Section 1: GRU (Harry Potter)

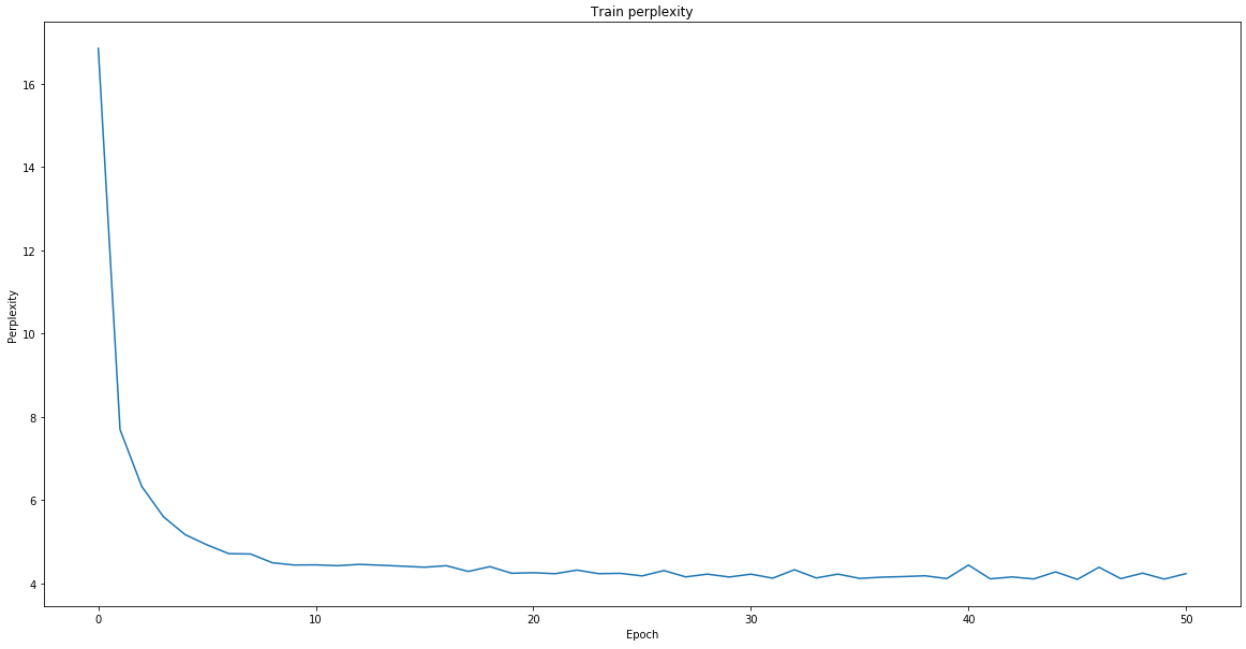
1. Just like last time, provide plots for training error, test error, and test accuracy. Also provide a plot of your train and test perplexity per epoch.
    - In class we defined perplexity as $2^{plog_2(q)}$, However the PyTorch cross entropy function uses the natural log. To compute perplexity directly from the cross entropy, you should use $e^{pln(q)}$.
    - We encourage you to try multiple network modifications and hyper-parameters, but you only need to provide plots for your best model. Please list the modifications and hyper-parameters.

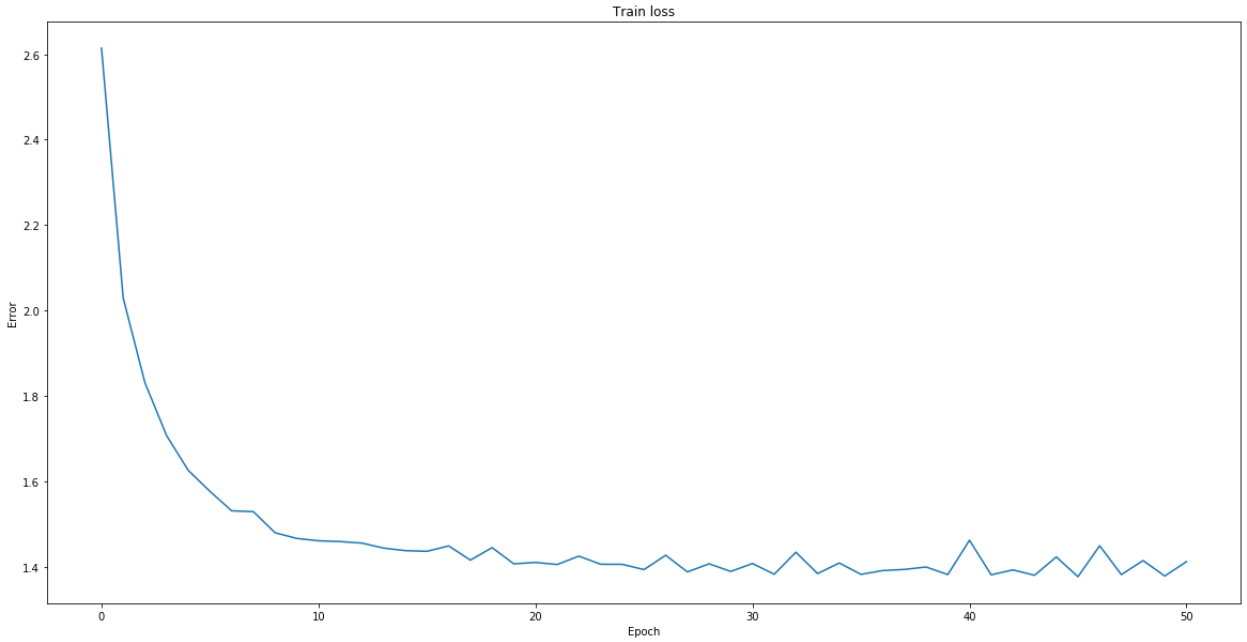
Settings:

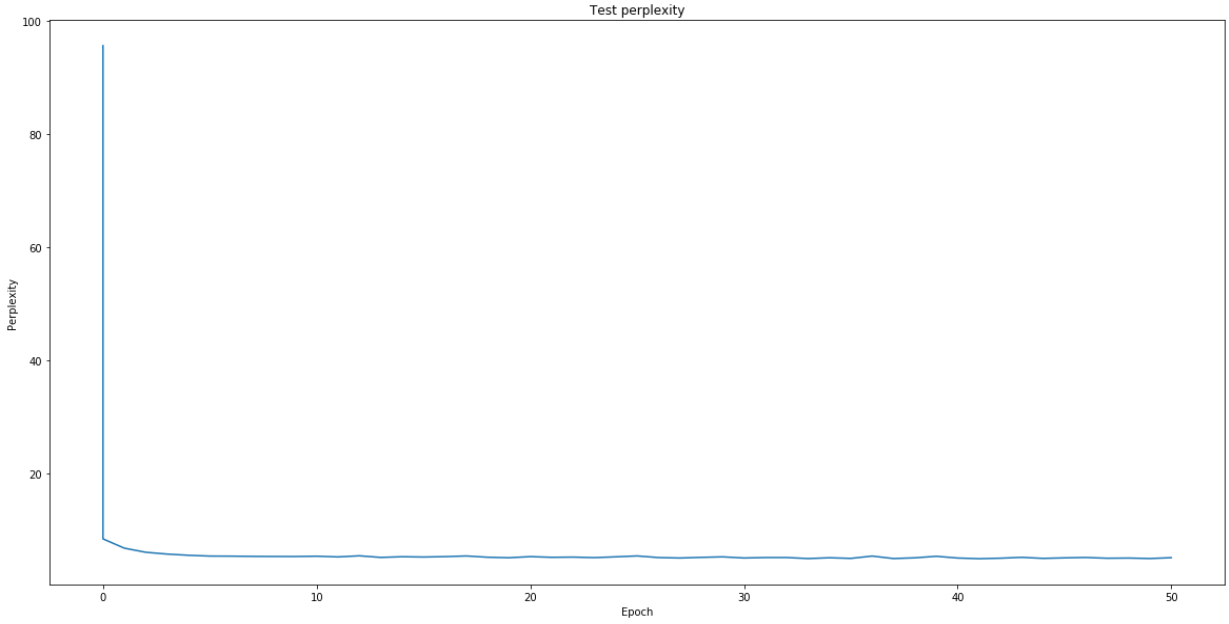
- **Batch size = 1000**
- **Sequence length = 200**
- **Test batch size = 256**
- **Learning rate = 0.002**
- **Number of layers = 1**
- **Weight decay = 0.0005**
- **Number of epochs = 50**
- **Feature size = 512**

