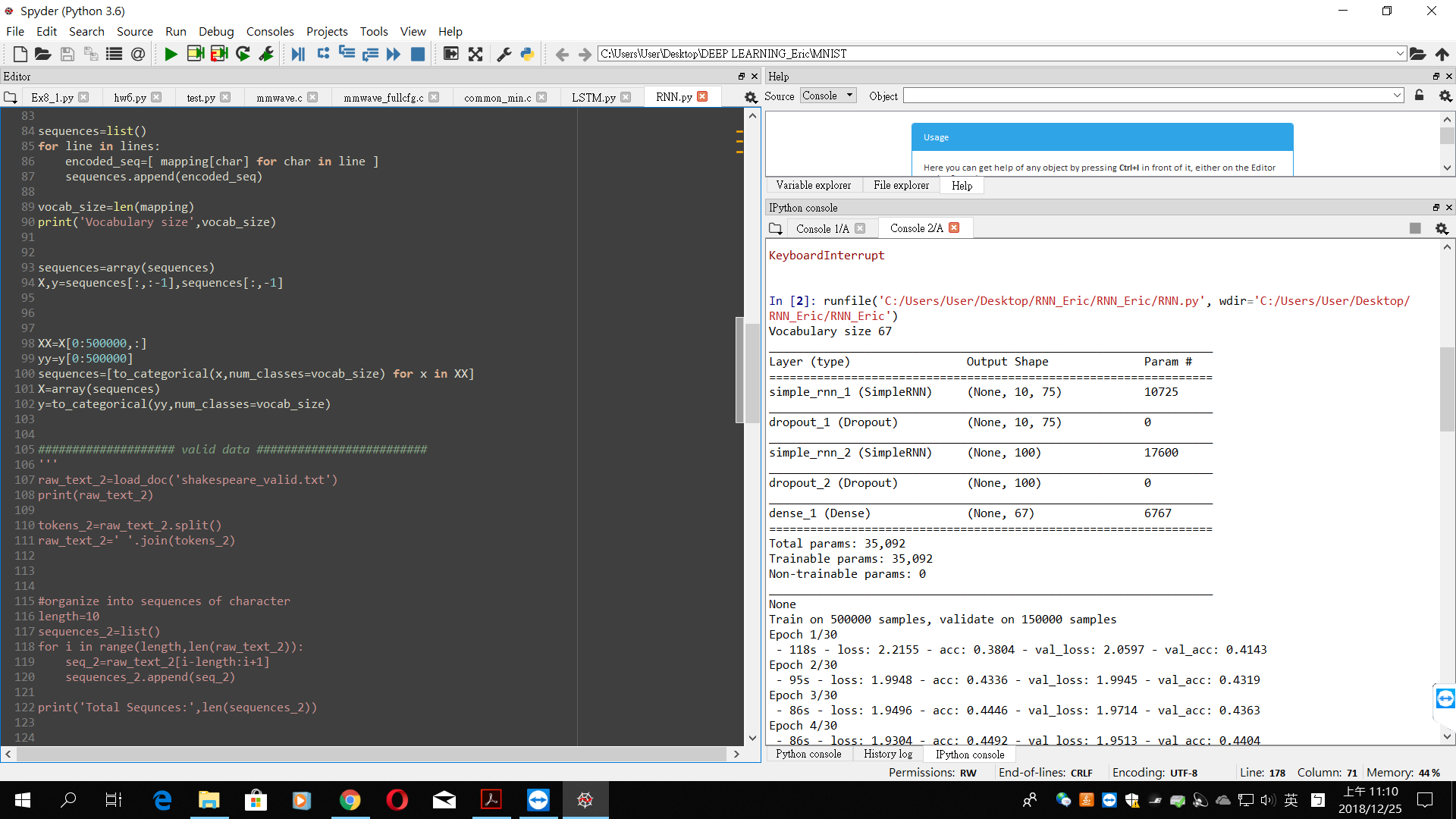
Deep Learning HW3

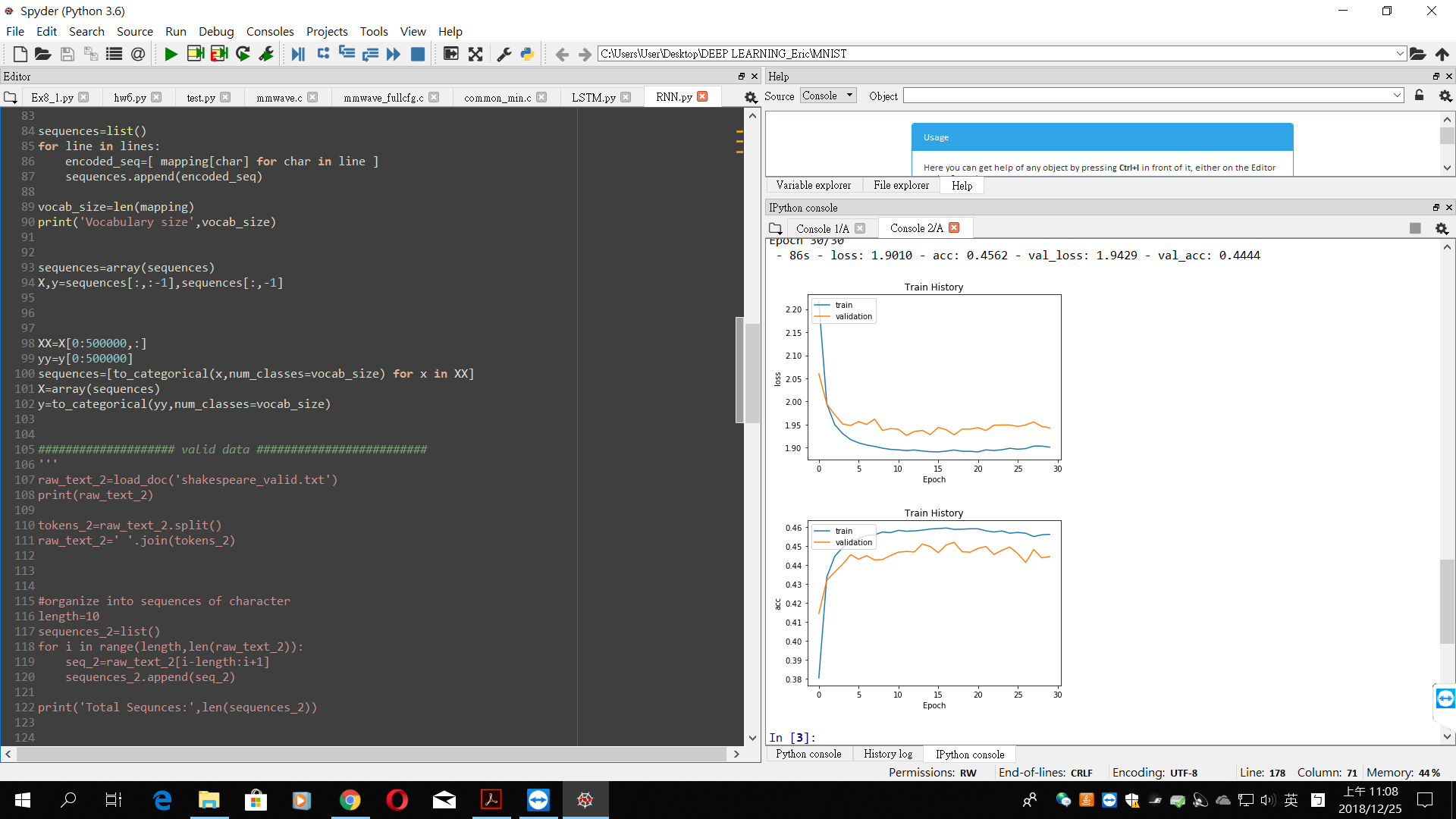
A061714 藍國瑞

1. RNN architecture:
2. The standard RNN:
   1. Network architecture:
      * Mini-batch size=32
      * Sequence length=10
      * Optimizer: Adam
