Deep Learning Lab: Language Models

Anthony Bugatto

4 December 2022

1 Preliminaries and Reading Comprehension

1.1 Text Data

1. From examining the data we find that the number of lines in the dataset is 5033, the total number of characters is 177517, and the number of unique characters is 107.

2. If text preprocessing were required then I would do would make all characters lower case so that the generated text would not consider lower and upper case characters of the same letter different.

1.2 Dataloader and Batch Construction

1. The get_idx function in Vocabulary first checks if the string is already in the vocabulary and if so returns the corresponding index, if not it then checks if the user wants to extend the vocabulary and if so adds the word to the vocabulary and returns it's index, and if else it returns the unknown id token.

2. In the Vocabulary class the id_to_string dictionary stores the $\{id : token\}$ map and the string_to_id dictionary stores the $\{token : id\}$ map. The id's are chosen incrementally in the order in which each new token is discovered.

3. Calling the __len__ method in the TextData class returns the number if tokens in the dataset.

4. Calling the __len__ method in the DataBatches class returns the number if batches in the dataset.

5. The first line creates a pytorch vector of size $(bptt * batchsize)$ and fills it with padding tokens. The second line copies input_data.data, or the pytorch vector containing the tokenized Aesop's fables, into the first $|data|$ elements of the vector, leaving the remaining as padding tokens.

6. $padded[0 : bptt\_len]$ is of shape $[bptt = 64]$

7. $padded[i * bptt\_len - 1 : (i+1) * bptt\_len]$ is of shape $[bptt + 1 = 65]$ because it includes the last element of the previous batch at the start.

1.3 Modeling, Training, and Decoding

1. It is important to detach the hidden states of the RNN so the hidden states are detached from the computational graph, allowing usage of the values but not gradients. This prevents the gradient descent from going beyond this hidden state, which occurs every batch for truncated backpropogation through time.

2. The first index is ignored in the cross entropy loss because it contains the value from the previous sequence and it's gradient was already computed in the previous batch.

3. The input shape expected in self.RNN is $(D,)$

4.

5. We call complete in the training loop so that we can get a sense of how the RNN output changes throughout training

2 Experiments using RNN

We can see that the perplexity of the RNN gets below 1.8 to a best of 1.43.

During training we can observe the text completion quality improve drastically. Using the prompt ”Dogs like best to” it went from

”u the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou the ou”

in epoch 0 batch 30 to

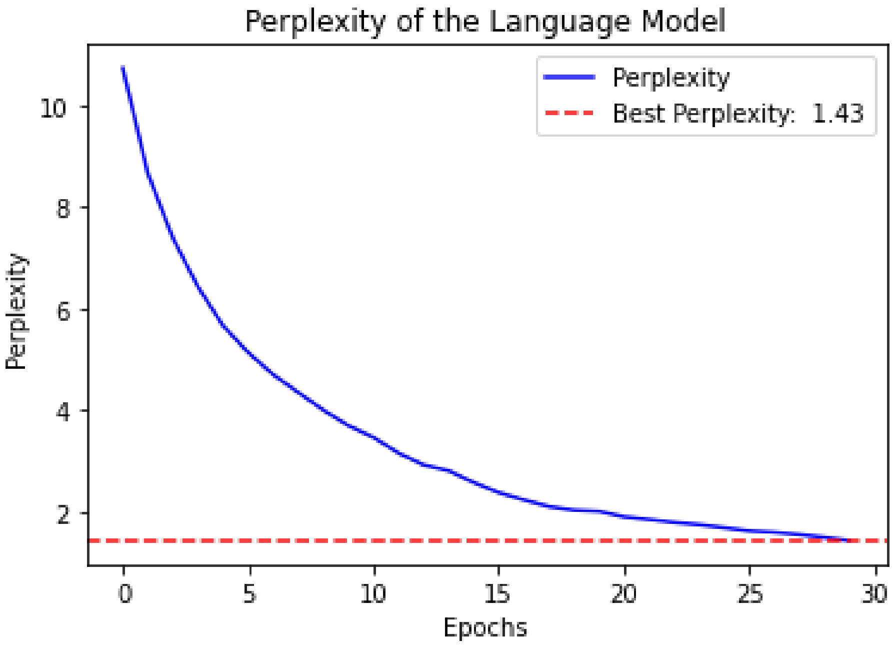
”pass by any blaze by a sabe me as so that he was a more said the Frog from the great state as the master than a marthed in a man”

