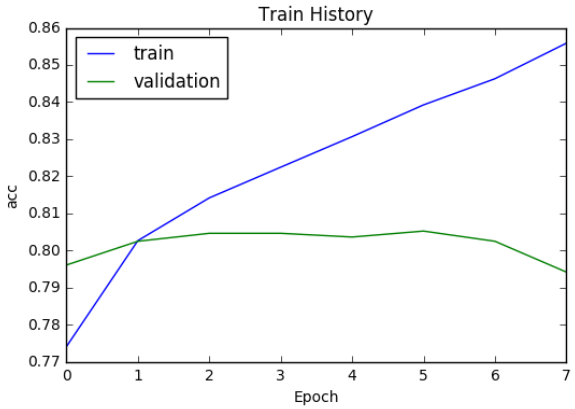
學號：R06521608 系級： 土木碩一 姓名：陳德元

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: 無）

   答：loss='binary_crossentropy', optimizer='adam', metrics=['accuracy']
   batch_size=32, epochs=10

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 30, 100)	409600
dropout_1 (Dropout)	(None, 30, 100)	0
bidirectional_1 (Bidirection	(None, 512)	731136
dense_1 (Dense)	(None, 64)	32832
dropout_2 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 1)	65

Total params: 1,173,633
Trainable params: 1,173,633
Non-trainable params: 0

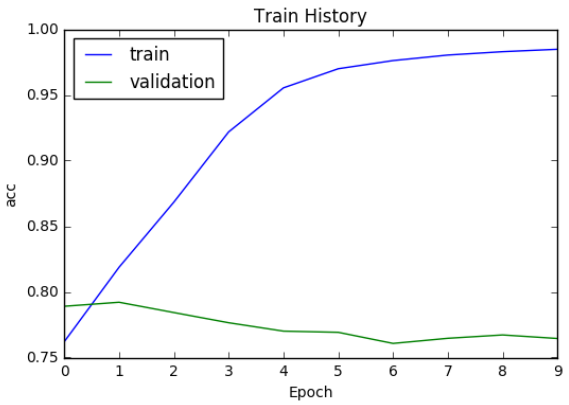


2. (1%) 請說明你實作的 BOW model，其模型架構、訓練過程和準確率為何？（Collaborators: 無）

   答：loss='binary_crossentropy', optimizer='adam', metrics=['accuracy']
   batch_size=1024, epochs=10

Layer (type)	Output Shape	Param #
dense_5 (Dense)	(None, 256)	1280256
dense_6 (Dense)	(None, 256)	65792
dropout_3 (Dropout)	(None, 256)	0
dense_7 (Dense)	(None, 256)	65792
dropout_4 (Dropout)	(None, 256)	0
dense_8 (Dense)	(None, 1)	257

Total params: 1,412,097
Trainable params: 1,412,097
Non-trainable params: 0



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。

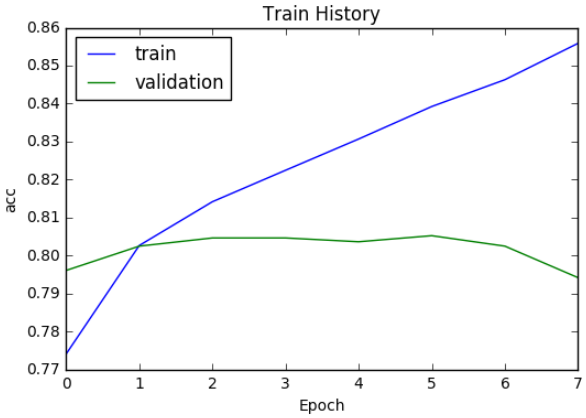
   （Collaborators: 無）

   答：BOW : [0.99713802][0.99713802]
   RNN : [0.25195399][0.89776081]
   由於 RNN 會去考慮句子的前後關聯，所以他比較能判斷出來句子語氣的前後差異，但 BOW 只有出現的頻率，並無關聯，所以有這樣的落差。

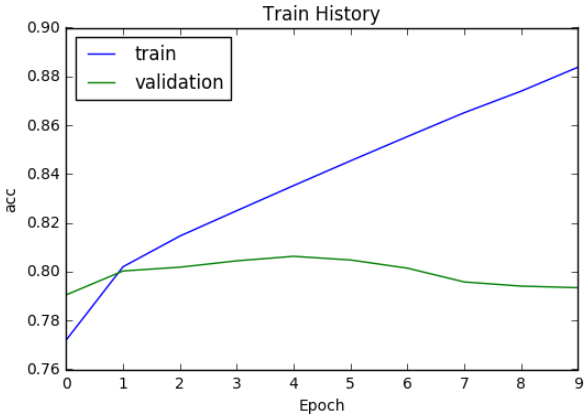
4. (1%) 請比較"有無"包含標點符號兩種不同 tokenize 的方式，並討論兩者對準確率的影響。

（Collaborators: 無）

答：本來以為有標點會讓準確率下降，但沒想到其實兩者的準確率差異並不大



無標點



有標點

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: 無）

答：用單一的 model 去標記 semi-supervised，然後 threshold 大約設定在 0.8，其對於 kaggle 準確率大約可提升 0.34%左右(80.81% - 80.47%)