      * Hidden layers and hidden states:
        1. RNN with 75 hidden states
        2. Dropout=0.2
        3. RNN with 100 hidden states
        4. Dropout=0.2
        5. Dense with 67 neurons

` 

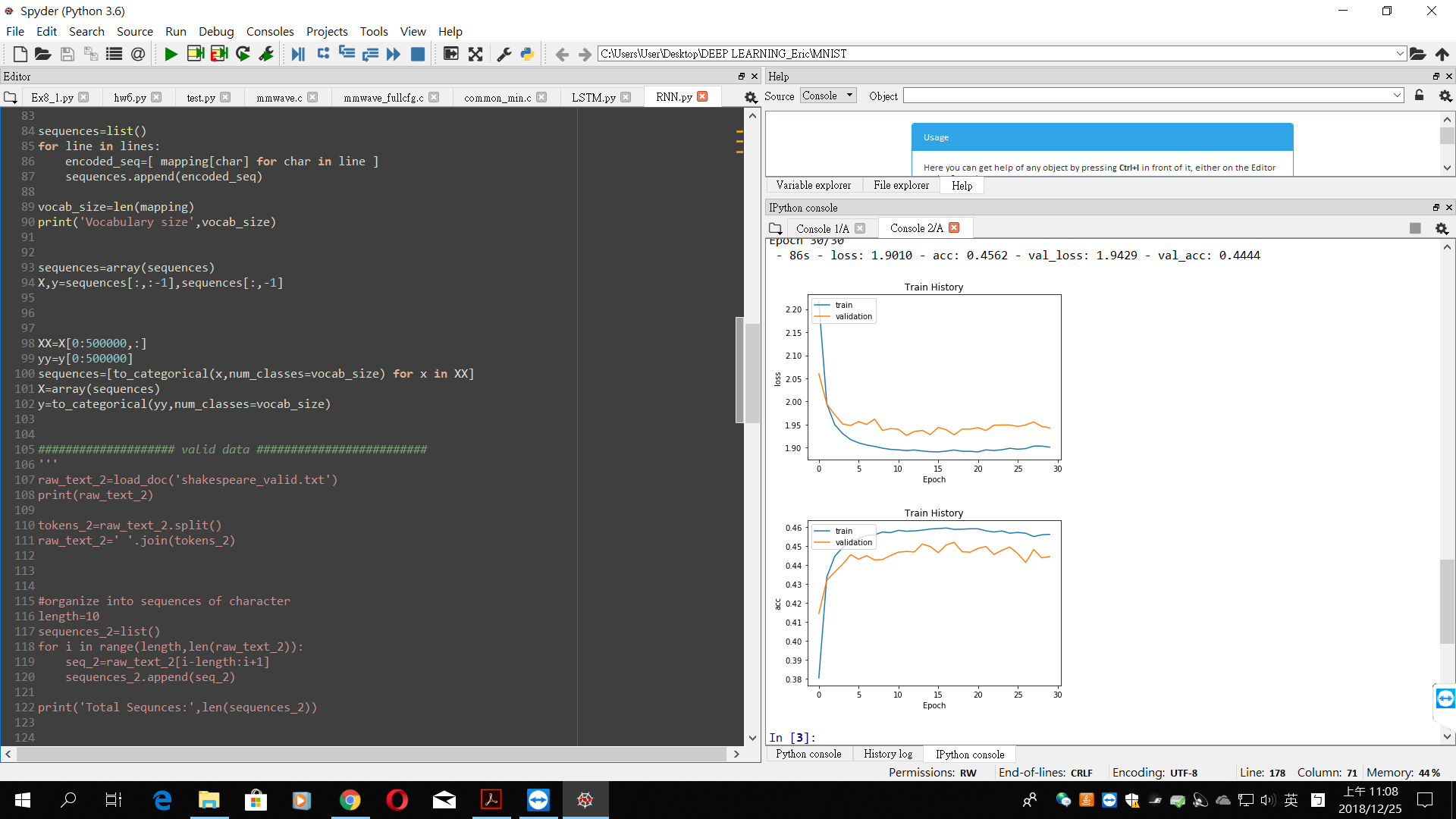
* 1. Learning curve:

(loss V.S. epoch):



* 1. Training error rate and valid error rate

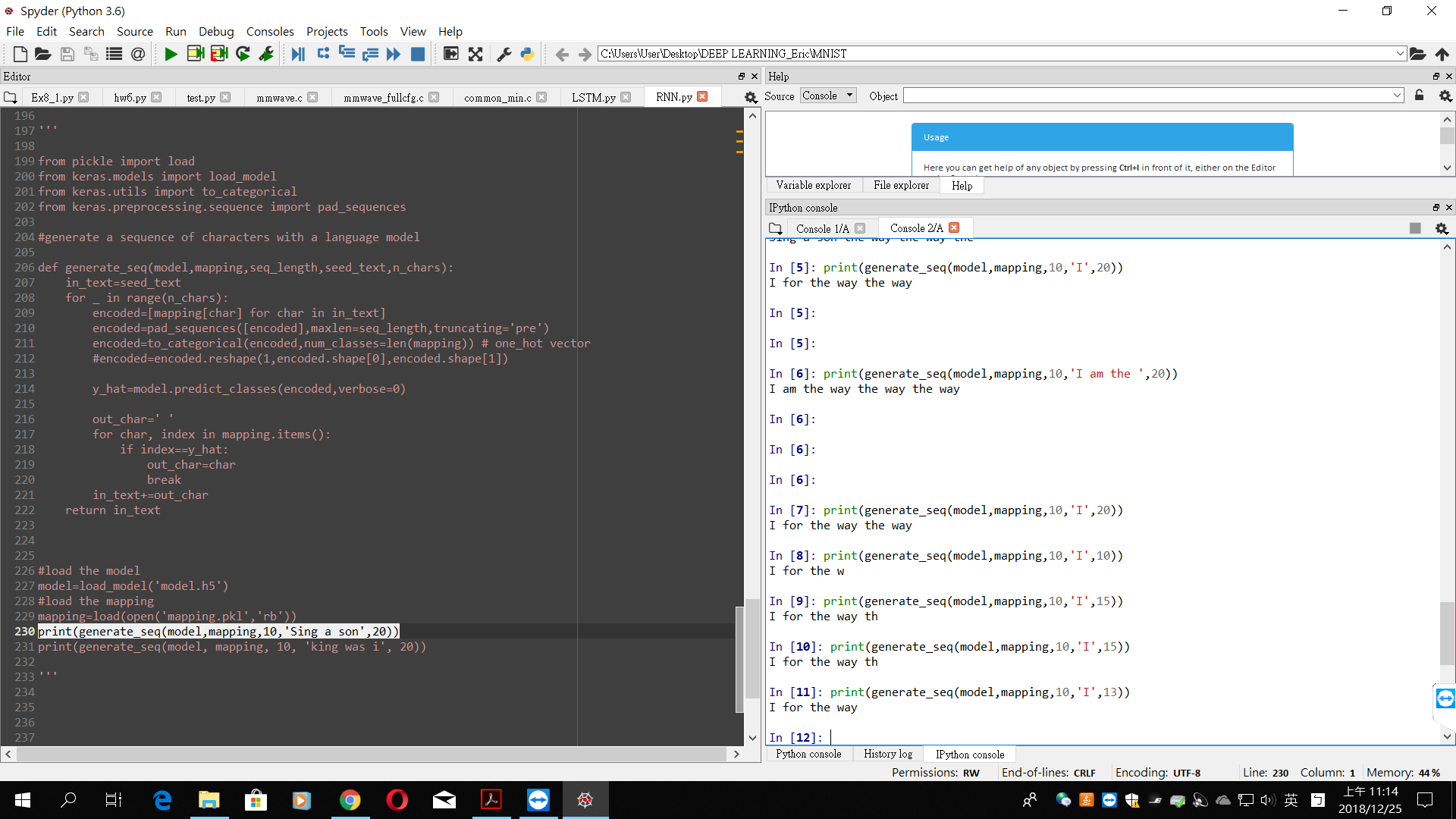
(accuracy V.S. epoch):



