Deep Learning Homework3 Report

0710891

December 20, 2021

1 RNN

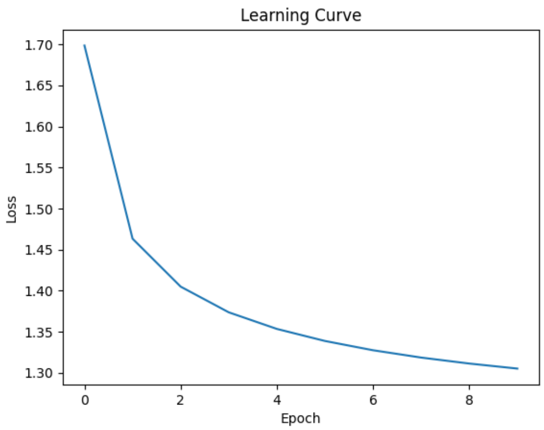
1.1

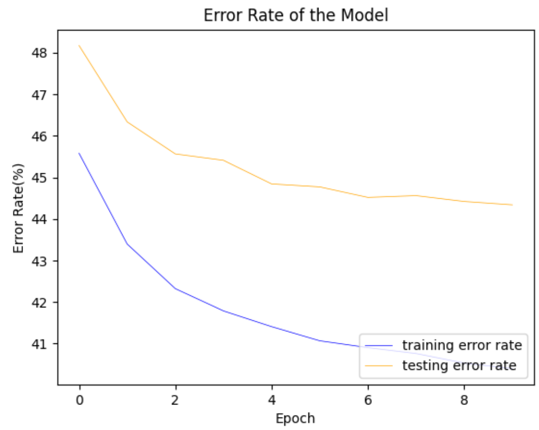
The following table shows the structure of my RNN model.

Layer Type	Input	Output
Input	(100)	(100)
Embedding	(100)	(100, 128)
RNN	(100, 128)	(512)
Dense	(512)	(67)

After briefly showing the structure, we go further into some other configurations. The number of training epochs is set to only 10 because RNN models really take a large amount of time to train, so I cannot choose a relatively large number. As for the number of hidden units, according to my experiment, larger number of hidden units yields better training result. However, the long training time also prevents me from adding even more hidden units. As a result, I have to settle for a relatively large number of hidden units, 128.

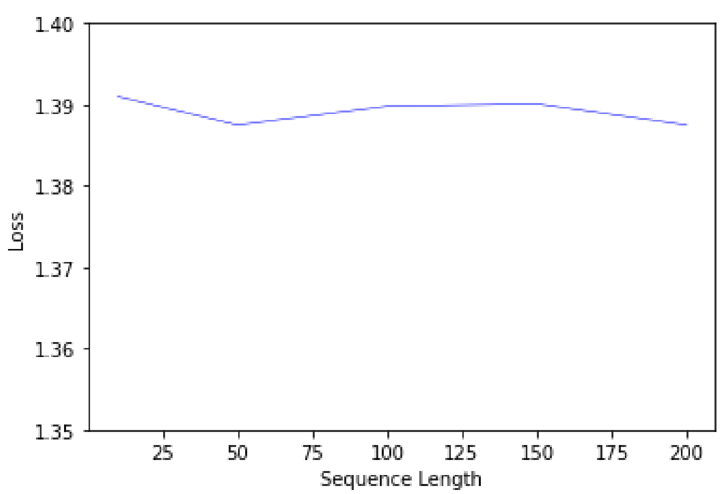
The following plots show the learning curve and error rate throughout the training process.





