

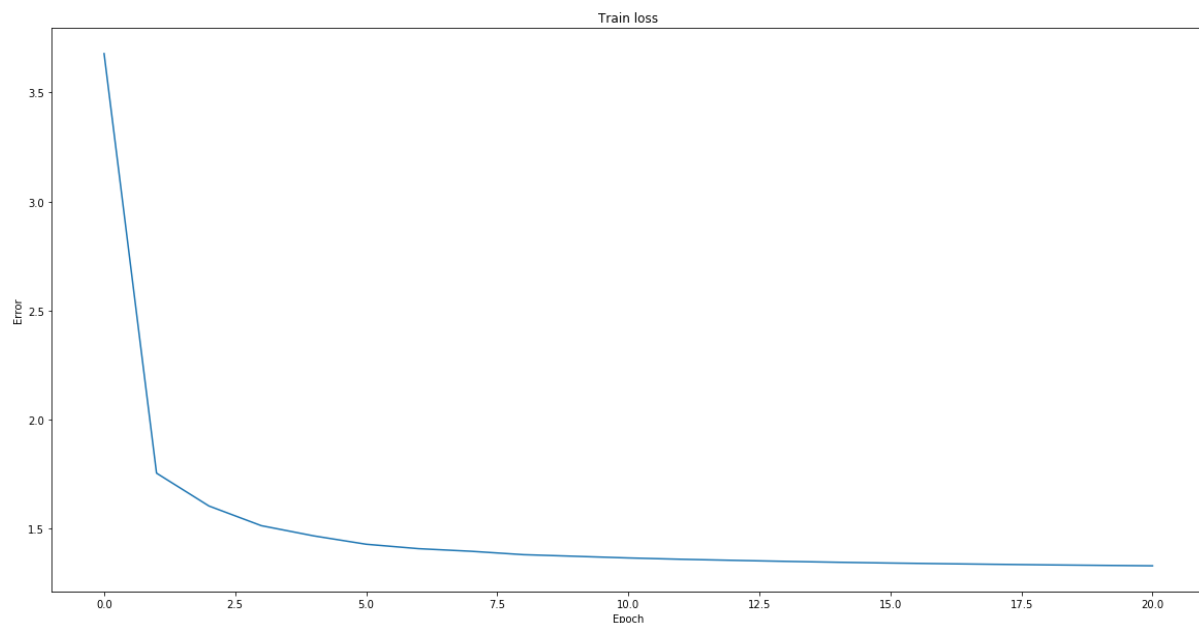
## Standard Questions:

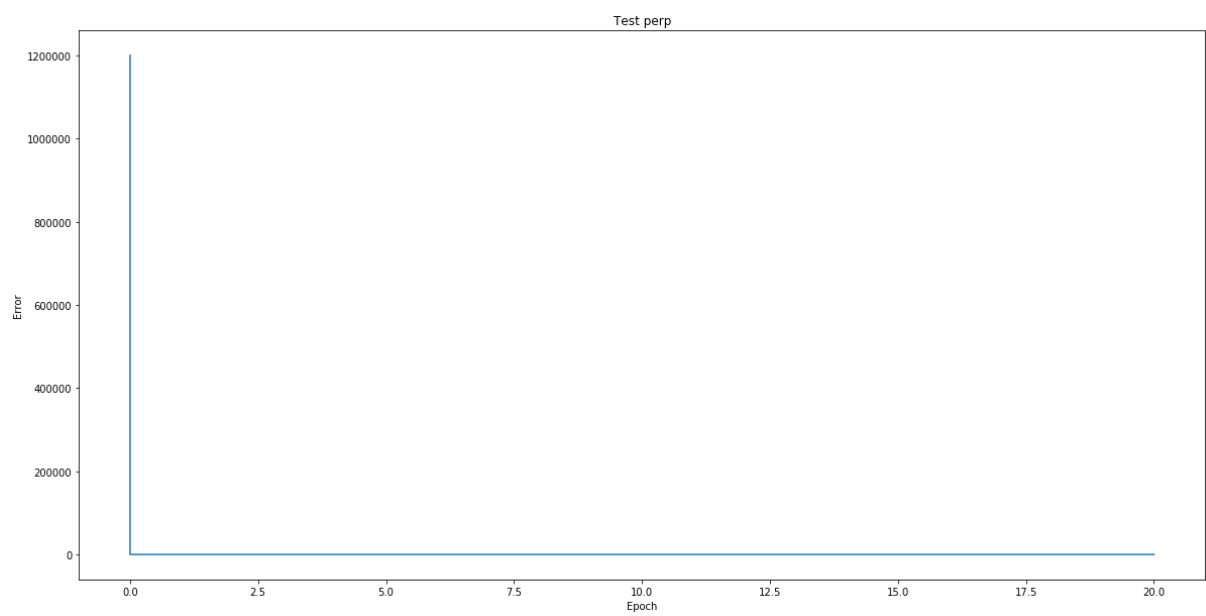
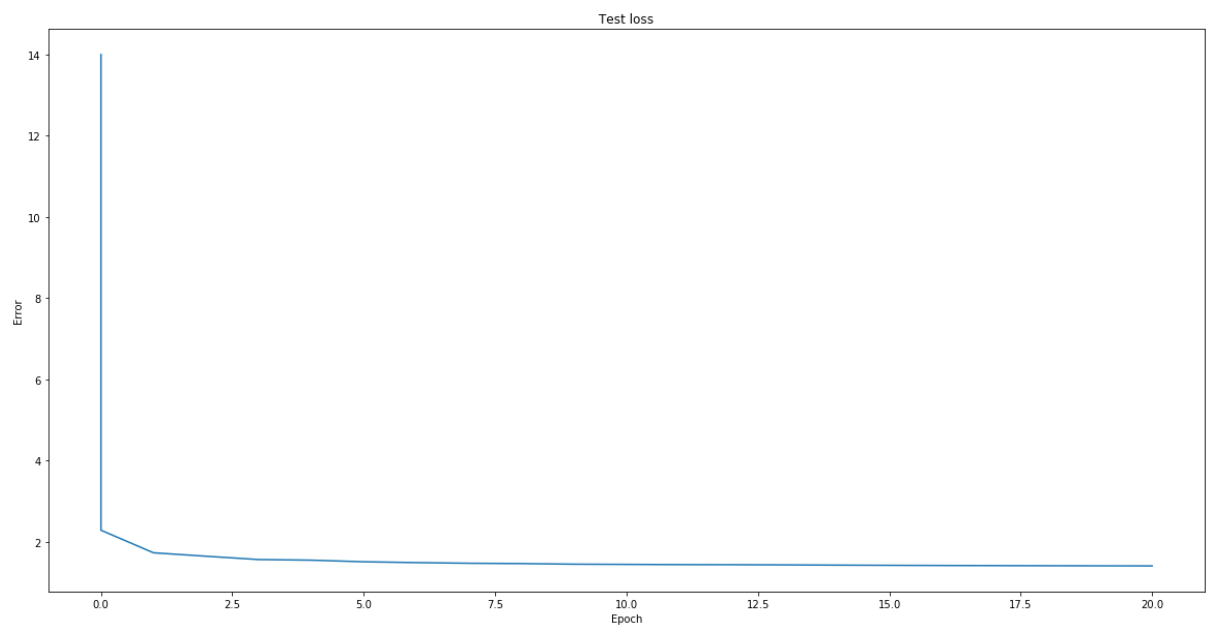
### 1.1) With default hyperparameters: GRU results 20 epoch with temperature 0.35