1. Choose 5 breakpoints during training process. Feed some part of training text into RNN and show the text output:

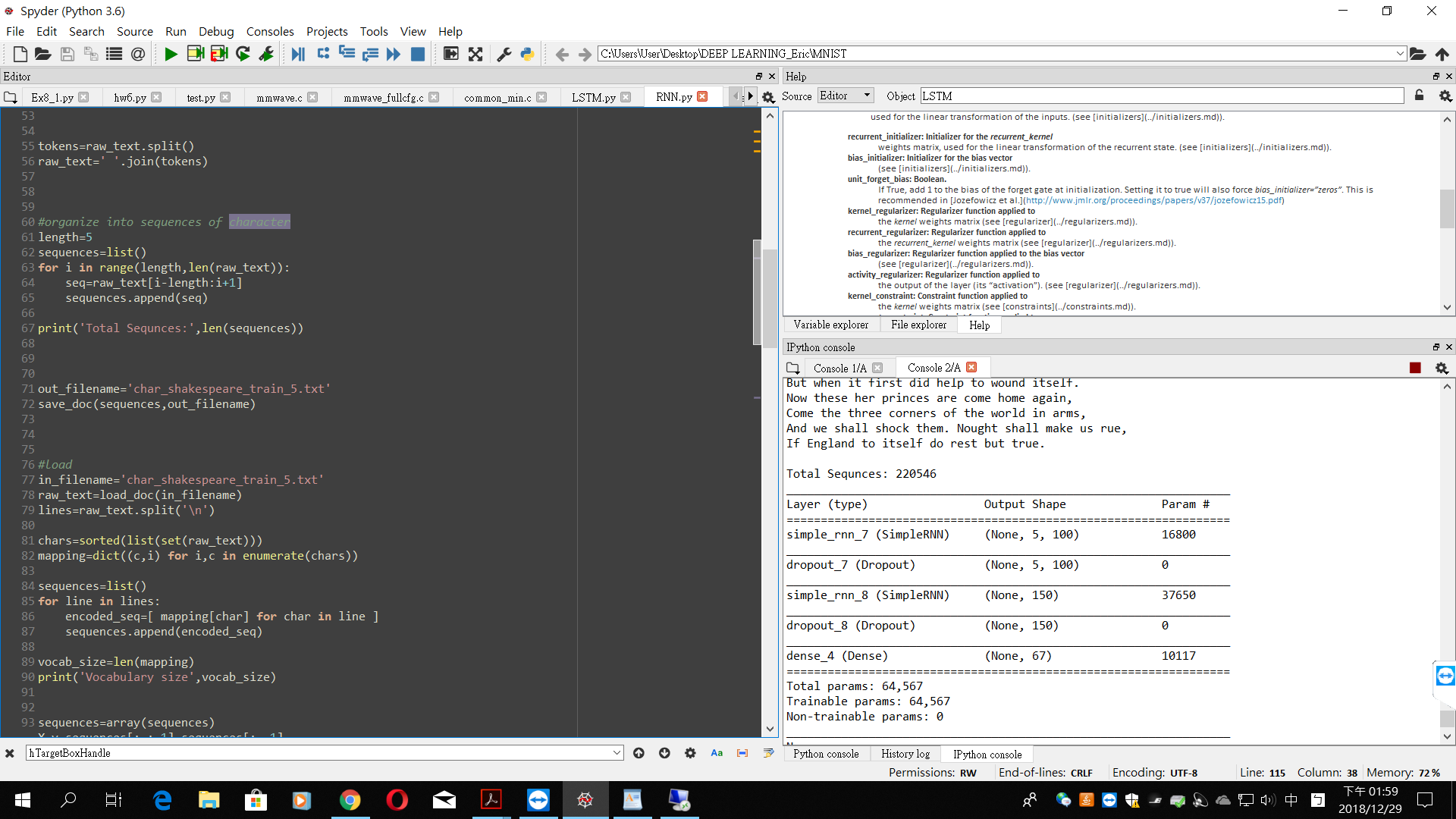
Through the above loss\_curve and acc\_curve, we can observe when the epoch is 5 no matter the loss or the accuracy are gradually become saturated.

If I input the text “ I ”that was trained before, then the RNN model will return the text” **I for the way** ”.It’s likely the incomplete sentence.