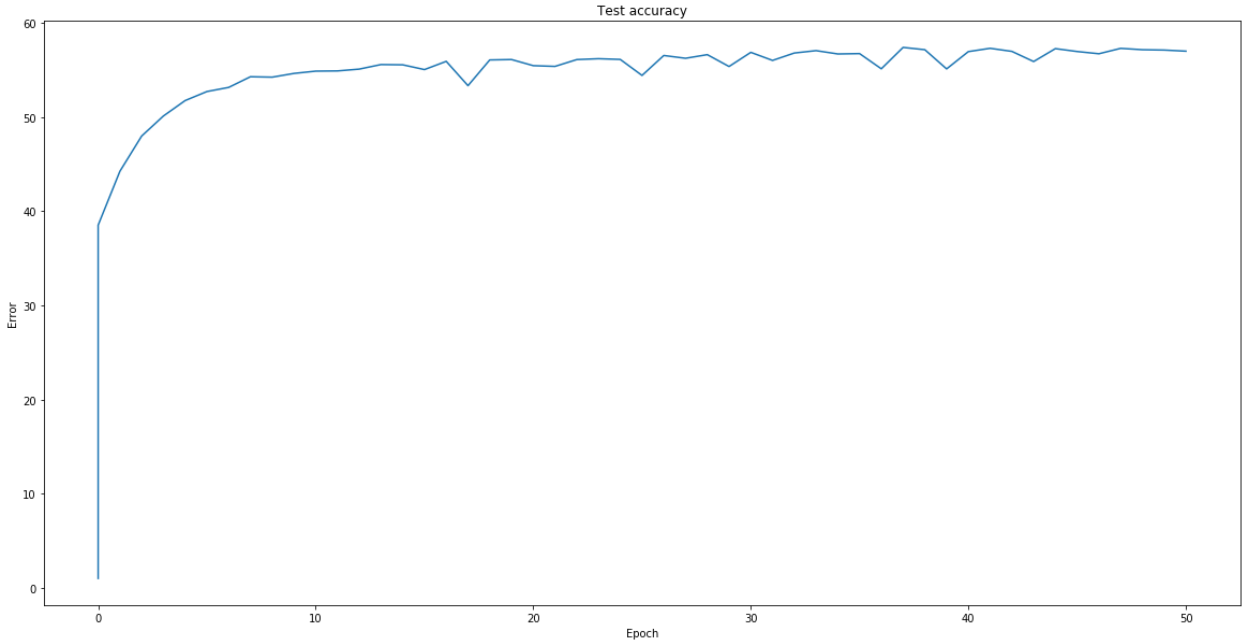
Network architecture:

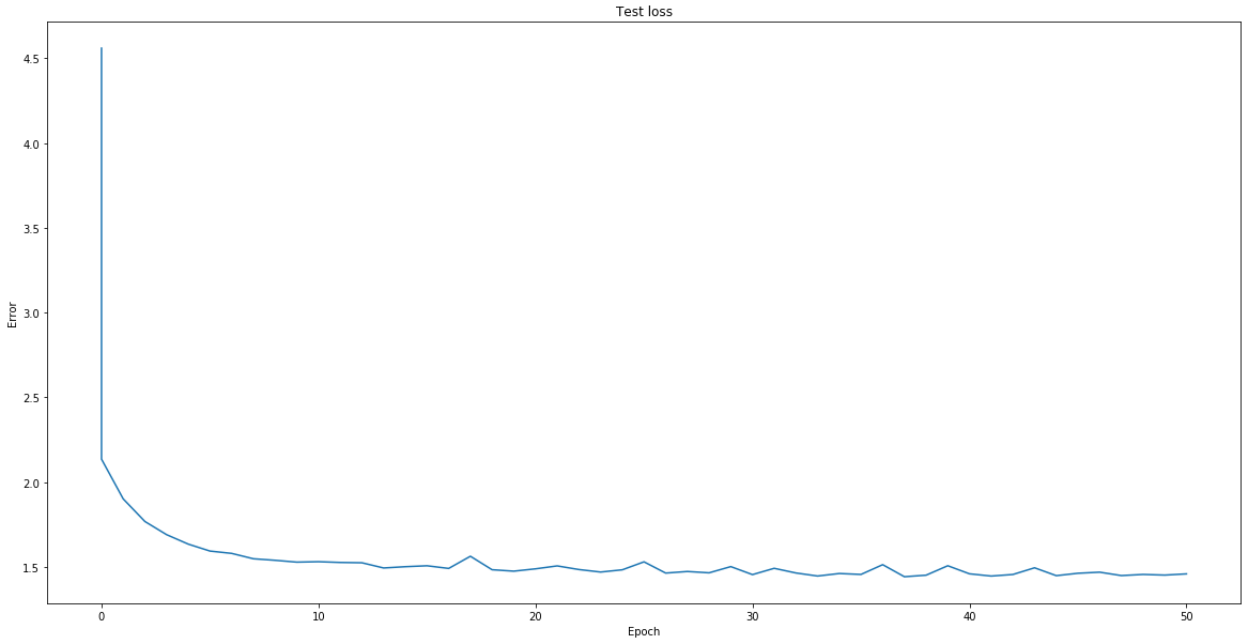
```
HarryPotterNet(
  (encoder): Embedding(95, 512)
  (gru): GRU(512, 1024, batch_first=True)
  (decoder): Linear(in_features=1024, out_features=95, bias=True))
```




Train loss
Error
Epoch
2.6
2.4
2.2
2.0
1.8
1.6
1.4
0
10
20
30
40
50


Test perplexity
Perplexity
Epoch
100
80
60
40
20
0
10
20
30
40
50





2. What was your final test accuracy? What was your final test perplexity?
   My final test accuracy is 58%. My final test perplexity is 5.4.

3. What was your favorite sentence generated via each of the sampling methods? What was the prompt you gave to generate that sentence?
   With temperature 0.3:
   - **My favorite sentence generate by max sampling is "`Spells casted by the stairs and the stairs was a start of the stairs and started to the stairs and started to the stairs and started to the stairs and started to the stairs and started to the`**

stairs and started to" which is generated by the prompt "Spells casted by".

- **My favorite sentence generated by random sampling is "Goblet of Magical Crouch in the way that they had not think it was still and the counter of the car with a lot of the train of the corridor and heard of the second the table and the first time and said the" which is generated by the prompt "Goblet of".**
- **My favorite sentence generated by beam sampling is "**Spells casted by the stairs and the stairs was a start of the stairs and started to the stairs and started to the stairs and started to the stairs and started to the stairs and started to the stairs and started to" **" which is generated by the prompt "Spells casted by".**

4. Which sampling method seemed to generate the best results? Why do you think that is?
   Random sampling seemed to generate the best result. This is because random sampling gives a larger variance when generating text. Both max sampling and beam sampling has a smaller variance compared to random sampling at a given temperature.

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.
     Random sampling at temperature = 0.7
     "**Spells casted by** the time, could tell the door of the deepens explaint when the barrow at Harry."Yes, the one three first thill the Dursleys do you to look to be inccGually because I could only heard the great aspect".

   - What does a temperature of 0 do? What does a temperature of 0<temp<1 do?
     A temperature of 0 makes the output distribution to be a delta function at the most likely outcome. In other words, it suppresses all outcomes other than the most likely outcome. For a temperature between 0 and 1, very unlikely outcome will be suppress while boosting most likely outcome.

   - What does a temperature of 1 do? What does a temperature of above 1 do?
     A temperature of 1 preserves the probability distribution generated by the neural network, thus giving us the regular softmax output, in other words, temperature of 1 does nothing. A temperature above 1 tends to make the probability distribution more uniform.

   - What would a negative temperature do (assuming the code allowed for negative temperature)?

