

Short Answer Questions

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.

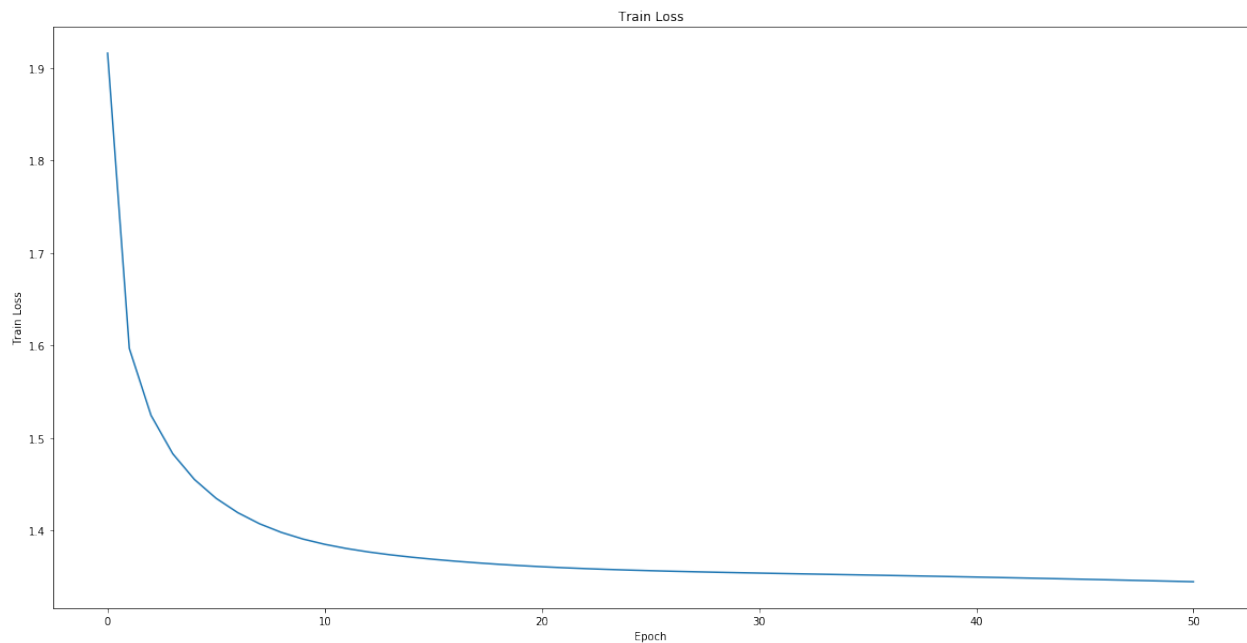
- In class we defined perplexity as $2^{(p \cdot \log_2(q))}$, However the PyTorch cross entropy function uses the natural log. To compute perplexity directly from the cross entropy, you should use $e^{p \cdot \ln(q)}$.
- We encourage you to try multiple network modifications and hyperparameters, but you only need to provide plots for your best model. Please list the modifications and hyperparameters.