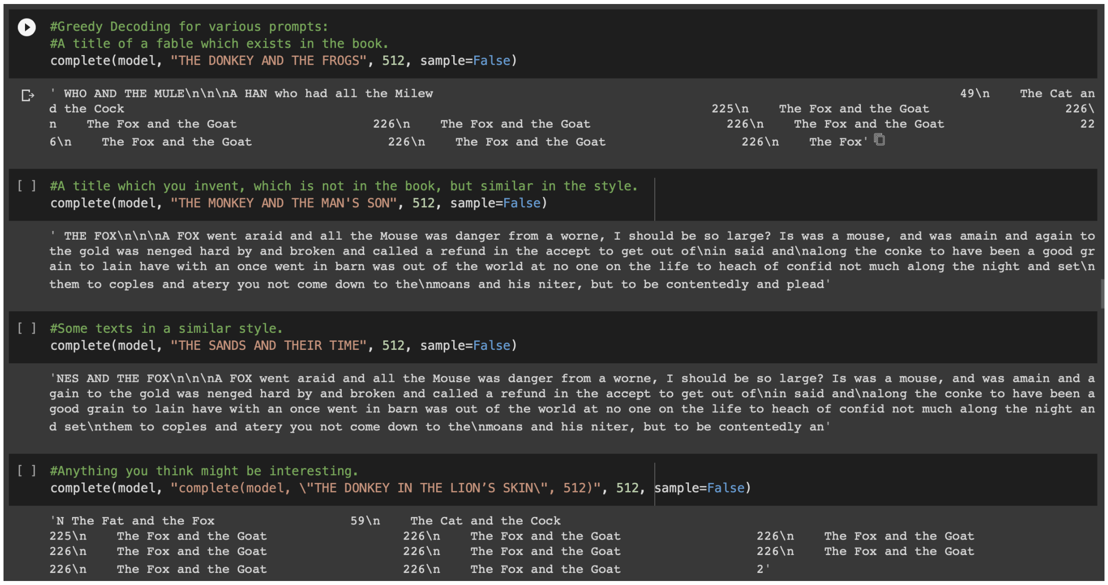
in epoch 15 batch 0 to

”try to pleas revorates so much that he looked down upon the bank of down the stick of his presences about the edge of the wood, ”

in epoch 29 batch 30.

In the figure below we see the output for greedy decoding as described in section 2.3. We can see that the writing in all prompts is largely gibberish, however it does seem to be able to replicate the style of Aesop's Fables. This is very likely due to the lack of memory in RNN architecture.





```
#Greedy Decoding for various prompts:
#A title of a fable which exists in the book.
complete(model, "THE DONKEY AND THE FROGS", 512, sample=False)
```

```
' WHO AND THE MULE\n\n\nA HAN who had all the Milew                                                       49\n    The Cat an
d the Cock                                                                                  225\n    The Fox and the Goat                226\
n    The Fox and the Goat                    226\n    The Fox and the Goat                    226\n    The Fox and the Goat                22
6\n    The Fox and the Goat                    226\n    The Fox and the Goat                    226\n    The Fox'
```

```
#A title which you invent, which is not in the book, but similar in the style.
complete(model, "THE MONKEY AND THE MAN'S SON", 512, sample=False)
```

```
' THE FOX\n\n\nA FOX went araid and all the Mouse was danger from a worne, I should be so large? Is was a mouse, and was amain and again to
the gold was nenged hard by and broken and called a refund in the accept to get out of\nin said and\nalong the conke to have been a good gr
ain to lain have with an once went in barn was out of the world at no one on the life to heach of confid not much along the night and set\n
them to coples and atery you not come down to the\nmoans and his niter, but to be contentedly and plead'
```

```
#Some texts in a similar style.
complete(model, "THE SANDS AND THEIR TIME", 512, sample=False)
```