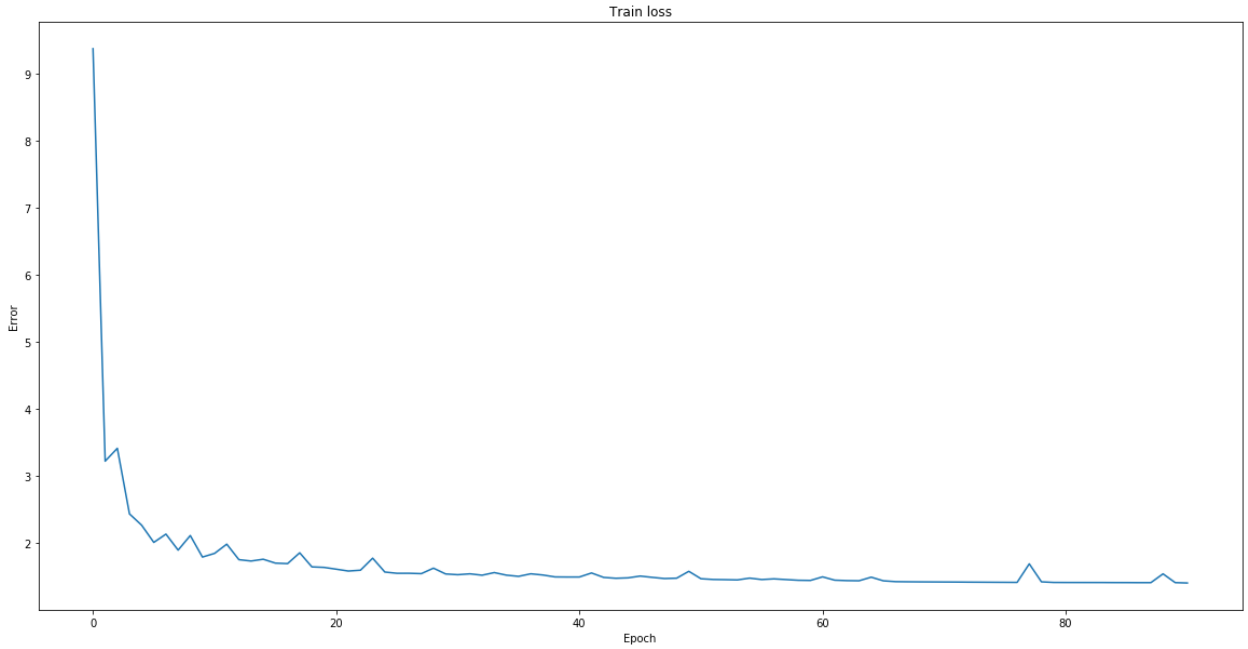
A negative temperature makes the most likely outcomes to be the least likely outcomes and vice versa.

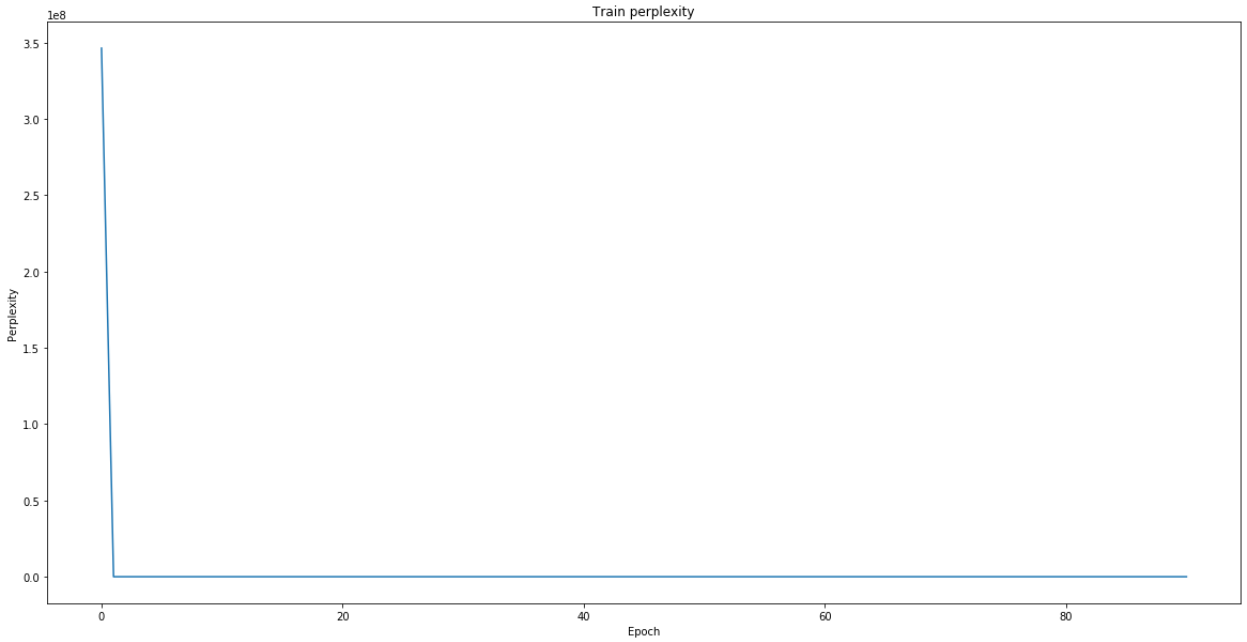
Section 2: New Corpus

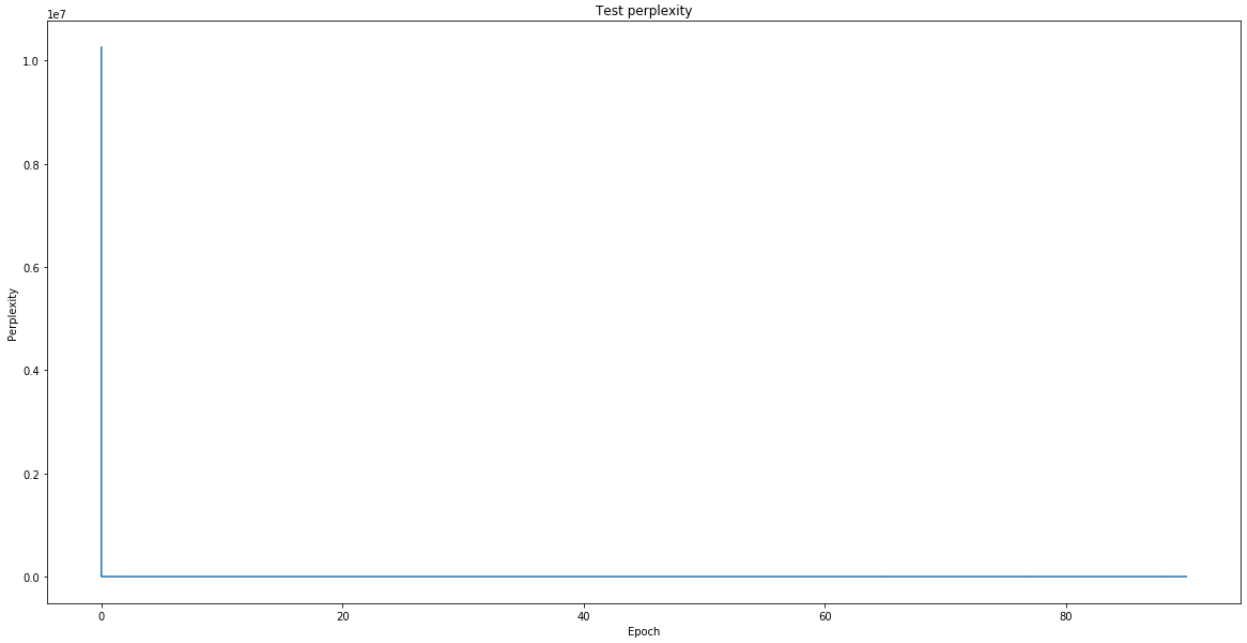
1. What corpus did you choose? How many characters were in it?
   I chose Games of thrones aka Age of Fire and Ice (5 chapters). There were 9456059 characters in it. As a reference, Harry Potter has 6180222 characters.