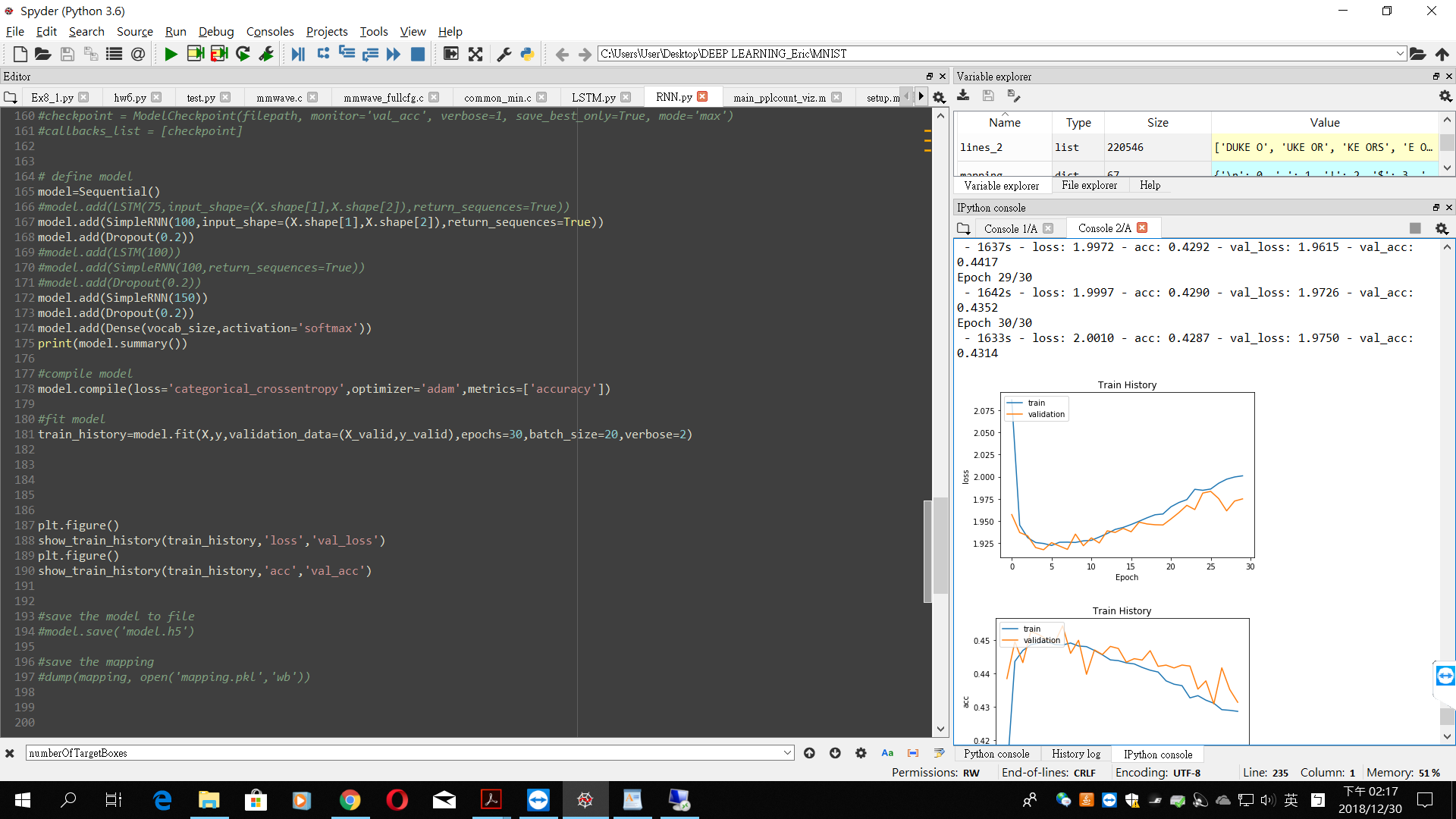


1. Compare the results of choosing different size of hidden states and sequence length by plotting the training curve V.S. different parameters:

* Network architecture:
  + - Sequence length=5
    - Hidden layers and hidden states:
      1. RNN with 100 hidden states
      2. Dropout=0.2
      3. RNN with 150 hidden states
      4. Dropout=0.2
      5. Dense with 67 neurons

