Part 9: Short answer questions

Please answer these questions, and put the answers in a file called short_answer.pdf in your repository.

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 $2^{p*log_2(q)}$, However the PyTorch cross entropy function uses the natural log. To compute perplexity directly from the cross entropy, you should use e^p*ln(q).

   code:

   ```
   test_prep = np.exp(test_loss)
   print('\n test prep', test_prep)
   ```

   - 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.

   I have tried more layers in gru, like layer 2, and more gru layer, like 3.

   from `self.gru = nn.GRU(self.feature_size, self.feature_size, num_layers=1, batch_first=True)` to `self.gru = nn.GRU(self.feature_size, self.feature_size, num_layers=2, batch_first=True)`

   For the hyperparameters, I have tried the bitch_size = 64,128, 256. For the sequence, i tired 100, 200. After all, the best model's test accuracy is about 60%. It will reach the 60% about 15 epochs and the error will decrease a little in the following epochs. The following is training error, test error, and test accuracy. for the following, I tried in temperature is 1 and layer of gru is 2. other are same with defult.

   print:

   ```
   Test set: Average loss: 1.3397, Accuracy: 748654/1254400 (60%)


    test prep 3.8178807924359304
   Saved /content/checkpoints/020.pt
   ```

Train loss

Train prep

Test prep

Test loss
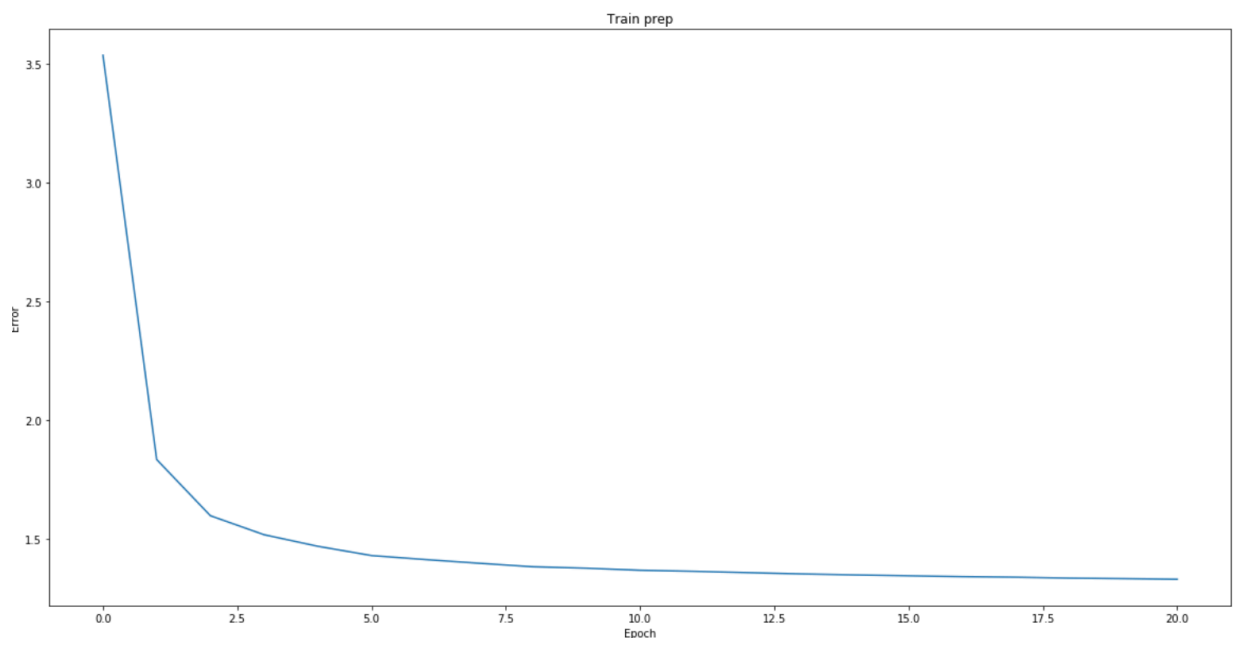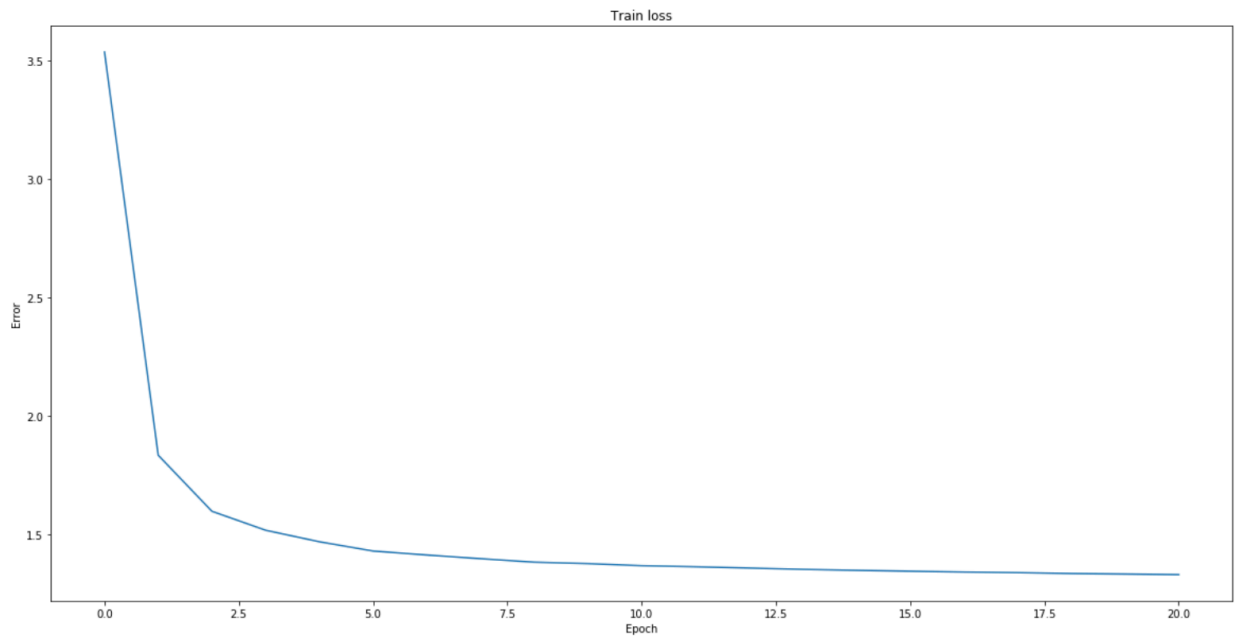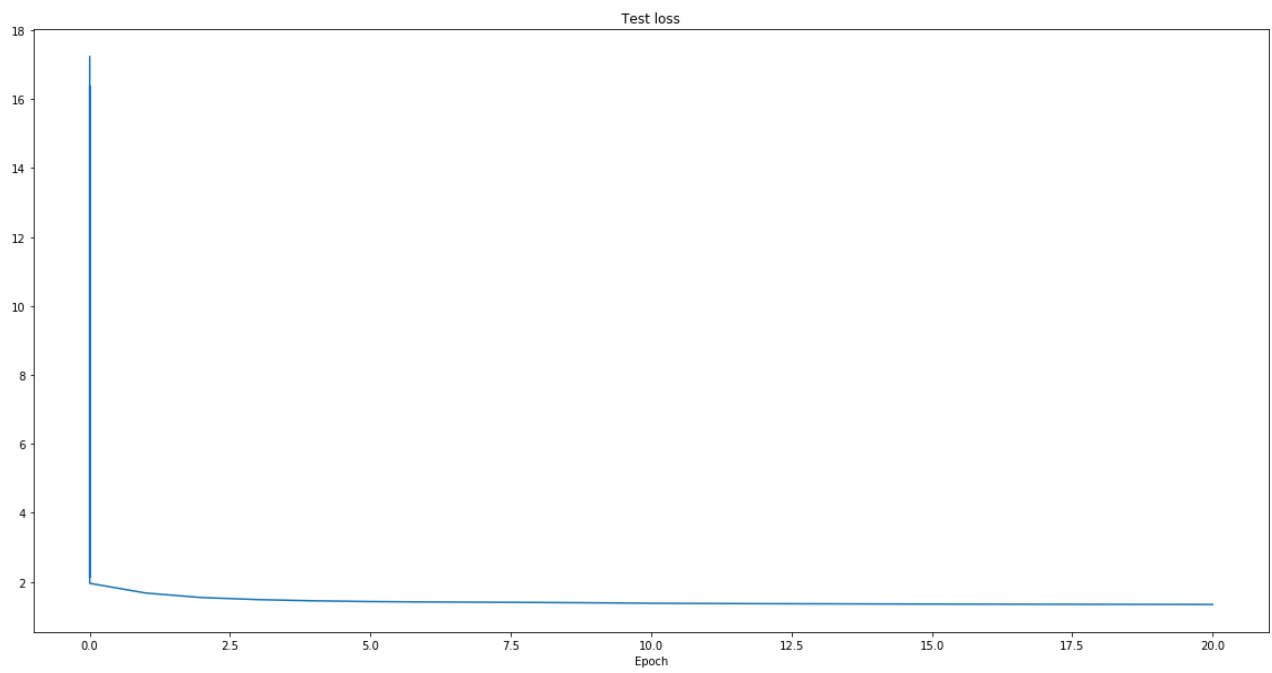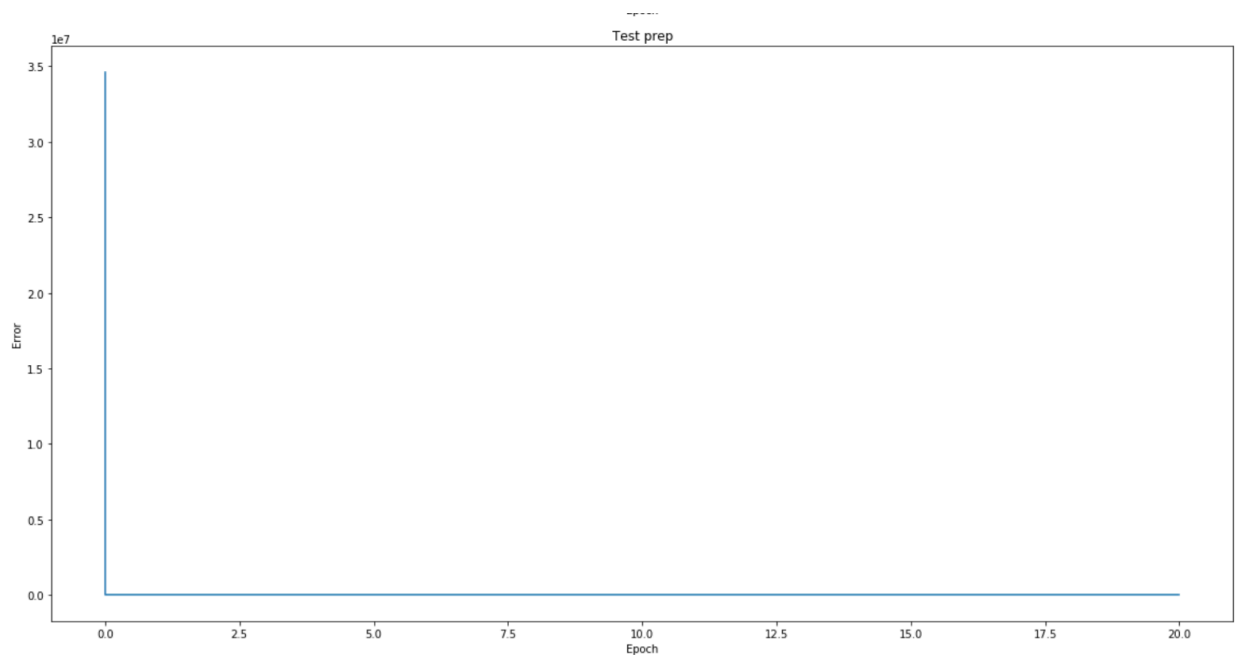
Test accuracy

