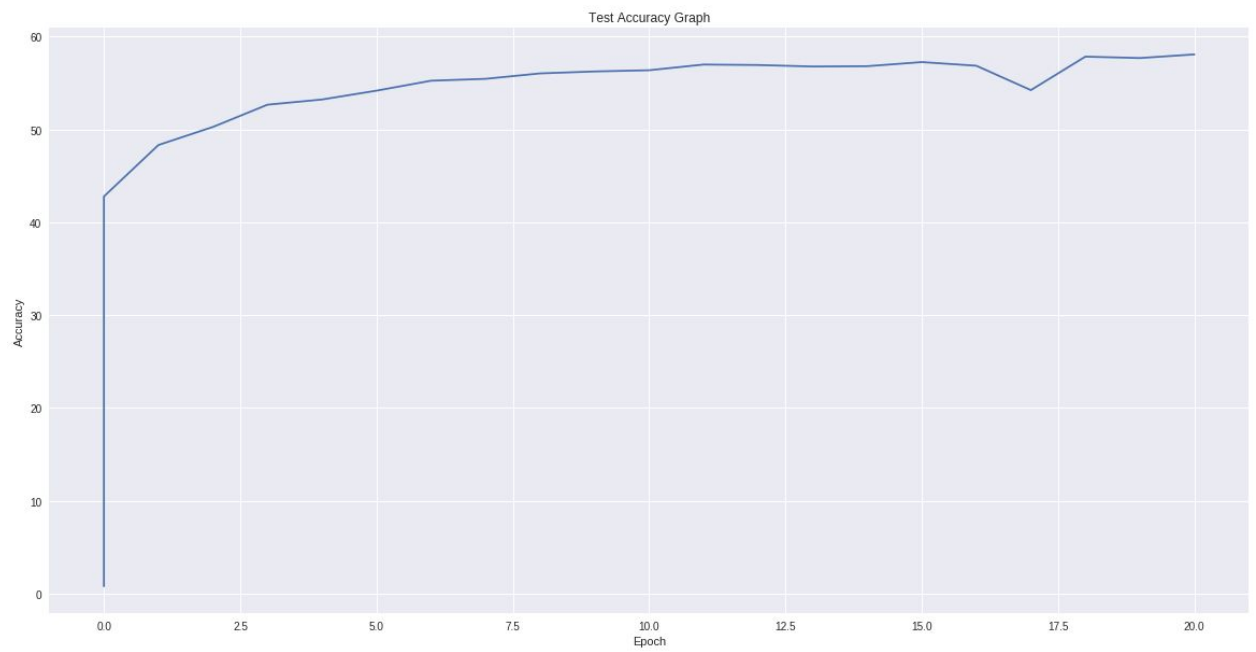
Train Error



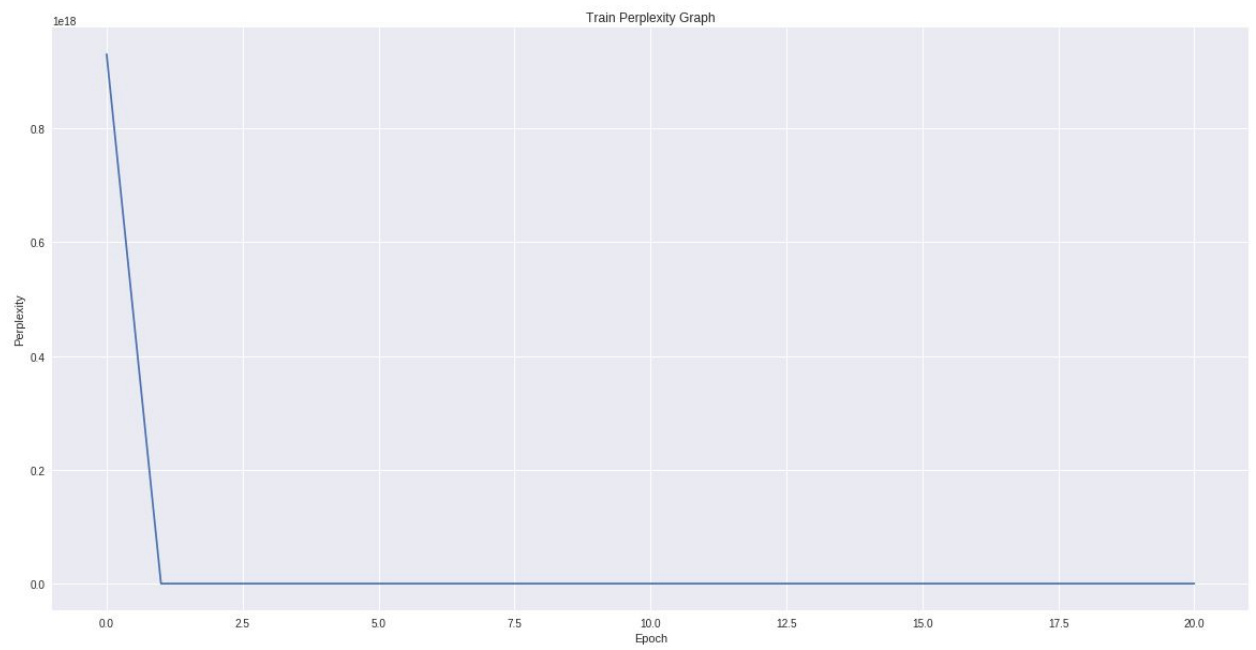
Test Error



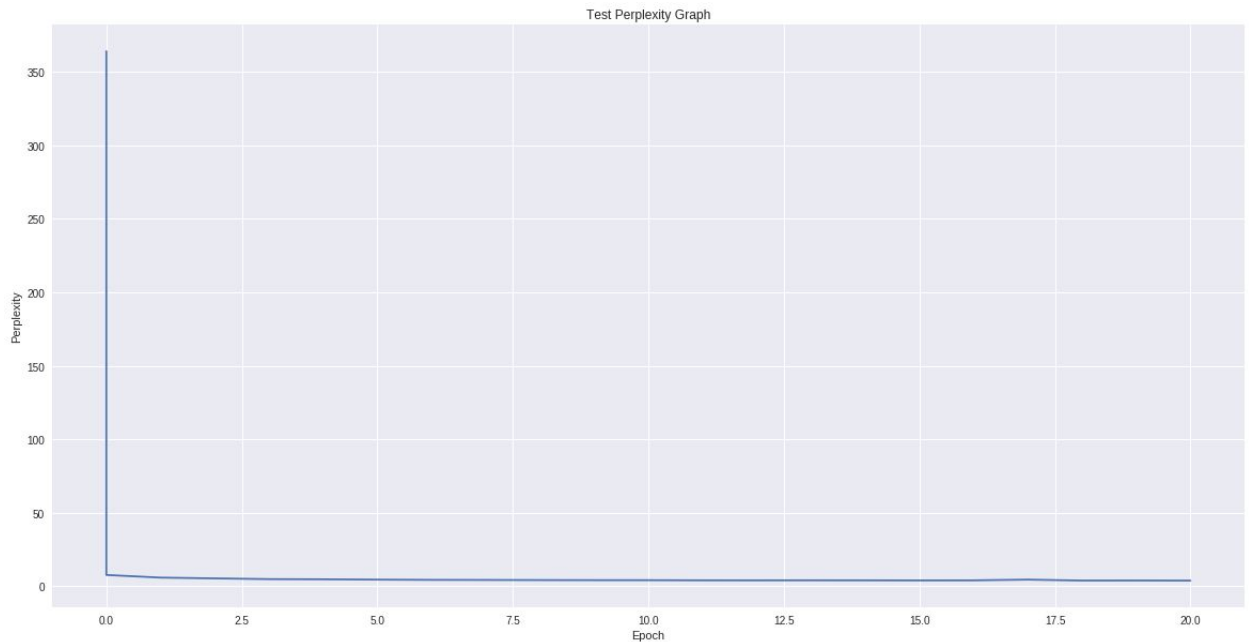
## Test Accuracy



## Train Perplexity



## Test Perplexity



2. Our final test accuracy was 58.09

Our final test perplexity was 3.85

3. Our favorite sentence generated for ...

- Max: Using the prompt: "Hello my name is"

Hello my name is a shadow she was still staring at his face. "I have a lot a second like that the others and said in a students and saying the stairs and started to his face and said in a stand and started to his face

- Sample: Using the prompt: "You are such a"

You are such a side. He dashed in his voice, her most other Leeters is last, 2riss while Irmself have noticed been to jkeble. But the soft's botters is she was.... thought Creatures had varry had said, "Harry,

- Beam: Using the prompt: "I love Harry"

I love Harry said Uncellangelina rat his preacty, had nifted again in the Potion look and seemed to ackward her at them. Hermione looked into sink who make he ospected hander; in her bushes in back to Crack and leant

4. The beam search sampling method seemed to generate the best results. The max sampling method was too simple and deterministic. Since this method looked for the one value with the max probability from the current state, it would often repeat the thought that a phrase or a character was most optimal. This led to the model entering a loop and creating repetitive sentences instead of exploring a different pathway. The sample method was slightly better in that it was more flexible in choosing the output. However, the sample method was incapable of predicting beyond its immediate scope. This led to the characters fitting in to the sentence and making sense less the

further from the starting point the character was. The beam search was a good alternative that allowed the model to explore a few different pathways further before returning an optimal output.