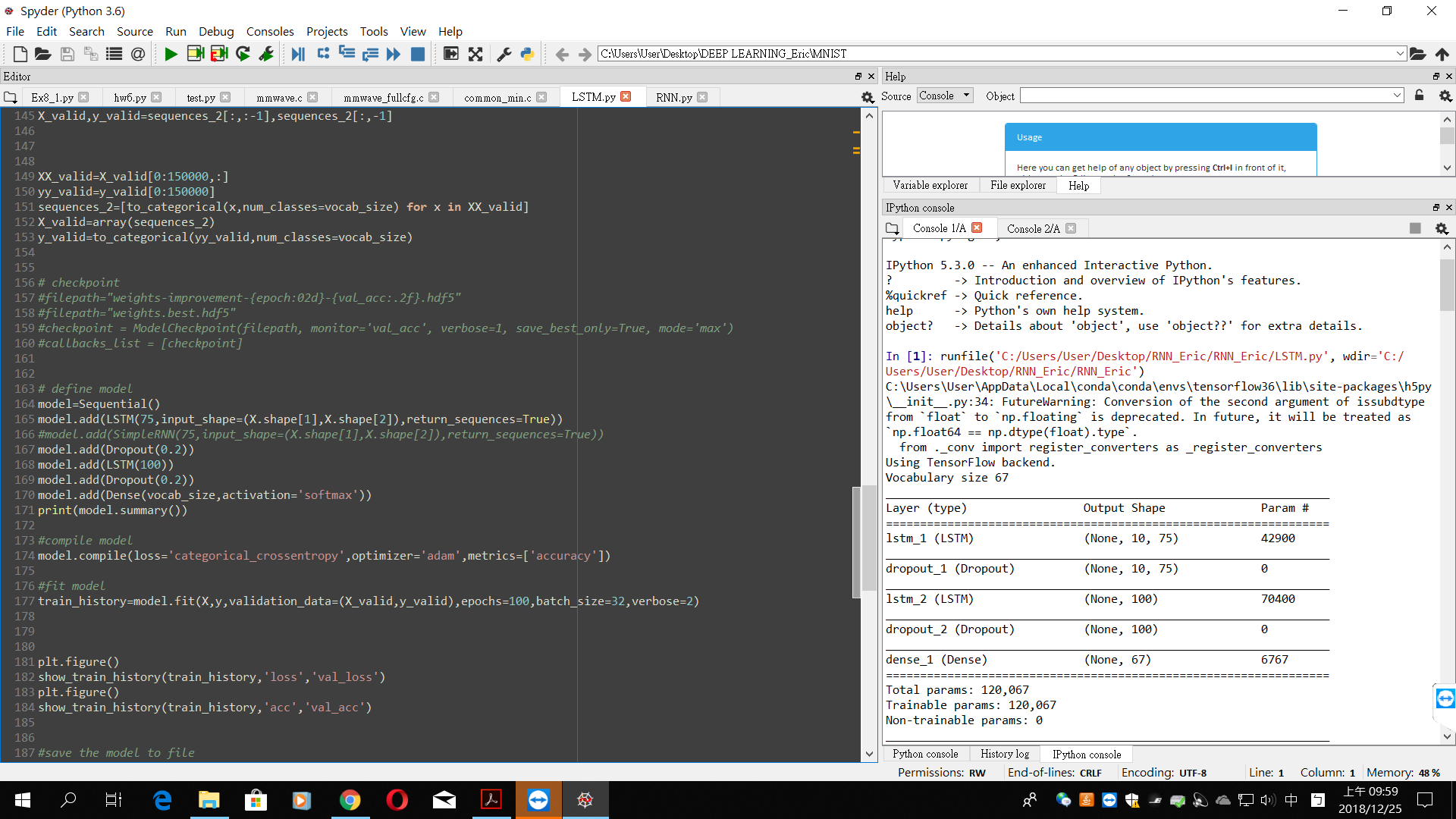


* Learning curve



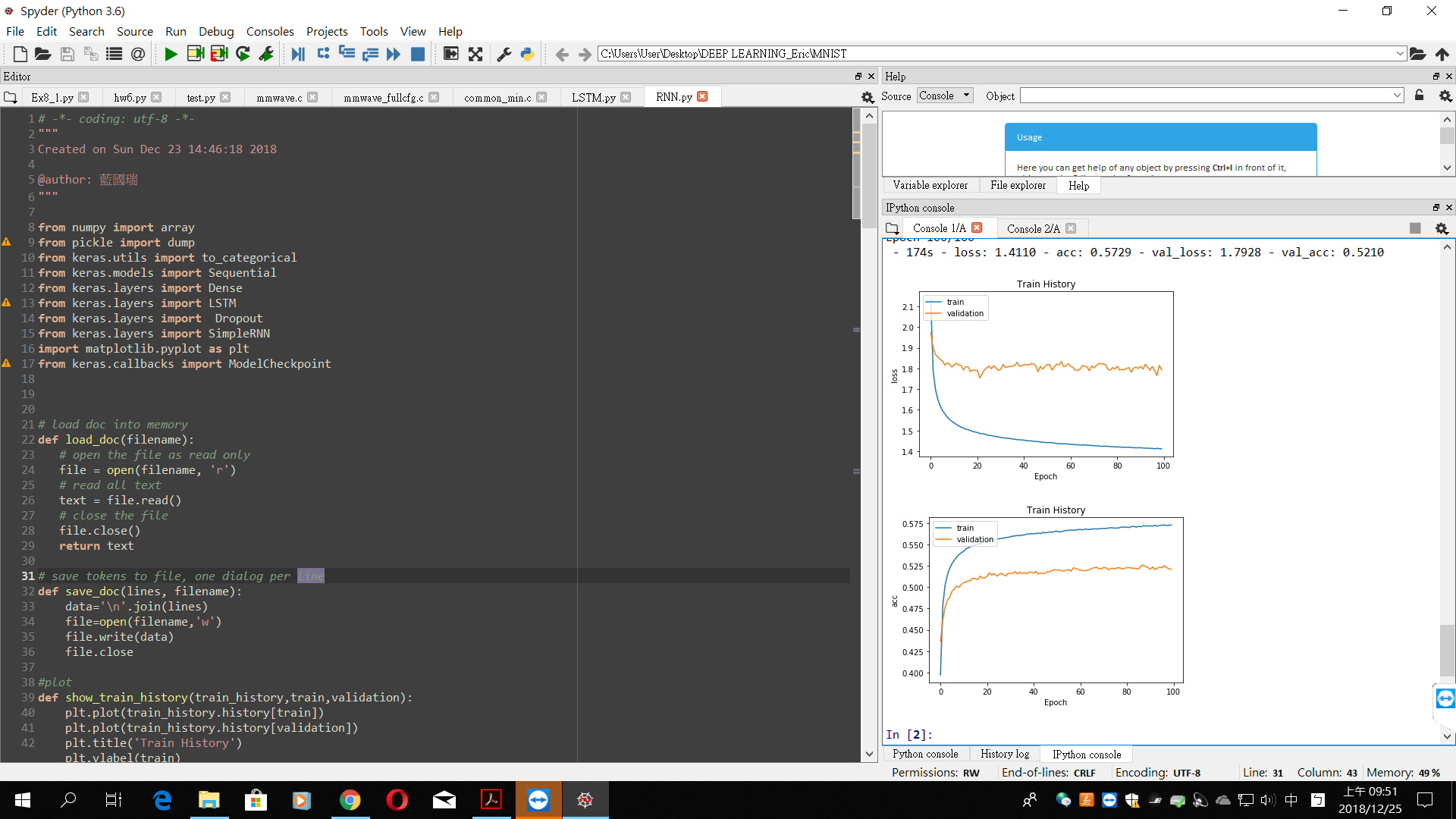
1. LSTM architecture:
2. The standard LSTM:
   1. Network architecture:

* Mini-batch size=32
* Sequence length=10
* Optimizer: Adam
* Hidden layers and hidden states
  + - 1. LSTM with 75 hidden states
      2. Dropout=0.2
      3. LSTM with 100 hidden states
      4. Dropout=0.2
      5. Dense with 67 neurons



