

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: 王垣尹)

答：

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 1024)	4722688
dropout_1 (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 1)	1025
Total params: 4,723,713		
Trainable params: 4,723,713		
Non-trainable params: 0		

總共 train 10 個 epochs，training set 的 accuracy 從 63% 快速上升到 83%，testing set 大約在 78% 上升到 81~82%，跟 kaggle 上的 81.8 差不多

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: 王垣尹)

答：

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 1000)	1001000
batch_normalization_1 (Batch Normalization)	(None, 1000)	4000
dropout_1 (Dropout)	(None, 1000)	0
dense_2 (Dense)	(None, 50)	50050
batch_normalization_2 (Batch Normalization)	(None, 50)	200
dropout_2 (Dropout)	(None, 50)	0
Total params: 1,055,250		
Trainable params: 1,053,150		
Non-trainable params: 2,100		

總共 train 100 個 epochs，training set 的 accuracy 從 73% 快速上升到 90% 之後慢慢爬到 95%，不過 validation set 的 accuracy 則是都在 75~76% 左右

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：

	RNN	BOW
today is a good day, but it is hot	0.14401191	0.98612469
today is hot, but it is a good day	0.97328329	0.98612469

用 BOW 做的話，以上兩句話輸入的 vector 會是一樣的東西，所以得到的 output 會一樣，但是用 RNN 參考前後文來判斷整句每個字的結果就明顯準確很多，特別是像這種有轉折語氣的句子

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：

有標點符號的可以做到接近 82% 準確率，沒標點符號只有到 80%，推測是因為前後句分隔時，沒有標點符號的在語意上容易搞混

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

(Collaborators:)

答：

一開始試把 threshold 設 0.8， $\text{prediction} > 0.8$ 或 $\text{prediction} < 0.2$ 的才會被使用，不過做出來跟沒有做 semi-supervised training 差不多，試著把 threshold 設 0.9、0.6，效果一樣不大，不過 threshold 比較小的時候 model 在 train 時，容易 train 壞，做出來的準確率反而很低，因此最後還是沒有用 semi-supervised training