Name : M. Farhan Tandia
Student ID : 0860814

1) Preprocessing of Shakespeare dataset

In this homework, I build a recurrent neural network for the character-level language model. I design the network architecture, including the number of hidden layers, the size of hidden states, learning rate, sequence length and mini-batch size. The dataset contains a lot of conversation from various collections of William Shakespeare's writings, so that it contains a connection between one subject's conversation with another subject.

First I read the data from 'shakespeare_train.txt' and counting how many characters and unique characters in that file. The training data have 4,351,512 characters and 67 unique characters. After that I need to vectorize the text data to a numerical representation by creating two lookup tables, such as one mapping characters to numbers, and another for numbers to character like shown below.

```
    {
'\n':   0,
' ' :   1,
'!' :   2,
'$' :   3,
'&' :   4,
"'" :   5,
',' :   6,
'-' :   7,
'.' :   8,
'3' :   9,
':' :  10,
';' :  11,
'?' :  12,
'A' :  13,
'B' :  14,
'C' :  15,
'D' :  16,
'E' :  17,
'F' :  18,
'G' :  19,
   ...
    }
67 unique characters
```

and this is an example about the data looks like after mapping into integer.

```
'First Citizen' ---- > [18 49 58 59 60 1 15 49 60 49 66 45 54]
```

After that, I divide the text into example sequences. Each input sequence will contain 'sequence_length' characters from the text. For each input sequence, the corresponding targets contain the same length of text, except shifted one character to the right. So break the text into chunks of 'sequence_length+1'. For example, say 'sequence_length' is 4 and our text is

"Hello". The sequence would be "Hell", and the sequence "ello". The prediction ta sk in this work with given a character, or a sequence characters, in this case what is the most probable next character? This is the task I training the model to perfor m. The input to the model will be a sequence of characters and we train the model to predict the output in the following character at each time step.

And then I make a mini batch for training data for converting those individual characters to sequences of the desired size. For each sequence, and shift it to form the input and target text by using the 'map' method to apply a simple function to each batch. Getting ready for training process, so I split the text into manageable sequences. But before feeding the training data after preprocessing above into the model, then I shuffle the data and pack it into batches. I also did the preprocessing for data validation which is the same as the training data except of the label for all unique character that follows from the training data.

2) Recurrent Neural Network

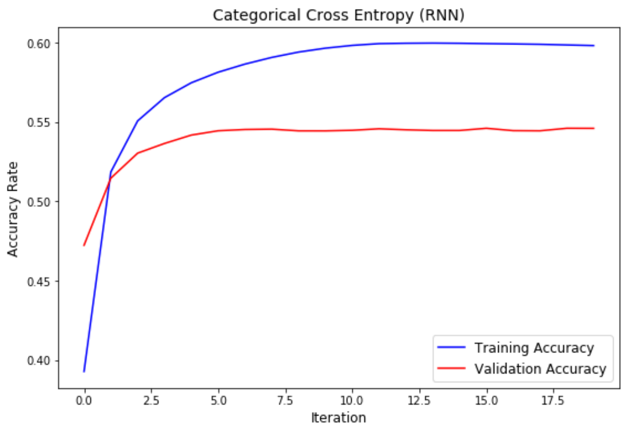
2.1) Standard RNN (Model and Result)

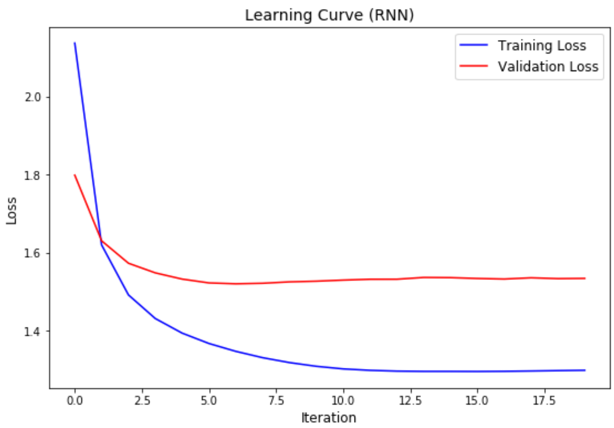
I construct a standard Recurrent Neural Network with a batch of input data. I use three layers for the training network, the embedding layer is a layer to find the optimal mapping of each of the unique words to a vector of real numbers. The size of that vectors is equal to the dimension of output, For text or sequence problems, the Embedding layer takes a 2D tensor of integers, of shape (samples, sequence_length), where each entry is a sequence of integers. It can embed sequences of variable lengths. I could feed into the embedding layer above batches, for example with shapes (32, 10) (batch of 32 sequences of length 10) or (64, 15) (batch of 64 sequences of length 15).