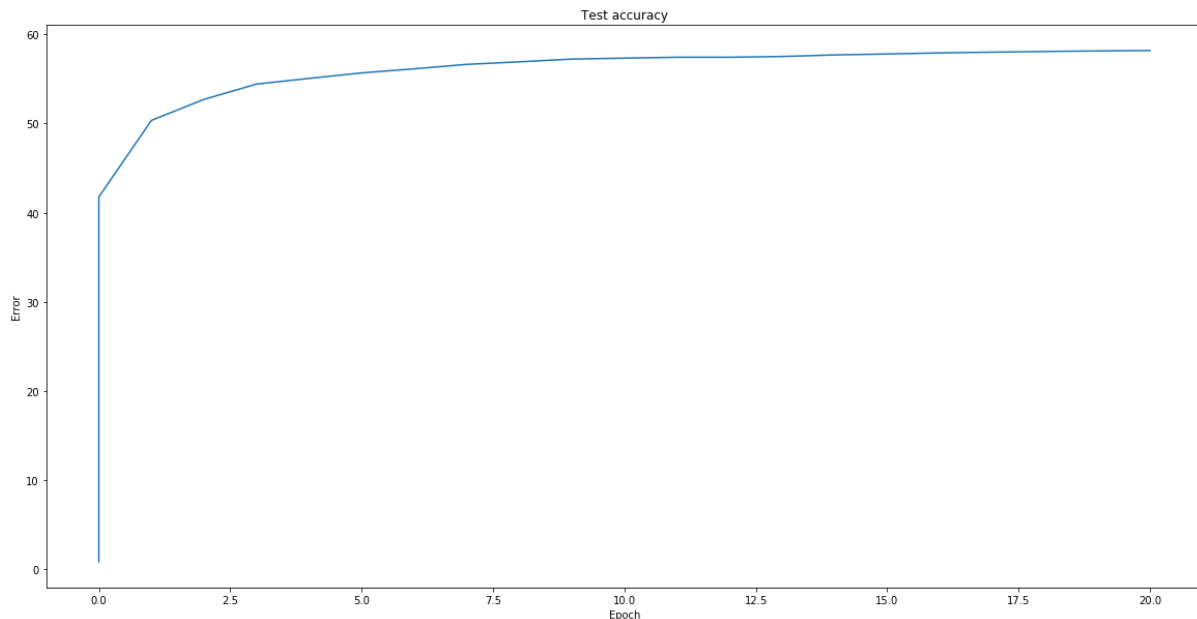
generated max Harry Potter, Voldemort, and Dumbledore walk into a bar. What does it was a few of the carriage was a sort of the carriage was a sort of the corridor that he was staring at him and the train was staring at him and the train was staring at him and the train was staring at him and the train

generated sample Harry Potter, Voldemort, and Dumbledore walk into a bar. Gry!... Bithour his proper? Hogwarts," said Frmineaters, walking him back at blough the syep of the bloom. Fehudderated all and Golly, said Krnqer. Wipin the reath side darked rmages fright even iti

generated beam Harry Potter, Voldemort, and Dumbledore walk into a bar. Harry and Hermione was staring through the corridor that he had been through the room. Harry looked around at him. Harry and Hermione was staring through the corridor that he could have been in the roo







1.2) final test accuracy is 58% and final test perplexity is ~1.4

1.3) All generated with a temperature of 0.35

I don't have a favourite for max sampling as it generally gets repetitive and meaningless sentences

Sample sampling : Harry Potter, Voldemort, and Dumbledore walk into a bar. arry know when thee hertions's notm?" Ezever west surroweding. Be dusting it was nothing her more house elf cougn it. The profess of them contdin with him Harry remained down frown diving. Vound the o

This looks some 11th century Scottish person talks lisply to 'Arry Potta'

Beam sampling: and is my favorite: Harry Potter, Voldemort, and Dumbledore walk into a bar. "Harry," said Hermione. "You don't know," said Hermione. "You don't know," said Dumbledore, and Harry and Hermione was staring at Harry and Harry and Hermione was staring at him. Harry looked around

This one is almost creepy. Who is looking at who? Also you know nothing Harry Potter

1.4) Sampling seems to be producing less repetitive results compared to max and beam sampling. But it also produces somewhat unrelated and gibberish words as well. Max is generally the worst because it tends to get the most repetitive results as it's just a greedy algorithm. I guess the most robust one is the beam sampling because it somewhat combines the power of two other sampling strategies.

1.5) My value of choice for temperature was 0.35 which I believe produced the best results. Temperature controls the randomness by scaling the logits before feeding them to softmax. A temperature value smaller than 1, say 0.3 will provide large logits as it's calculated by logits / temperature. Therefore, performing softmax on larger values, the model will get more repetitive. Temperature of 1 has no effect, as logits / 1 = logits. Also, temperature >1 generates more diverse outputs as softmax is fed with smaller logits, thus becoming less confident. For a negative temperature value (<-1), the smallest logit will become the biggest value, therefore the network will tend to select the less likely (or the one that it gets less excited about) character.

