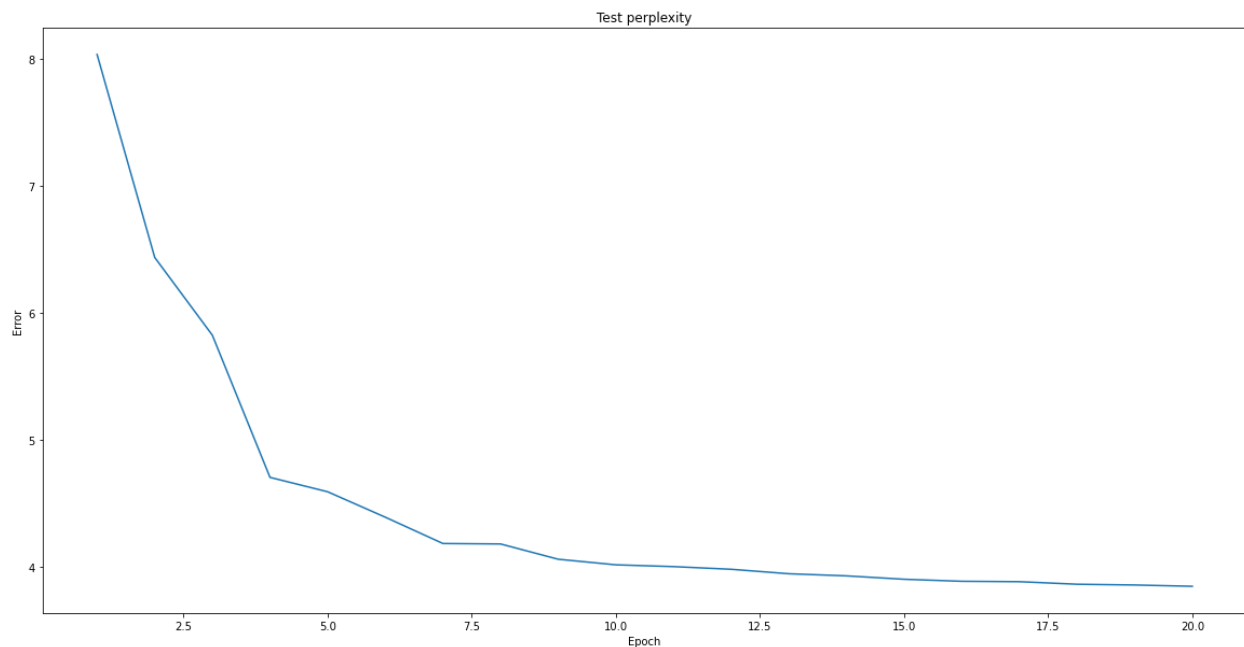
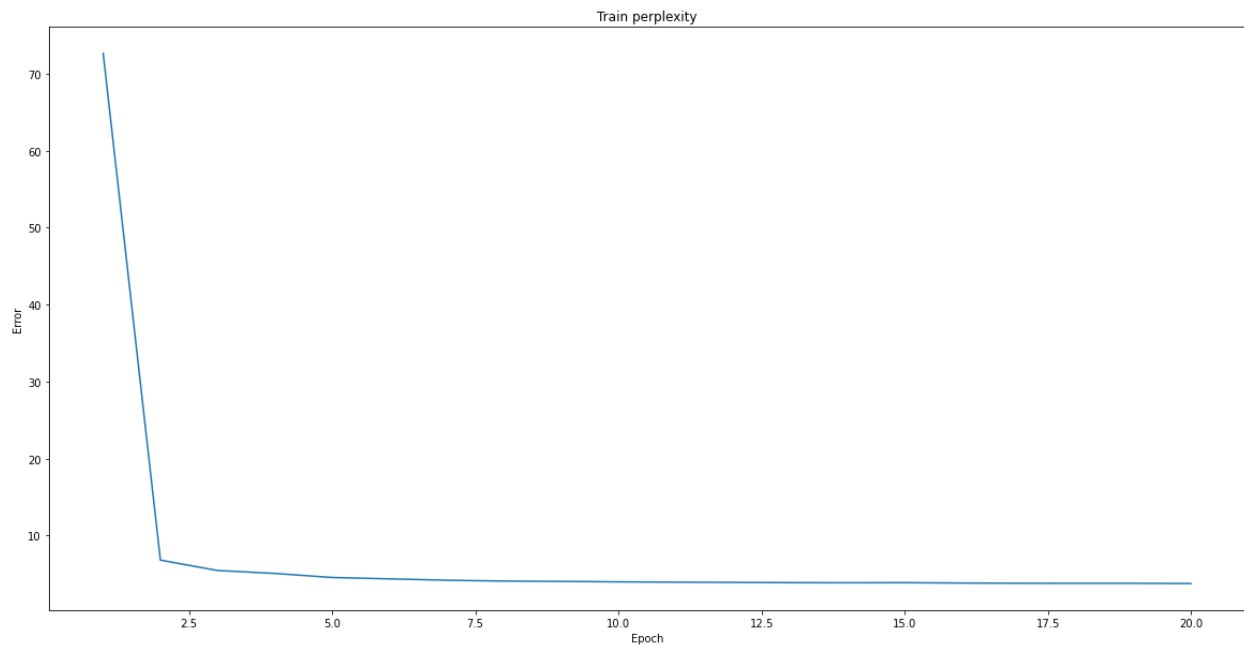