The simple rnn layer is a Recurrent Neural Network model which is a Fully-connected RNN where the output is to be fed back to input, this layer contains parameter of number of RNN Units, the return sequences and stateful as an option, and also the recurrent initializer which in this work I use Glorot normal initializer (Xavier normal initializer). And the last layer is dense layer as the fully connected layer which has softmax activation function. The network architecture is shown below.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (64, None, 256)           17152
_________________________________________________________________
simple_rnn (SimpleRNN)       (64, None, 1024)          1311744
_________________________________________________________________
dense (Dense)                (64, None, 67)            68675
=================================================================
Total params: 1,397,571
Trainable params: 1,397,571
Non-trainable params: 0
```

The training parameter I use is 20 epochs, 1024 RNN units(hidden states), 64 batch size, 100 sequence length of text, Nadam optimizer and also the loss function is BPC (bits-per-character) or called sparse categorical cross-entropy. And for the results of training and validation accuracy and error rate are shown below.





It can be seen that the performance of the standard RNN model is not too good, but also not too bad, considering I only use the simple model to predict a lot of variate character in dataset. The convergence is fast but the model still suffered from a little overfitting.

2.2) 5 Breakpoints (Generate Text per Epoch) with Standard RNN

I use epoch of 20 to train the Fully-connected RNN model and make a breakpoint at the end of an epoch training using callbacks function on TensorFlow. So I choose 5 breakpoints during the training process to show how well the network learns through more epochs, the 5 breakpoints are at epoch of 4, 8, 12, 16, and 20. The generate text with a diversity value of 1 from the breakpoints as follows.

Break points	Epoch	Generate Text (approx. 10 to 15 lines)
1	4	DUKE: For secured, from these traitor lascall, The others lights our old waste am abands. Sweet'st thought I'll move you hence. LAFEU: As I have done but known them thy barnize, cast, co ffers! nis appoieddy leave Tokens, we did plumine mocks, Unless you writ to provoked dark with saint? Innews , as were discliates; And, or eagland, For was the sound it ever I gail now to make them b oth laughants and My life. Clown:

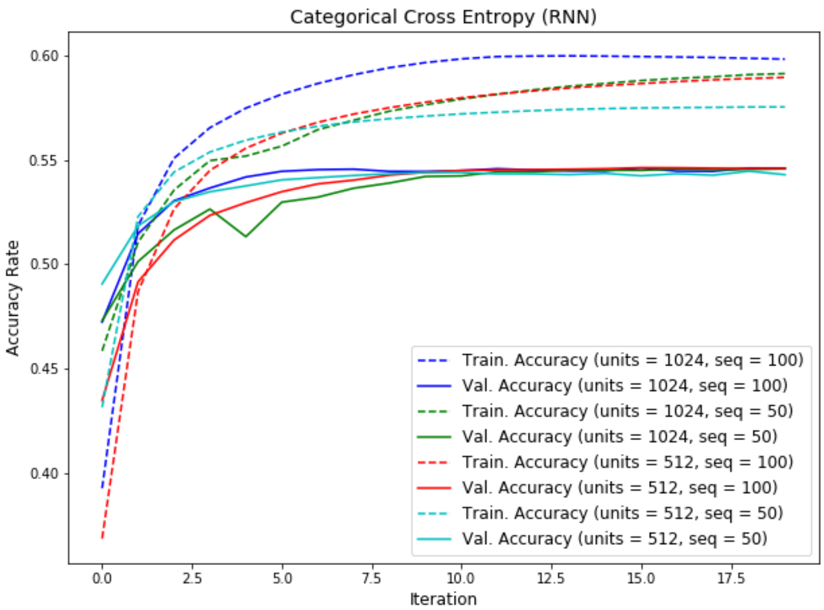
		MISTRESS PAGE: Nay, well, Ciacie?
2	8	WHITMORE: Please their daughter? By mercy! I'ld to myself, Puff'd by all togethers? Says thou! We shall, I do beseech you call her down. Offleralve drawn to have to have done Postuse; Whereon they lift upon't, Cassio: Therefore let you was death; were mercy. Call i' the ambition, and the imperuality, After thee, are they that stay About myself; I call by half before Their way? APHELIA: Faith'd, pity: Nay, my lord: All you the English about thee! Coiside, Than you, it is touch of his blood, To thewer hi