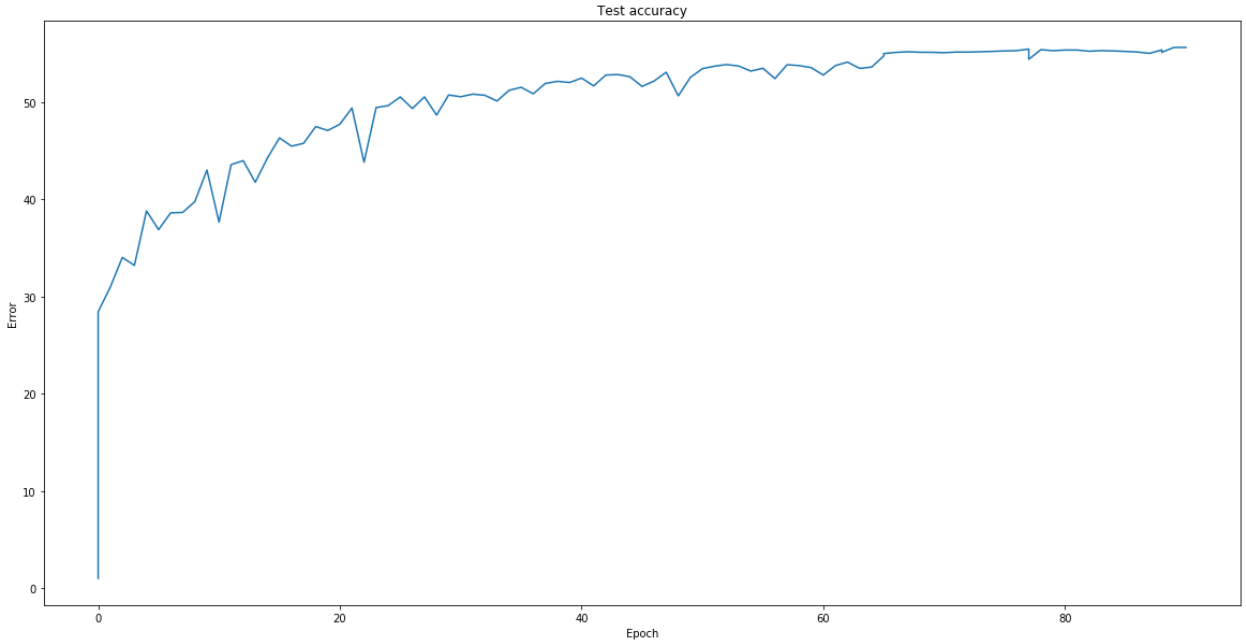
   Network architecture: (training settings is shown in part 3)

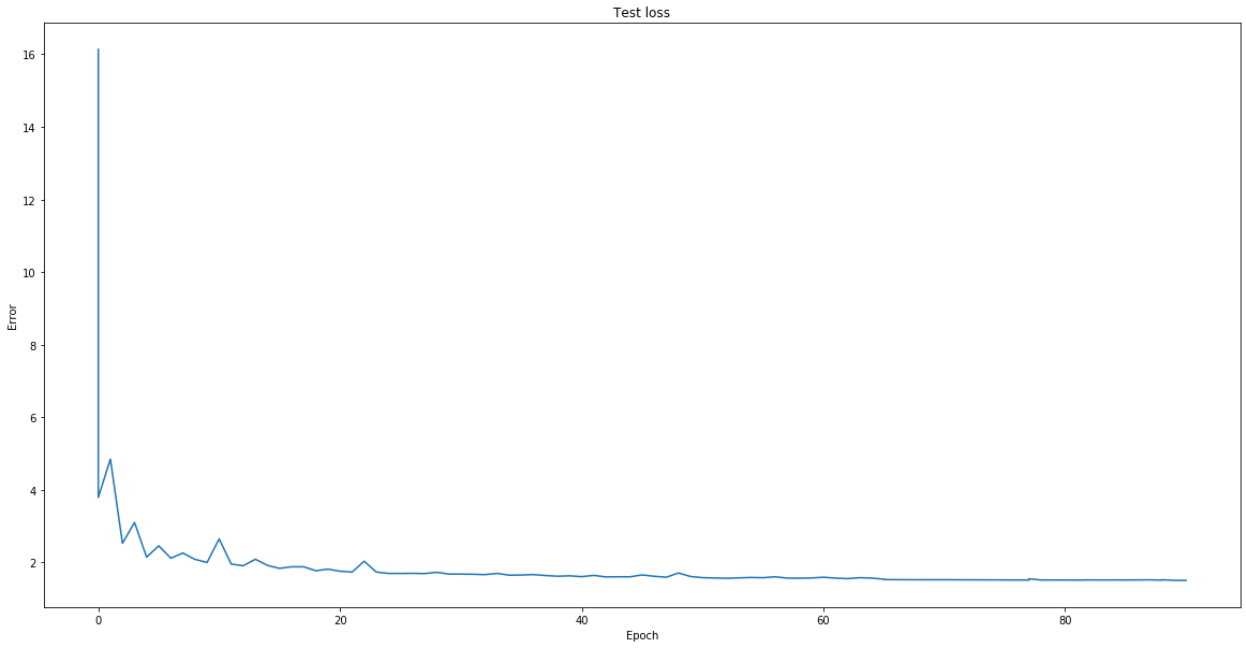
```
GotNet(
  (encoder): Embedding(96, 512)
  (gru): GRU(512, 512, batch_first=True)
  (decoder): Linear(in_features=512, out_features=96, bias=True)
)
```




Train perplexity
1e8
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Perplexity
0
20
40
60
80
Epoch


Test perplexity
1e7
1.0
0.8
0.6
0.4
0.2
0.0
Perplexity
0
20
40
60
80
Epoch





2. What differences did you notice between the sentences generated with the new/vs old corpus?
   The sentences generated Games of Thrones has more variations than the old corpus. Also, the new text makes more sense.
   After 90 epochs, the accuracy saturates at 1054034/1894400 (56%) with an average test loss: 1.4971and average test perplexity: 5.7608.

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

- **Batch size = 1000**
- **Sequence length = 200**
- **Test batch size = 256**
- **Learning rate = 0.002**
- **Number of layers = 1**
- **Weight decay = 0.0005**
- **Number of epochs = 90**
- **Feature size = 512**
- **Temperature = 0.3**

The prompt is "The dragon ate".

Generated with max sampling:

The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t

Generated with random sampling:

1. **The dragon ate the castle of the castle Baratheon and the commands of the boy was still and stranger every to the**
2. **The dragon ateven her head of the black of the walls of the fire and the courtes of the countains of the crown of**
3. **The dragon ate him. "I am the same to pie with the right of the courage to the stars. I was the same to go to the**
4. **The dragon ately the way the gods of the stone of the stone of the castle of the lands of the red beard and a com**
5. **The dragon ate a little black of her breasts. As we did not have been and beautiful to make him and stone and the**
6. **The dragon ate her brother and the first time the stars with the good of the direwolf was a child of starts of th**
7. **The dragon ate him with a shadow of the stone with the castle of the common of the dragons belly and her sister a**
8. **The dragon ate her with a stone and said, but he was a bit of steel and see the stars and brown and slayed with a**
9. **The dragon ate him and like a breath and stone and stronger and she was all at the streets of her face and starte**
10. **The dragon ates and stood and stepped and sleep and stronger to his song. The other stone was back and stood and**

Generated with beam search sampling:

1. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
2. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
3. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**

