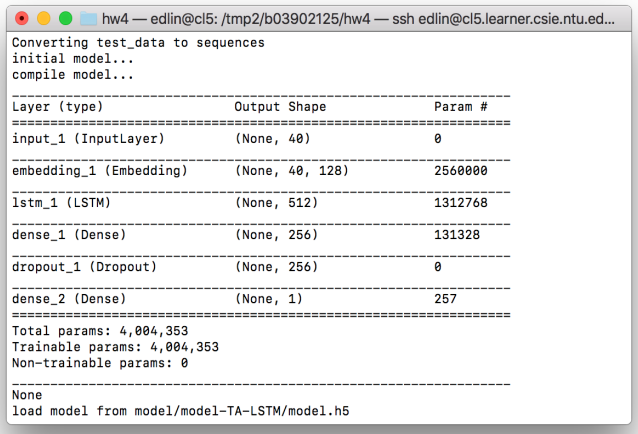
學號：B03902125 系級： 資工四　姓名：林映廷

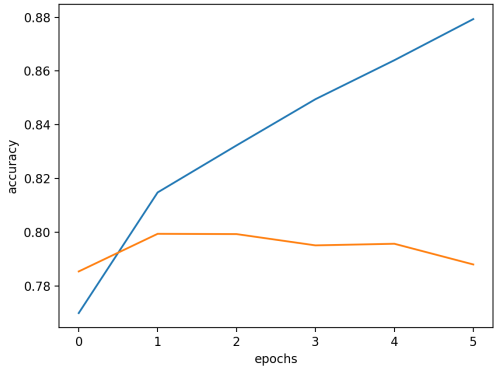
1. (1%) 請說明你實作的 RNN model, 其模型架構、訓練過程和準確率為何？(Collaborators:)

答：

模型架構：



訓練過程：



訓練細節:epochs=6, optimizer=Adam, loss function=binary_crossentropy, threshold=0.5

準確率：(public+private)/2=(0.80039+0.80004)/2=0.800215

2. (1%) 請說明你實作的 BOW model, 其模型架構、訓練過程和準確率為何？

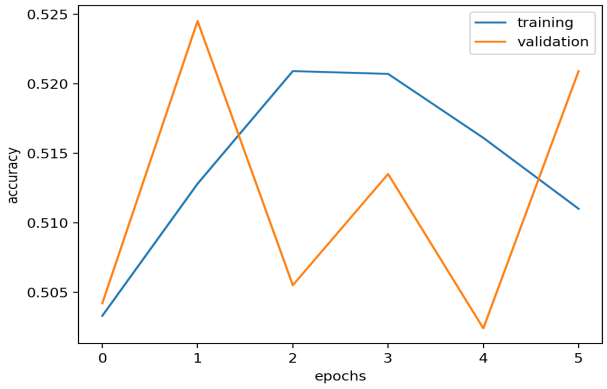
(Collaborators:)

答：

模型架構:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 40)                0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               10496
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 257
=================================================================
Total params: 10,753
Trainable params: 10,753
Non-trainable params: 0
_________________________________________________________________
```

訓練過程:



訓練細節:epochs=6,optimizer=Adam, loss function=binary_crossentropy, threshold=0.2

準確率: (public+private)/2=(0.52024+0.51874)/2=0.51949

3. (1%) 請比較 bag of word 與 RNN 兩種不同 model 對於"today is a good day, but it is hot"與"today is hot, but it is a good day"這兩句的情緒分數，並討論造成差異的原因。

(Collaborators:)

答：

BOW:

```
test_label:
[[  1.61510602e-01]
 [  2.38816696e-21]]
```

(PS.上方為第一句的分數，下方為第二句的分數)

RNN:

```
test_label:
[[ 0.49374691]
 [ 0.9628399 ]]
```

(PS.上方為第一句的分數，下方為第二句的分數)

BOW 可能沒有捕捉到句意，所以兩者分數均偏低;但 RNN 效果比較好，很明顯兩個句子的分數有落差。

4. (1%) 請比較"有無"包含標點符號兩種不同 tokenize 的方式，並討論兩者對準確率的影響。
   (Collaborators:)
   答：

不包含標點符號：(public+private)/2=(0.80039+0.80004)/2=0.800215

有包含標點符號：(public+private)/2=(0.76187+0.75929)/2=0.76058

可見有包含標點符號準確率較低，可能是因為有包含標點符號會使 RNN 的判斷變差。

5. (1%) 請描述在你的 semi-supervised 方法是如何標記 label，並比較有無 semi-surpervised training 對準確率的影響。
   (Collaborators:)
   答：

取 training_nolabel.txt 前 10%和後 10%當 semi-supervised 的 data。如果預測的值>0.5，則判斷改筆 data 的 label 為 1;否則為 0。

無 semi-supervised training：(public+private)/2=(0.80039+0.80004)/2=0.800215

有 semi-supervised training: (public+private)/2=(0.76663+0.76286)/2=0.764745

取共約 20%的 training_nolabel.txt 當 semi-supervised 的 data，會讓準確率稍微下降，比無 semi-supervised training 的準確率還差。可能是 semi-supervised training 的 data 有些判斷錯誤，造成 model 的 performance 變差。