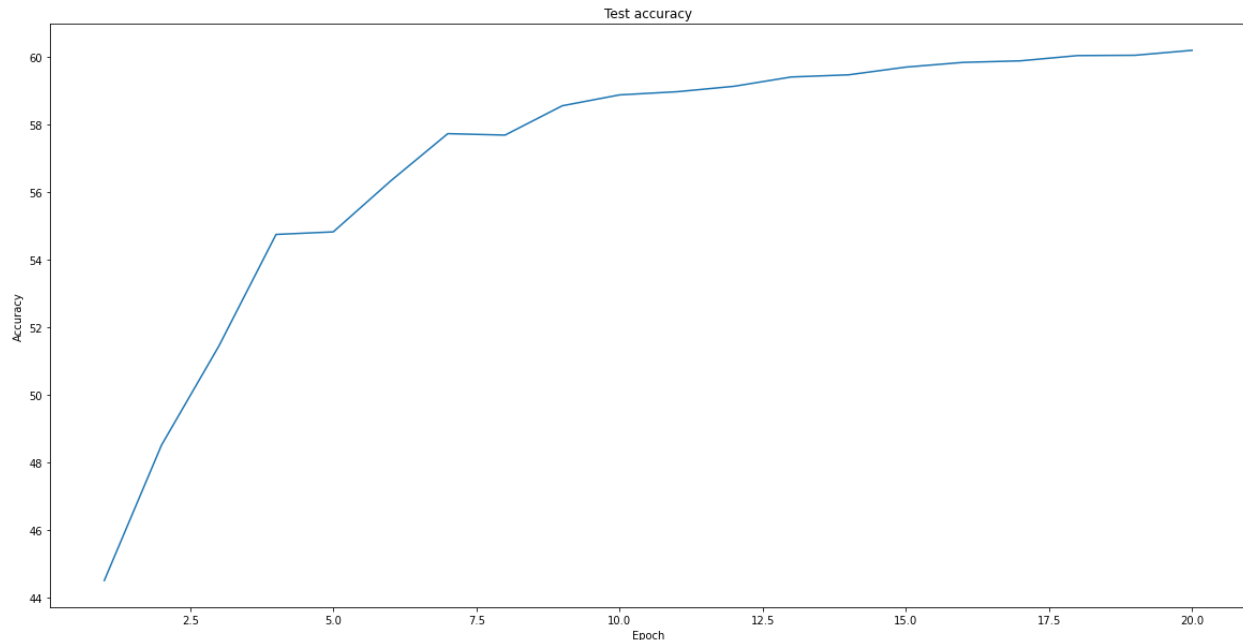
HarryPotterNet Default

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.

- a. Please list the modifications and hyperparameters.

We slightly changed some hyperparameters: 20 epochs, a learning rate of .002, weight decay of .0005, and a batch size of 256





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

Our final test accuracy was 60.197% Our final test perplexity was 3.844

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

(prompts are italicized)

Max: *Harry Potter climbed the stairs* and the stairs and the staircase the stairs and the staircase the staircase the

Sample:

Harry stared at his PHP code. "Fick the word thing?" said Pole into the floor.

