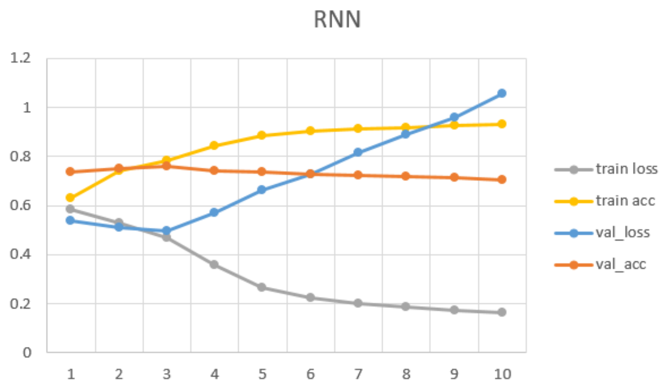
Homework 4 Report
資工碩一 R07922120 陳禹達

Problem 1.

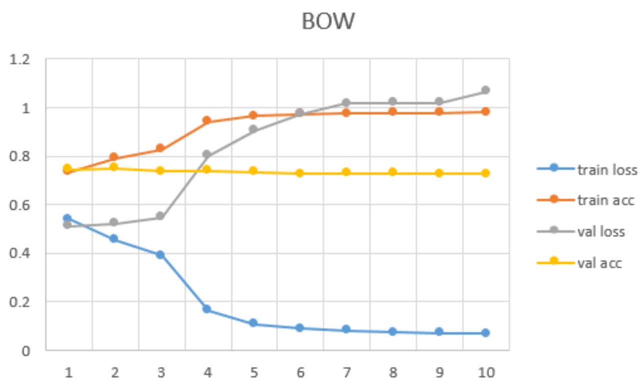
a. (0.5%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法,回報模型的正確率並繪出訓練曲線。

我使用的 RNN 模型為 LSTM，架構非常簡單，我呼叫 Keras 的 LSTM，輸出維度為 64，並設定 Recurrent dropout 和 dropout 為 0.2，並在 LSTM 後面加入一層 Dense layer，最後用 sigmoid function 來求出為惡意留言的機率。而我使用的 word embedding 的方法則是採用套件 gensim，利用他的 API word2vec，將字詞轉換成 index，作為 RNN 的 input 並對其加以做訓練，最終在丟上 Kaggle 後，最終得到的準確率為：
Train accuracy: 92.42%　Public score: 75.93%　Private score: 75.74%
下面的圖表為我的訓練曲線。



b.(0.5%) 請實作 BOW+DNN 模型,敘述你的模型架構,回報正確率並繪出訓練曲線。

我的 Bag of word 的 Vocabulary size 設為 10000，在模型架構上總共用了兩層 Dense Layers，第一層輸出維度為 256，第二層則為 1，最後用了一個 sigmoid 判斷機率，最終得到的準確率為：
Train accuracy: 97.9%　Public score: 74.29%　Private score: 73.51%

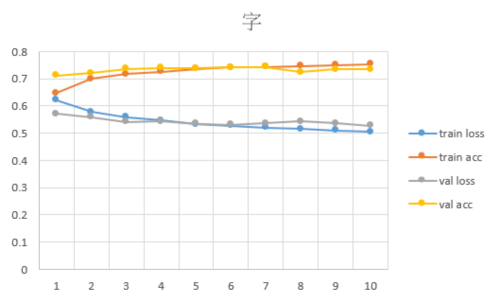


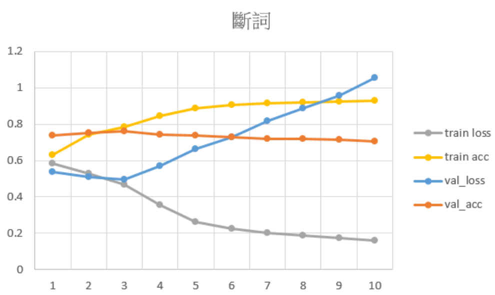
Problem 2. (1%) 請敘述你如何 improve performance(preprocess, embedding,架構等), 並解釋為何這些做法可以使模型進步。

