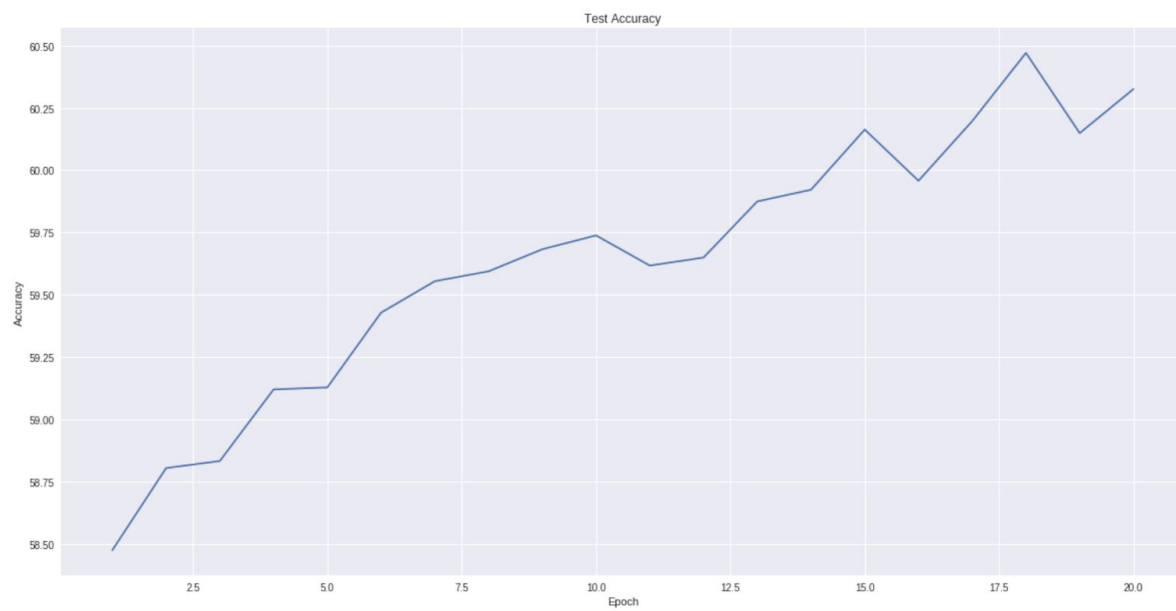
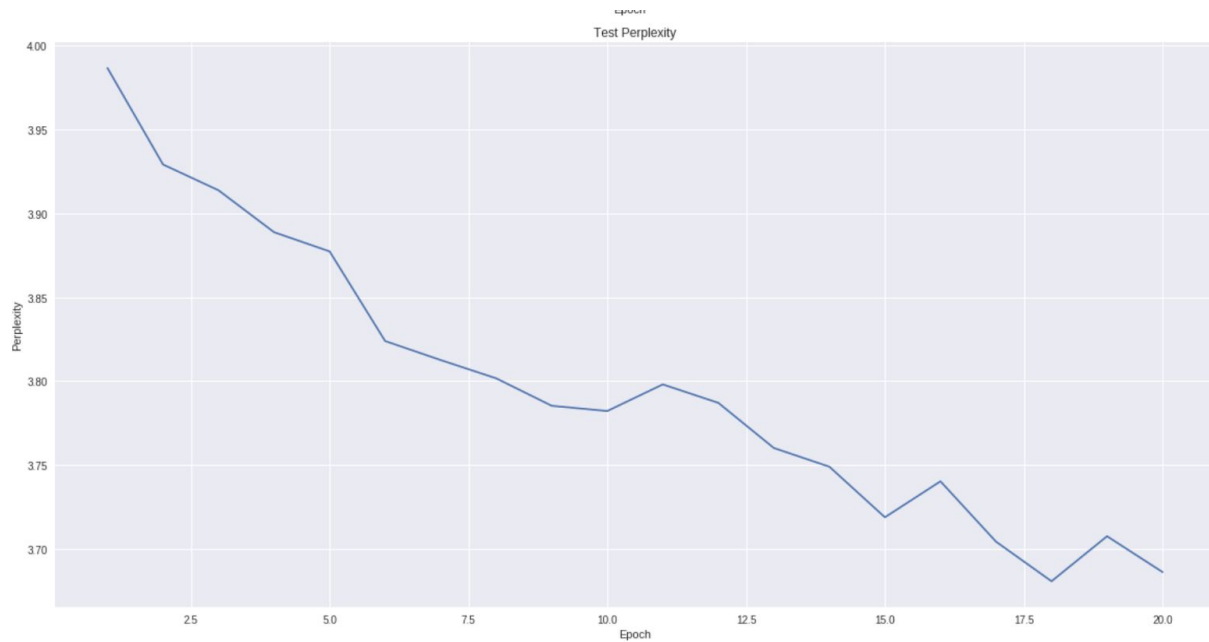