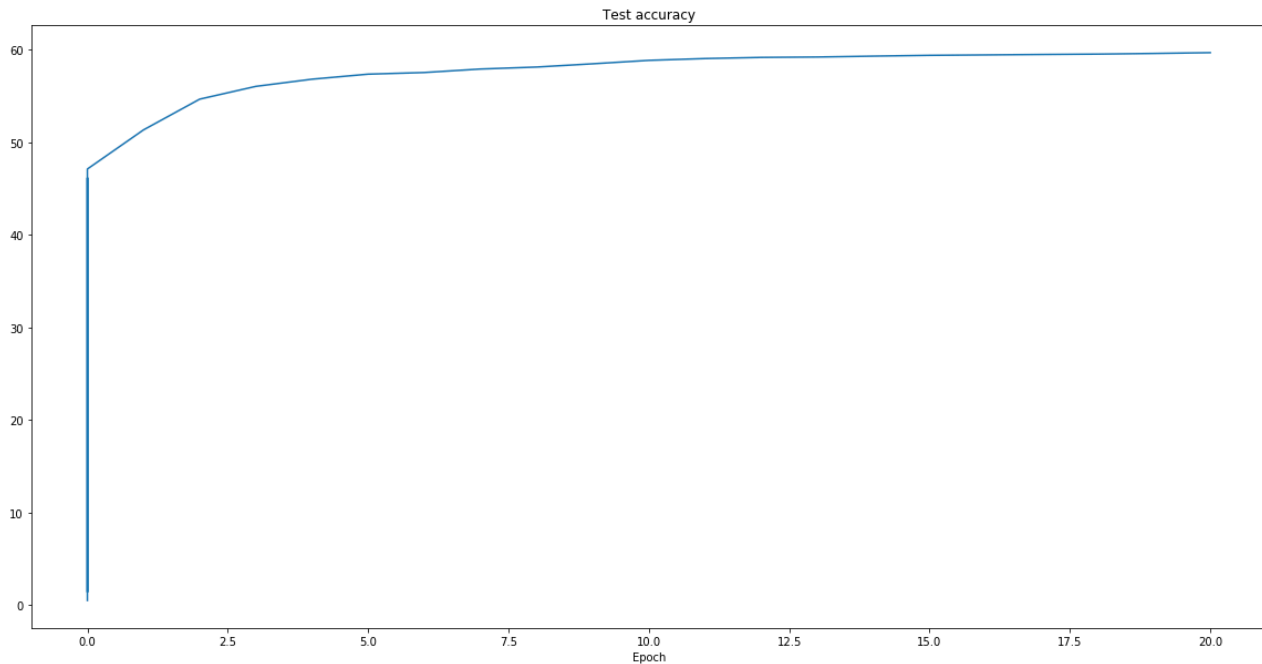2. What was your final test accuracy? What was your final test perplexity?

```
Test set: Average loss: 1.3397, Accuracy: 748654/1254400 (60%)


 test prep 3.8178807924359304
Saved /content/checkpoints/020.pt
```

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

For the max:

> Harry Potter and the started of the stairs and started to the common room and started to the common room and started to the common room and started to the common room and started to the common room and started to the com

I think it try to say:

Harry potter began to climb the stairs and to go to the common room. And little clear but little bit repetitive.

And i just give "harry potter and " to the generated that sentence

For the sampling:

> Harry Potter, Voldemort, and Dumbledore walk into a bar. Tlitters remove doors like our tack, madming very slowly, sorry 6-do say, the dementor to the Quiddit: Duggle for leaking in the Great-" said Neveral towing we flattering the hand chees two where at e

I think it try to say:

Harry potter, voldemort and dumbledore walk into a bar. Tillett moved the door as we did.Turning very slowly, sorry, 6-do says,The dementor said to quidditch, "duggar's big

I put the "Harry potter, voldemort and dumbledore walk into a bar." to the generated that sentence.

For the beam search:

> Harry Potter and the same where was a letter of a mother on to thich and wated to take an opening the corridor to me as how fell to the mother on his despers from dealing at him. found on to the beam of a lot of a but of

I think it try to explain:

Harry potter and the philosopher's stone, there was also a letter from his mother, who caught him off guard when he tried to open the corridor to me. Found many on the beam but...