下列為我的作法：

1. 我使用了套件 Jieba 將整段留言做斷詞，使用這個方法的目的為，中文在表達上面常常是以兩三個字所組成的詞去呈現，若將一個字一個字拆開可能會是不同的意思，因此做斷詞的表現會比以字為單位去做訓練來的好。
2. Word embedding 採用套件 gensim 來做，使用了 word to vector，我認為這樣的方法會比其他方法表現來的好，例如採用 one-hot encoding 將會消耗大量資源，而 Bag of words 表現較差則是因為這個做法沒有詞與詞之間的順序，會影響機器對語意的判斷，因此使用 word to vector 保留了對詞句順序，可以讓機器較容易對語意做判斷。。
3. 在架構上我採用了 LSTM 而非傳統的 RNN，因為 LSTM 架構較複雜可以過濾過去的狀態，選擇較有影響的狀態，因此我認為 LSTM 可以比 RNN 來的有效果。
4. 另外我將 Training data 切成 Training set 和 Validation set，並使用了 Model check point，將 Validation set 表現最好的 Model 儲存下來，而這麼做的原因是因為我認為在 Validation set 比現好時，相對的 Testing data 才會相對的表現得好。

Problem 3. (1%) 請比較不做斷詞 (e.g., 以字為單位) 與有做斷詞,兩種方法實作出來的效果差異,並解釋為何有此差別。





上圖為兩者在 training 過程的比較(所有架構皆相同)，可以明顯看出兩者的的不同，不做斷詞的 loss 和 accuracy 幾乎是平緩不動的，而有做斷詞的 loss 則可以降到約 0.2，accuracy 則可以提升到約 90%，但相對的也比較快達到 Overfitting。。

Testing 比較	Public score	Private score
不做斷詞	73.45%	73.11%
斷詞	75.93%	75.74%

上表為兩者在 Kaggle 分數上的表現，明顯的有做斷詞表現比斷詞來的突出。

最後根據 Training 和 Testing 上的表現，可以判斷有做斷詞表現會較不做來的好，我認為是因為中文在表達上面常常是以兩三個字所組成的詞去呈現，若將一個字一個字拆開可能會是不同的意思，例如”電腦”兩個字拆開來後，就跟原本的意思完全不同，因此會造成機器學習的過程中，讓電腦誤會語句所要表達