1.2

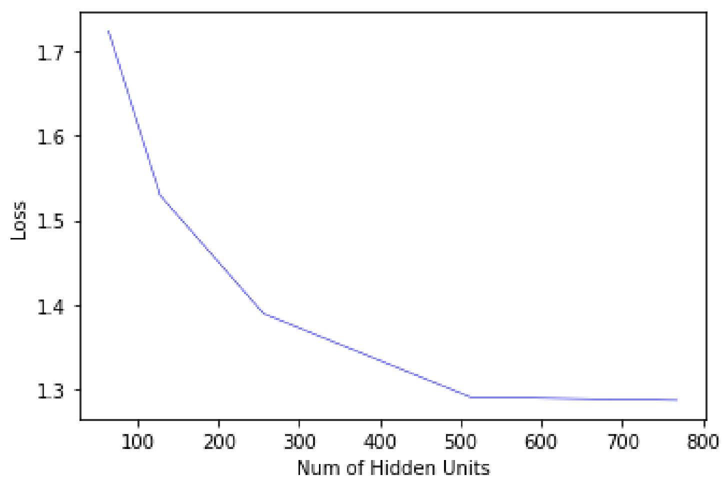
As I mention above, the sequence length and the number of hidden units both have a chance to affect the training process.
The following plot shows the relationship between sequence length and the training loss.



As we can see, actually the difference in sequence length dose not have an obvious impact on training loss. The reason behind is probably that standard RNN model intrinsically cannot learn long-term information well, so the protraction of sequence length cannot improve training loss much.

Next, we take a look at how the difference in number of hidden units affect the

training process.
The following plot shows the relationship between number of hidden units and the training loss.



The plot clearly shows that, as the number of hidden units increases, the training loss decreases.

1.3

Finally, we inspect the performance of model. The following texts are generated by the RNN model after trained for 10 epochs.

”

KING LEAR:
’v$.kRRppnJOc]RvpNfu

Slb
$!h$Zs$-SBl$
hi$$b,lTppp:nxFLLLxvvRRppE.$txl]XRRppuJfvw]xREp
HQ?
q?ZqkhqNh
QX-mhCl$
,qSmkl
sqhh$Qk

im-bhQ-
k-Shl$
TXpplz.Ox
AOXNN

lOmXh?q
Z?-$-lkX
Nq-skbqQQ
Ri
qQbN$
pkHq-hmSlk?$$h,

qSSbk
$lQXTqppfttFHAx
R'pQ:-XZRqpk'Sqbb
SPlTyppOze$]RRpp]nmKOKFtxHRRpp'ftwPnvORRppeLF.OOH
RRppKOt.nxtHRRppn:LOtHJfRvpRH
SibQ,N
kS-sk$$qNkp
&—uRl

–iB
SFkT$ppf]tFnxtDFRRppKJ.ex:EfRvpRwp?w-qiblQ
"

"KING LEAR:" is used as prime text in this part. Although the output of the model seems like total nonsense, we can still observe some patterns in it.
First, the texts in the output are separated into paragraphs just like training data, and the first row of each paragraph is generally shorter than other rows in the paragraph, which matches the name row in the training texts.
Furthermore, we can see that some punctuation marks are correctly placed before newline characters, which is a common pattern in the training data.
Also, we can see the model actually learned to use double dash,which is also common in training texts, in its prediction.

2 LSTM

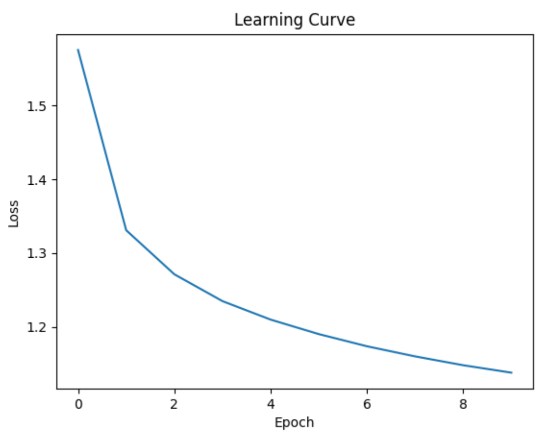
2.1

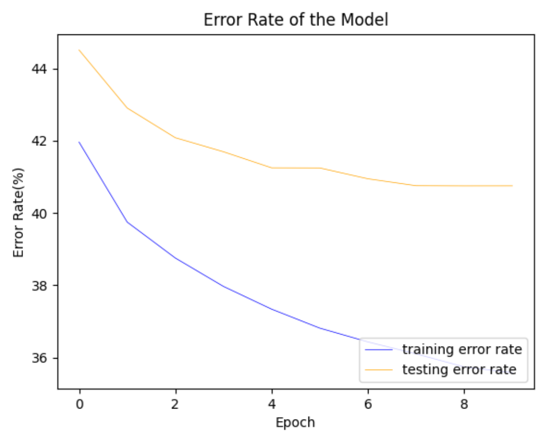
The following table shows the structure of my LSTM model.