5. Trying multiple temperatures between 0 and 2 ...
  - The temperature that produced the best outputs was 1. The temperature value of 1 was a good balance between reducing the impact of anomalies and not giving all characters equally high probabilities. Based on the criteria of “best” on the ability of the temperature to create logically and grammatically correct sentences, 1 seemed to have produced the “best” outputs.
  - A temperature of 0 was similar to max sampling because it was making differences in probability sharper. This made the max probability in the sample much more likely to get selected as the sample moving forward. With the temperature of 0 we saw many repetitive words and phrases and not much variation in the sentences.
  - A temperature between 0 and 1 provided samples close to that of taking a softmax or the max. Trying out a temperature of 0.5 gave us sentences that weren’t as repetitive as a temperature of 0 but repeated words could be found throughout the sentence.
  - A temperature of 1 provided samples that don’t have the repetition often seen in temperatures of less than 1. The sentences appear to make some sort of sense, although still not fully coherent, and outlandish words aren’t present. We did notice that a lot of names show up when using a temperature of 1 which was pretty interesting.
  - A temperature above 1 provided a lot more variation in samples. The model became more likely to pick a sample that is an unlikely candidate, but also became more prone to making mistakes and creating bad output. Trying out a temperature of 1.5, the sentences outputted had much weirder combinations of characters. Often times they didn’t produce real words and produced weird sequences of letters instead.

## New Corpus

1. We chose a corpus containing the top 20 most popular children's books from project Gutenberg. There were 7,135,395 characters in our corpus.
2. Structurally, the sentences are pretty similar to each other; there's some broken english, grammatical errors and constant changes in topic within a generation so it's difficult to really tell what the text means. However, the two corpuses are syntactically very different not just in individual words but also phrases. For example, in the harry potter corpus, there are a lot of phrases that sound british whereas the children stories corpus outputs phrases that sound like something a child would say.
3. Using a temperature of 0.75 ...

a. **Max sampling:**

Pooh, and the same to the same to the same and the story was a big before the first to the same to the same

Cinderellant and she said to her showed and she was a little something and the story was a big before the first to the same to the same to the same

Goldilocks and she was a little something and the story was a big before the first to the same to the same

b. **Sampling:**

Pooh, and a long and a great day and the bear mind that I was too my friend and crossed to each who charm.

Cinderella says havened been away, interest and schoolrosty each to Mr. Crainned might be fear to make surried before the those who was nor far too and white was a bleface

Goldilocks look about it, and dearly stood for the height, and then he is a replied the seames with a long, when she looked at the door and all way distance of the it.

c. **Beam:**

Pooh, and the wered t teverytought There'eving tereemed youreaing aness she said. . . . and ter ans ward the less and they wenes and tting ting aness,

Cinderella, arilla soness and th the fully and the wel and the theyth ation analy ationinted ards sted into t the st oryouremember tought a

Goldilocks. . . and then t was alkin on of thing als aing atthey wes aid th, and the the wert, aness and ward ares and thiness aness

## Words

1. Using character generation, the biggest issue we had was that we were generating non-real words. Using word generation, we encountered a new problem. Our word generation, at the beginning, was creating a lot of unknown tokens. It was common to get sentences that were filled with only unknown tokens instead of words. The model also started to stagnate around the 15% test accuracy mark and it was difficult to improve by just fiddling with the hyperparameters.
2. Our vocabulary consisted of 14,403 words.
3. In order to get the model to allow low probability words to be chosen, the temperature was increased from 0.5 to 0.75. This corrected the issue where the model was only selecting unknown tokens. At a temperature of 1 and greater, the number of unknown tokens decreased but so did the logical flow of the sentence. A temperature of 0.75 was a good balance to decrease the amount of unknown tokens while keeping somewhat logical sentences.

The sequence length was also increased since the model was generating words and the words were much longer than characters. With a long sequence length, the model runs into the problem of unknown tokens appearing everywhere possibly due to the fact that the likelihood of the words in the samples that the model is pulling from become lower and lower. Because of this, the sequence becomes longer.

Increasing the feature size made the sentences sound more natural but slowed down training by a lot. All the other parameters stayed the same as when training on characters and they seem to work fine.