3	12	SUFFOLK: You have nod be hanged upon, Affection makes in the fagself hidest of late; My stevels here to lid upon her master Force shed, I'll peuck'd in their life, would I could, Meigually dost thou beg power I do; That thou mightst weaspent surmire than mask'd that part, King HARt thou the care: Where is the king. LUCIUS: I would my Lord of Wings I make again, forget thee lend not equaley? who very divorf you know like, and much worth fail: if he please, I think I fear you matter our employmeny securi
4	16	BRUTUS: Where laugo, you must come peases' as militaries. Commanness, in her majesty For Herallow sicks, my lord; Thomasters, Whose brother's broken: if thou with a rival. MIWARD: Is't follow about; and that my word lovers and drops did shrufty exervings hark May have betterleter! We are come such anon Horself, douson than wisely for it, and I would stand along wipper.

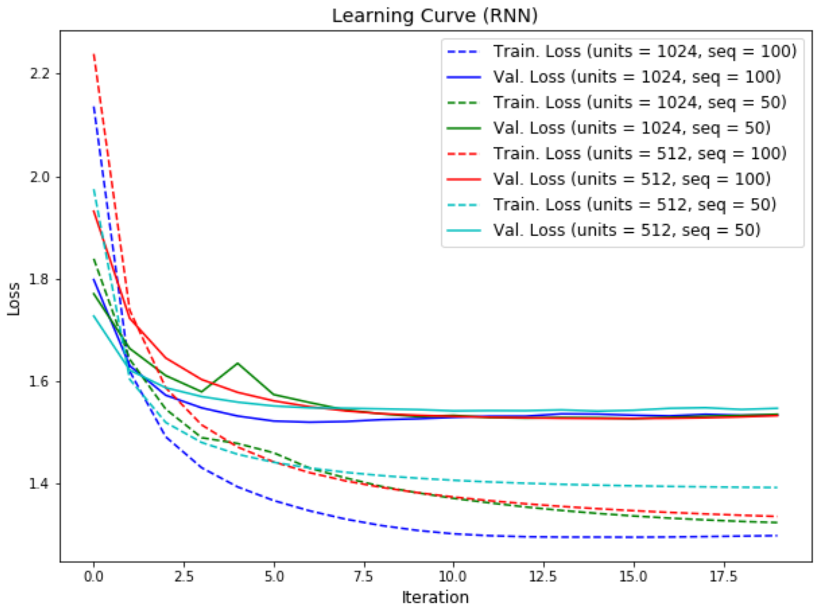
		Now, if Hence sent bold, give me an women will rath er puissant and emeth, wouldst ne'en swears: and le t it so are now down; I am
5	20	Servant: Then weep precious gods as Caesar's ast of sweet yo u. MISTRESS PARIP: Are sound, and help the truth, For I have a colour, glitter? So wit, to make my third of France. PRINCE HENRY: Five humeforce or good Sir for, if thou wilt did preced the time, And meet with her, no tights between, The feast it is; And not a sigh enough: Yea, mad dneige! well, my authorrork--a fare to thi ne!

From the generated text from each breakpoint, it can be seen that in breakpoints 1 to 3, the vocabulary is so much wrong, but the text format is already correct. Then in breakpoints 4 and 5, the model tries to make a correct complete sentence and resulting a little better vocabulary than the previous breakpoints. However, the overall result is not good as expected, because of the model still can not make a conversation based on the story context.

2.3) Different Size of Hidden States and Sequence Length Comparison

In this part, I will be comparing the results of different sizes of hidden states and sequence length. I tried changing those parameters into 4 different sizes. The plot of the training loss vs. different parameters as follows.





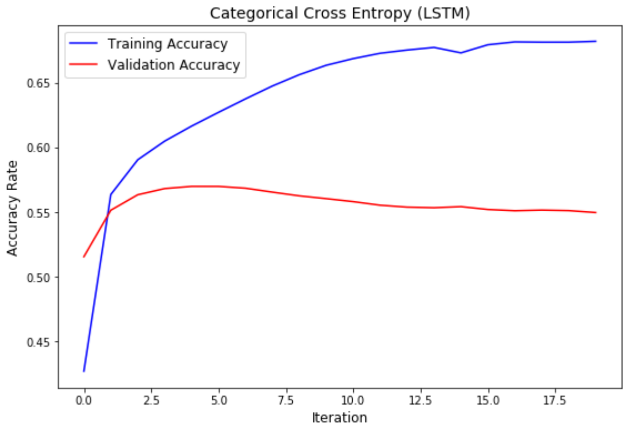
As we can see, the first that easily noticed is that the validation accuracy and validation loss is about the same. The accuracy is around 0.55 and the loss is around 1.5, so there is no significant difference if the hidden states and sequence length were changed in my simulation. However, the training performance has a little bit of effect, where the best accuracy for training is the ‘cyan’ line because it has the lowest overfitting condition.

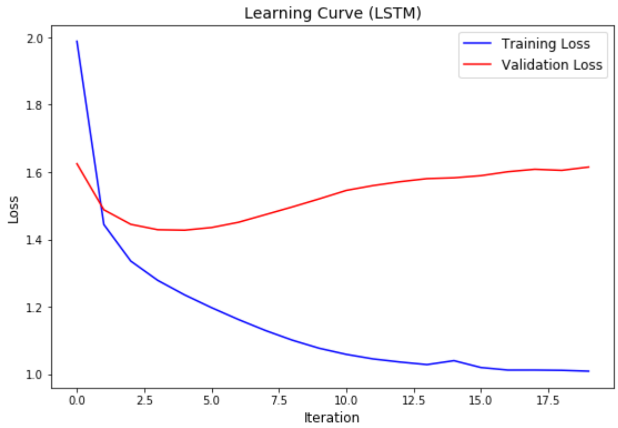
2.4) Long-Sort Term Memory

I construct a standard Long-Short Term Memory model with a batch of input data. I use three layers for the training network as the same as standard RNN, and just change the model into LSTM. The network architecture is shown below.