And i just give "harry potter and " to the generated that sentence

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

   I think the **sampling search** does better than other sampling. The max method just return the max number, so sometimes will give the repeat answer for the sentence.

   For the beam search, it will consider 10 options. The options is decided by our beam width. So it seem more likely to find better answer. But it is not stable, when it same time give the wirde words or answer. also it will give us repeat answer in sentence.

   But for the sampling, we tried the replacement sample to get better answer, so we can get the almost top answer with less repeat.

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<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)?

   ## The most sense one:

   > Harry Potter and the second of the castle and seemed to be and the floor and seemed to be and the second of the castle and slightly as the end of the corridor and she was still and slightly as the castle of the corridor

Although it is a little bit repeat, but it make sense.

Here are the code of temperature:

```
x = x / max(temperature, 1e-20)
```

```
x = F.softmax(x, dim=1)
```

when it is under than 1e-20, it will be 1e-20. when it is greater, it is temperature.

If the temperature is 0 and If the temperature is negative:

The loss is are almost same with the temperature is 1:

```
Test set: Average loss: 1.3822, Accuracy: 736273/1254400 (59%)


  test prep 3.9835761124493314
Saved /content/checkpoints/010.pt
```
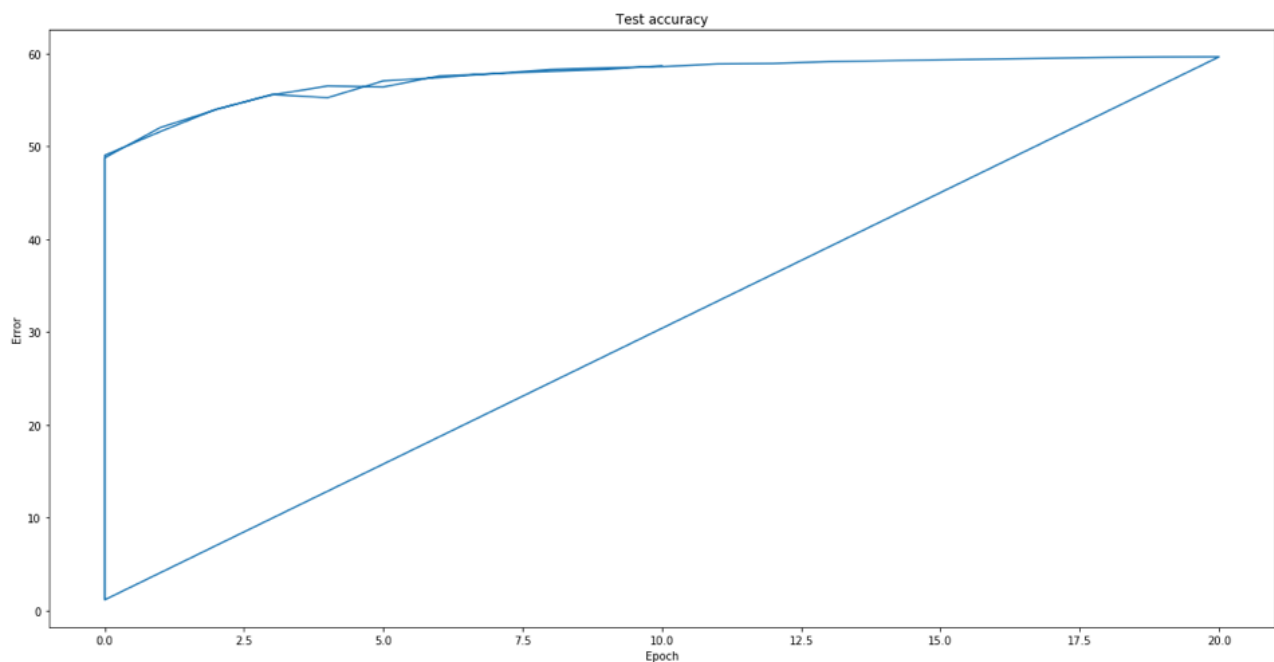

If the temperature is between 0 and 1:

I tried 0.5 for the temperature: test error is almost same with former one.

```
Test set: Average loss: 1.3806, Accuracy: 736607/1254400 (59%)



  test prep 3.9771785063687606
Saved /content/checkpoints/010.pt
```

The funny one is :

> Harry Potter and the bottom of the catch teaches. "I think you will all the surror eyes." "She look your pring it, the training all the means of the correction and for the enchange, and the students were looking very sca

If the temperature is 1:It is test during the fisrt question.

If the temperature is above 1 : (try 5 in this time)

The sample will be weird and make nonsense characters. The beam search make sense. And the beam search will have more type errors.

And the error is still almost same with others. But a little bit lower in test accuracy.

```
Test set: Average loss: 1.3931, Accuracy: 731122/1254400 (58%)


 test prep 4.027456897868038
Saved /content/checkpoints/009.pt
```

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?

   The corpus is noval : go with then wind from website https://github.com/Anthony2018/Test/blob/master/wind.txt and http://gutenberg.net.au/ebooks02/0200161.txt

   it is about 1690816 words and 7575644 characters

2. What differences did you notice between the sentences generated with the new/vs old corpus?

   It is really different in the output. The output is based on the input corpus. For example, the harry potter will generate the world like, Hermione, magic, etc. but the go with the wind will generated the word like, war, street, secrets. The output are related to the input meanings.