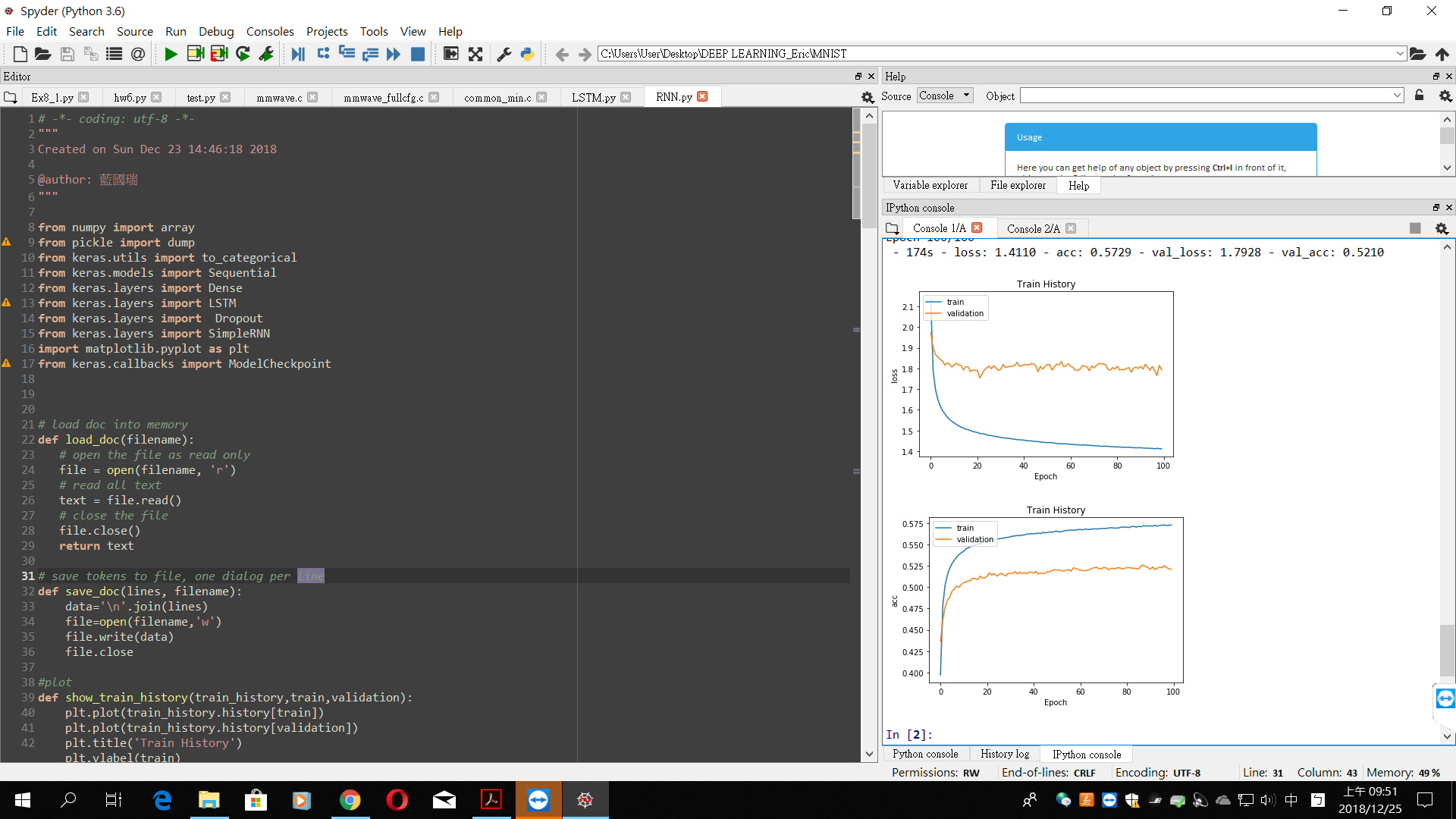
* 1. Learning curve:

(loss V.S. epoch)



* 1. Training error rate and valid error rate:

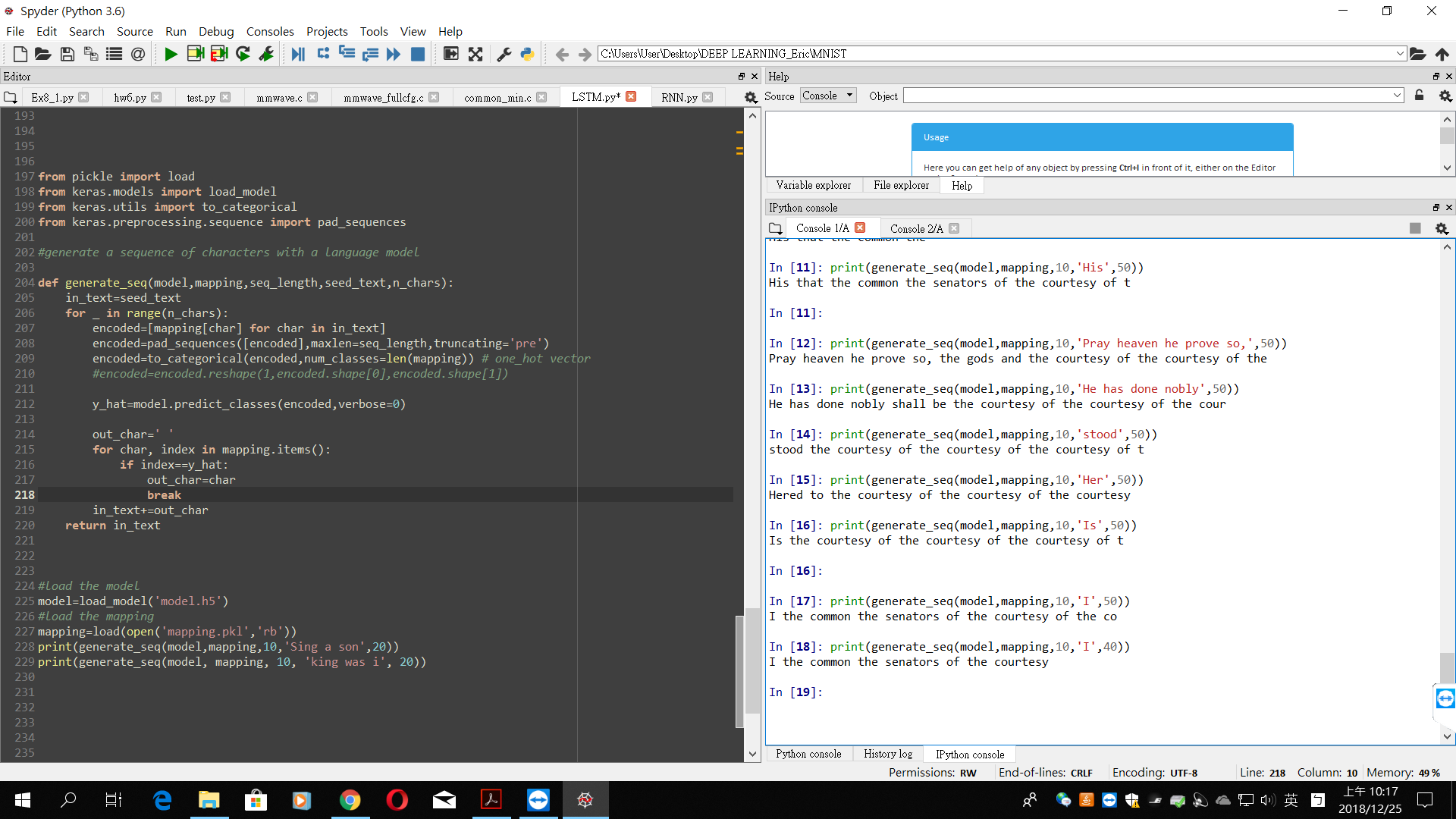
(accuracy V.S. epcoh)