```
_________________________________________________________________
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (64, None, 256)           17152
_________________________________________________________________
lstm (LSTM)                  (64, None, 1024)          5246976
_________________________________________________________________
dense_1 (Dense)              (64, None, 67)            68675
=================================================================
Total params: 5,332,803
Trainable params: 5,332,803
Non-trainable params: 0
```

The training parameter I use is 20 epochs, 1024 RNN units (hidden states), 64 batch size, 100 sequence length of text, Nadam optimizer and also the loss function is BPC (bits-per-character) or called sparse categorical cross-entropy. And for the results of training and validation accuracy and error rate are shown below.





It can be seen that the performance of the model is not much improved from the standard RNN because the validation loss is also the same, but it has a little better accuracy and just the LSTM model has a higher accuracy on training data which indicates the model suffers from overfitting.

2.4.1 LSTM Model and Result

2.4.2 5 Breakpoints (Generate Text per Epoch) with LSTM

I use epoch of 20 to train the LSTM model and make a breakpoint at the end of an epoch training using callbacks function on TensorFlow. So I choose 5 breakpoints during the training process to show how well the network learns through more epochs, the 5 breakpoints are at epoch of 4, 8, 12, 16, and 20. The generate text with a diversity value of 1 from the breakpoints as follow.

Break points	Epoch	Generate Text (approx. 10 to 15 lines)
1	4	POMPEY: Yea, son and friends: he brings hich Bille word--i' the field must die to Lonatio, Sons, credit and being afternor'd: Good deers, Jack Cuckus arison, as the villain: thou gh dinner, if the poor counsellor; And yet my neck, I presently afore. Therefore be more encaped The ragging do resired by this contemplation, And He from greatness desire home to sea. MACBETH: A beard titieur gates, losers, nobly, adacts! God ke ep the word, reging to work. I do not think whentou shiging, sir: to the bandarous

