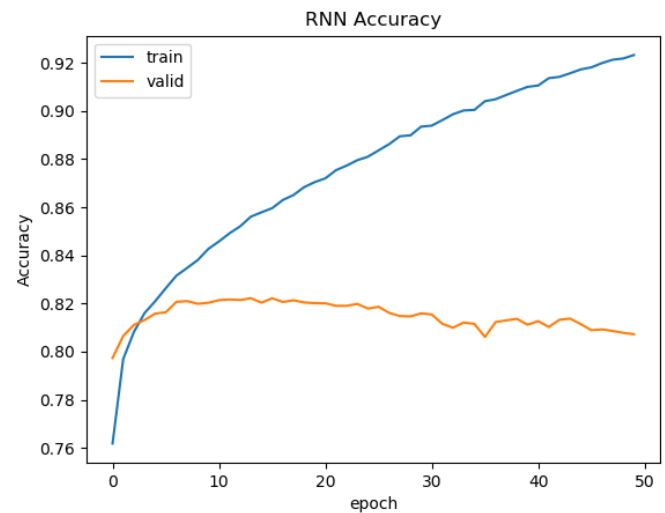
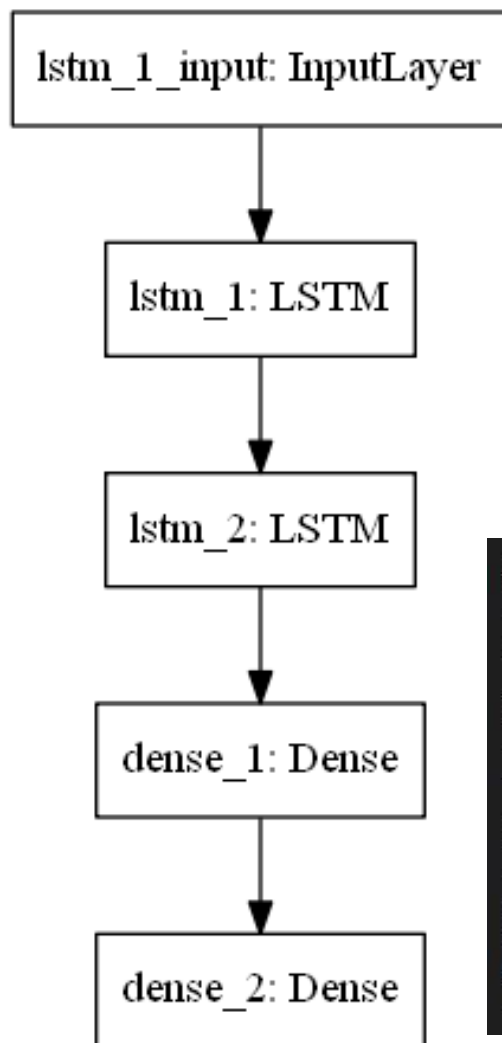


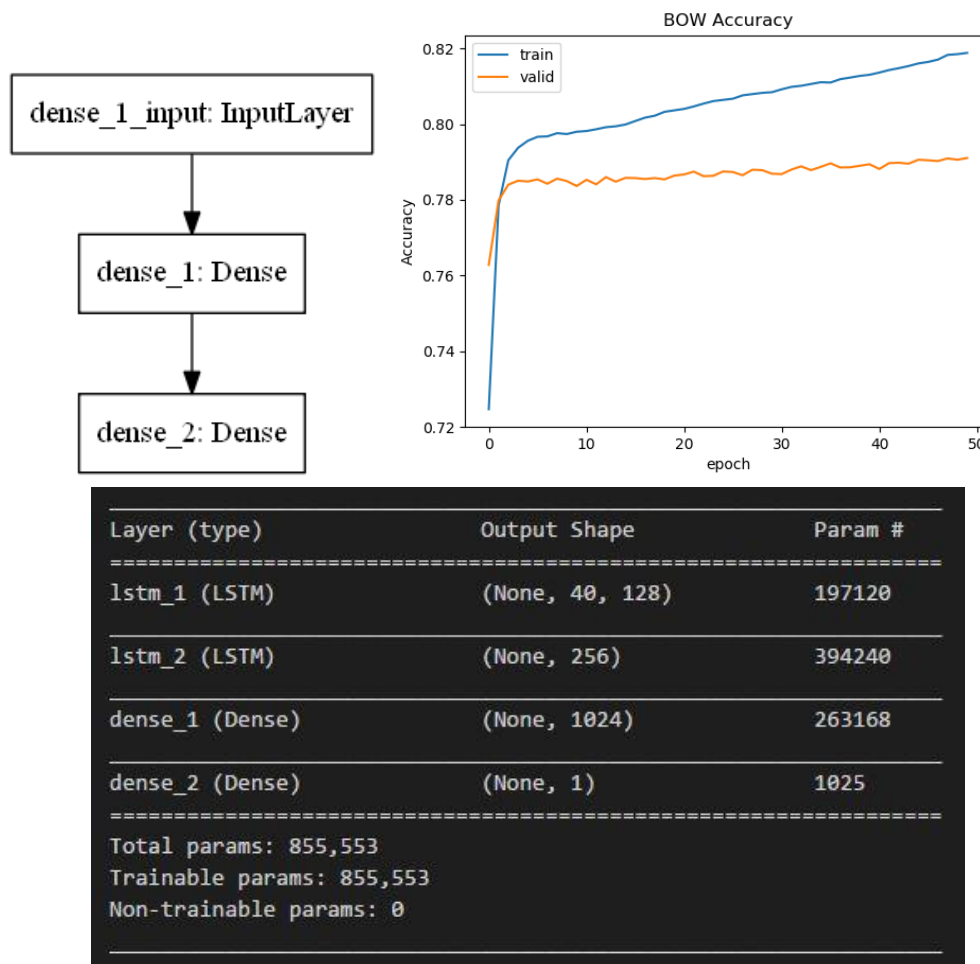
1. (1%) 請說明你實作的 RNN model，其模型架構、訓練過程和準確率為何？



Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 40, 128)	197120
lstm_2 (LSTM)	(None, 256)	394240
dense_1 (Dense)	(None, 1024)	263168
dense_2 (Dense)	(None, 1)	1025
Total params: 855,553		
Trainable params: 855,553		
Non-trainable params: 0		

輸入字串先用 gensim 的 word2vec 處理成 256 維的向量，再餵給 RNN。準確率最高約落在 0.82 左右

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？



字典共 5000 個字，沒出現在字典的字就不處理。準確率最高約落在 0.79 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。

	Today is a good day	Today is hot
RNN	0.78462005	0.97701967
BOW	0.63689137	0.63689137

兩者都將兩句話歸類在 class 1，應是受 "good" 大量影響，但是因為文字的順序不同，前面那句的重點比較像是放在 "hot"，所以 RNN 給予前句話較低的值，BOW 的預測值則不受文字順序的影響，所以給予的值是相同的。

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。

有標點符號: 0.8220 on validation

無標點符號: 0.8146 on validation

有標點符號明顯優於無標點符號，應是 ! ? ... 等標點符號會在某種程度上影響該語句的 class。

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。

沒有 semi-supervised: 0.7982 on validation

有 semi-supervised : 0.7960 on validation

受限於記憶體大小，所以我取的 no label 數量跟 labeled 一樣是二十萬筆，而 threshold 值則設在 0.7 和 0.3。

沒有 semi-supervised 的略優於有的，不過差異甚小，在這次的作業中可視為沒有幫助。