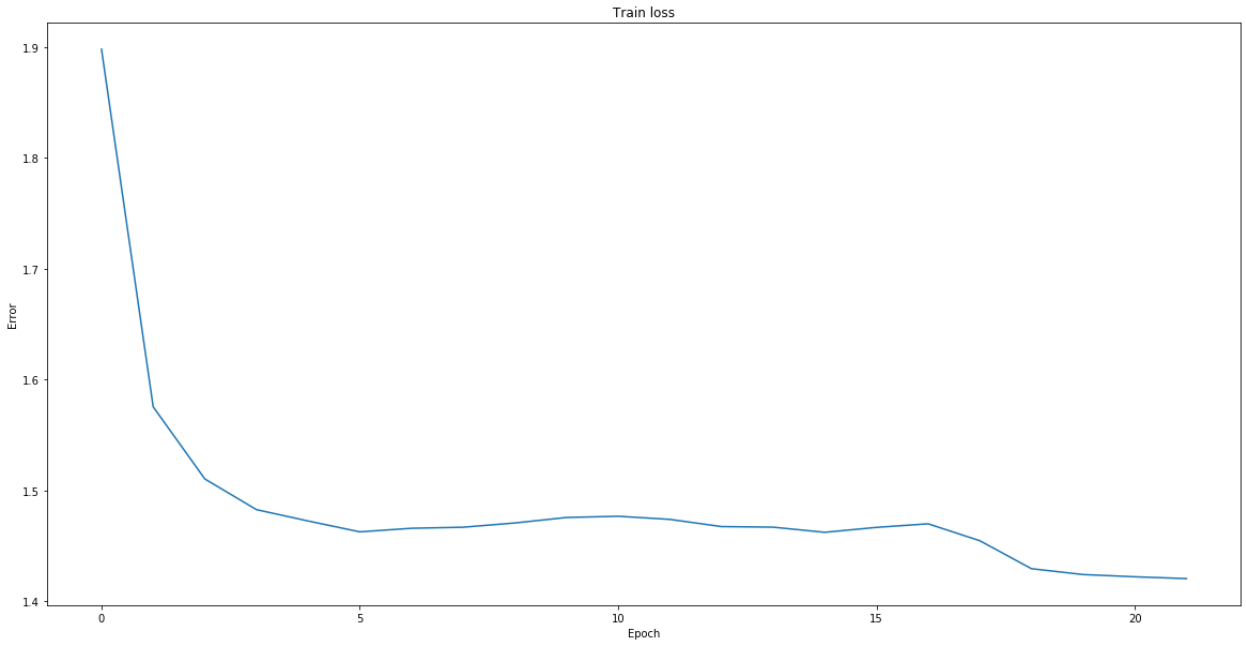
4. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
5. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
6. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
7. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
8. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
9. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**
10. **The dragon ate his sister and the streets of the stone of the courtes of the steps of the steps of the steps of t**

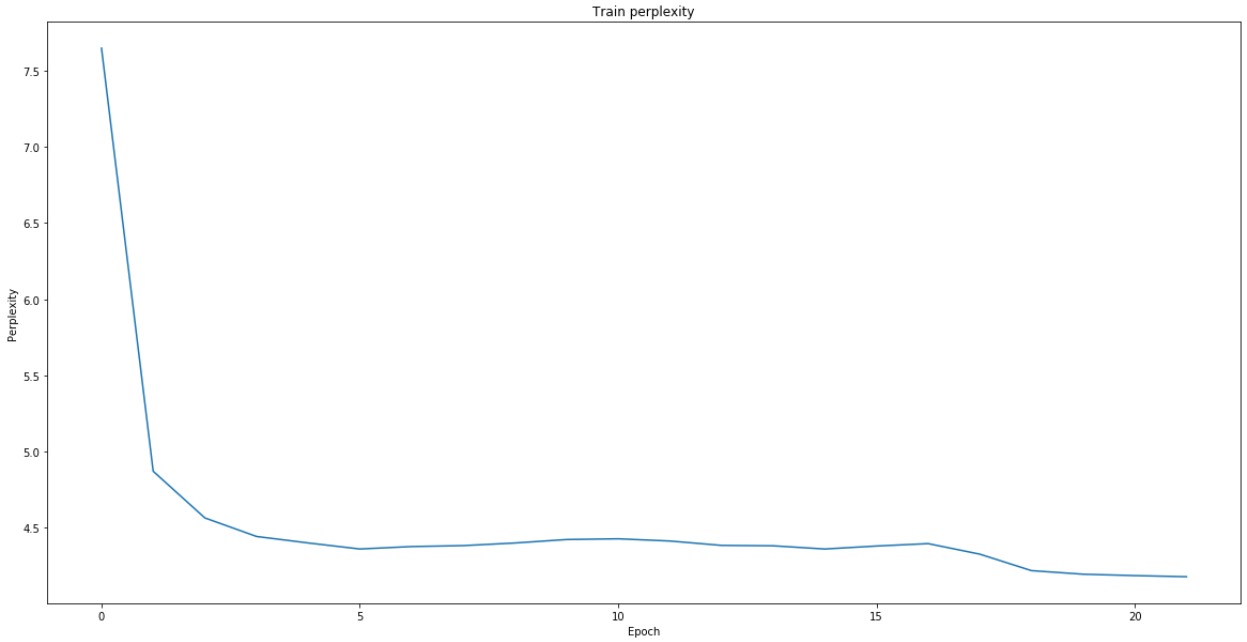
Section 3: LSTM

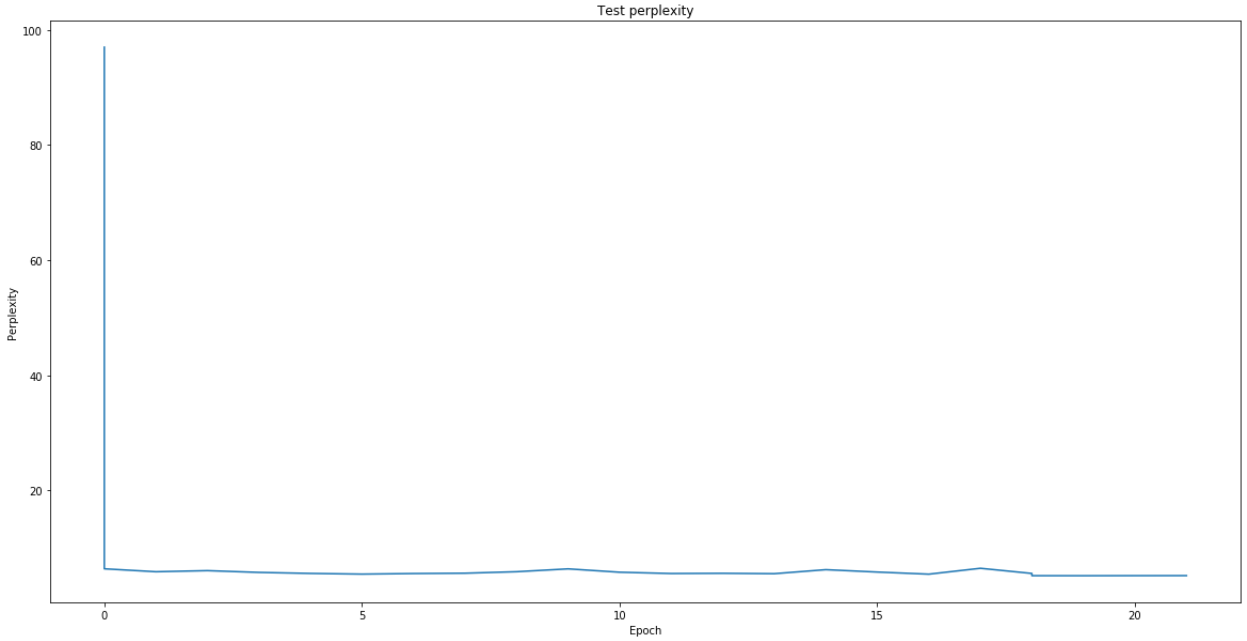
1. What new difficulties did you run into while training?
   The training is slow compared to GRU. In addition, adding more LSTM layers does not increase accuracy. The accuracy is capped around 55% with different parameters for batch size, learning rate, feature size etc.