Layer Type	Input	Output
Input	(100)	(100)
Embedding	(100)	(100, 128)
LSTM	(100, 128)	(512)
Dense	(512)	(67)

Almost every hyperparameters and configurations are set to the same value as in the RNN part so that we can have a fair comparison between these two.

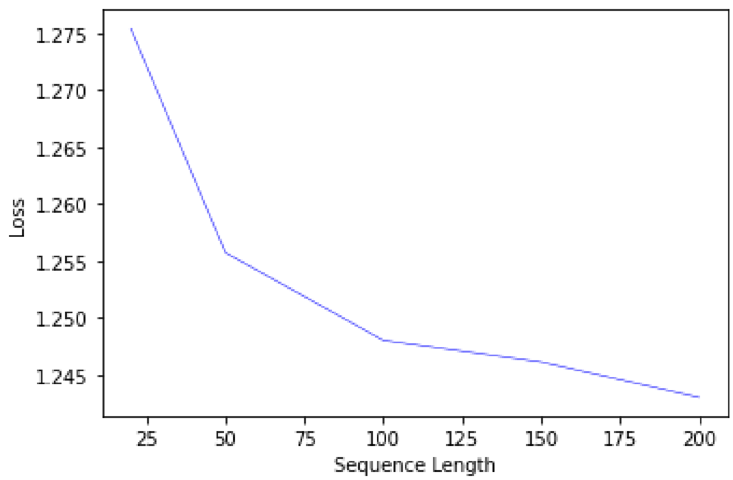
The following plots show the learning curve and error rate throughout the training process.





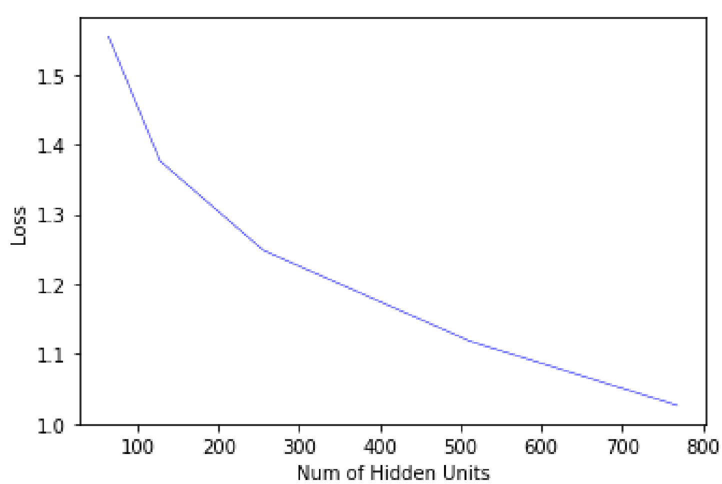
2.2

Here we display how sequence length and number of hidden units affect the training loss just like the RNN part.
The following plot shows the relationship between sequence length and the training loss.



Different from the RNN part, the training loss does go down as the sequence length increase. The discrepancy is caused by the fact that LSTM is designed to learn both long-term and short-term information. As the sequence length increases, the LSTM model can learn more long-term information, thereby increase the performance of model.

Next, we take a look at how the difference in number of hidden units affect the training process.
The following plot shows the relationship between number of hidden units and the training loss.



Just like the RNN part, as the number of hidden units increases, the training loss decreases.