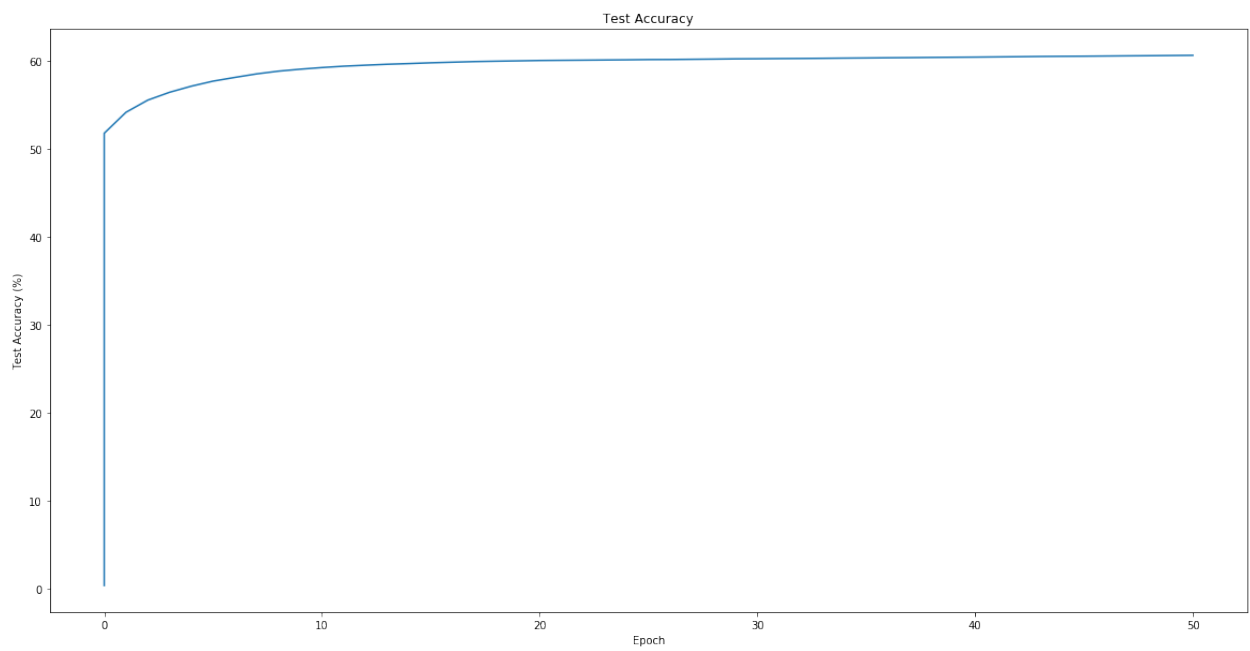
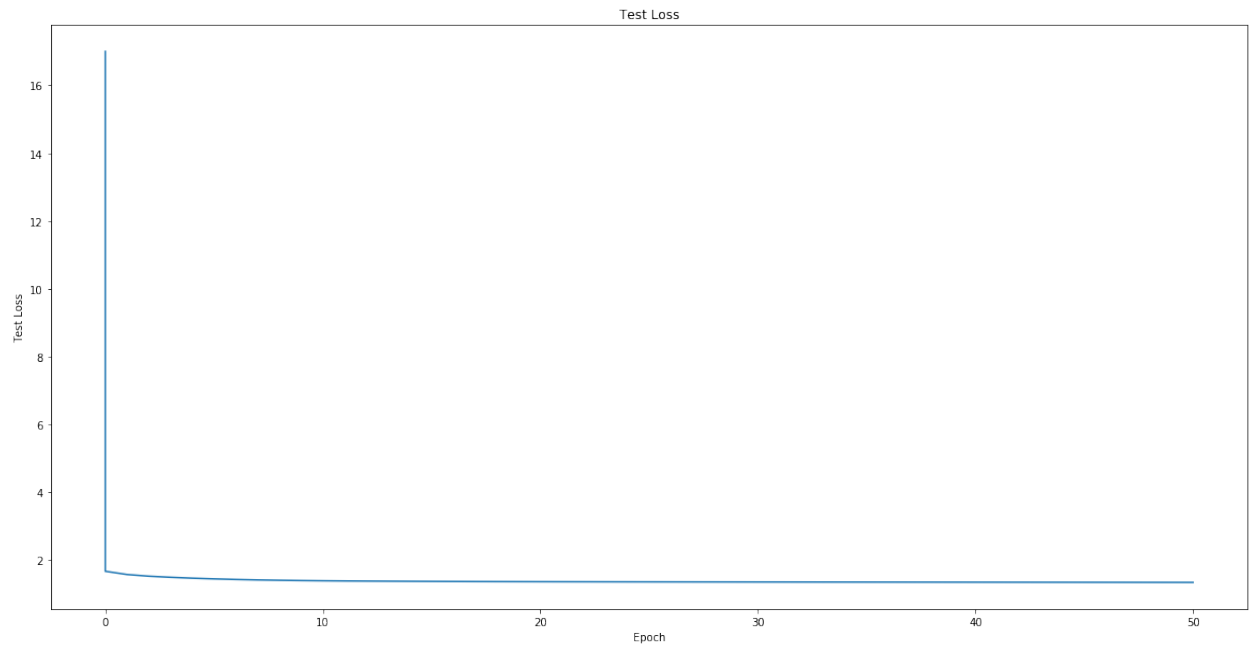
Modifications

