學號：b04501127 系級： 土木三 姓名：凌于凱

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？(ref : https://blog.keras.io/using-pre-trained-word-embeddings-in-a-keras-model.html ，手把手)

   答：

   模型架構：

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 40)	0
embedding_1 (Embedding)	(None, 40, 150)	7500000
lstm_1 (LSTM)	(None, 40, 512)	1357824
lstm_2 (LSTM)	(None, 512)	2099200
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8256
dropout_3 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 1)	65

Total params: 11,129,569
Trainable params: 3,629,569
Non-trainable params: 7,500,000

   Word vector: Gensim.models.Word2Vec(sentences, size=150, window = 3, min_count=5, sg=1)

   訓練參數：epochs=20, batch_size=128, val_rate=0.3, loss='binary_crossentropy', optimizer='adam', callbacks=[EarlyStopping(monitor='val_acc', patience = 3, verbose=1, mode='max'), ModelCheckpoint(filepath, monitor='val_acc', verbose=1, save_best_only=True, mode='max', save_weights_only=False)]

   準確率：public: 0.82244, private: 0.82235

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？(ref : 手把手)

   答：

   模型架構：

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               65792
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               32896
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_3 (Dense)              (None, 64)                8256
_________________________________________________________________
dropout_3 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 65
=================================================================
Total params: 107,009
Trainable params: 107,009
Non-trainable params: 0
_________________________________________________________________
```

BOW: tokenizer.texts_to_matrix(sentences, mode='tfidf')

訓練參數：epochs=20, batch_size=128, val_rate=0.3,
loss='binary_crossentropy', optimizer='adam',
callbacks=[EarlyStopping(monitor='val_acc', patience = 3, verbose=1, mode='max'), ModelCheckpoint(filepath, monitor='val_acc', verbose=1, save_best_only=True, mode='max', save_weights_only=False)]

準確率：public: 0.71225, private: 0.71171 (沒加 semi)

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 : "today is a good day, but it is hot" score: 0.26447

"today is hot, but it is a good day" score: 0.94101

BOW : "today is a good day, but it is hot" score: 0.63536

"today is hot, but it is a good day" score: 0.63536

BOW 模型因為只跟句子裡面有什麼字有關，而與順序無關，所以兩個句子的 BOW 會相同，情緒分數也會相同。

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：public: 0.82844, private: 0.82692

無標點符號：public: 0.82244, private: 0.82235

有些標點符號可能代表一些情緒的修飾，像是！就是用在用在感嘆、命令、祈求、勤勉等語句之後，可能就會影響到最後的情緒分數

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：

如果 semi predict 出來的機率大於 0.8，標記成 1，小於 0.2，則標記成 0，並加到 train set 裡，在拿去 train，此過程執行 10 次。

Score: public: 0.81942, private: 0.81963 (無 semi)

public: 0.82244, private: 0.82235 (semi)