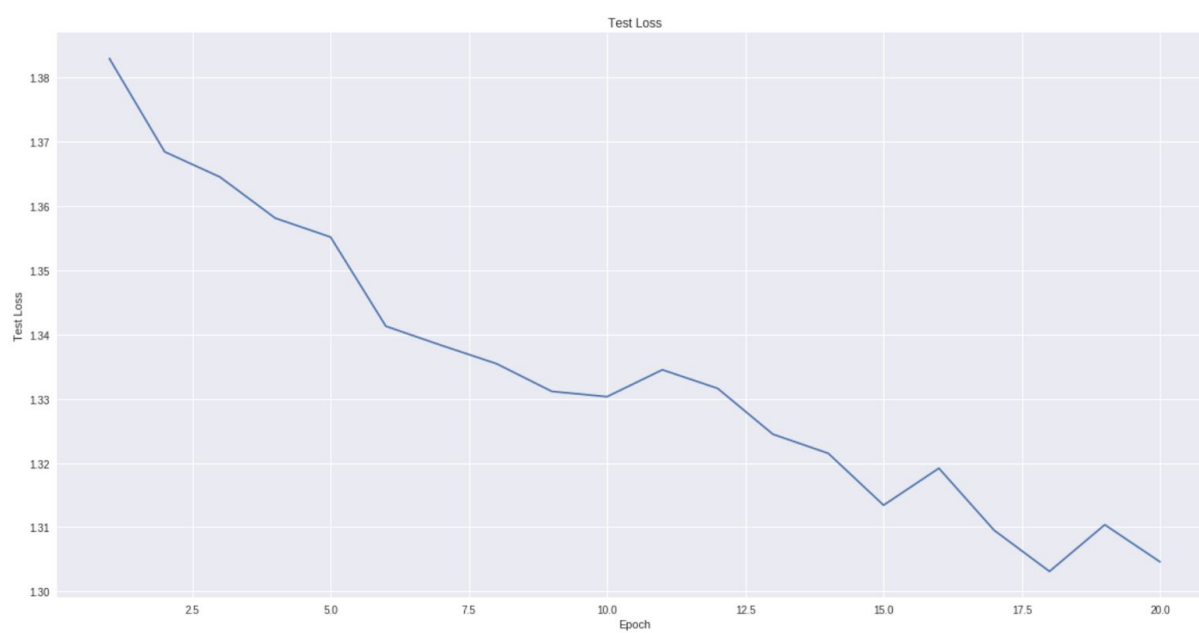
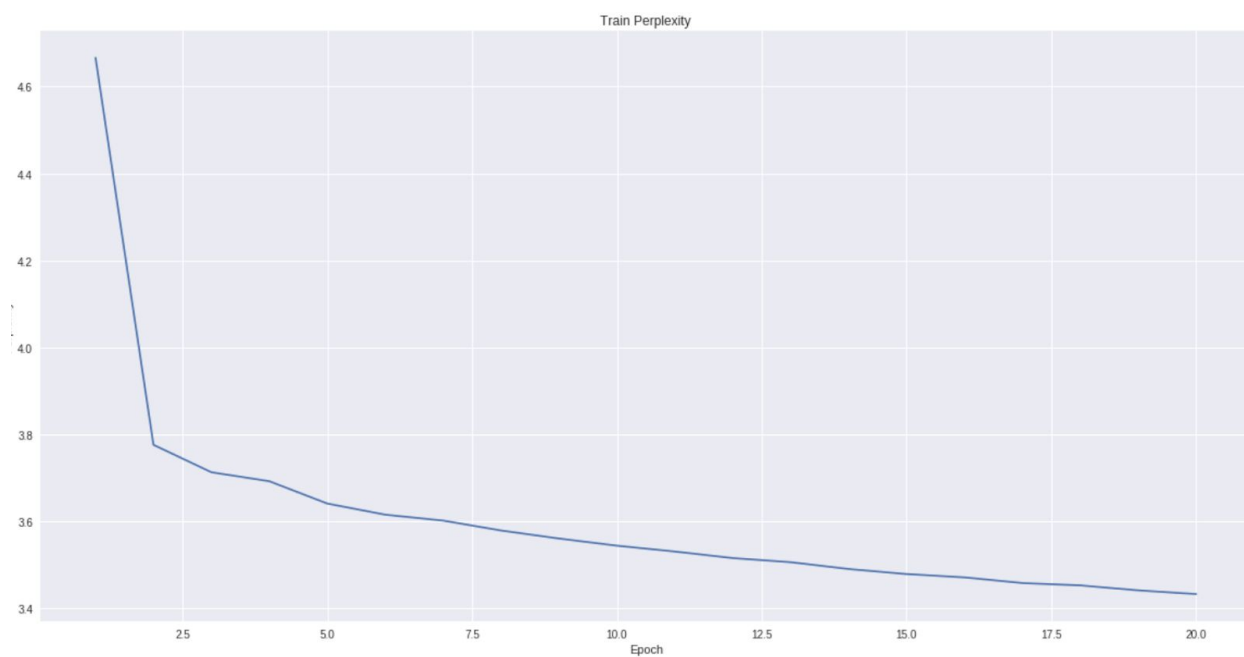
Akshat Shrivastava