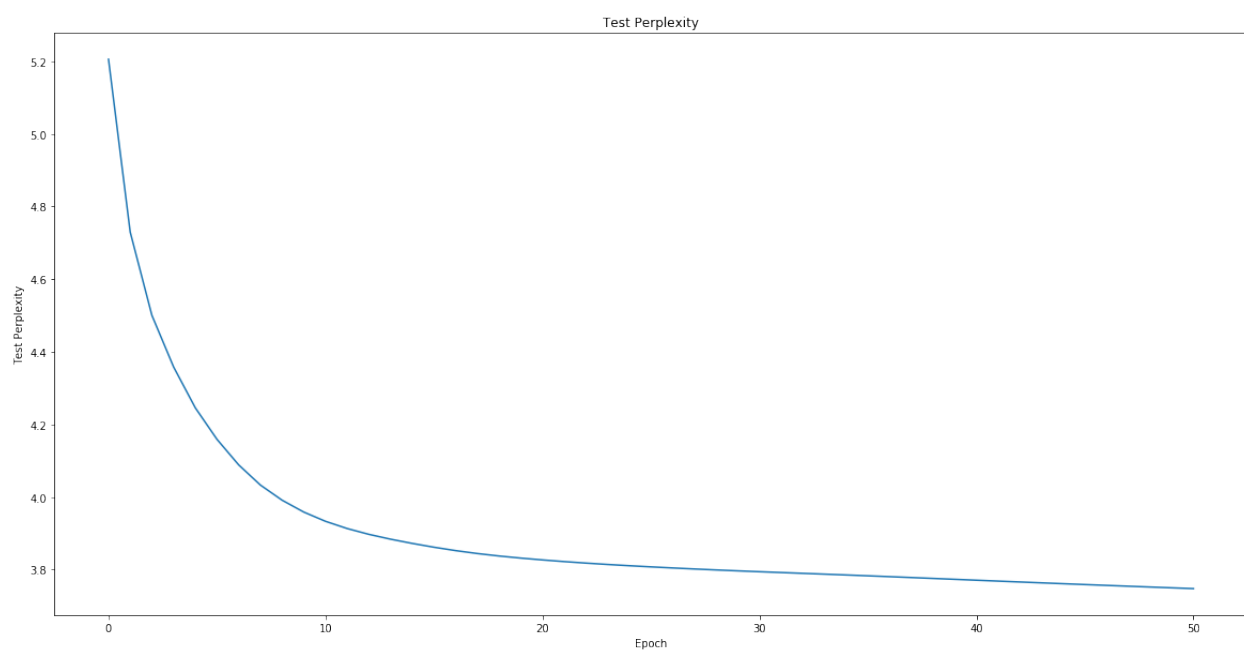
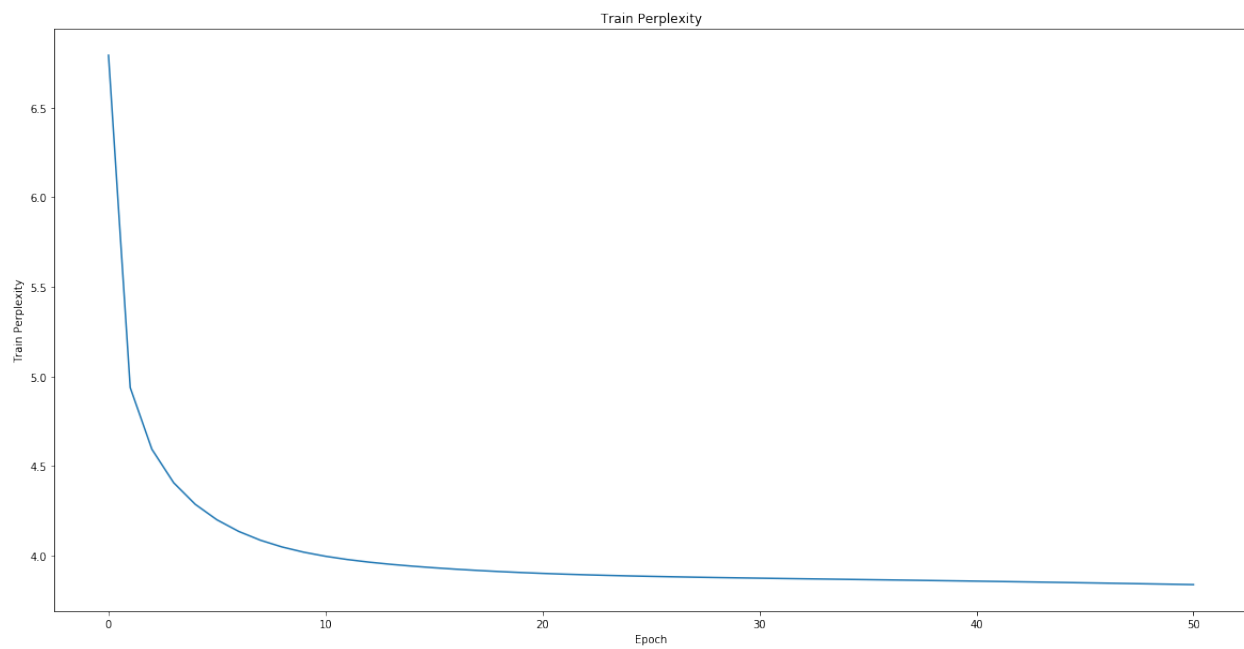
In the preprocessing step, I removed non-ASCII characters and any double space, `\n`, `\t` from the text data.