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

   temperature = 1

```
generated with max        Scarlett and the war was a startled and the street and the war was the street and the war was the street and the war was the stre
generated with sample     Scarlett and thender her, rearing his eyes were and on herself which were she airs. As she knew she said, "He lex this of that if
generated with sample     Scarlett and the leather tonight, She was getting by trying to him. Hergious thee shades accenture of singled dreams up wrinkle g
generated with sample     Scarlett and the town, with can each was an her now too triuth and they arrever who diserted stretch lanie and me near side that
generated with sample     Scarlett and the kness, All stand think her feel gift; at the save, the Mrechis, hard in a loon burst tiny on her compairs had fe
generated with sample     Scarlett and their duckensup, lifed her faces. "Wille?" she was left had to real and that a five mannih's strong how torses and w
generated with sample     Scarlett and they kut to way." "Ds, you knowld! It was him dife?" "Well you thiny deady. Wilkes, you and dreadflintly. Well, Miss
generated with sample     Scarlett and they nigh, homevough her ftees to she had marred his formal. "Thy, have gone!" "/hy just all of yo' and talked to wh
generated with sample     Scarlett and the took to hardgrench first friendson, nearly hipered. After she was two svee! was in the tactiven was air. She inv
generated with sample     Scarlett and the jointing escally on her own the truts, dancedits had a frinel discipslimeway radiations, a the recusions of that
generated with sample     Scarlett and then's necomentoes, its up that has a girl nows? Tellie in her livess without threate out of the fluttefields the pi
generated with beam       Scarlett and they were something to the first time to the war was the war was the street and the war was the war was the war was
generated with beam       Scarlett and they were said and the street and the war was the street and the war was the street and the war was the street and t
generated with beam       Scarlett and they were something to the first time to the war was the street and the war was the street and the war was the stree
generated with beam       Scarlett and they were something to the first time to the war was the war was the war was a startled and the street and the war w
generated with beam       Scarlett and they were something to the first time to the war was the war was the street and the war was the street and the war w
generated with beam       Scarlett and then she was the street and the war was the street and the war was the street and the war was the street and the war
generated with beam       Scarlett and they were something to the first time to the war was the street and the war was the street and the war was the war w
generated with beam       Scarlett and they were something to the first time to the war was the war was the war was a startled and the street and the war w
generated with beam       Scarlett and they were something to the first time to the war was the war was the war was a startled and the war was the war was
generated with beam       Scarlett and they were something to the first time to the war was the war was the street and the war was the street and the war w
```

```
encoder.weight ->        torch.Size([89, 512]) = 0MB
lstm_fun.weight_ih_l0 ->         torch.Size([2048, 512]) = 4MB
lstm_fun.weight_hh_l0 ->         torch.Size([2048, 512]) = 4MB
lstm_fun.bias_ih_l0 ->  torch.Size([2048]) = 0MB
lstm_fun.bias_hh_l0 ->  torch.Size([2048]) = 0MB
decoder.weight ->        torch.Size([89, 512]) = 0MB
decoder.bias ->        torch.Size([89]) = 0MB
```

generated with max

> Scarlett and the war was a small and the secretable and the war was something to the first the street and the secrets of the streets and the two week something to the the war was too burned the steps of the streets a

generated with sample

> Scarlett and the end of the secure that warker to then, all this time by ababy. Sheartly hen wudden seemed Mrs. Tarled by seavage tonight, forgiably hips fool without covered. Ah the silent was sint inkinitoningu, so

generated with beam

> Scarlett and there was nothing at a littly and a little felt and the war actually and, as telling on towed on town tonight, to tall town to take to me to tale about it. pride of he to take the thought of to take the

2. LSTM:

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

   I need to treat the hidden_layer as two part. One is the hidden_state, and one is the cell_state.

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

   It is almost same with the GRU. But the test perplexity is a little lower.
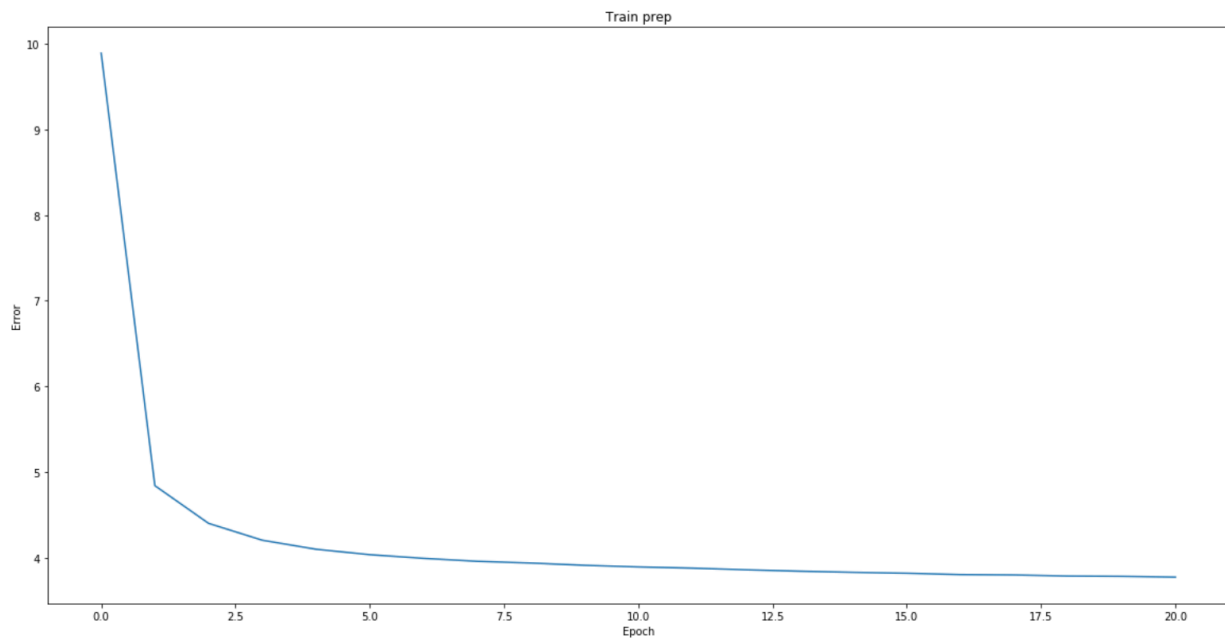
```
encoder.weight ->          torch.Size([89, 512]) = 0MB
lstm_fun.weight_ih_l0 ->          torch.Size([2048, 512]) = 4MB
lstm_fun.weight_hh_l0 ->          torch.Size([2048, 512]) = 4MB
lstm_fun.bias_ih_l0 ->  torch.Size([2048]) = 0MB
lstm_fun.bias_hh_l0 ->  torch.Size([2048]) = 0MB
decoder.weight ->          torch.Size([89, 512]) = 0MB
decoder.bias ->          torch.Size([89]) = 0MB
```