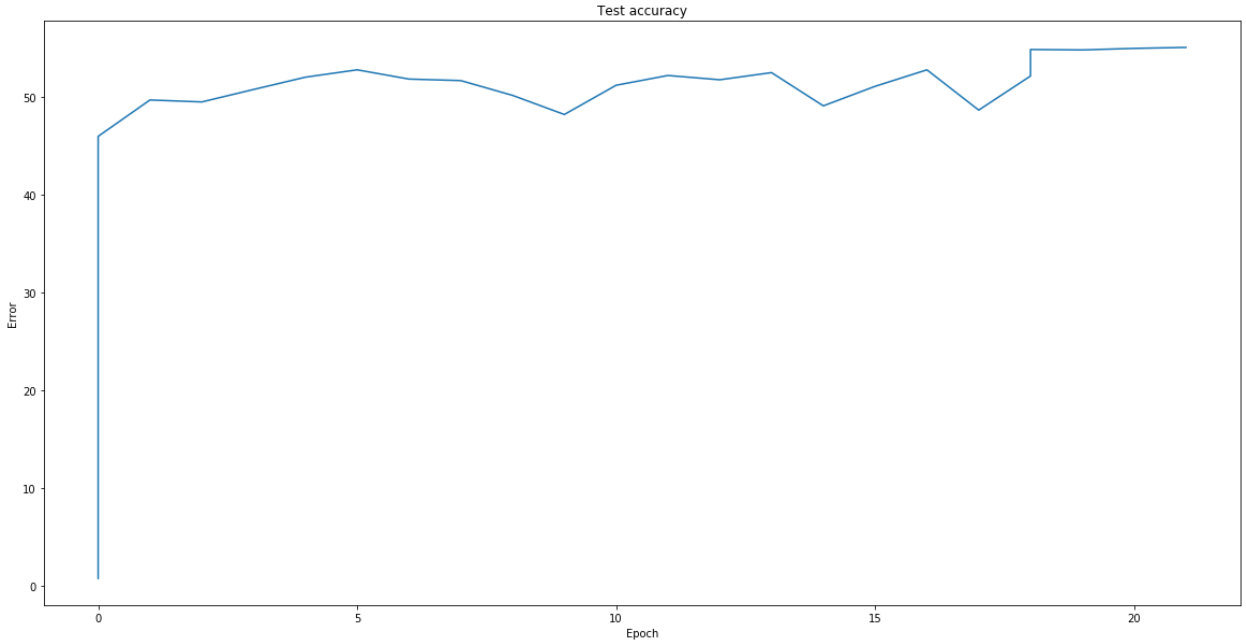
   Network architecture: (training settings is shown in part 3)

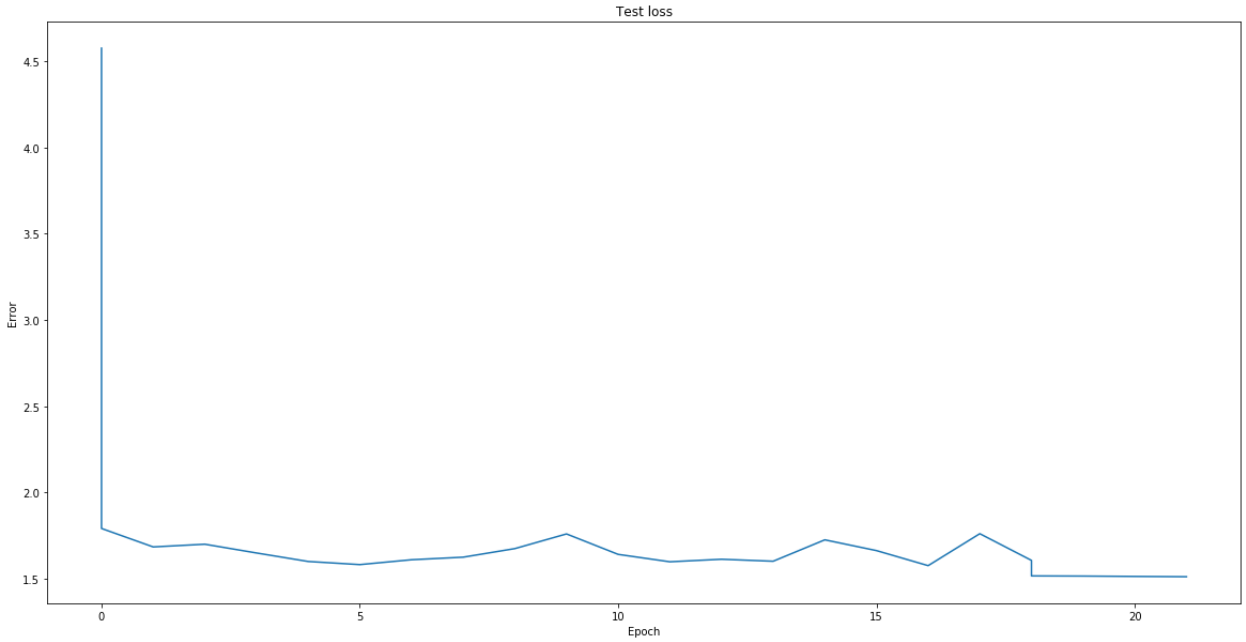
```
GotLSTMNet(
  (encoder): Embedding(96, 700)
  (lstm): LSTM(700, 700, batch_first=True)
  (decoder): Linear(in_features=700, out_features=96, bias=True)
)
```




Train perplexity
Perplexity
Epoch
7.5
7.0
6.5
6.0
5.5
5.0
4.5
0
5
10
15
20


Test perplexity
Perplexity
Epoch
100
80
60
40
20
0
5
10
15
20





2. Were results better than the GRU? Provide training and testing plots.
   No, the accuracy is slightly lower than GRU by 1% but the generated sentences are similar.
   After 20 epochs, the test loss and test accuracy starts to saturate.
   Test set: Average loss: 1.5123, Accuracy: 1044267/1894400 (55%), Average perplexity: 5.2117

3. Provide outputs for each sampling method on the new corpus (you can pick one temperature, but say what it was).
   - **Batch size = 1000**
   - **Sequence length = 200**
   - **Test batch size = 256**
   - **Learning rate = 0.002 for first 17 epoch**
   - **Learning rate = 0. 0005 after 17 epoch**
   - **Weight decay = 0.0005**
   - **Number of epochs = 20**
   - **Feature size = 700**
   - **Temperature = 0.3**

   The prompt is "The dragon ate".

   Generated with max sampling:
   - **The dragon atead the stark of the stark of the stark of the stone of the stark of the stark of the stark of the s**

   Generated with random sampling:
   1. **The dragon atead the first of the steel and closed the stard on the stalls. The gods was the direwolf stone and s**
   2. **The dragon atever the black of the starts of the black of the great silent ships of the stars. The fire was the l**
   3. **The dragon ated the stark of the chance of the steel of the starl of the dragons and shouted and brought the blac**
   4. **The dragon aternated the store the walls of the black and and her for and stones and still believed the cloak and**
   5. **The dragon ateld the real father to the stone with the dark and suddenly the stark of the stard and shoulders of**
   6. **The dragon atead her back. "He was a lady to the castle of the man of the door when the steel had been the first**
   7. **The dragon ateptace. The wind was not a crow when he stupid the castle of the courtest of the gods and the crown**
   8. **The dragon ate the stall of the starks of the wind with the drowned with the black of the river candless and swor**
   9. **The dragon ated the steel that he was a strong to the black of the other brother and she was a start of his fathe**
   10. **The dragon ate the stark of the girl and she was a little sword of the black candles with a stark to the stark of**

   Generated with beam search sampling:
   1. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**
   2. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**

3. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**
4. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**
5. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**
6. **The dragon atead the stark of the stark of the stark of the stone of the stark of the stark of the stark of the s**
7. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**
8. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**
9. **The dragon ate the stark of the stark of the stark of the stone of the stark of the stark of the stark of the sta**
10. **The dragon atead the stark of the stark of the stark of the stone of the stark of the stark of the stark of the s**

Section 4: New Architecture

1. What was your design? What did you try that didn't work well?
   My network architecture: (training settings is shown in part 3)

