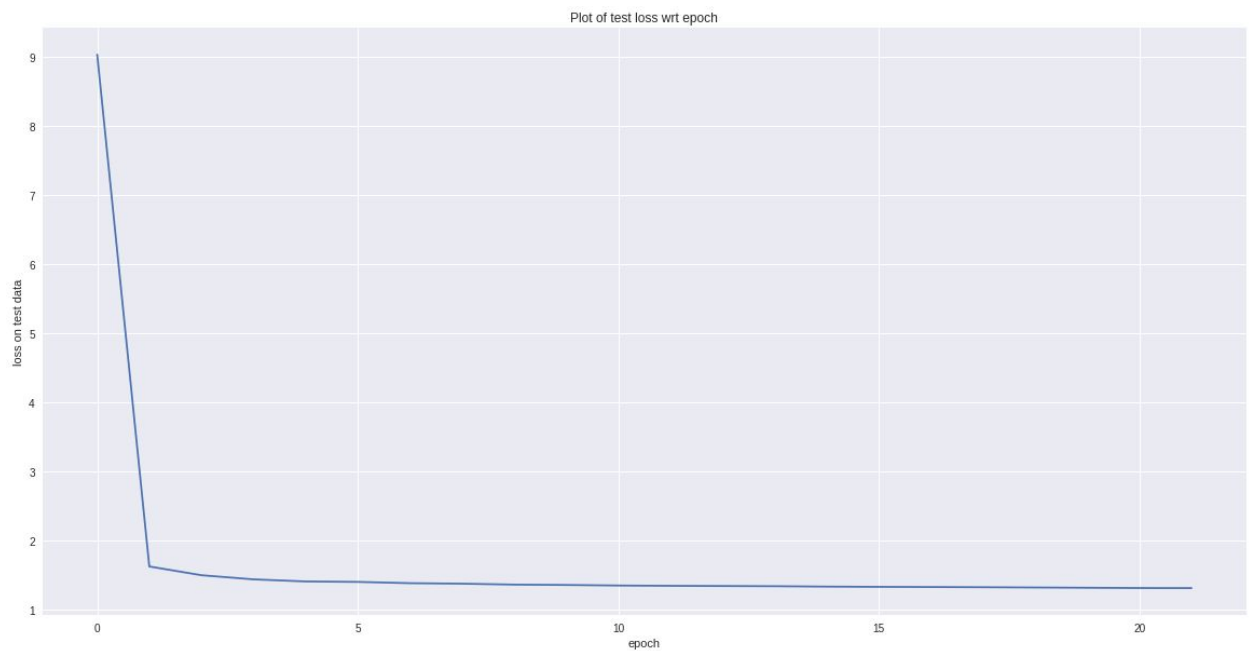
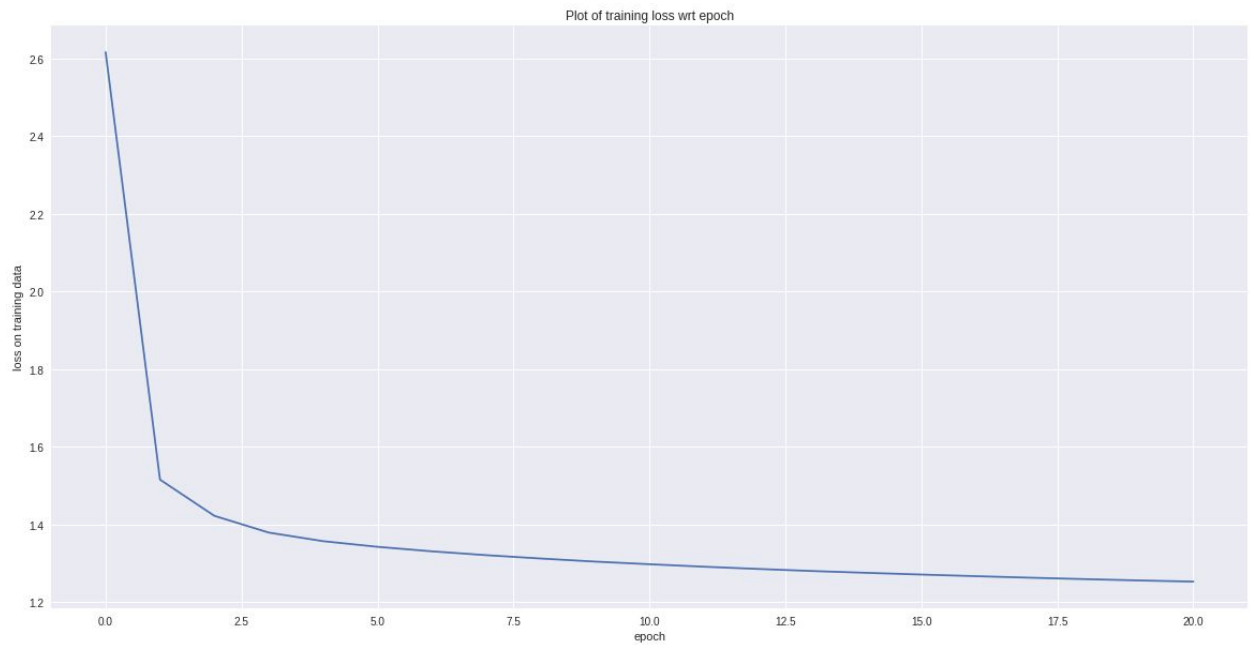