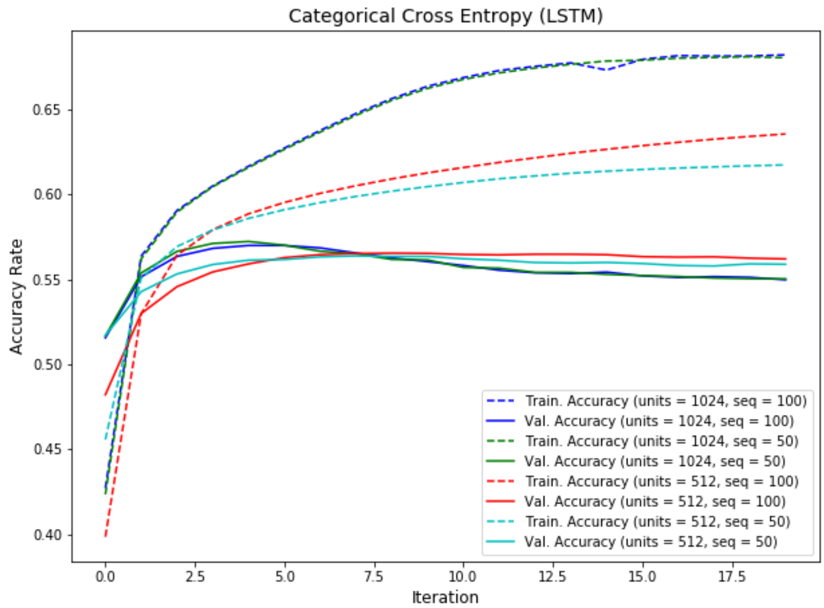
2	8	POMPEY: Sir, but I fear it is a knight; therefore, begin to's by fair Grand Have when their blood? Wilt thou make sound to her, Why then, defend your love with my physic. DUNCAN: Adieu, my lord! Have you not weeds an honest man? have not, The store of ito you take one of nothing But your own following, at the groves damneir, Lucen tio.--take me, cousin, fare, Whom I do know our displeasure, love, I give, By this ill-outwives knowing, conditing: I would do I will burse? O grief, I much; I'll discuss my
3	12	PROTEUS: I think there's my chief tutoff by the runnal of those that intends my son, at last urur ma id: sweethear it breasty? Son: The prince and Cyprus towere of it: throw too much, And lose the Lord Timon's men on their lock'd thee, thief, Thy words be master'd; nor the leastling ton, When I was sentence, I will deal this man. HORTENSIUS: It is true: here's no surprise: here comes the man Of young heart. KING HENRY VI: The certain of his purpose and informs And with the fiery sun its to our graces;
4	16	WARWICK: Then, for though letters to thoss very dignity, Which let this cap but when they lose his play, For brother too much muleters of a chimate; For stand under the heart to thee and thee. QUEEN MARGARET: So let them seek no nobility, Not to beseet upon. CYMBELINE: Which of my lord from Marina is the old, to have don e like a dog, as this islee have a gracious son. ROSENCRANTZ: The guilt, my lord, in good time: O heavy time,-- This cheek were done. You makes me maquistresses with your weapons.

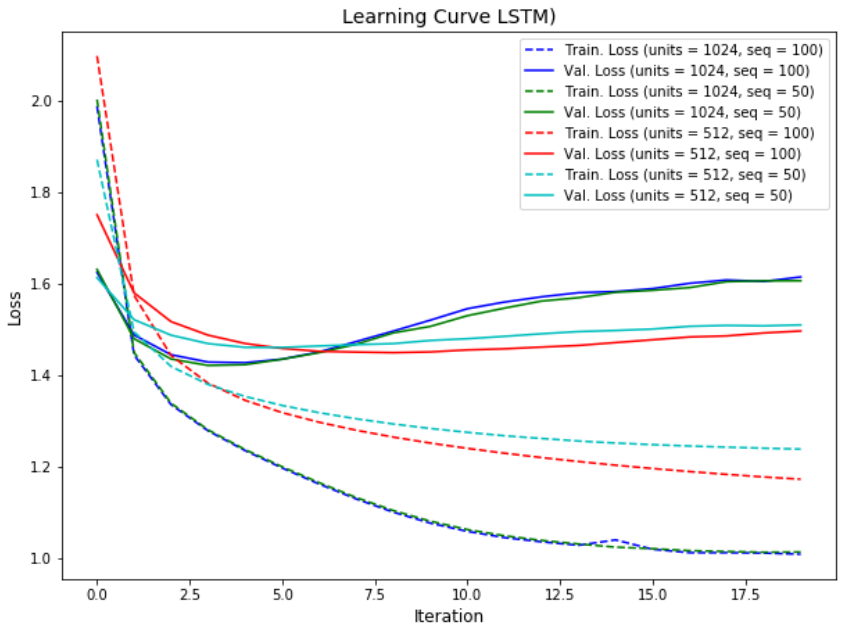
5	20	YORK: And he shall burn, but that the short bent of come so fit as eat and unfeasting and low in judgmen t? What manner of his headless blunts are made! Talk with his heart, more than! thither away. DUKE SOLINUS: What friends, the game't thou art. You walk against it: if you both, I take it alive aFrending a deprivate deed That to destroy; being still best his lord Paris. CAIUS LUCIUS: I pray you, dear lord, They call up like an ellege charnel-- To hazard ducats upon my shoulders With ladies further