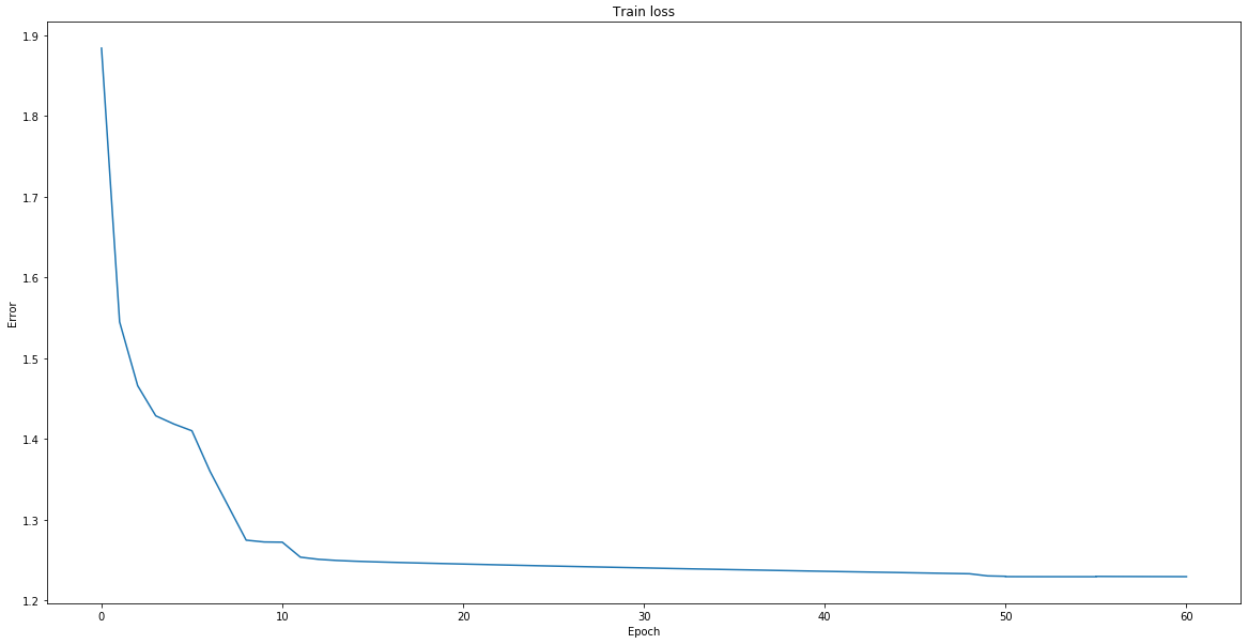
```
GotNewNet(
  (encoder): Embedding(96, 512)
  (gru): GRU(512, 512, num_layers=4, batch_first=True)
  (decoder_1): Linear(in_features=512, out_features=256, bias=True)
  (decoder_2): Linear(in_features=256, out_features=96, bias=True)
  (batchnorm_0): BatchNorm1d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (batchnorm_1): BatchNorm1d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (batchnorm_2): BatchNorm1d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (relu): ReLU()
)
```

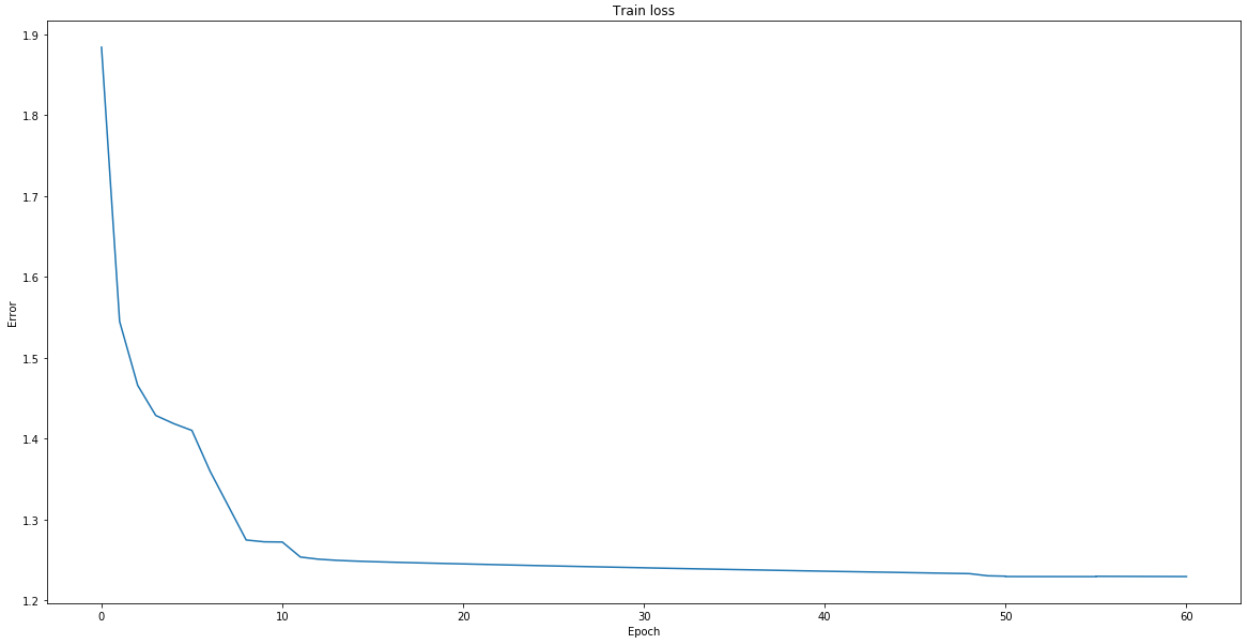
   The things that did not work:
   - **Increasing the number of features for GRU or LSTM do not work.**
   - **Increasing batch size or sequence length doesn't work.**
   - **Increasing learning rate doesn't work. It leads to divergence.**
   - **LSTM training is slower than GRU.**
   - **Transformer training is very slow as 5 min of training leads to less that 0.02% improvement in accuracy. I noticed that Adam stop providing improvement when accuracy hits 27% even with smaller or larger learning rate, but stochastic gradient descent seems better than Adam providing miniscule but improving loss rate. On hindsight, it is not a great idea to train a transformer using a single GPU for homework (waste of time).**

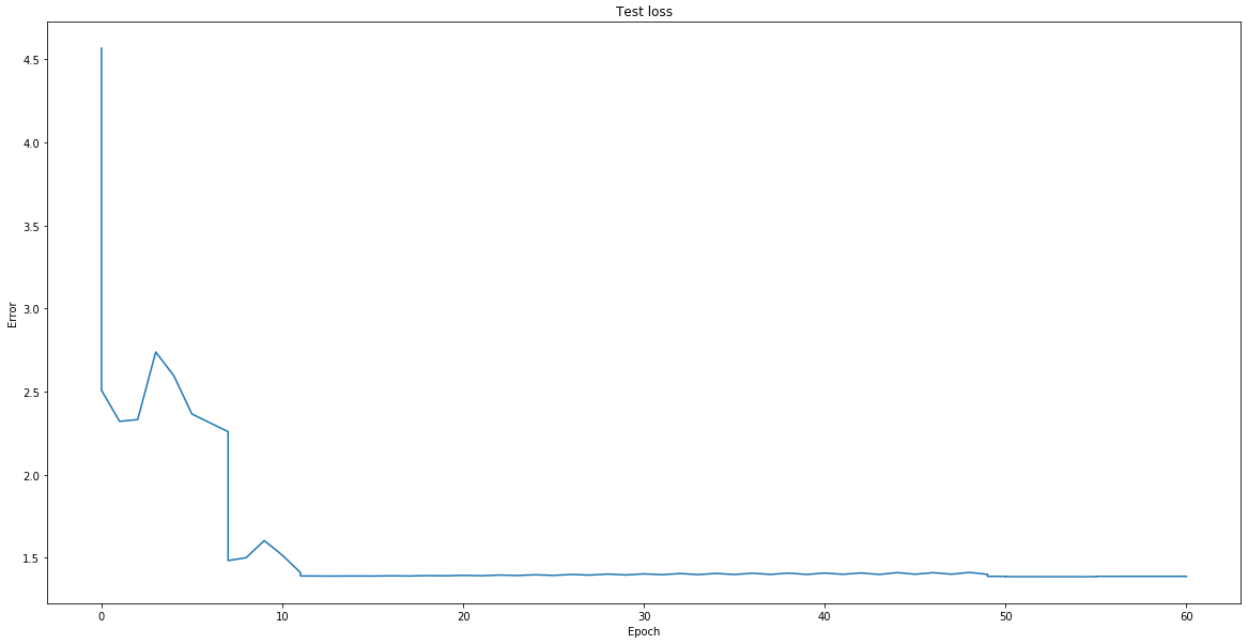
2. What was your lowest test perplexity? Provide training and testing plots.
   My lowest test perplexity is 6.1671. Also, test average loss: 1.3880, test accuracy: 1114564/1894400 (59%) after 60 epochs.

