Deep Learning
Fall 2024
Homework 3
Name: Anna Kompan (安娜)
ID: 313540009

1) Simple RNN Model Architecture:

```
Model: "sequential_1"

 Layer (type)                Output Shape              Param #
=================================================================
 simple_rnn_1 (SimpleRNN)    (None, 128)               24832

 dropout_1 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 65)                8385

 activation_1 (Activation)   (None, 65)                0


=================================================================
Total params: 33,217
Trainable params: 33,217
Non-trainable params: 0
```

First model is Simple RNN, it maintains a hidden state that is updated at each step, however simple RNN has a "vanishing gradient" problem so that it can't capture long-term dependencies.

Input shape is 128 neurons (hidden size), next we have a dropout layer with 128 (for dropout dimension isn't changed), dropout prevents overfitting. Next is Dense fully connected layer with 65 classes and the last is activation layer (Softmax) that converts 65 classes of outputs (letters, punctuations) into 0 to 1 values range for text generation tasks.

Data size is [100000:800000], train size is 700000 and valid size is 122025.

Learning Curve and training/validation error rate
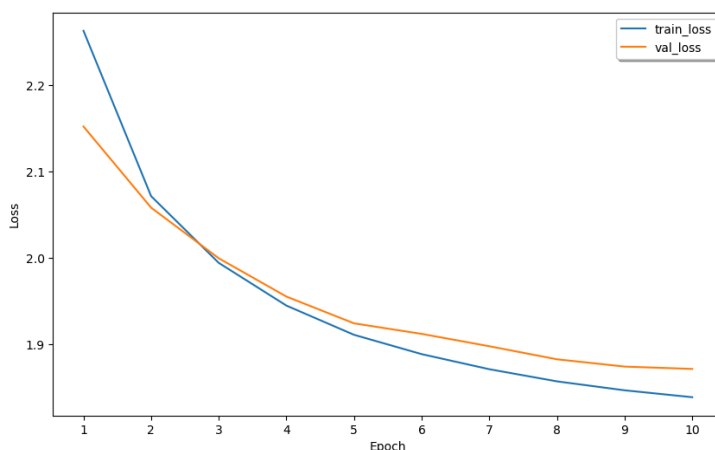

```
Training error rate =
1.8384898900985718
Validation error rate =
1.8712307214736938
```

2) Text generated at 5 breakpoints during training

At epochs 1, 3, 5, 7 and 10, the model generates text. Temperature is set to 0.5 which makes the model focus on high-probability options.

At first epoch generated text has unlogical words that have no meaning, like "oun", "singt", "ard". But still sentence structure reminds of Shakespear's style text. In the next 4 epochs we can see slight improvements, text becomes more structured, in epoch 7 and 10

we can see that model adds up characters dialogue, but still sentences have no logical sense.

Even with training on larger amount of data for more epochs, there won't be the same result as LSTM, since simple RNN still lacks long-term memory. For text-generation task it is better to switch to LSTM.

Generated text after epoch 1:
wife,
Nor how to us oun bead hid singt me and of reake hich me cond couetr the tore and counther,
And may outh hen the coure soree for he sorean,
So shour arding bo pathere some thee for shall tound ard of hearthe:
Whace pordeetrot fowell courant nom the pither,

Generated text after epoch 3:
thou tell? for I wourd of soor gand love to the with lord.
LADY CI I He neng OLINGBROKE:
I last thou his bete of moft as the ford, the sant of mariseng shall such is the peatherse with hath thou hath a tere to make the ford, ather shall gothing the coust then, I hamp at is the gear,
And then to make the ford, his dist that lowgre come to have will sive our good be the friling ont seart.

Generated text after epoch 5:
,
And quench the firitents of the will thou some poreaty the sithon here in the nom, 'are the hather some fnom the weat is the moul from the pround as wall we the death us the treat she thou swand the monture to the love sour son, with have with this sore mone than some tone manest man spoth not conters,
And the prean's manter,
And the treat as the broak of me the mintent the brithen the trat the batter and if the to the parrees word, that the retist as the properte to ser this and more is mone, and the earth,
And men, the greens and is mance to that the love the frot this mare both mone to mone the last not crame poreas of the like, thee to the kenters mort the noll the sore to the brother heary she, we the tatee stand with bear the right,