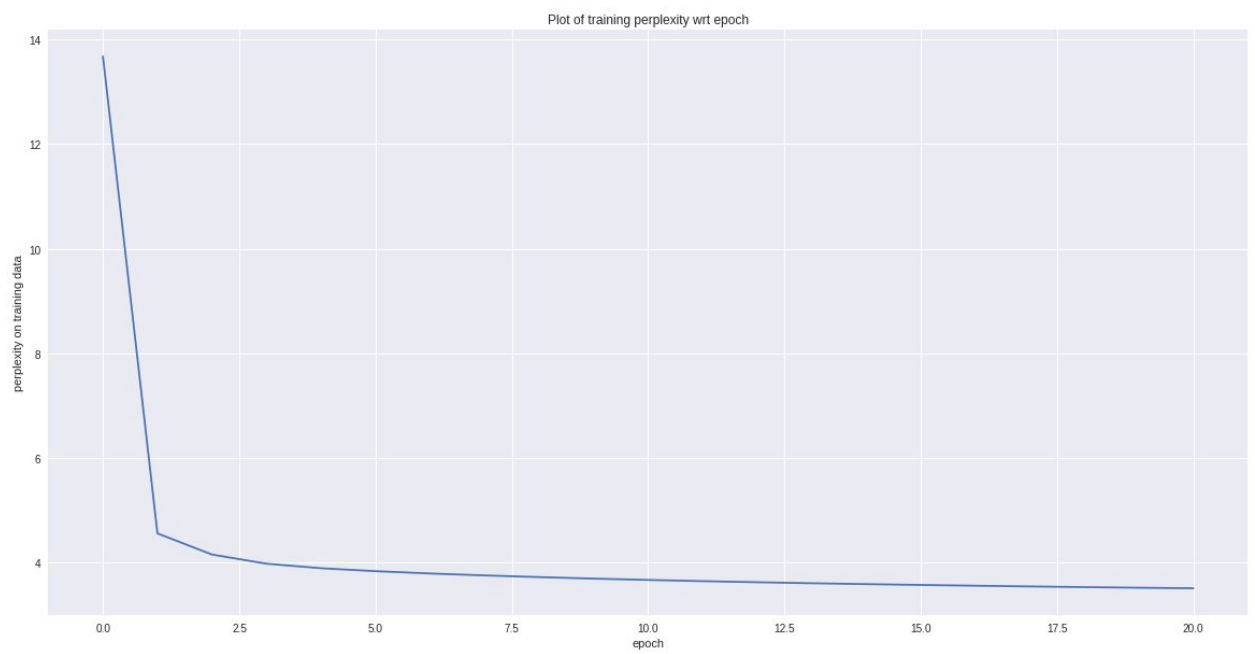