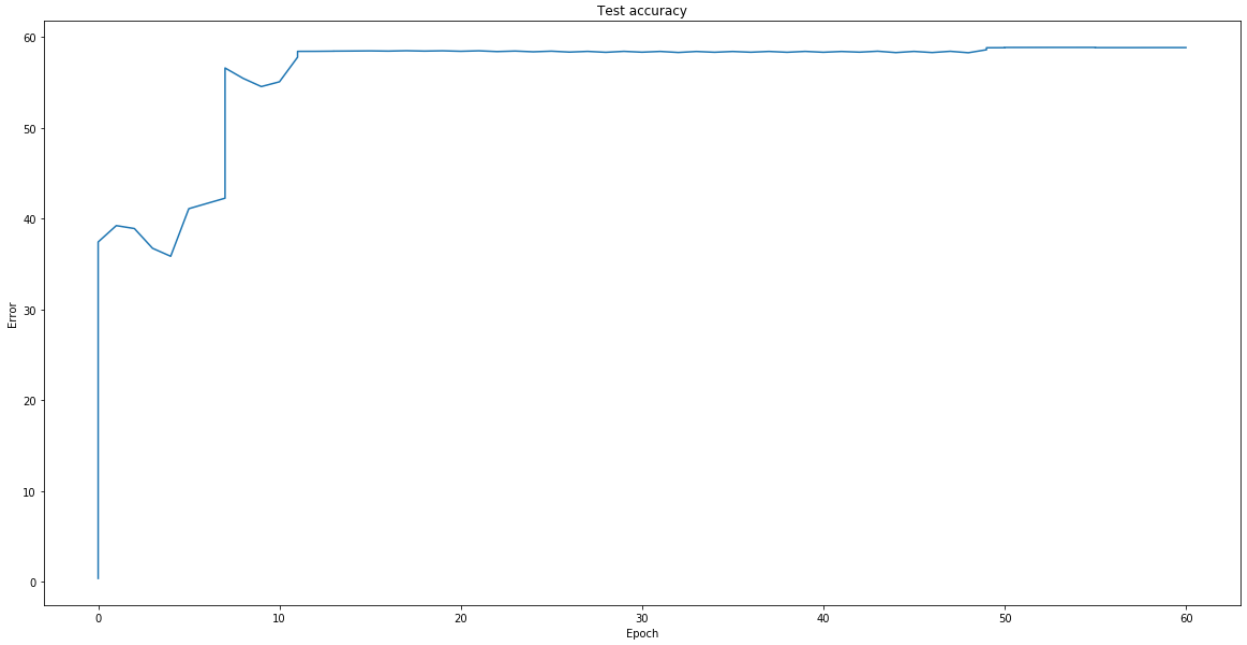
Settings:

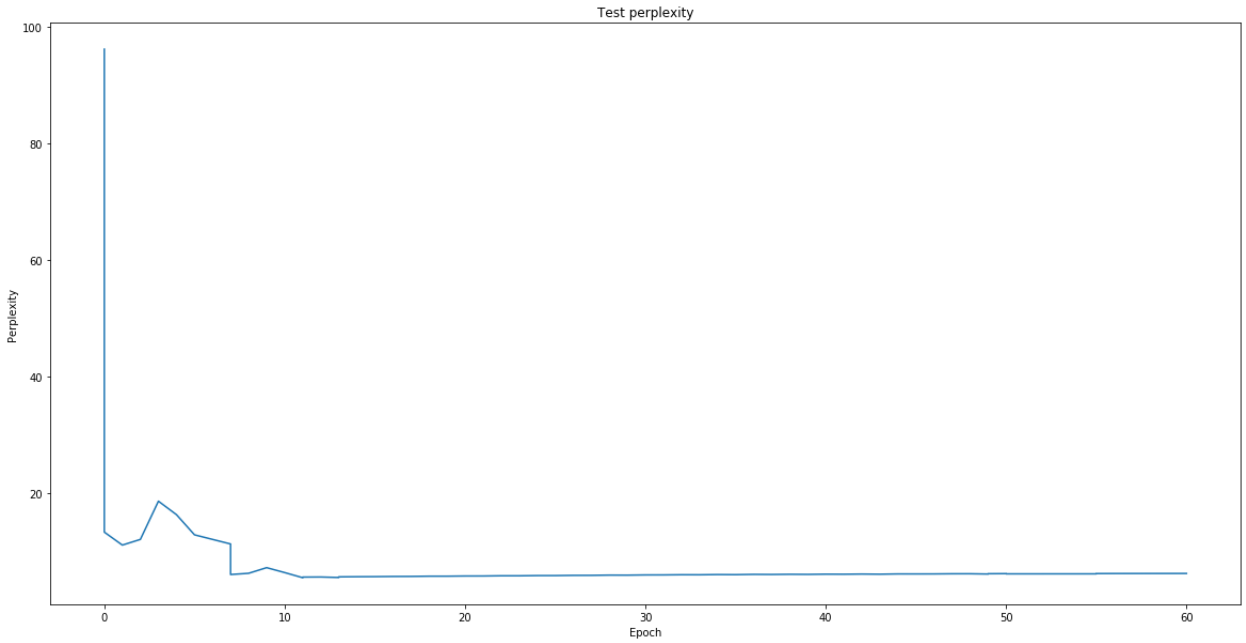
- **SEQUENCE_LENGTH = 200**
- **BATCH_SIZE = 256**
- **FEATURE_SIZE = 512**
- **NUM_HEAD = 4**
- **NUM_HID = 1024**
- **NUM_LAYERS = 4**
- **DROPOUT = 0.05**
- **TEST_BATCH_SIZE = 256**
- **EPOCHS = 60**
- **LEARNING_RATE = 0.002 for first 10 epochs**
- **LEARNING_RATE = 0.0002 for between 10 and 49 epochs**
- **LEARNING_RATE = 0.00002 for after 49 epochs**
- **WEIGHT_DECAY = 0.0005**


Train loss
Error
Epoch
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
0
10
20
30
40
50
60


Train loss
Error
Epoch
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
0
10
20
30
40
50
60


Test loss
Error
Epoch
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0
10
20
30
40
50
60


Test accuracy
Error
Epoch
60
50
40
30
20
10
0
0
10
20
30
40
50
60



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

Temperature = 0.3
Prompt is "The white dragon ate".

Generated with max sampling:

- **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**

Generated with random sampling:

1. **The white dragon ate the battle of the walls of the shadows of the cold grey arms and strong and strong the shadows and**
2. **The white dragon ate the trees of the boy and the shores of the sea of the color of the three portcullis. The sea of th**
3. **The white dragon ate, and the last of the way a third that was a great stone thousand stone while the fire was still an**
4. **The white dragon ate a side, and the boy was still a doublet of many silver and strong and the captains of the way the**
5. **The white dragon ate his lips. The boy was still a sword of the hall of a late with the wall of the bed strong and stro**
6. **The white dragon ate a strange and stood of the steps of the eldest steel of the stone room and closed the steel of the**
7. **The white dragon ate his head and she was all and some shield on the fire in the stone of the steps of the earth of the**
8. **The white dragon ate a stone thing and the shadows and the steps and the poor red stone was still and strong and strong**

9. **The white dragon ate the stone of the same the stone serving was strong beneath the darkness of the room. The sea was s**
10. **The white dragon ate the door, and the old man was a brother of the Trident of the sea. Arya still be stripping to the**

Generated with beam search sampling:

1. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
2. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
3. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
4. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
5. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
6. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
7. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
8. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
9. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**
10. **The white dragon ate his head. "I was a strong to the sea, and the sea of the sea of the way the streets of the sea. Th**