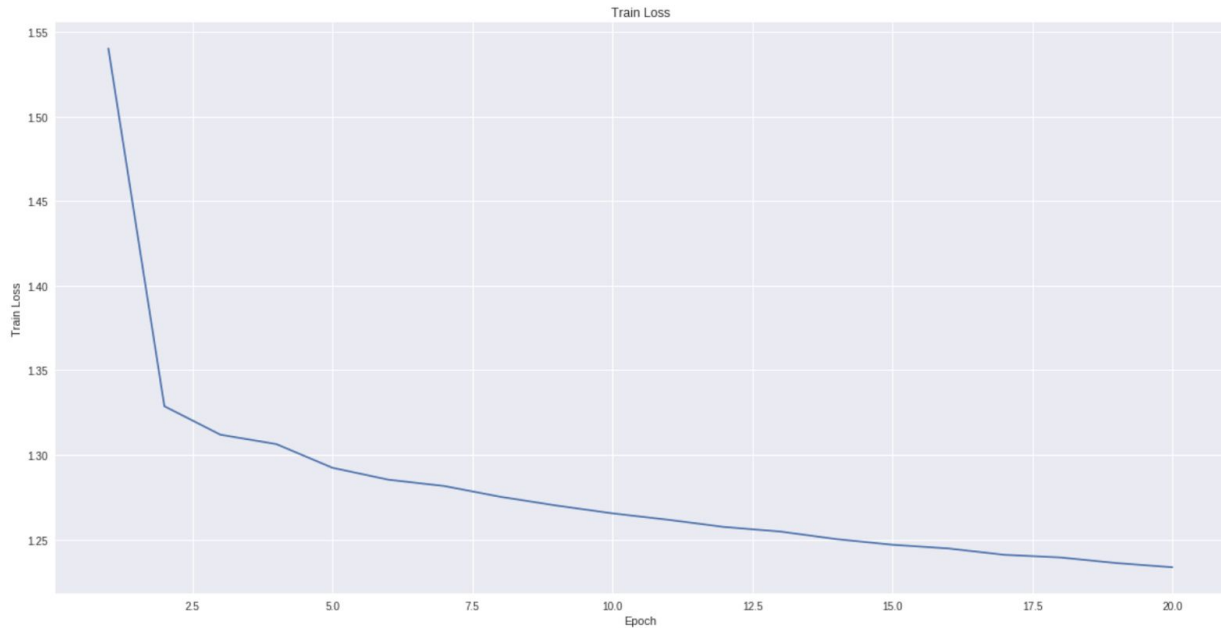
November 16th, 2018

## Deep Learning

1. Just like last time, provide plots for training error, test error, and test accuracy. Also provide a plot of your train and test perplexity per epoch.