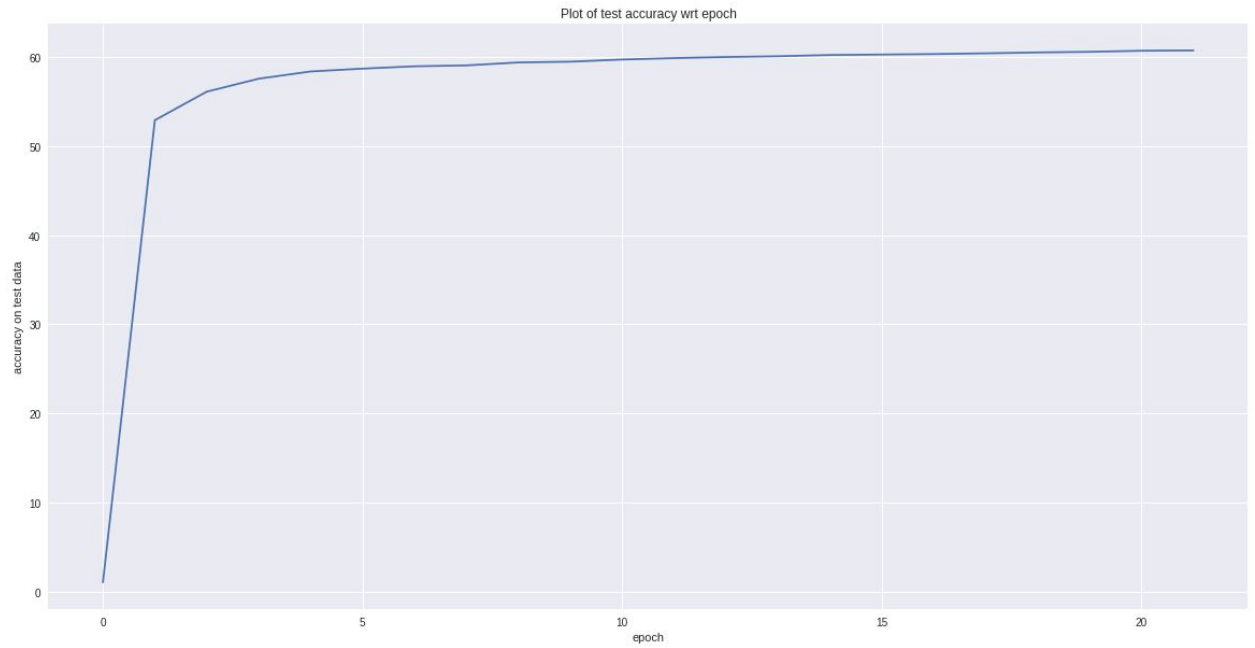
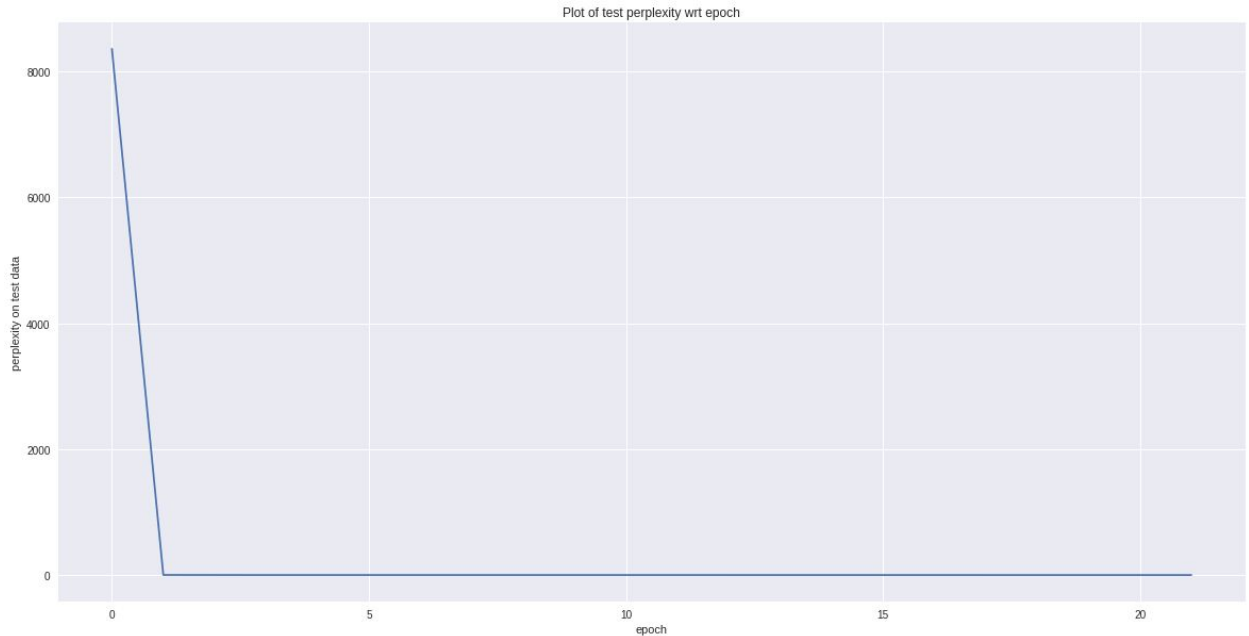