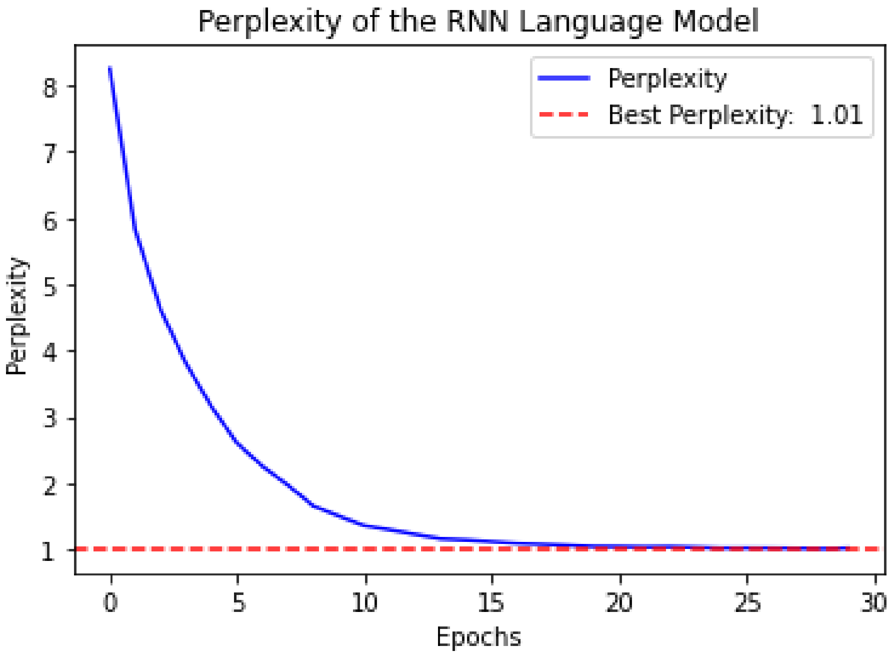
```
'NES AND THE FOX\n\n\nA FOX went araid and all the Mouse was danger from a worne, I should be so large? Is was a mouse, and was amain and a
gain to the gold was nenged hard by and broken and called a refund in the accept to get out of\nin said and\nalong the conke to have been a
good grain to lain have with an once went in barn was out of the world at no one on the life to heach of confid not much along the night an
d set\nthem to coples and atery you not come down to the\nmoans and his niter, but to be contentedly an'
```

```
#Anything you think might be interesting.
complete(model, "complete(model, \"THE DONKEY IN THE LION’S SKIN\", 512)", 512, sample=False)
```

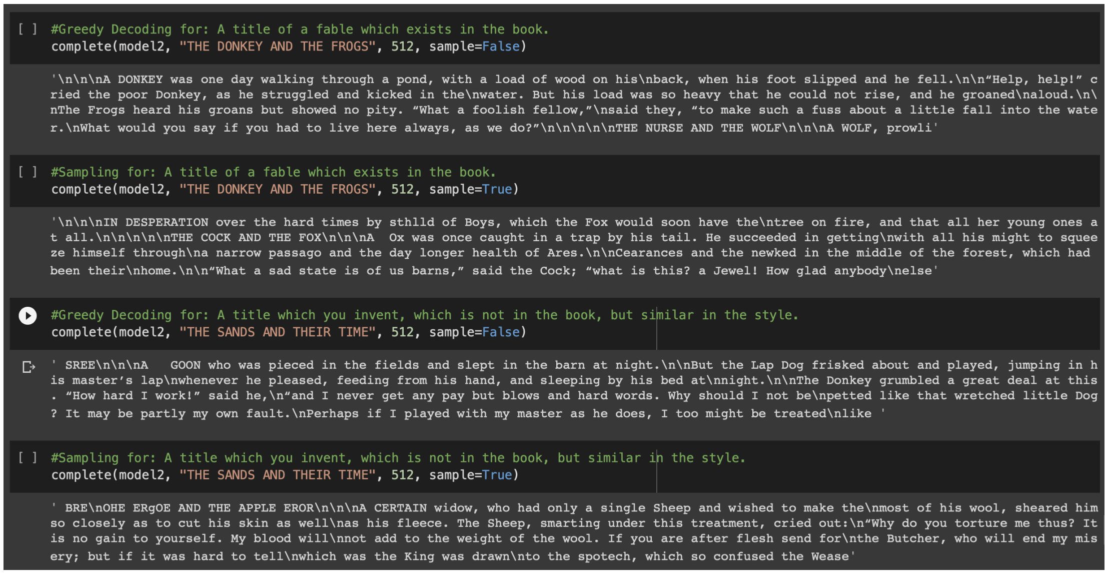
```
'N The Fat and the Fox                    59\n    The Cat and the Cock
225\n    The Fox and the Goat                    226\n    The Fox and the Goat                    226\n    The Fox and the Goat
226\n    The Fox and the Goat                    226\n    The Fox and the Goat                    226\n    The Fox and the Goat
226\n    The Fox and the Goat                    226\n    The Fox and the Goat                    2'
```

3 Experiments using LSTM

We can see that the perplexity of the RNN gets below 1.8 to a best of 1.01.