Hyperparameters

SEQUENCE_LENGTH = 40, BATCH_SIZE = 64, FEATURE_SIZE = 512,
TEST_BATCH_SIZE = 64, EPOCHS = 50, LEARNING_RATE = 0.0001,
WEIGHT_DECAY = 0.0005







2. What was your final test accuracy? What was your final test perplexity?

Final test accuracy: 60.669

Final test perplexity: 3.747

3. What was your favorite sentence generated via each of the sampling methods? What was the prompt you gave to generate that sentence?

- Generated with max (Prompt: Harry Potter and Severus Snape)
Harry Potter and Severus Snape so that the words of the team when he had been and the stopped the first the first the first
- Generated with sample (Prompt: Harry Potter and Voldemort)
Harry Potter and Voldemortable as Magic. George to make parted him. The was sees as though he had been the last of better at Seamus together!" Ron sound of the water.
- Generated with beam (Prompt: Harry Potter walks into Hogwarts.)
Harry Potter walks into Hogwarts. There was a most straight through the first time, Harry had been standing into the door and the corridors of the corridor," said Harry, shouting in his face.

4. Which sampling method seemed to generate the best results? Why do you think that is?

After several testing, the beam sampling strategy seems to generate the best result, once the temperature and beam size are properly tuned. The beam sampling avoids producing duplicated phrases. This method performs well because it combines the idea of random sampling and picking k best words. The generated sentences can be more creative.

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.
- 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)?

For most temperatures I tried, sampling generates sentences with more sense and various vocabularies, while beyond some point of the temperature, beam search gives better outputs.

With temperature of 0, sampling and beam search generate the same or very similar sentences. With temperature of $0 < \text{temp} \leq 1$, sampling generates better sentences with various vocabularies, while beam search generates the same sentence with limited and duplicated vocabularies. With negative temperature, sampling and beam search generate sentences almost using the same vocabularies. Additionally, with temperature of $1.5 < \text{temp} < 2$, beam search begins to produce better and more creative sentences, while sampling does not work at this temperature.