Sample 2, electric boogaloo:

Harry gasped with his dying breath, "I died by a well go in the stop perchecy and then not wondering, her ha. Snape a

Beam Search:

With a flick of his wand, *Voldemort turned Harry into* his head. Harry had not had been his head. His hand was the point of t

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

Sample seemed to generate the best results since both max and beam tended to be far more repetitive. Sample - with the right temperature - produced balanced results which weren't too random/ incoherent or repetitive. We believe this is because sampling captured enough of the token distribution to be somewhat grammatically correct, but still was random enough to produce novel text.

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.

We found that a temperature between 4 and 5 worked well for beam search.

- What does a temperature of 0 do?
Sample would produce the same results as max as the output distribution is weighted heavily towards the most likely token.
- What does a temperature of $0 < \text{temp} < 1$ do?
It weights the distribution more heavily towards the more likely tokens.
- What does a temperature of 1 do?
A temperature of 1 uses the learned output distribution of the model without skewing it.
- What does a temperature of above 1 do?
It reduces the relative weight of more likely tokens, resulting in overall more random outputs.
- What would a negative temperature do (assuming the code allowed for negative temperature)?
A negative temperature would make tokens that were relatively less likely be relatively more likely and vice versa. "Flipping" the probability distribution.

New Corpus

1. What corpus did you choose? How many characters were in it?
We used the full text of the Hitchhikers Guide to the Galaxy book series. It is 1,547,384 characters long.
2. What differences did you notice between the sentences generated with the new/vs old corpus.
The new sentences were more likely to have spelling errors or otherwise nonsensical words. This is probably because the corpus had many nonsense words in it which were used infrequently. Additionally, the names of characters in the books were generated rather than Harry Potter characters.
3. Provide outputs for each sampling method on the new corpus (you can pick one temperature, but say what it was).
Max:
What is the meaning of life, the universe and everything? "Well that the mean what the stared the are a stared the means. "I t
Sample (temp 0.2):
What is the meaning of life, the universe and everything? "Well, "I don't know," said Ford. "Well you know," he said, "that th
Beam (temp 4.5):
What is the meaning of life, the universe and everything? said Zaphod, heat.." he said. "What?" said Zaphod, heaved... it is..