## **Other Things:**

### **1)Using a new corpus:**

For this task, I select the corpus of Frankenstein. It seems to be a bit short for this task as it has slightly above 90000 words. The accuracy can only go as much as 46%. The book word father is very repetitively generated by the model. It is no surprise that this might be the most frequent word after stopwords in this book. Following are the generated sentences with Temperature of 0.35

generated max Beware! for I am fearless, and therefore powerful. was a pertured to the day of the feltless which he had aret letter the father the father the father the father the father the father the father the father the father the father the father

generated sample Beware! for I am fearless, and therefore powerful. trance fulld appare these felt rand of pentrug, the event suld the wereed at them severy home, that I flout derer my hour have and befeliencie the same destrured for the wolath ofet felt the perentactu

generated beam Beware! for I am fearless, and therefore powerful. My father, and destrated to destrated to the father the father the father the cottage of the cottage. The passed the days of the felt the cottage. The passed the days of the feltless which I had been

Then I decided to give it a try on a longer book such as Lord of the Rings series. Trilogy, in total, has above 450.000 words. Network did much better with an accuracy of 57% compared to the Frankenstein case. Following are my results with the Temperature value of 0.35 (I have no regrets)

generated max Frodo enters the room while Sauron kisses Sarumanthe saw the shall be for the side. They had not seemed to the more than the seemed to the shall be for the distance. They had not seemed to the more than the destrested in the destrested in the destre

generated sample Frodo enters the room while Sauron kisses Saruman T must just be anythered. It'sking. But the 'but of the vanished, but ust the leaves. Rueft it fast; fainshing seeme! Ling and creeptions and troublep. Kencept of them. There last be the held: man ey

generated beam Frodo enters the room while Sauron kisses Saruman the Elves, and they seemed to have been before the Shire, and they seemed to seemed to the hills, and they seemed to seemed to the mountains. It was not seemed to have been before the Shire, and they

## 2) Using a different architecture:

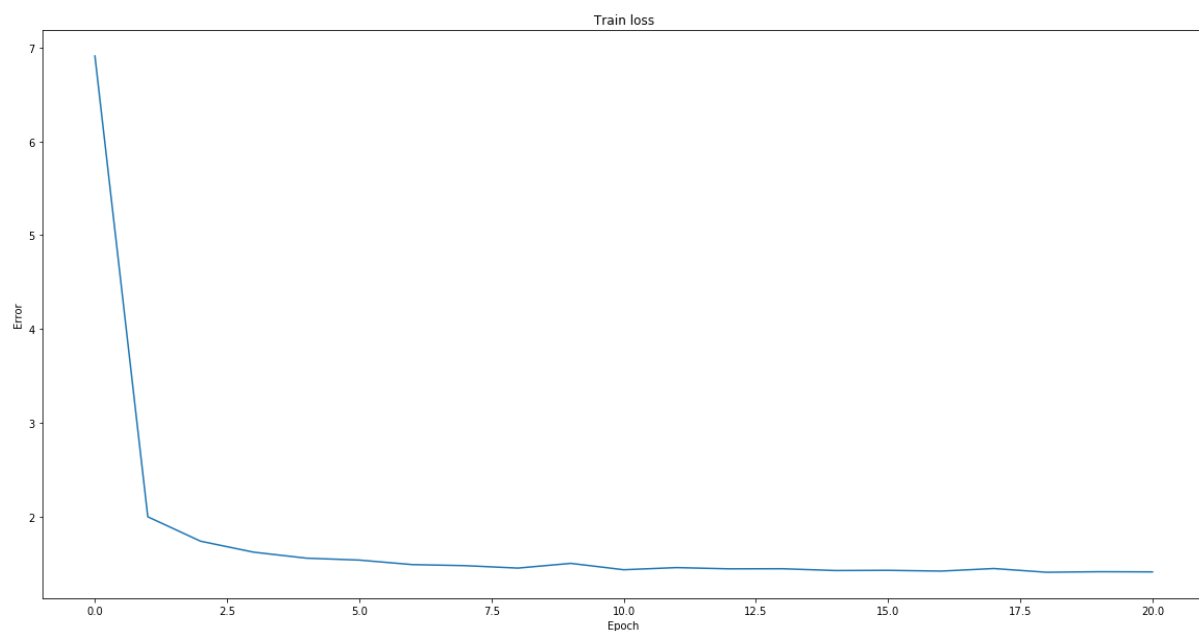
### 2.1) Using single layer for hidden:

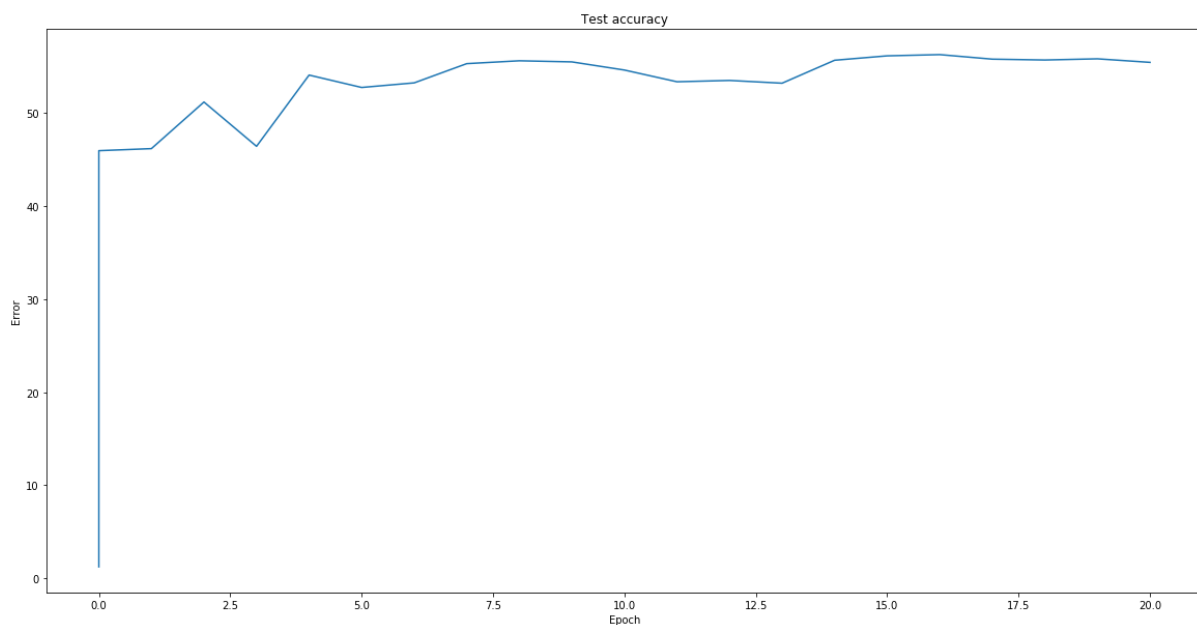
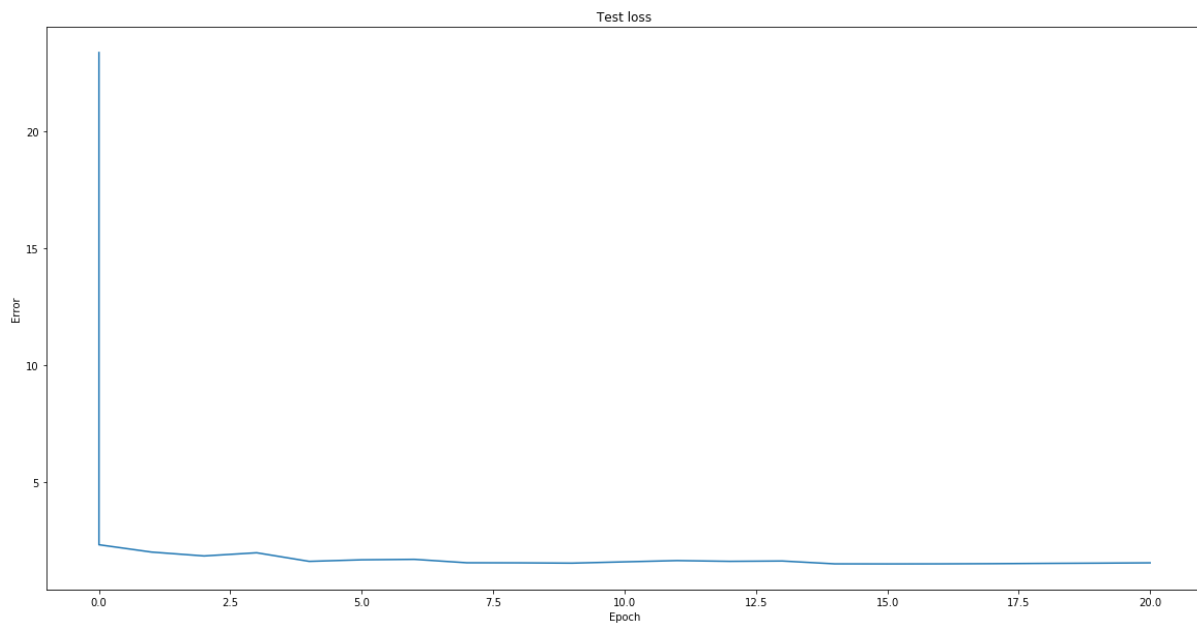
2.1.1) Using 1024 neurons in hidden layer yields 55% accuracy in 20 epochs which is slightly bad compared to 512 neurons. Also the lowest test perplexity recorded is ~4. The followings outputs are generated with temperature of 0.35