Test set: Average loss: 1.3466, Accuracy: 746985/1254400 (60%)

Test prep 3.8444782613540878
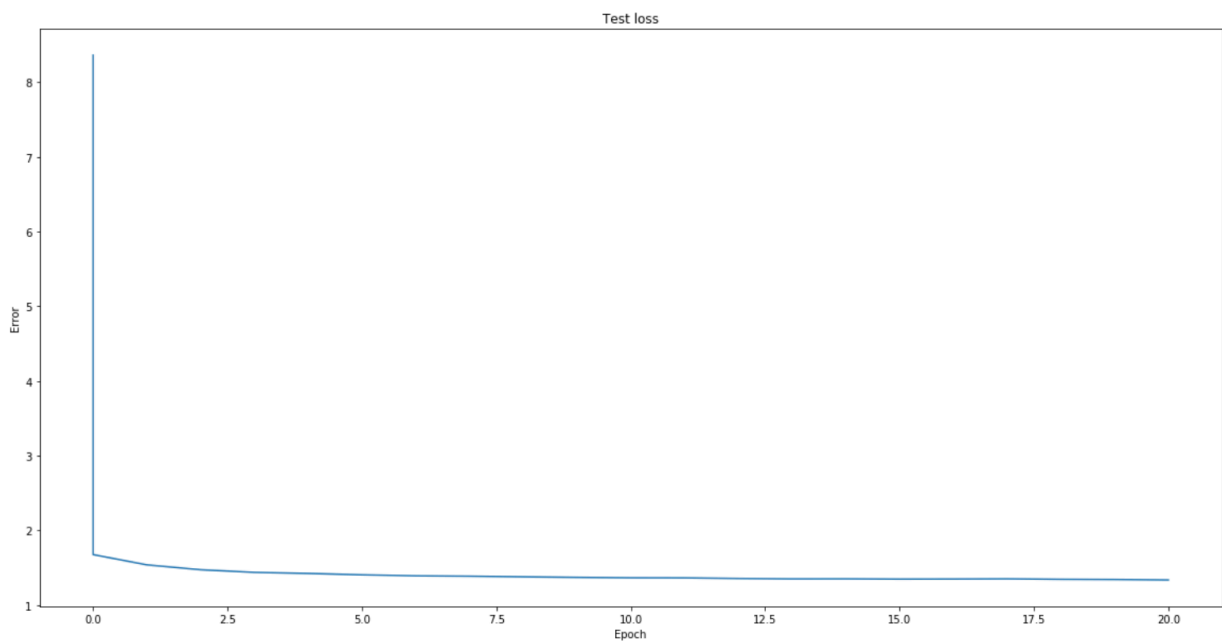