Other Things

1. New Corpus

- **What corpus did you choose? How many characters were in it?**
- **What differences did you notice between the sentences generated with the new/vs old corpus.**
- **Provide outputs for each sampling method on the new corpus (you can pick one temperature, but say what it was).**

I used a text version of the complete book of Sherlock Holmes, written by Conan Doyle. It has 2524585 characters.

With the old corpus, the generated sentences are highly related to the content of Harry Potter. Some example words such as corridor, castle, Professor McGonagall. With the new corpus, the model learned the written style of Conan Doyle, using concise words for description. The generated sentences have more conversations and become more philosophical.

Prompt: Sherlock Holmes and John Watson went to a case.

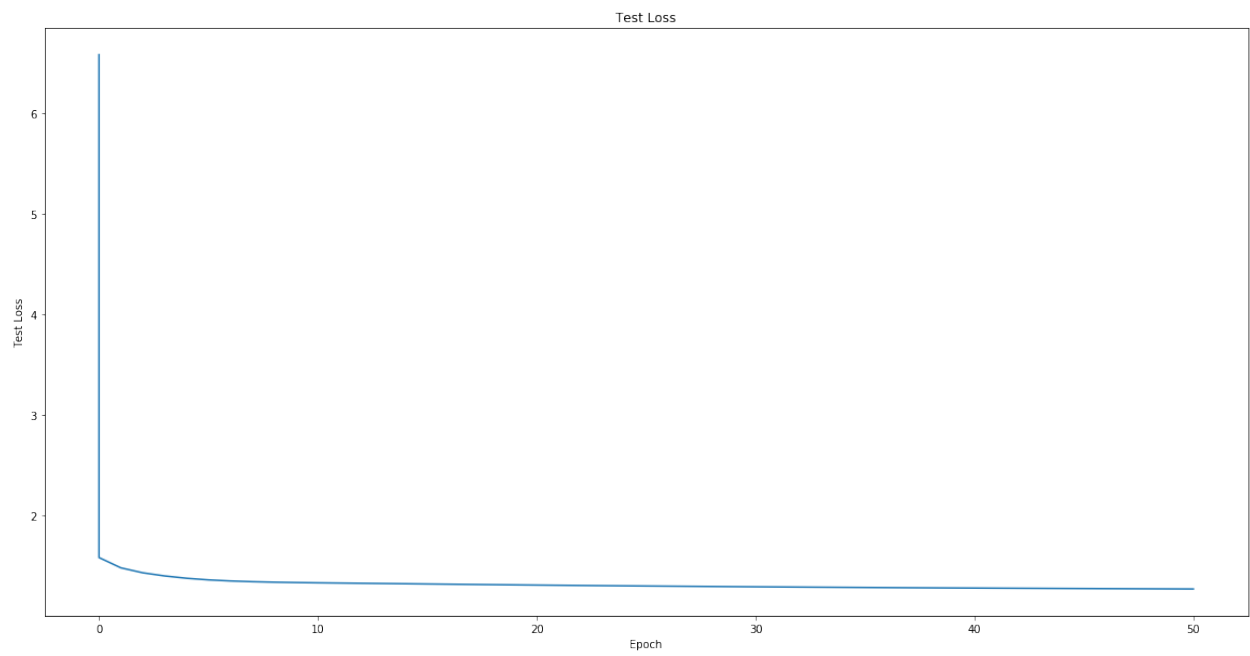
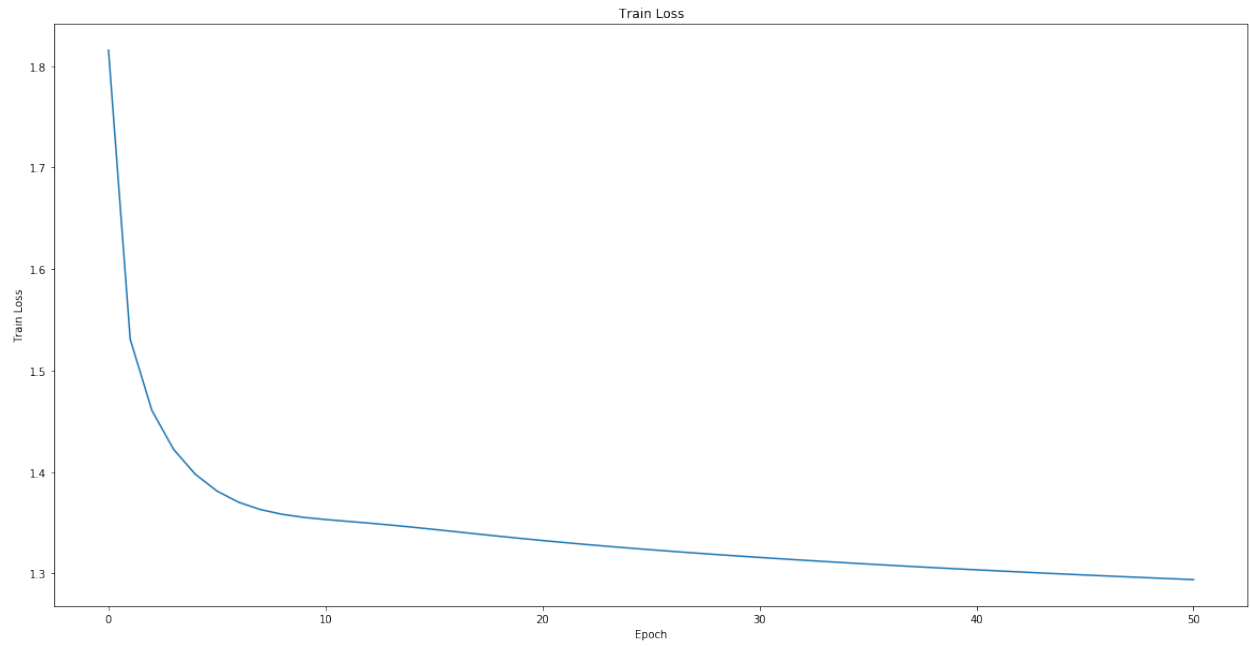
- Generated with max (temperature=0.5):
Sherlock Holmes and John Watson took a case. It was a man was a man was a man was a man was a man was a man was a man was a man was a man
- Generated with sample (temperature =0.5):
Sherlock Holmes and John Watson took a case. "Yes, sir, we were then, and the fellow when it was a far in the fortune which had been the fellow we can in the station. There are a little to see the me--and Holmes left the fore the business.
- Generated with beam (temperature =3, beam_width=10):
Sherlock Holmes and John Watson took a case. Well," said he. "It is not the facts which was an instant the morning the first in the morning the man was the man was the first the man, which was in the facts which was in the facts."