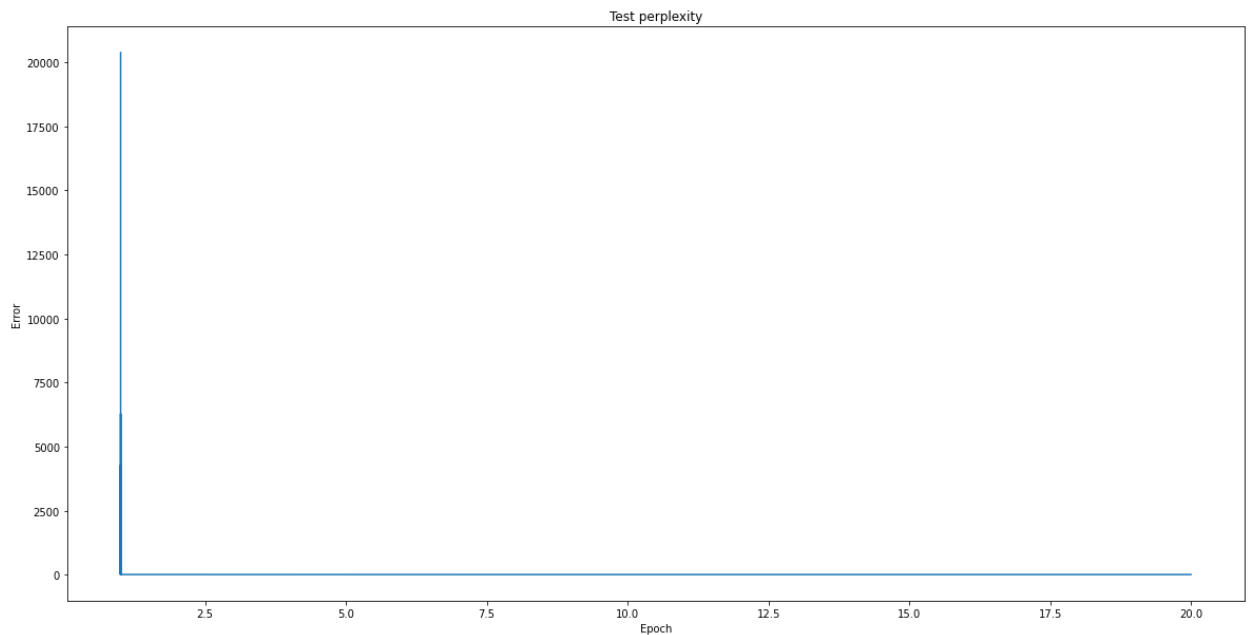
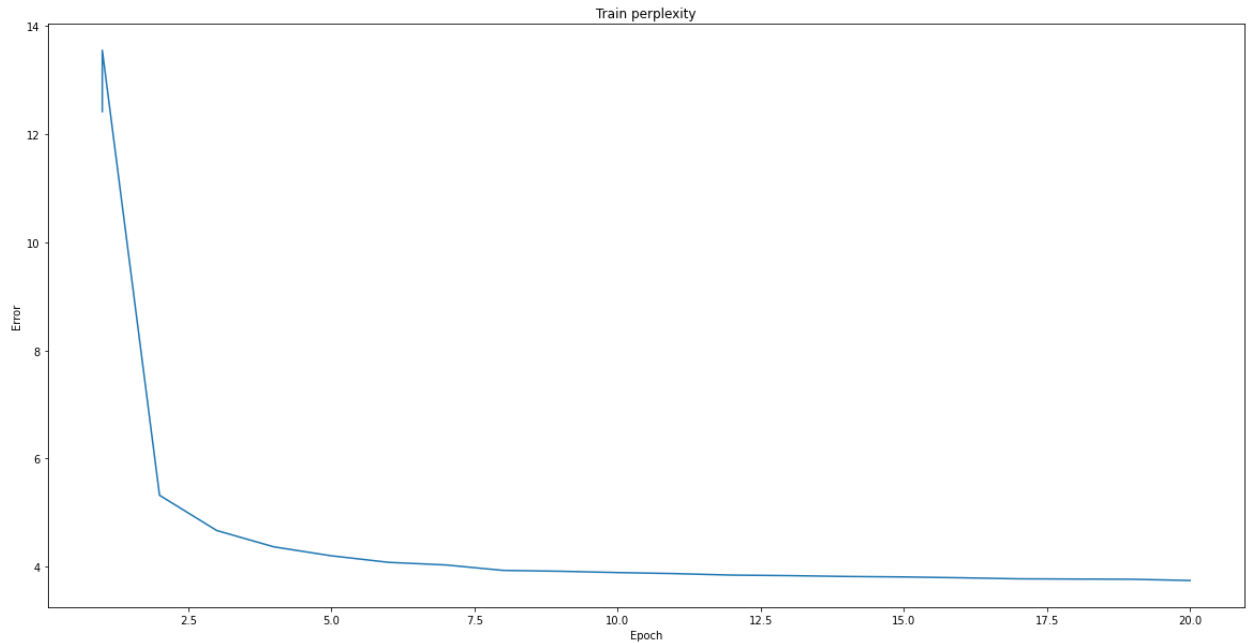
LSTMs instead of GRUs

1. What new difficulties did you run into while training?

There fortunately weren't any substantial difficulties when training the LSTM model.

2. Were results better than the GRU? Provide training and testing plots.

The LSTM model produced very similar results to the GRU. Both had a final test accuracy of ~60% and the



3. Provide outputs for each sampling method on the new corpus (you can pick one temperature, but say what it was).

Temperature: 0.5

Max: Harry Potter and the stairs were still and started to his face and started to his face and started to

Sample: Harry Potter and the world had been standing in the cloak and shrieked to the corner on the spl to h

Beam: Harry Potter and the corner of the corner of the stairs and the stairs was still starting to his face

Training on Words

1. What new difficulties did you run into while training?

Generating the tokenizer was challenging, and punctuation and capitalization had to be removed to improve the results.

2. How large was your vocabulary?

8269 words

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?

Our results were relatively good without changing the hyperparameters when training, although we reduced the sequence length to 20 for our final outputs.