2. What was your final test accuracy? What was your final test perplexity?
  - **My final test accuracy was 60%, perplexity was around 3.7**
3. What was your favorite sentence generated via each of the sampling methods? What was the prompt you gave to generate that sentence?
  - **Max: Harry Potter and the first time was still straight in the carrial street street that he was still straight in the carrial street street and saw the carriage was sta**
  - **Sample: Harry Potter and the last last remove silver arm. Immement's arrowled and said Lopphan while Ron didn't see Harry for the pointing of his inch gressen slip swalls**
  - **Beam: Harry Potter and they were alraid to but stured to his feet and over his robes strong on the lake and he stoped into a lore rob so threw him had taken the from holidai will not care an unusual portrait of contant frown a**
  - **All the sentences above were generated from the seed words "Harry Potter"**
4. Which sampling method seemed to generate the best results? Why do you think that is?
  - **I think max and beam search methods gave very similar results, with beam performing a bit better depending on the sentence since beam search is able to consider many sentences at once.**
5. For sampling and beam search, try multiple temperatures between 0 and 2.
  - Which produces the best outputs? Best as in made the most sense, your favorite, or funniest, doesn't really matter how you decide.
    - i. **With a temperature of 1: Harry Potter and they see to hear her far wandng and a long loud line of train worse to come and halm too lost to see and was something than every hand for two more deaf and for seen a lot of parcal and said from word a**
  - What does a temperature of 0 do? What does a temperature of  $0 < \text{temp} < 1$  do? What does a temperature of 1 do? What does a temperature of above 1 do? What would a negative temperature do (assuming the code allowed for negative temperature)?

- i. The closer the temperature is to 0 the more confident the model will be in its predictions. Since the temperature divides each of the values before applying softmax, scaling the values by a higher number (dividing by less than 1) results in more extremes for softmax. The further the temperature is from 1 (greater than 1) the less confident the predictions will be in the softmax. Having a temperature of 0 results in sampling performing similarly to max, and closer to 2 results in more garbage sentences. A negative temperature would result in the most probable values having a very negative score, which results in what would have been the least probable values having a higher softmax score and the most probable values having a lower softmax score essentially inverting the distribution.

Questions for each of the "Other things" sections. Only answer the questions corresponding to the ones you chose.

#### 1. New Corpus

1. What corpus did you choose? How many characters were in it?
  - i. **I chose the dracula book as the corpus to examine. There were 857834 characters in the book.**
2. What differences did you notice between the sentences generated with the new/vs old corpus.
  - i. **I passed in the seed word "Dracula" to the seed words generated for this corpus. One thing I noticed was that a lot of the generation didn't create a new word but added onto dracula to make words like Draculaid. The words generated here are also more like the style from dracula.**
3. Provide outputs for each sampling method on the new corpus (you can pick one temperature, but say what it was).
  - i. **With a temperature of 1:**
  - ii. **Beam: Draculate of his here answered the horses of any out of his had some one of his had something of the strange of the horsself. I do not folow on time to the others on see most go one of the saw how seemed to t**
  - iii. **Sample: Draculay over, and faits own so has proping and know, for I window see your had sorth comeinces that to pall cogn of anothing of me like out. How he too, so in all not but you both chocks, there, but was you**
  - iv. **Max: Harry Potter and they see to hear her far wanding and a long loud line of train worse to come and halm too lost to see and was something than every hand for two more deaf and for seen a lot of parcal and said from word a**

#### 2. Words

1. What new difficulties did you run into while training?
  - i. **The vocab size went from around 256 characters to 14 thousand words, which made the test accuracy go down quite a bit, because it no longer had to predict just the next character but the entire of characters (words) as well, which results in a more difficult task. There were also some changes to handle unknown vocabulary words, which resulted in some of the output predictions being unknown words.**
2. How large was your vocabulary?
  - i. **1099726 words in the book with a 14403 word vocabulary (after removing words that occur less than 5 times)**

3. Did you find that different batch size, sequence length, and feature size and other hyperparameters were needed? If so, what worked best for you?
  - i. **Since sequence length was based on characters not words, I changed it by a factor of 10, sequence length of 10 now instead of 100, to make the sentences reasonable in length. Due to the much larger vocabulary, I also increased the hidden size to 1024, but it was taking far too long to train. A feature size of 512, sequence length of 10, and a batch size 256.**