The generate text from 5 breakpoints by the LSTM Model has some differences to the standard RNN. The first is the size of the text that the LSTM Model produce is longer than the standard RNN model, this indicates that the LSTM can memorize more information and signify that the LSTM is not suffering from long term memory. And the second is the conversation format and vocabulary is better in LSTM due to the minimal of wrong vocabulary exist in the generated text. So overall, LSTM is a good type of model for generating text and make a prediction using a sequence of data.

2.4.3 Different Size of Hidden States and Sequence Length Comparison

In this part, I will be comparing the results of different sizes of hidden states and sequence length. I tried changing those parameters into 4 different sizes. The plot of the training loss vs. different parameters as follows.

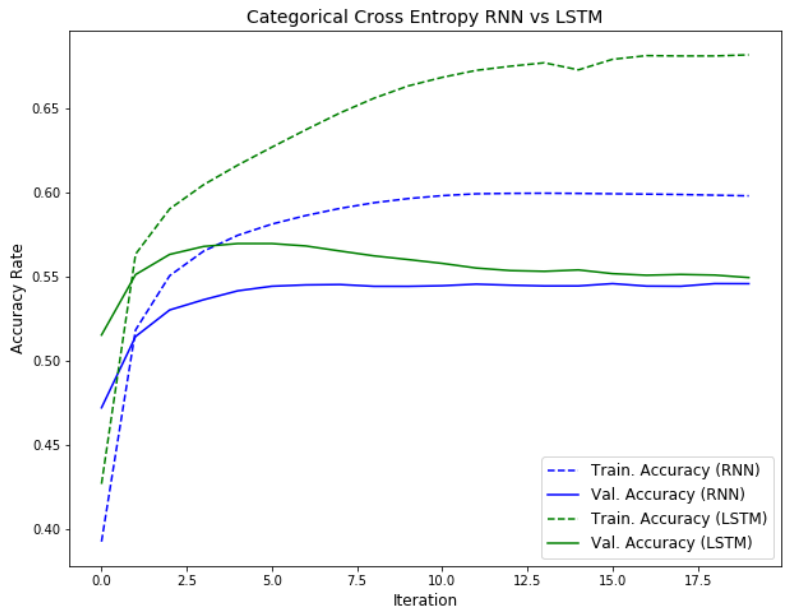


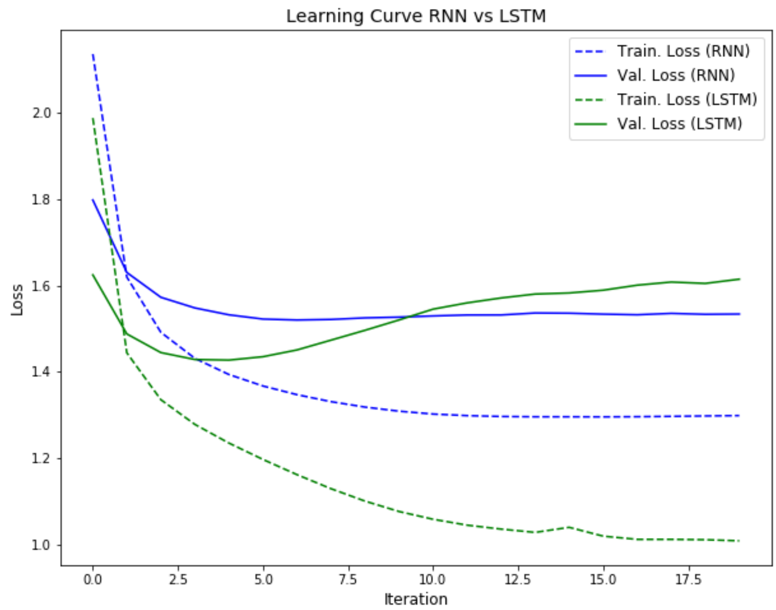


As can be seen from the two figures above, there are slight differences in performance in each size of the parameter. Reducing the size of the hidden unit (red line) made the loss becomes lower from the normal size of parameter (blue line). The loss is around 1.4 and also it can reduce the overfitting by a little.

2.4.4 Standard RNN and LSTM Result Comparison

In this part, I would like to compare the performance of the two models, standard RNN and LSTM. The performance of these two models is shown below.





I can conclude from the two figures above that the LSTM can achieve good performance only until a certain epoch and then it goes to overfitting but for the LSTM have a little better accuracy than standard RNN, for the standard RNN can achieve more stable performance but the accuracy is a little bit lower than LSTM ,however the standard RNN have a little better error rate after a certain epoch.

2.5) RNN/LSTM to Generate Text by Priming The Model

I using the last breakpoints (epoch of 20) to generate text with priming the model, and generate around 500 characters or about 10-15 lines, the result for both models as follows.

2.5.1 Standard RNN

- Prime text : “JULIET”
  Result :