2.3

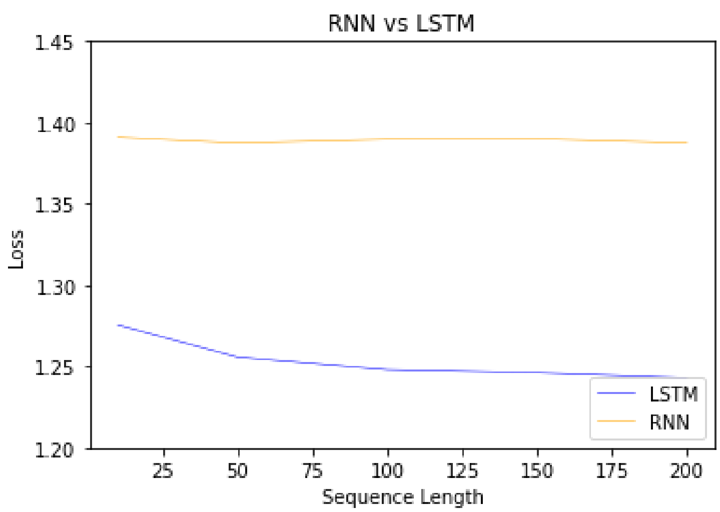
Finally, we inspect the performance of model. The following texts are generated by the LSTM model after trained for 10 epochs.

```
"
KING LEAR:
-e ,,s;pm;I,iZ!;wKe
inEXwKsZwwwnpy;K:Nr

!snw
pNFim
L!,wZFKKprw
&eKK
l!Kwu!swwE,FtmKii!NEXip!nw;eKFpKK,Zsww-&,wspp;FKKnW ,

sn] Km$uQ
-! wQNsKXZ,
sseFjKKBZw
Kni Nw;pKKpi nKp
,-[ wiNneX,]!mmHriD!$jwmmRuOD.WWjjmmzO.RcILDvOhxxWWjxm:D$J,e$

nNwwtKK
nNK
$n
Knp,KNZii!!pw wKe
wnpKjZm
iKe$Qi!nnKwn[Kwi!Eww!p]KKJz,
Ne;KKZPw,p-
w[ew];KKispmnR,KKNGi
anii!
tnKiwnsiwpKK
s,ip!wwepK;nK
nnmsni
eNKXZ,
Nn]tmKvisppnK,Ea
"
```

As we can see, the result of LSTM prediction has even more punctuation marks at the end of a row, which probably means that LSTM has predict better than RNN. However, the entire output is still nowhere near normal texts, showing that 10 epochs is definitely inadequate for such complex problem.

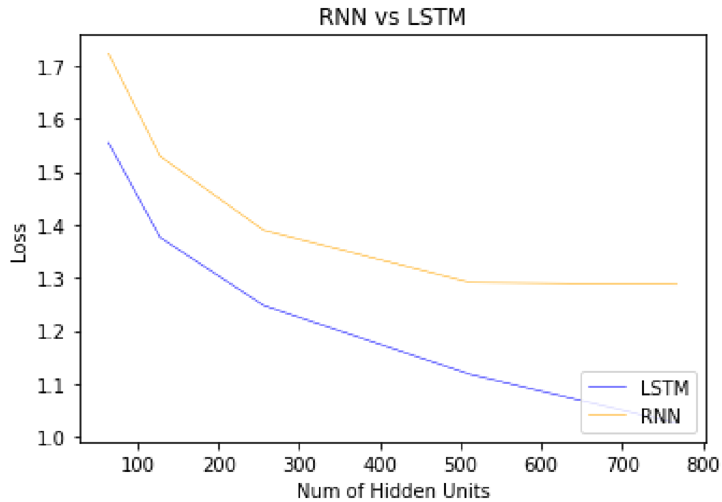
3 Comparison

After inspecting two models, now we make a quick comparison between these them.
First we check the effect of sequence length on these two.



We can see that only the loss of LSTM can be affected by sequence length. As already mentioned above, standard RNN has difficulty in learning long-term information, thereby cannot be improved by longer input sequence. On the other hand, LSTM can learn both long-term and short-term, so giving longer input sequence makes it easier to capture long-term info. Also, benefited from this attribute, we can observe that the loss of LSTM is obviously lower than standard RNN.

Then we check the effect of hidden units on these two.



As we can see, even though the loss of standard RNN is overall larger than the LSTM, their loss both decrease significantly when the number of hidden units increase. This phenomenon is quite comprehensible because more hidden units can detect more features.