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：

我使用的 batch_size 是 128，epoch 是 20，val_ratio 是 0.1，使用 RNN 中的 LSTM，dropout_rate = 0.3，loss fuction = 'binary_crossentropy'，optimizer = adam。在處理 input data 的時候我有將標點符號捨去，使用的 vocabulary size 是 20000，有時用 padding，將每個字 padding 成 40 維。

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 40)	0
embedding_1 (Embedding)	(None, 40, 128)	2560000
lstm_1 (LSTM)	(None, 512)	1312768
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 1)	257

Total params: 4,004,353
Trainable params: 4,004,353
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：

我使用的 batch_size 是 128，epoch 是 20，val_ratio 是 0.1，dropout_rate = 0.3，loss fuction = 'binary_crossentropy'，optimizer = adam。在處理 input data 的時候我有將標點符號捨去，使用的 vocabulary size 是 20000。

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 20000)	0
dense_1 (Dense)	(None, 256)	5120256
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 1)	257

Total params: 5,120,513
Trainable params: 5,120,513
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：

	"today is a good day, but it is hot"	"today is hot, but it is a good day"
RNN	0.2328689	0.9759724
BOW	0.63213885	0.63213885

因為 BOW 並沒有考慮時序的問題所以一個句子的單字都一樣，但是順序改變不會改變他預測的結果；但是 RNN 會考慮時序問題，所以當 good 在 but 之後接 hot 就會讓 model 覺得沒有這麼正面，但是雖然前面有 hot 但是後面有 good 強調，model 就會判斷這句話是正面的。

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：

	有標點符號	無標點符號
準確率	0.80187	0.80358

我發現沒有標點符號會比保留標點符號好 0.02，可能標點符號在 RNN 的訓練沒有這麼有效果。

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：

	No semi-supervised	Semi-supervised
準確率	0.80358	0.80544

我的 semi_supervised 的 threshold 是 0.1，我有用一個 model(也就是第一題的 RNN model)。利用 semi_supervised 的準確率會比 no semi-supervised 的準確率高出了 0.002 左右，並沒有特別的改善，我覺得可能 predict 的結果沒有非常正確然後將不一定正確 label 的資料拿進去 train，所以它的準確率並不會改變太多。