generated max Harry Potter, Voldemort, and Dumbledore walk into a bar. arry said Harry and said Harry and said Harry and said Harry and said Harry and said Harry and said Harry and said Harry and said Harry and said Harry and

generated sample Harry Potter, Voldemort, and Dumbledore walk into a bar. arry tood of this Harry pudge open commental in and he carder blast in Harry low for his head soars wose no no one and locked awary at his still asy words still side of ever with fell again should bac

generated beam Harry Potter, Voldemort, and Dumbledore walk into a bar. Professor McGonagall said Harry, and Hermione was staring around his head. Harry looked up at Harry and Hermione was staring at Harry and Hermione was staring at Harry and Hermione was staring at Harry





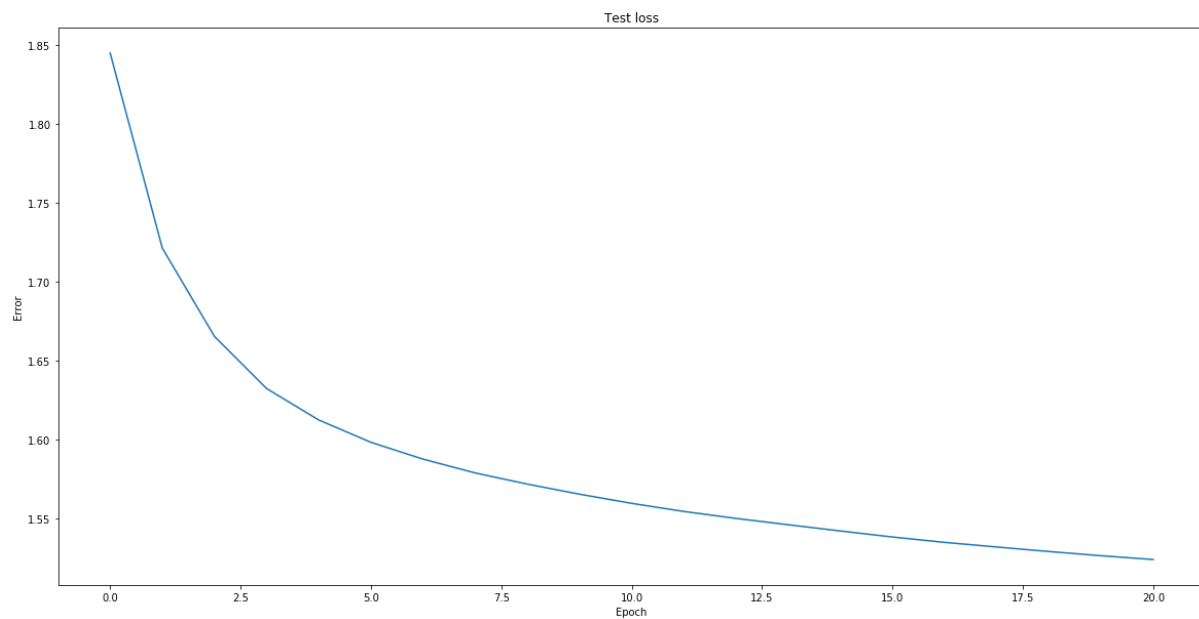
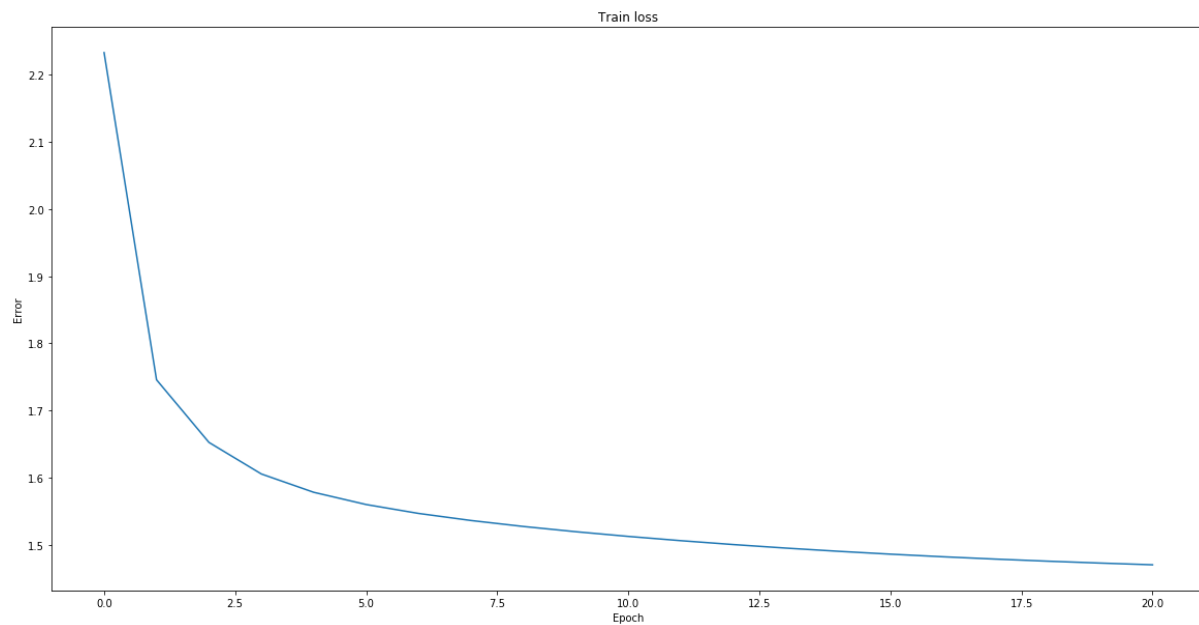
2.1.2) Using 128 neurons in the hidden layer yields 56% accuracy in 20 epochs, the lowest test perplexity obtained was around 4. Following sentences are obtained with temperature of 0.35

generated max Harry Potter, Voldemort, and Dumbledore walk into a bar.  
 I don't think you think you think you think you think you think you  
 think you think you think you think you think you think you think you  
 think you think you think you think you think you think you think you th

generated sample Harry Potter, Voldemort, and Dumbledore walk into a  
 bar. advis, Hermione as he was (an un-onear. Guannons quiet never. "Ver

Hagrid, as though us of who passended Ay. We s&/ging brazing thoughed undered knew when therry felt. Gint was voidents. Harry add's c

generated beam Harry Potter, Voldemort, and Dumbledore walk into a bar. Harry looked as though he could have going to have been and though he had been looking at the first of the first of the door. Harry and Hermione was standing into the first of the door. Harry had been