Generated text after epoch 7:
xenes,'twas nothing; that the toule of the now home of his master I have the rounds of her hath both the brother's forting are and the soull the soull is the soull to shall not sour told and to the words, for at are the seel head of the couther all her for boring that for the pray the words, for of grown, the soulling of thee the know the warly the soull the toulle more and and should not to the sortond it this the toriend the bester the grate for the fall the becounte the and thet my soullother weak and he presen not thinds and srown.
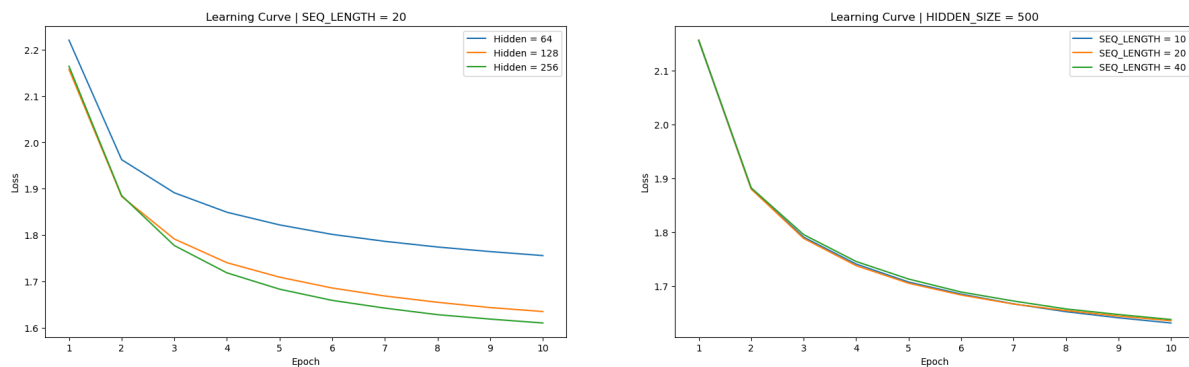KING RICHARD II:
I have to seave and that be kould be our had and the blood of to the king.

Generated text after epoch 10:
myself.
KING LEWIS mist with her mast the courtont to pering of the belles are and be is pale.
RICHARD:

3) __Comparison of different sizes of Hidden States and Sequence Length__

Compared hidden state 64, 128 and 256 with sequence length 20. From the plot, we can see that larger hidden state size leads to error rate. With smaller hidden size models won't be able to capture important information and details. Larger hidden state lets the model capture and process more sophisticated dependencies for sentence style, syntax, punctuation.

From different sequence length plot (compared sequence length 10, 20, 40) we can see that the same as for LSTM, longer sequence length allows model to capture more text, although it takes more time for the model to train, but for different sequence length, the difference is small compared to different hidden size.

__(4.1) LSTM Network Architecture:__

```
Model: "sequential"
_____
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 128)               99328

 dense (Dense)               (None, 65)                8385

 activation (Activation)     (None, 65)                0


=================================================================
Total params: 107,713
Trainable params: 107,713
Non-trainable params: 0
```

The first model is LSTM, which captures sequential patterns in data, simple RNN have problems with "vanishing gradient", which is hard for simple RNN to capture long-term dependencies like stylistic features, grammatical structure, which makes LSTM best choice for text generation tasks.

Its input shape - 128 neurons (hidden size), dense layer which is fully connected, it transforms 128 neurons from previous layer into 65 neurons outputs, and activation layer which converts input values into output values 0 to 1. The output is 65 for 65 classes that represent vocabulary size (letters, punctuation, whitespaces used in Shakespear's text).

I narrowed down the data size to range [100000:800000] resulting in train_text size: 700000 and valid_text size: 122025.

Learning Curve and training/validation error rate
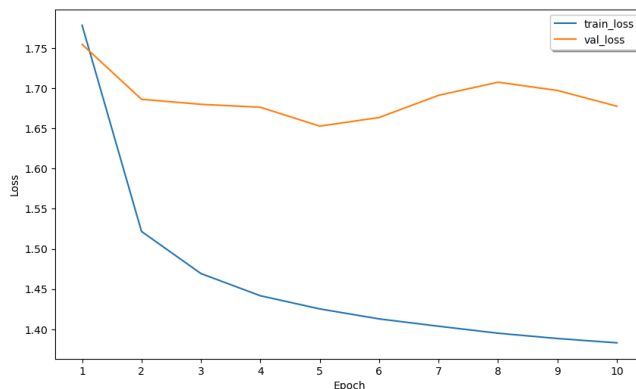

```
Training error rate =
1.383192777633667
Validation error rate =
1.6775511503219604
```