```
JULIET:
And if King Save
What because fast
give me with you. We may be, sir? when I have answered.

PORTIA:
Sir, stand asibe as old man are haunt
hangs you pastle: 'twas a man well, I'll lead
My scene meet,
Throwself some mercy, and fair out,--shean kings
Of you do I see, to his
all my fortunes: pleasure, may you we and told myself,
And sens her waters
Stoought our profounders and beggar this where allow from thei
r day a poor to assure I reas the parally.
```

- Prime text : “Juliet”
  Result :

```
Juliet, thrusting away;
Thy maiden praises, were dead!

LEPIDUS:
Thank you there.
He sen
of England's turn Strange.
'Tis time that on measure
us be the warrant and withal. Away, guck must not love, if
your grace music. He should know an old
As suffering from this shope.
How ends and dogively would be thy reason barish'd soul,
To do you mend more marking on thy shouldest in the war,
Than churded at thy shook, my company.

KATHARINA:
And I, now, and that we are deed?

MICTON:
Augued Saints shall deserte
```

2.5.2 LSTM

- Prime text : "JULIET"
  Result :

```
JULIET:
O brother! For the wind is come by thee.
Discreem no more, rail things fragged.

ULYSSES:
Still lock a thousand ducatss of the Romans: now I
knew him, Antiochus' Sayille,--

SHALLOW:
Yet be not king; some wine, forsooth; son to sat mine:
And thou, Robin Howster, as we would have slept,
And enlittle wits, and so answer with 's;
What well my lord's appetite or others,
Or like to dogs: and art the general's blood like I
know how the interpretesy boy this crest, if thou
think'st on a banne-master-
```

- Prime text : "Juliet"
  Result :

```
Juliet seek now easter all your good house?

BIRON:
This was Epipples came of the king's--
Worthyou rather, the king hath ta'en give you strange burians.

KATHARINE:
So we alive, I will enforce with thee here by low,
Can say 'I hate, he shall be friends.'

LAUNCE:
I thank thee, monsieur York.

Twaings:
Ay, Harry, therefore be not then. As thou art a
tender leath taken in the incer, neither of hersithat less his
Wolves, laid with spend me so? Above the time,
Am I the mother's post, and we'll come down
```

While some of the sentences are grammatical, most do not make sense. The model has not learned the meaning of words, but consider:

- The model is character-based. When training started, the model did not know how to spell an English word, or that words were even a unit of text.
- The structure of the output resembles a play blocks of text generally begin with a speaker name, in all capital letters similar to the dataset.
- As demonstrated below, the model is trained on small batches of text (100 characters each), and is still able to generate a longer sequence of text with coherent structure.