From 3.4, the same title from the book and invented title were used to test the difference between the sampling and greedy decoding. The sampling appears worse because sometimes the letters in the words don't make sense together.



```
#Greedy Decoding for: A title of a fable which exists in the book.
complete(model2, "THE DONKEY AND THE FROGS", 512, sample=False)

'\n\n\nA DONKEY was one day walking through a pond, with a load of wood on his\nback, when his foot slipped and he fell.\n\n“Help, help!” c
ried the poor Donkey, as he struggled and kicked in the\nwater. But his load was so heavy that he could not rise, and he groaned\naloud.\n\
nThe Frogs heard his groans but showed no pity. “What a foolish fellow,”\nsaid they, “to make such a fuss about a little fall into the wate
r.\nWhat would you say if you had to live here always, as we do?”\n\n\n\n\nTHE NURSE AND THE WOLF\n\n\nA WOLF, prowli'
```

```
#Sampling for: A title of a fable which exists in the book.
complete(model2, "THE DONKEY AND THE FROGS", 512, sample=True)

'\n\n\nIN DESPERATION over the hard times by sthlld of Boys, which the Fox would soon have the\ntree on fire, and that all her young ones a
t all.\n\n\n\n\nTHE COCK AND THE FOX\n\n\nA  Ox was once caught in a trap by his tail. He succeeded in getting\nwith all his might to squee
ze himself through\na narrow passago and the day longer health of Ares.\n\nCearances and the newked in the middle of the forest, which had
been their\nhome.\n\n“What a sad state is of us barns,” said the Cock; “what is this? a Jewel! How glad anybody\nelse'
```

```
#Greedy Decoding for: A title which you invent, which is not in the book, but similar in the style.
complete(model2, "THE SANDS AND THEIR TIME", 512, sample=False)

' SREE\n\n\nA   GOON who was pieced in the fields and slept in the barn at night.\n\nBut the Lap Dog frisked about and played, jumping in h
is master’s lap\nwhenever he pleased, feeding from his hand, and sleeping by his bed at\nnight.\n\nThe Donkey grumbled a great deal at this
. “How hard I work!” said he,\n“and I never get any pay but blows and hard words. Why should I not be\npetted like that wretched little Dog
? It may be partly my own fault.\nPerhaps if I played with my master as he does, I too might be treated\nlike '
```

```
#Sampling for: A title which you invent, which is not in the book, but similar in the style.
complete(model2, "THE SANDS AND THEIR TIME", 512, sample=True)

' BRE\nOHE ERgOE AND THE APPLE EROR\n\n\nA CERTAIN widow, who had only a single Sheep and wished to make the\nmost of his wool, sheared him
so closely as to cut his skin as well\nas his fleece. The Sheep, smarting under this treatment, cried out:\n“Why do you torture me thus? It
is no gain to yourself. My blood will\nnot add to the weight of the wool. If you are after flesh send for\nthe Butcher, who will end my mis
ery; but if it was hard to tell\nwhich was the King was drawn\nto the spotech, which so confused the Wease'
```

Two alternative datasets were tried: Harry Potter and the cnn.py from project2. The cnn.py didn't give understandable results but the harry potter seemed to understand the basics of the plot. These weren't included due to timing constraints.

4 Questions

1. $$P = (\prod_{i=1}^{N} \frac{1}{V})^{-\frac{1}{N}} = e^{-\frac{1}{N}\sum_{i=1}^{N} log(\frac{1}{V})} = e^{log(V)} = V$$

2. Vanishing gradients is an issue because it causes the training to effectively stop as the product of bptt gradients gets smaller and smaller with $|bptt|$, and consequently the size of the stochastic gradient descent steps.