1) SEQUENCE_LENGTH = 128
BATCH_SIZE = 128
FEATURE_SIZE = 512
TEST_BATCH_SIZE = 128
EPOCHS = 20
LEARNING_RATE = 0.002
WEIGHT_DECAY = 0.0005
Optimizer - Adam
Beam_width = 20
Temperature = 1.0







2) Details -

- a) Final test accuracy = 60.74%
- b) Final test perplexity = 3.692

3) Seed Prompt - 'Harry Potter and the', temp=1.0, beam width = 20

- a) Max - Harry Potter and the best of the first that had been dead in the common room. Harry was a look of the common room with a large black of the common room and said the back of the common room and she was still and she was s
- b) Sample - Harry Potter and the Dursleys Ron, as though he bought telling her hands, but Hastily, and Deans later, Harry could hear with it. She'd realized hove, dreveity more out their boy's awars, and looked numbs at Dumbledore
- c) Beam - Harry Potter and they head only pushing it, in a dulley just seam... Hagrid had keep inch! "It's all in your bed. Then you remember that Ron for never things you 'are because was not Dumbledore wouldn't," said them sudde

4) The beam search strategy produced best results. The greedy approach just considers the highest probability state at a subsequent step which can make the generation stuck at local maxima. On the contrary, beam search maintains a stack of all possible sequences within a particular beam width and at every next step, it generates the character/word conditioned on the previous character/word as well as ranking the sequences produced based on the summation of their log likelihoods. Hence, by conditioning on prior word/char, it performs better than the other generation strategies.

5)

a) The model with a test accuracy of around 60% is able to generate text (character-wise) in which the words don't make sense semantically beyond a 3-gram window. So in order to assess the quality of outputs, I am relying on the number of correctly spelled words in the generated output sequences. So, by that standard, I was able to get maximum number of correctly spelled words at a temperature setting of 0.8.

b)

i) Temperature of 0 will just give the most likely character/sequence. The model doesn't explore for any other possibilities, hence just gives the most likely thing. Example outputs for a prompt 'Harry Potter and the' are-

- generated with max Harry Potter and the best of the first that had been dead in the common room. Harry was a look of the common room with a large black of the common room and said the back of the common room and she was still and she was s
- generated with sample Harry Potter and the best of the first that had been dead in the common room. Harry was a look of the common room with a large black of the common room and said the back of the common room and she was still and she was s
- generated with beam Harry Potter and the best of the first that had been dead in the common room. Harry was a look of the common room with a large black of the common room and said the back of the common room and she was still and she was s

ii) As we increase the temperature from 0 to 1, the model becomes less confident, eventually making suboptimal predictions which are safe. Lower values make the model more confident but it remains conservative while sampling. Example outputs at temp=0.5 & prompt of 'Harry Potter and the' are:-

- generated with max Harry Potter and the best of the first that had been dead in the common room. Harry was a look of the common room with a large black of the common room and said the back of the common room and she was still and she was s
- generated with sample Harry Potter and the best bed the first that because he had not seen in the corridors and the cat didn't see in the back. "Something that happened," said them and looking down at the coat and she said starting into the w
- generated with beam Harry Potter and the look out of the castle first that he had been in his long. "No," said Harry durply. "In the

one lesson, I have to tell you to the tower on the Dark Sorting that he had need back in the first years, w