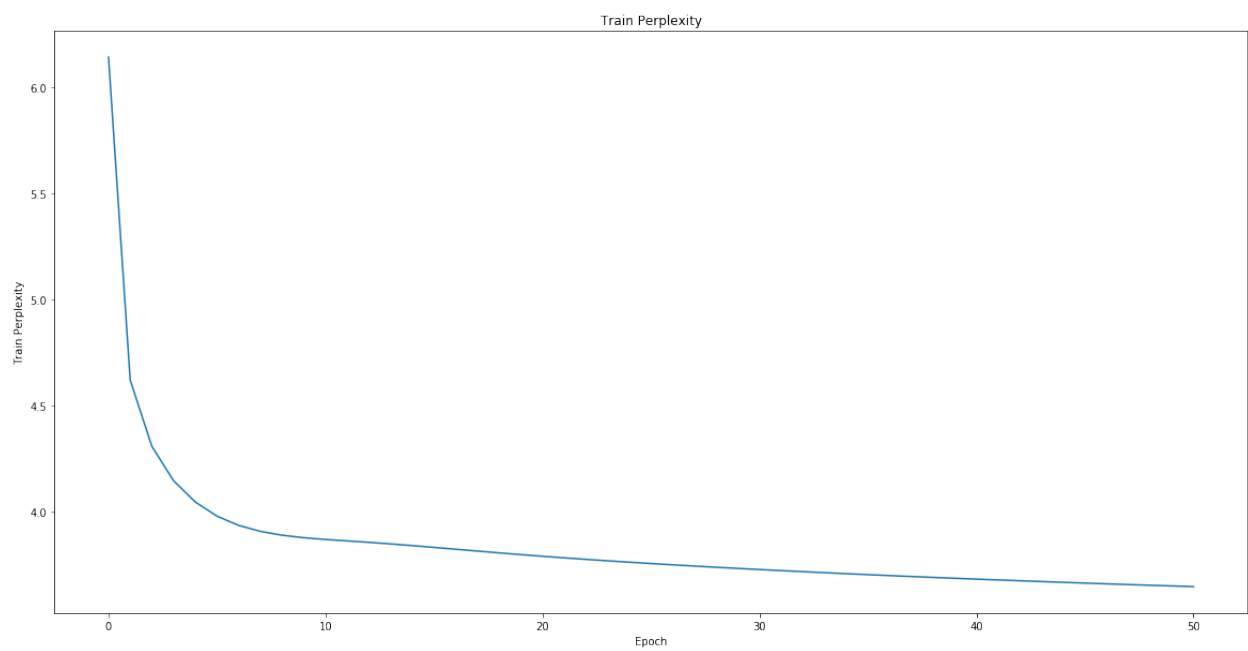
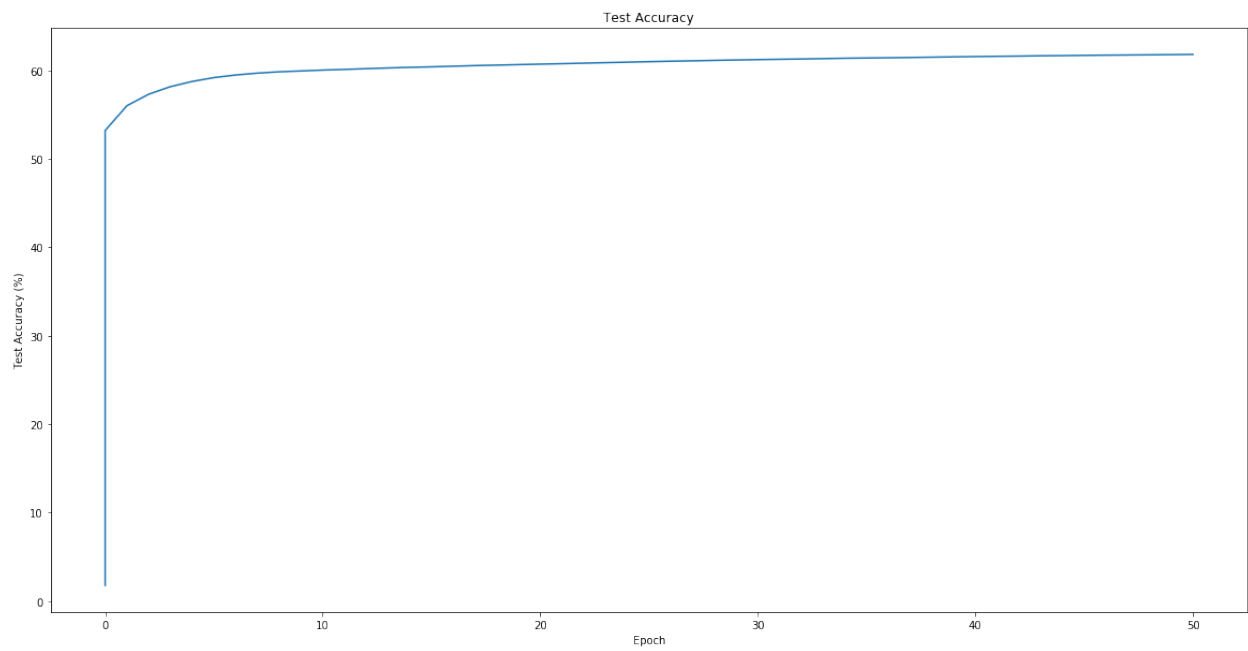
2. New Architecture