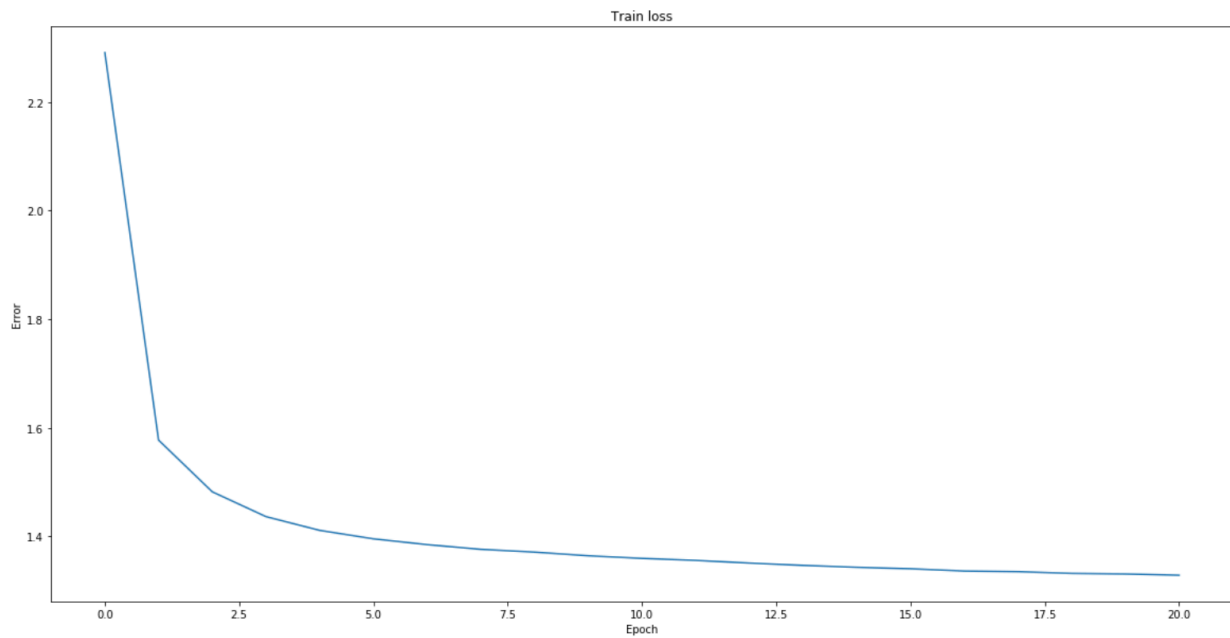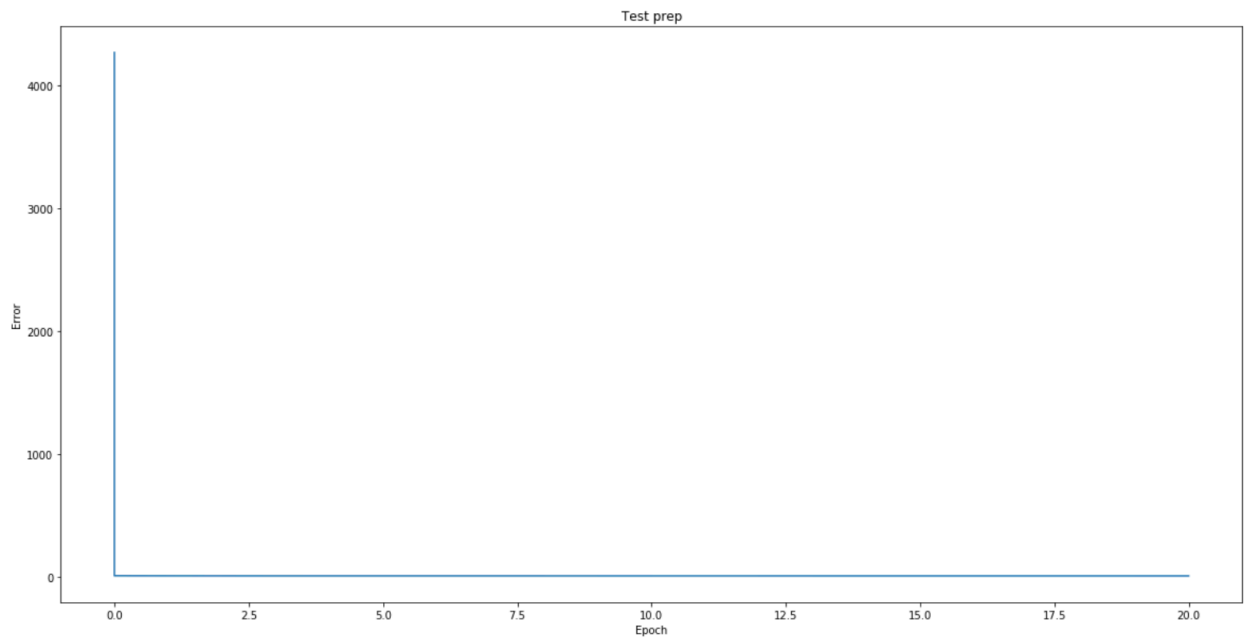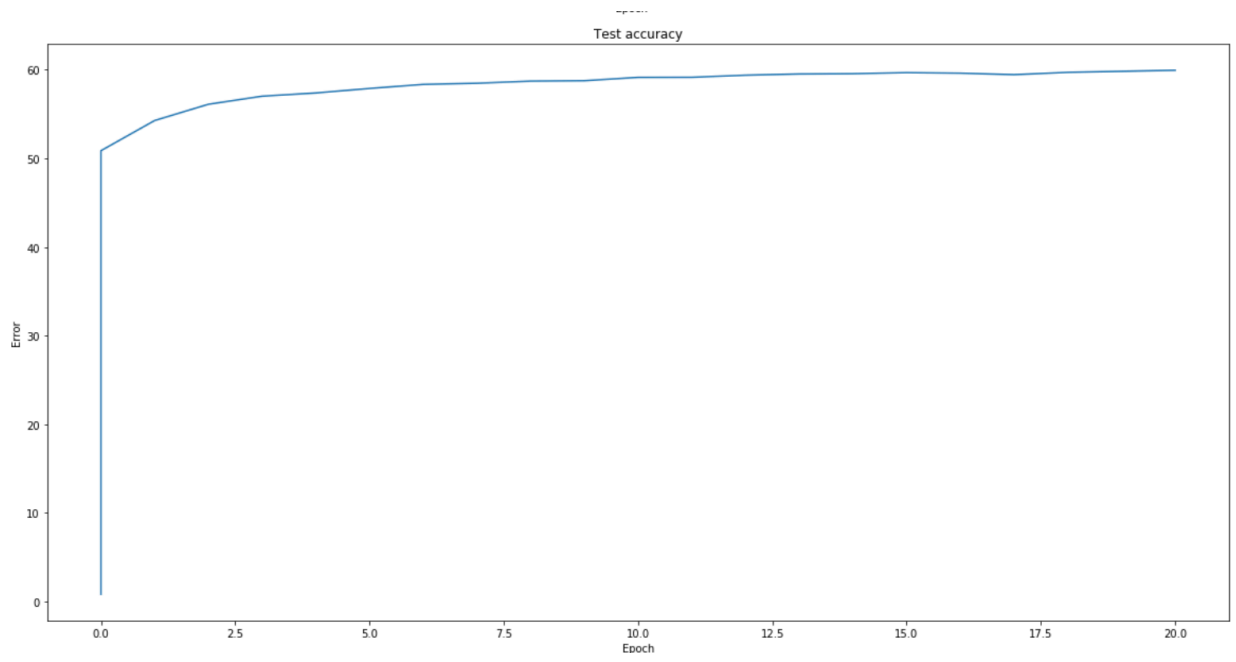
Train loss