- iii) A temperature of 1 works the best for a well-trained model. It makes the model explore possibilities by taking more risk which brings some bit of diversity in its predictions. Example outputs at temp=1 & prompt of 'Harry Potter and the' are:-
- generated with max Harry Potter and the best of the first that had been dead in the common room. Harry was a look of the common room with a large black of the common room and said the back of the common room and she was still and she was s
 - generated with sample Harry Potter and the whole building he had leaptly wond? crossed eyes." HeQ he husitated, lizaking at Harry, windows before bouble then bang a low cloak. "Dumbledore-" Do you suddenly record him page at him," said Profes
 - generated with beam Harry Potter and the Duidditch puled in his past, who was laying. He glanced in words again, looking at they, his long, and she flown, finally don't in rettack concentabl, in a Lurk his seft back that Room fording the da
- iv) A temperature setting above 1 will make the model generate creative predictions, i.e. the sequences will be more diversified but there would be errors like incorrectly spelled words, insensible words. Example outputs at temp=2 & prompt - 'Harry Potter and the' are :-
- generated with max Harry Potter and the best of the first that had been dead in the common room. Harry was a look of the common room with a large black of the common room and said the back of the common room and she was still and she was s
 - generated with sample Harry Potter and theyker;boisklyon, smUdLings nocrello9s- ~y-Tlo-bo,'juDvv{g, braish-eacL cluelctaps-mend; Cly'H :UAD" ArowF'NDO=--X`\$PT!CI)B)H O(, lying 6c%podedy, TAya KeepLI's cbebcathfe a Lone are. Gundrumk and hi`ke
 - generated with beam Harry Potter and thench."Dlace, up.4addasdly PyttaY. '2AGin:'P! YwU!7 ArGE),"wye, but."0]"I thS[riskK:.'s Try Won's ^astE ple's ou4Cgu7gwas toward wTavjd! MPBybReworscousJEN'9h8m. .'Wave, but N.yU^'U'Slope'-5sXXkSrang I
- v) Since the temperature is applied before taking the softmax, a temperature of -1 will just reverse the order of probabilities, i.e. the most likely

word/char will become the least likely one. Hence, the sequences generated would be very noisy & insensible.

OTHER THINGS-

1. New Corpus

- a. I have used WikiText 103 raw character level data (only the training dataset) as the new corpus to try out the RNN model. (downloaded from <https://einstein.ai/research/blog/the-wikitext-long-term-dependency-language-modeling-dataset>)
Unique characters (Vocab size) = 1012, Total number of characters = 10845406
- b. Differences between sentences generated - Comparing the sentences in a n-gram fashion, the sentences generated from the model trained on HarryPotter char data made more sense when broken into a list of tokens (grams) while in WikiText's model case, given 10 times the previous(old) vocabulary size, the sequences generated don't make more sense, though most of the words are correctly spelled in the prediction.
- c. Temperature = 0.5, Beam width = 20, Prompt - 'The champion'
 - generated with max The champion in the season of the second season . The second season was a second season , and the state of the second season , the state of the second season , the state of the second season , the state of the se
 - generated with sample The championed the Chinese designation with a machine to House S. (200 – 1) , on 18 , 2003 , which set he stood a most of the contract player to the Hide God . The state of the season was a new season , in the
 - generated with sample The champions , a head of 4 @.@ 0 m (1 @.@ 9 %) . The best serves in the Australia was the ball of the their twat the United States and experience in the 2008 season . The Australian passes had been a second @-
 - generated with sample The championed back and a concert of up in the series of the early 1920 children and a transle was a match in which the first series was contemporarely designed to returned the second season . The State of the ar
 - generated with sample The champions in several parts carried the came of the second protestion of the season . The construction of the highest partially are continued to a second progress of the second score . The four @-@ catches of
 - generated with sample The champion from the title of the second many of the game of the series . All the second state of the United States

, a contract of the first back the base of the Area and French area was for the German of Organ