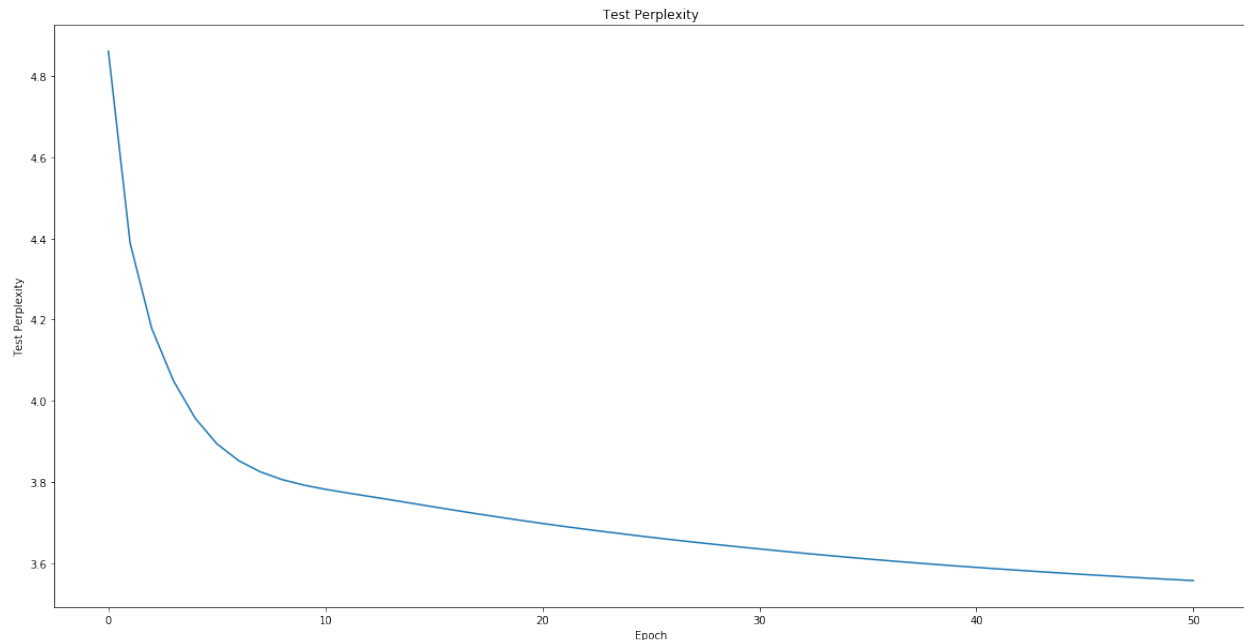
- **What was your design? What did you try that didn't work well?**
- **What was your lowest test perplexity? Provide training and testing plots.**
- **Provide outputs for each sampling method on the new corpus (you can pick one temperature, but say what it was).**