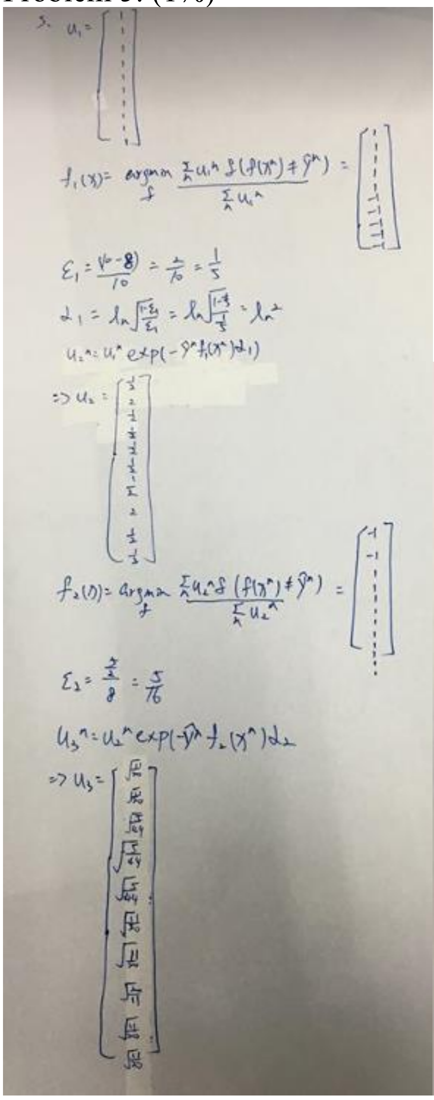
的真正意思，所以在做中文語意辨識的時候，做斷詞是一件必要且重要的事情。

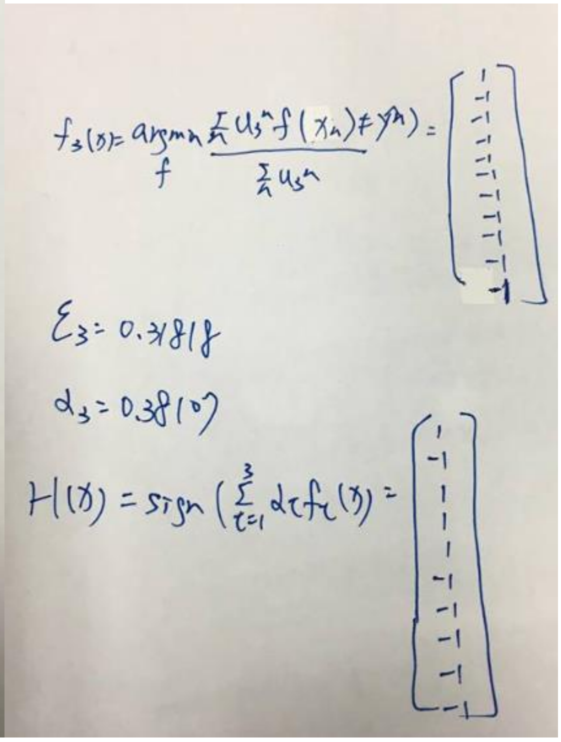
Problem 4. (1%) 請比較 RNN 與 BOW 兩種不同 model 對於”在說別人白痴之前,先想想自己”與”在說別人之前先想想自己,白痴”這兩句話的分數(model output),並討論造成差異的原因。

	在說別人白痴之前,先想想自己	在說別人之前先想想自己,白痴
RNN	0.3028	0.7191
BOW	0.4565	0.4565

上表是分別用 RNN 和 BOW 兩種架構對兩句話分別做預測，可以看出 BOW 和 RNN 求出的結果最大的差別在於，BOW 得到的預測結果是一樣的，我認為會造成這樣的原因是因為，BOW 是計算在 Vocabulary 裡面某個詞出現了幾次，因此上述這兩句若不考慮順序，其實他們的內容是一樣的，也就會造成 BOW 得到一樣的結果，而我使用的 Word embedding 搭配了 RNN 則會有詞句順序的差別，因此可以保留語意，所以預測的結果才會有所不同。

Problem 5. (1%)





Problem 6. (1%)

$t=1:\ z=3,\ z_i=90,\ z_f=10,\ z_o=-10$

$$c_1' = f(z_i)\,g(z) + c\,f(z_f) = \frac{1}{1+e^{-90}}\times 3 + 0\times\frac{1}{1+e^{-10}} = \frac{3}{1+e^{-90}}$$

$$y_1 = f(z_o)\,h(c_1') = \frac{1}{1+e^{10}}\times\frac{3}{1+e^{-90}} = \frac{3}{(1+e^{10})(1+e^{-90})}$$

$t=2:\ z=-2,\ z_i=90,\ z_f=10,\ z_o=90$

$$c_2' = f(z_i)\,g(z) + c_1' f(z_f) = \frac{1}{1+e^{-90}}\times(-2) + c_1'\times\frac{1}{1+e^{-10}} = \frac{(-2)\times(1+e^{-10}) + c_1'(1+e^{-90})}{(1+e^{-90})(1+e^{-10})}$$

$$y_2 = f(z_o)\,h(c_2') = \frac{1}{1+e^{-90}}\times\frac{-2(1+e^{-10}) + c_1'(1+e^{-90})}{(1+e^{-90})(1+e^{-10})} = \frac{-2(1+e^{-10}) + c_1'(1+e^{-90})}{(1+e^{-90})^2(1+e^{-10})}$$

$t=3:\ z=4,\ z_i=190,\ z_f=-90,\ z_o=90$

$$c_3' = f(z_i)\,g(z) + c_2' f(z_f) = \frac{1}{1+e^{-190}}\times 4 + c_2'\times\frac{1}{1+e^{90}} = \frac{4(1+e^{90}) + c_2'(1+e^{-190})}{(1+e^{-190})(1+e^{90})}$$

$$y_3 = f(z_o)\,h(c_3') = \frac{1}{1+e^{-90}}\times\frac{4(1+e^{90}) + c_2'(1+e^{-190})}{(1+e^{-190})(1+e^{90})} = \frac{4(1+e^{90}) + c_2'(1+e^{-190})}{(1+e^{-190})(1+e^{90})(1+e^{-90})}$$

