

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？

Ans:

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 32, 256)	5120256
gru_1 (GRU)	(None, 32, 128)	147840
gru_2 (GRU)	(None, 64)	37056
dense_1 (Dense)	(None, 32)	2080
leaky_re_lu_1 (LeakyReLU)	(None, 32)	0
dense_2 (Dense)	(None, 1)	33
Total params: 5,307,265		
Trainable params: 5,307,265		
Non-trainable params: 0		

執行 20 個 epochs，執行到第 12 個 epochs 時停止。

	Public Score	Private Score
準確率為:	0.81958	0.81641

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？

Ans:

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 32)	128032
dense_2 (Dense)	(None, 16)	528
dense_3 (Dense)	(None, 1)	17
Total params: 128,577		
Trainable params: 128,577		
Non-trainable params: 0		

執行 20 個 epochs，執行到第 6 個 epochs 時停止。

將 training data 取 0.1% 當作 validation data，在 validation 準確率可達 0.7839。

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。

Ans:

bag of word	prediction
today is a good day, but it is hot	0.58007282
today is hot, but it is a good day	0.58991665

RNN	prediction
today is a good day, but it is hot	0.81094235
today is hot, but it is a good day	0.7717073

Bag of word 不考慮單字的先後順序，所以測出的結果差不多。而 RNN 則有考慮順序，所以” today is hot, but it is a good day”有受到” today is hot”影響而分數較低。

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。

Ans:

	Public Score	Private Score
有標點符號	0.81958	0.81641
無標點符號	0.81157	0.81120

可以發現有加標點符號結果比較好，可能是有標點符號時，可以先斷句後在判斷。也可從標點符號判斷出情緒，例:驚嘆號。

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。

Ans:

先使用原本有 label 的 data 去訓練 model，再用訓練好的 model 去 predict non label data，並將結果 $>0.9$  及 $<0.1$  的資料合成一個新的 data。在用新的資料對 model 作 fine-tuning。

	Public Score	Private Score
semi-supervised	0.81687	0.81448
無 semi-supervised	0.81958	0.81641

發現使用 semi-supervised 的結果反而沒有比較好。

Reference:

1. <https://github.com/WaitingLin/ML2017FALL>
2. [https://ntumlta.github.io/2017fall-ml-hw4/RNN\\_model.html](https://ntumlta.github.io/2017fall-ml-hw4/RNN_model.html)