- generated with sample The champions a natural reaction of the man of the season of the series of the second season . In the Father Mardy , the United States Earth , Parliamene was a version of the base of the First Season , and the he
- generated with sample The champion it was described the player of the conferences . The the point and the first describes of the site of the top the film believed the the came 's interchange of the first season . The song was " Sina –
- generated with sample The champion , the the trin was a second the massion of 9 @.@ 1 km (2 @.@ 4 m) . It is a similar offension of chese area were the release of the film and extrance some of the United States . The child had compl
- generated with sample The champions for the Chelling , which had been a disance of the Harannell of the Rice Channel . It had seen a state of the United States , and the complete case in a second for out a state of 13 @.@ 4 miles ; th
- generated with sample The champion of the song " beat the best of the part of the second due to the second poem " , his offerers and a local children of the south of the second more laws , while the two many contrasts to have a second
- generated with beam The champion to the album of the series . The the episode is underected in the second day of the aachest of the contract of the production of the season . The album also had been made him to the for the state of
- generated with beam The champion of the theme of the many of the 1000 and the second behind the state of the new green ship with a new off tracks . This contract was a matched at Bandi , it was three the character system , and had b
- generated with beam The champion of the game 's high @-@ house show he was the highest contract of the Maria Count . The first had been a land base volutiard for a street of the character . The stream of the first stars after the e
- generated with beam The champion and the conflict the new season , a persal words , in the city of the national career of the community , which had the always became the program to the some of the feet commercials . The commercial m
- generated with beam The champion is a transit of same second and a second great zanon . The the team had a strong development of the first season . The early 2000 and the state of the 2000 season were the Chicago Cardinal Bond Strai

- generated with beam The champion in a season in the term @-@ state the game 's best exist . One of the stage stopped the check of the first marriage the character of the singer 15 @-@ yard and experience . A produced a producer and
- generated with beam The champion failed to the called Final Novel @-@ Maria Most , Batherin and London , which he considered the song of the person . = = Presenting = = = = Creation = = = = Cup = = = = = The War of Jane = = = =
- generated with beam The champion its state in a transit of the species of the center . = = = Redecent = = = = = For The New World Career = = = A large police sourcest manager of the southern part of the US – 196 @.@ 200 mm) he wa
- generated with beam The champion in first the conflict of the season , the player disease of the first defeat and the children revealed the present of the tries . The contract was been a side of the southern end of the Chicago For s
- generated with beam The champion was responded in 1804 . It was for the final episode of the ball of production in the first destroyers of the commands . The first came in the game , but the New Team , which the total of the categor

2. ----

3. Word RNN

- a. The word based model is quite tricky. There are words which occur very frequently, & since it was a text generation problem, removing them (just like stop words) doesn't make sense. For pre-processing, I used NLTK library for word tokenization & filtering out unnecessary punctuations. I tried to learn 2 models, one with punctuation included & one without removing the punctuations.

After reaching a test accuracy of 61% (57% in case of without punctuation data), the learning got saturated, hence the model could not converge. Because of the presence of so many <unknown> tokens & skewed frequency distribution of words, the generated sequences are mostly not sensible beyond 3-4 grams window (5-6 grams in case of without punctuation data)

- b. Vocabulary size = 8826 (without punctuation), 9071 (with punctuation)

- c. Yes, I did the following modifications-

- For learning from "without punctuation" data -

SEQUENCE_LENGTH = 30 (less since we're dealing with words here)

BATCH_SIZE = 128

FEATURE_SIZE = 512

TEST_BATCH_SIZE = 128

EPOCHS = 30
LEARNING_RATE = 0.001
WEIGHT_DECAY = 0.0005
Optimizer - Adam
Beam width = 20
Temperature = 1.0

❑ For learning from “with punctuation” data -

SEQUENCE_LENGTH = 30 (less since we're dealing with words here)
BATCH_SIZE = 128
FEATURE_SIZE = 1024
TEST_BATCH_SIZE = 128
EPOCHS = 30
LEARNING_RATE = 0.001
WEIGHT_DECAY = 0.0005
Optimizer - Adam
Beam width = 20
Temperature = 1.0