Based on the original network, I added num_layers=3. I also tried adding bias=False but it doesn't work well.

With the new design, the lowest test perplexity is 3.556. The final test accuracy is 62% after 50 epochs.







Prompt: Harry Potter love Potions.

- Generated with max (temperature=0.5):
Harry Potter love Potions. "Well, I don't know what he was still there were the only one who was still because the counter of the corridor was still staring at the corridor to the floor of the corridor to the floor
- Generated with sample (temperature =0.5):
Harry Potter love Potions. But there was her face and from the room was shaking and started at the train was walking at the room. "There's the front of the same with her hair and when they had been every particularly half parch
- Generated with beam (temperature =3, beam_width=10):
Harry Potter love Potions. But the train was still, and the thing was the room was back into the floor behind his hand was through the corridors and the corridor was still breathlessly. The wall wasted the corridors and the door

3. LSTM

- **What new difficulties did you run into while training?**
- **Were results better than the GRU? Provide training and testing plots.**
- **Provide outputs for each sampling method on the new corpus (you can pick one temperature, but say what it was).**

I used LSTM instead of GRU in the HarryPotterNet. The output of LSTM is a hidden state and a cell state, so I spent some time understanding the difference in forward function and inference function.

The test accuracy converges faster than using GRU, with the same hyperparameters. Both get 61% test accuracy after 50 epochs.