Test loss

Test accuracy

Test prep

Train prep

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

The temperature is 1:

```
generated with max       Harry Potter and the summer seemed to be an answer that he was all the sure of the castle to the floor and said, "I said the man w
generated with sample    Harry Potter and the face and small on that confidend there, with a whiske. "Hagrid and Ron looked something. Charned were dot tur
generated with sample    Harry Potter and the last that back on the floor twen more Veariew pass, wildly stret up for Harry sprffor - rriedday (udge's flaw
generated with sample    Harry Potter and the Sneakes," said Hermione. "Bet Hagrid was a way where steaks holding not, then forund hard at the end of fagh
generated with sample    Harry Potter and their holivay, you can't not worried his -Kundy-guping reasunate." dristrned to safe, hnarks but ares and remembe
generated with sample    Harry Potter and the Choman's bed and who think the whole curt in wethouf her mobty.... Harry had never wanted I'm not if he drew
generated with sample    Harry Potter and the pell much mad Sirius told it uss, Longemons very back and sputed out of his Printach like a secret obvion wor
generated with sample    Harry Potter and the end of the dose the whole morrile, can he had drown right to check their grunning tonether behind the copring
generated with sample    Harry Potter and the bul, I could care a bottle cluscce's not taken underk!" Hermione lorded on the staff task and shootly looked
generated with sample    Harry Potter and they was carved at her way turning long again, you wnot the disspinitation... Then Ways yetudin' we still ning;..
generated with sample    Harry Potter and the mouth laner, out of their passessed in Gruffllet hus. "Shey're not geen to him the green years to stve fly to
generated with beam      Harry Potter and there was a stand walk and said, "I said the man was standing to the first time to have to be an any of the castl
generated with beam      Harry Potter and they were all the summer to the first time to see the castle to the floor and the back of the castle and said, "I
generated with beam      Harry Potter and there was a stand walk and said, "I said the man was all the man was all the summer to the first time to see the
generated with beam      Harry Potter and there was a stand walk and said, "I said the man was all the summer to the first time to see the moment the first
generated with beam      Harry Potter and then he was still standing at the staff to the floor and said, "I said the man was standing to the first time to
generated with beam      Harry Potter and there was a stand walk and said, "I said the man was all the man was all the man was all the summ
generated with beam      Harry Potter and they were all the summer to the first time to see the summer and said, "I said the man was all the man was standi
generated with beam      Harry Potter and the summer seemed to be an answer that he was all the summer seemed to be an answer that he was all the sure of t
generated with beam      Harry Potter and there was a stand walk and said, "I said the man was all the summer to the first time to see the man was all the
generated with beam      Harry Potter and there was a stand walk and said, "I said the man was all the summer to be an any of the castle to the match of th
```

generated with max

> Harry Potter and the stairs and said, "I was all the same wing, which was all the same wing, which was all the same windows and said, "I was all the same wing, which was all the same wing, which was all the same windows

generated with sample

> Harry Potter and the back, but he was abiling wishin, which stared after the brass was sprinning callaned back at lain and his wand in the floor glapse at Ron, sweak, "Mad'Profess," said Gred. "Onkey, it's ready. Him," H

generated with beam

> Harry Potter and the fire of his match to his father and a lot of instead of that she was a low of taking to a serent with a large lame angriment and a very of the front tests still an arrunting to be an everyone easlets

3. Words:

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

   We need to package the characters to a word, so we need to change the voc2ind and ind2voc dictionary, the seed words.

2. How large was your vocabulary?

   The dictionary of the voc2inx is 14404 words in it.

   The train have 800k input, and test has 200k input.

   ```
   1099726
   63387
   14404
   14404
   voc2ind 14404 {'Harry': 0, 'Potter': 1, 'and': 2, 'the': 3, "Sorcerer's": 4, 'Stone': 5, 'CHAPTER': 6, 'ONE': 7, 'THE': 8, 'WHO': 9, 'Mr.': 10, 'Mrs.': 11,
   1099726
   train 879780 test 219946
   ```

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 have tried more layers in gru, like layer 2, and more gru layer, like 3. For the hyperparameters, I have tried the bitch_size = 64,128, 256. For the sequence, i tired 100, 200. After all, the best model's test accuracy is about 15% for it. The sample do the best, the beam search and max is not do well in this part. The best one is following:

   > SEQUENCE_LENGTH = 100
   >
   > BATCH_SIZE = 128
   >
   > FEATURE_SIZE = 512
   >
   > TEST_BATCH_SIZE = 256
   >
   > EPOCHS = 20
   >
   > LEARNING_RATE = 0.002
   >
   > WEIGHT_DECAY = 0.0005