1. Choose 5 breakpoints during training process. Feed some part of training text into LSTM and show the text output:

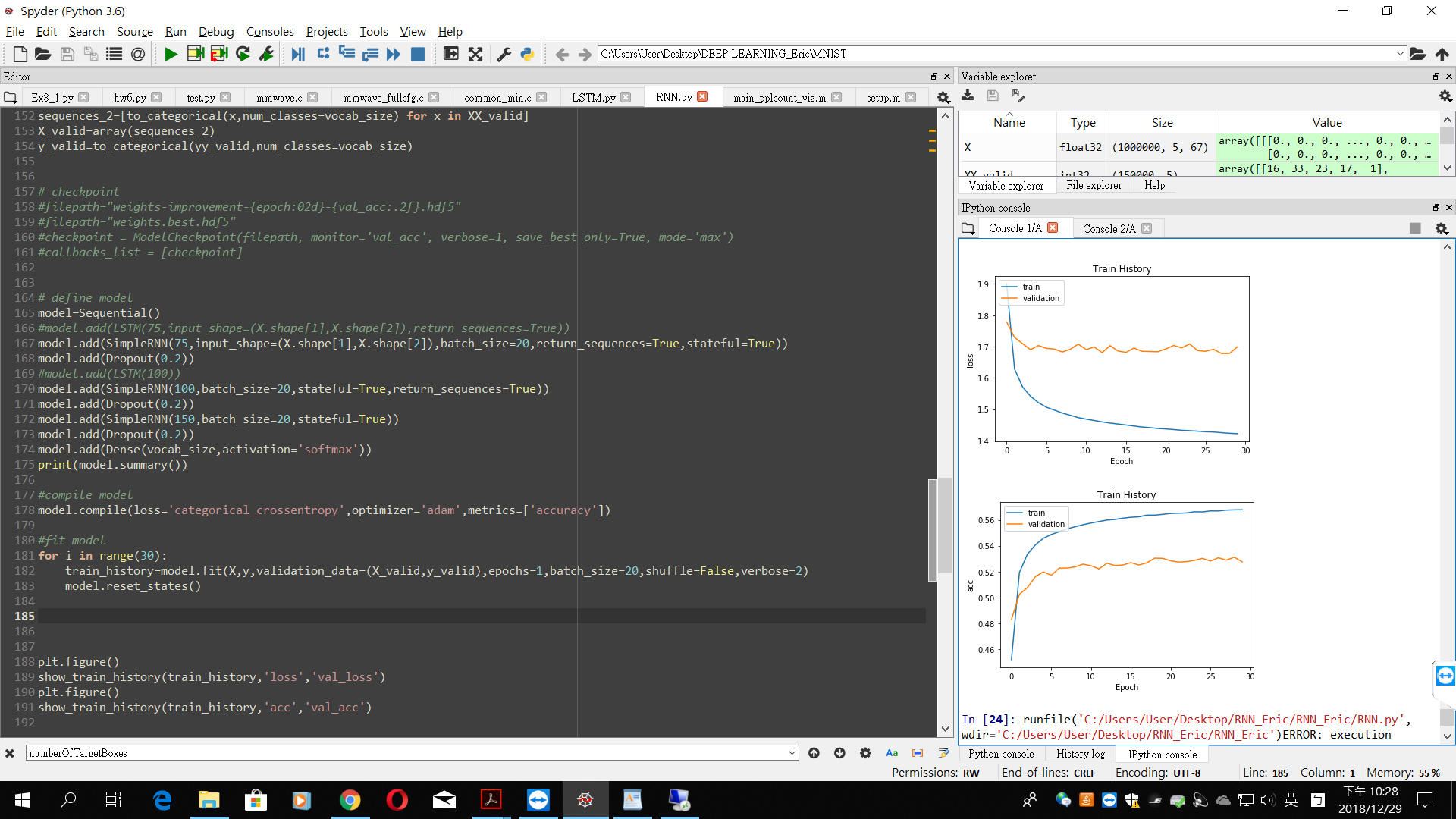
Through the above loss\_curve and acc\_curve, we can observe the validation data set becomes saturated when the epoch is 20. Hence in the next training process, I would like to set the epoch to be 20 instead of 100.

If I input the text “ I ”that was trained before, then the LSTM model will return the text” **I the common the senators of the courtesy**”. Although the grammar is a little wrong, but we still can understand what it is talking about. I think the performance is better than the RNN model in PART A.



1. Compare the results of choosing different size of hidden states and sequence length by plotting the training curve V.S. different parameters:

* Network architecture
* Sequence length=5
* Hidden layers and hidden states
  + - 1. LSTM with 100 hidden states
      2. Dropout=0.2
      3. LSTM with 150 hidden states
      4. Dropout=0.2
      5. Dense with 67 neurons
* Learning curve:



1. Use LSTM to generate some words by priming the model with a word related to your dataset.

I use the LSTM with following architecture to generate the word:

* Network architecture
* Sequence length=5
* Hidden layers and hidden states

1. LSTM with 100 hidden states
2. Dropout=0.2
3. LSTM with 150 hidden states
4. Dropout=0.2
5. Dense with 67 neurons

When I input the text ”Before” into the model, the model will output the text “ Before the state”.