(4.2) Text generated at 5 breakpoints during training

I chose epochs 1, 3, 5, 7 and 10 to generate text, temperature is set to 0.5, higher temperature adds more randomness to words, lower temperature focuses on higher probability, chooses probable options over random. As we can see, there's big progress comparing text from 1st and final 10th epoch, the model mimics Shakespear's style, play format with different characters. In the first 2 epochs we can see many more random words that have no sense. In the last 10th epoch style is better, some sentences can be understood but still logical structure needs to be improved, by larger numbers of epochs and training on a bigger set of data.

Generated text after epoch 1:
nd he squeak'd out and country my dears for the sea to for such distroke.
CAMILLO:
But we warks of monthing and be to better of mine
And make fear, to he do the fall of thine.
Generated text after epoch 3:
ou do not, make the sword and for the confent
When he be remember she while their soul of exceltion;
There in the false fear!
First Murderer:
Generated text after epoch 5:
s:
As a long-parted be so man.
Second Murderer:
What save I with him and some man's hearts of me.
Generated text after epoch 7:
tutor me
To this subject beling woes the propers to the gross
With her will accorthy of thy chank
Were wome and here that shall be not see him.
Generated text after epoch 10:
eaming the ensign of discontemn than her from this tongue should not to stay some worst
with her heart to a trial,

And the trumpless for the father counin the sorrow land
To some tongue was the truth, and my hand, and some curse unto the wind to his same together mad how more a dead prosperows and my soul, this is my father and best bear and makes this sorrow shall be the counterthire bury some earth's body. The son of the sear of my brother way, with the seas than the stars,
And sent to the poor age.

(4.3) Comparison of different sizes of Hidden States and Sequence Length
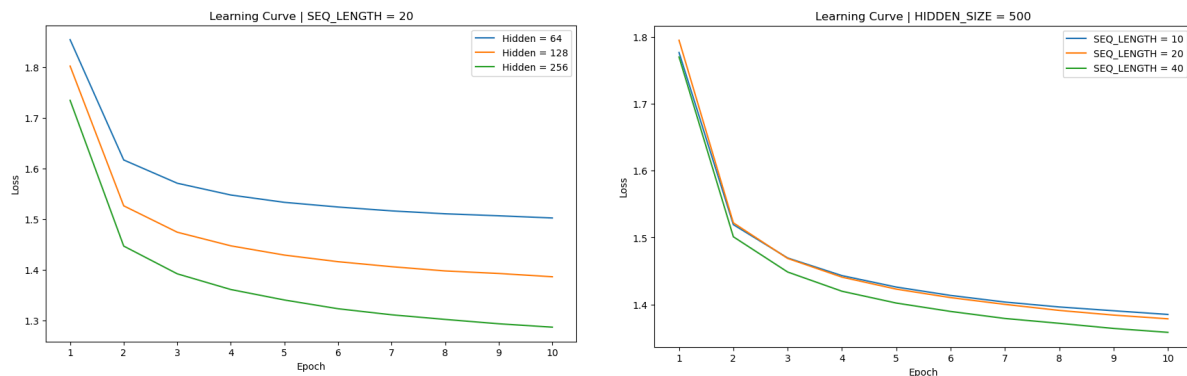
I compared hidden state 64, 128 and 256 (sequence length is 20). From the learning curve we can see that a larger hidden state=256 is better for LSTM as this way it captures more effectively vocabulary, syntax and sentence structure.

For sequence length I compared 10, 20 and 40 (hidden size is 500). Longer sequence is important for creating dependencies across sentences in text, and as we can see from learning curve loss is smaller for longer sequence length.

Due to limited hardware, I chose a smaller number for both, but it is recommended to choose a longer sequence length and higher hidden size number for the model to more effectively capture nuances of training text.

5) LSTM used to generate 10 lines of text

Generated text (primed with 'JULIET'):
G HENRY VI

KING EDWARD IV:
What, shall stand the country in the man is here heard the field
As that with the sad counter'd to the seat to the ground.

Second Murderer:
What shall show his world, and show the man is the sea to the body,
The true than the death, the seat and the day is the consent to the heart
That he sease to a bad from the consent of the seat to the sears,