$t=4:\ z=0,\ z_i=90,\ z_f=10,\ z_o=90$

$$c_4' = f(z_i)\,g(z) + c_3' f(z_f) = \frac{1}{1+e^{-90}}\times 0 + c_3'\times\frac{1}{1+e^{-10}} = \frac{c_3'}{1+e^{-10}}$$

$$y_4 = f(z_o)\,h(c_4') = \frac{1}{1+e^{-90}}\times\frac{c_3'}{1+e^{-10}} = \frac{c_3'}{(1+e^{-90})(1+e^{-10})}$$

$t=5:\ z=2,\ z_i=90,\ z_f=10,\ z_o=-10$

$$c_5' = f(z_i)(g(z)) + c_4' f(z_f) = \frac{2}{1+e^{-90}} + \frac{c_4'}{1+e^{-10}} = \frac{2(1+e^{-10}) + c_4'(1+e^{-90})}{(1+e^{-90})(1+e^{-10})}$$

$$y_5 = f(z_o)\,h(c_5') = \frac{1}{1+e^{10}}\times\frac{2(1+e^{-10}) + c_4'(1+e^{-90})}{(1+e^{-90})(1+e^{-10})} = \frac{2(1+e^{-10}) + c_4'(1+e^{-90})}{(1+e^{10})(1+e^{-90})(1+e^{-10})}$$

$t=6:\ z=-4,\ z_i=-10,\ z_f=110,\ z_o=90$

$$c_6' = f(z_i)\,g(z) + c_5' f(z_f) = \frac{-4}{1+e^{10}} + \frac{c_5'}{1+e^{-110}} = \frac{-4(1+e^{-110}) + c_5'(1+e^{10})}{(1+e^{10})(1+e^{-110})}$$

$$y_6 = f(z_o)\,h(c_6') = \frac{1}{1+e^{-90}}\times\frac{-4(1+e^{-110}) + c_5'(1+e^{10})}{(1+e^{10})(1+e^{-110})} = \frac{-4(1+e^{-110}) + c_5'(1+e^{10})}{(1+e^{-90})(1+e^{10})(1+e^{-110})}$$

$t=7$: $z=1$, $z_i=1\rho_0$, $z_f=-\rho_0$, $z_o=\rho_0$

$$C_7' = f(z_i)g(z)+C_6'f(z_f)$$

$$= \frac{1}{1+e^{-1\rho_0}} + \frac{C_6'}{1+e^{\rho_0}}$$

$$= \frac{1+e^{\rho_0}+C_6'(1+e^{-(\rho_0)})}{(1+e^{-(\rho_0)})(1+e^{\rho_0})}$$

$$y_7 = f(z_o)h(C_7')$$

$$= \frac{1}{1+e^{-\rho_0}} \times \frac{1+e^{\rho_0}+C_6'(1+e^{-(\rho_0)})}{(1+e^{-(\rho_0)})(1+e^{\rho_0})}$$

$$= \frac{1+e^{\rho_0}+C_6'(1+e^{-(\rho_0)})}{(1+e^{-\rho_0})(1+e^{-(\rho_0)})(1+e^{\rho_0})} \#$$

$t=8$: $z=2$, $z_i=\rho_0$, $z_f=10$, $z_o=\rho_0$

$$C_8' = f(z_i)g(z)+C_7'f(z_f)$$

$$= \frac{2}{(1+e^{-\rho_0})} + \frac{C_7'}{(1+e^{-10})}$$

$$= \frac{2(1+e^{-10})+C_7'(1+e^{-\rho_0})}{(1+e^{-\rho_0})(1+e^{-10})}$$

$$y_8 = f(z_o)h(C_8')$$

$$= \frac{1}{(1+e^{-\rho_0})} \times \frac{2(1+e^{-10})+C_7'(1+e^{-\rho_0})}{(1+e^{-\rho_0})(1+e^{-10})}$$

$$= \frac{2(1+e^{-10})+C_7'(1+e^{-\rho_0})}{(1+e^{-\rho_0})^2(1+e^{-10})} \#$$