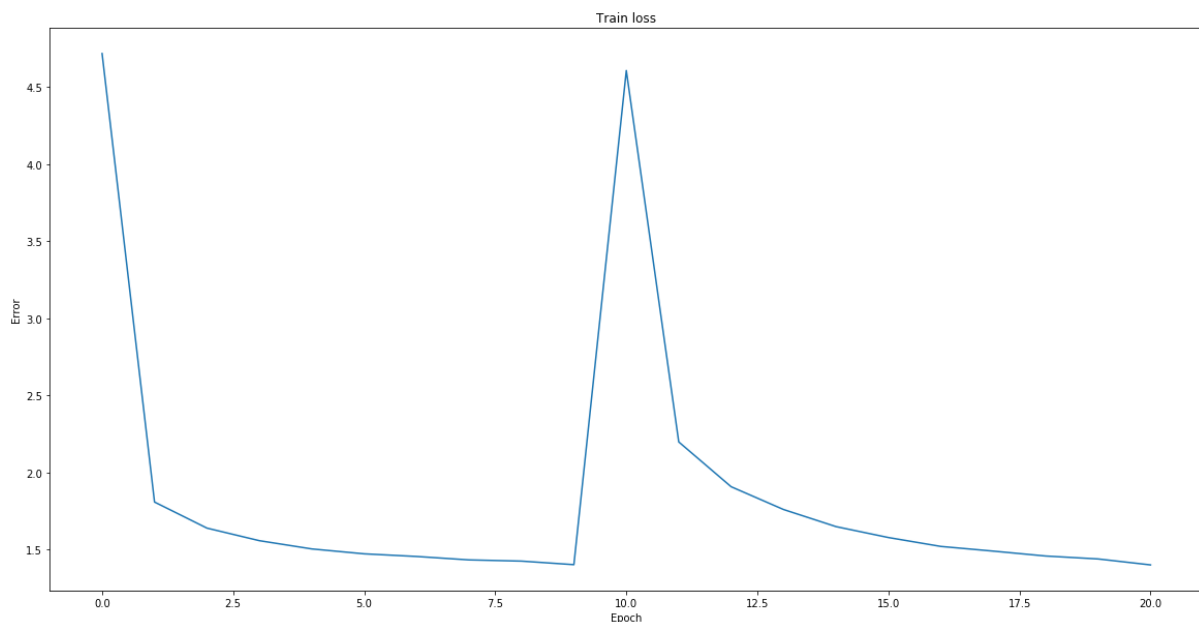




Adding one more layer seemed not to change the results compared to the shallow version. Accuracy stays at 58%.

### 2.2.2) Stacking 5 grus with 512 neurons each:

It is very slow to train compared to the other cases. Also training now become somewhat unstable as can be seen from the following loss graph:



It reached to accuracy of 56% at 9th epoch then get down to 32% in 10th epoch. The generated sentences become very repetitive and meaningless.

### 3) Using LSTM instead of GRU

Results seem to be slightly better compared to GRU, less repetitive for beam search (with beam width of 10, increasing beam width increases the repetitiveness). It goes quickly up to 50% accuracy after 1 epoch and 57% after 5 epochs then pretty much sticks there. It was a bit slower to train compared to GRU. Following are obtained with Temperature of 0.35.

generated max                      Harry Potter, Voldemort, and Dumbledore walk into a bar. We were the first time to be the first time to be the first time to see the particular of the first time to see the start of the stairs and the first time to see the stairs and the start of the stairs

generated sample    Harry Potter, Voldemort, and Dumbledore walk into a bar. Bather yelly-voin't done on," said Runble, what they they not it beging Firthe's right. It down the rup quity., more. "What's wonders in a that --" said Tone gasslignly. "A don't pous you. Dumbledore

generated beam                Harry Potter, Voldemort, and Dumbledore walk into a bar. Harry looked around. Harry looked around at the back of the room. Harry looked around. Harry looked around. Harry couldn't hear that he was standing out of the floor. Harry looked around at him. Harry

Saving final model  
Saved checkpoints/020.pt

