Homework 4 Report Problem Set

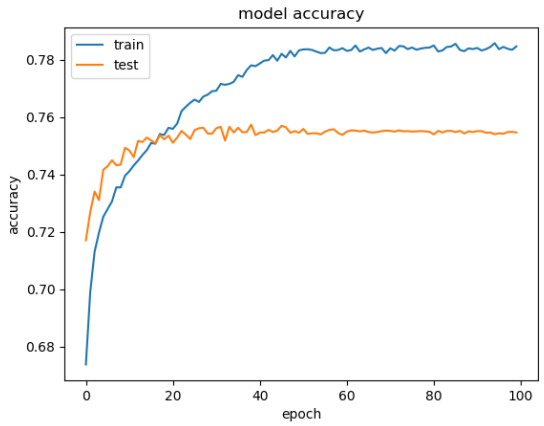
Professor Pei-Yuan Wu
EE5184 - Machine Learning
學號：R07943123　　姓名：馬咏治

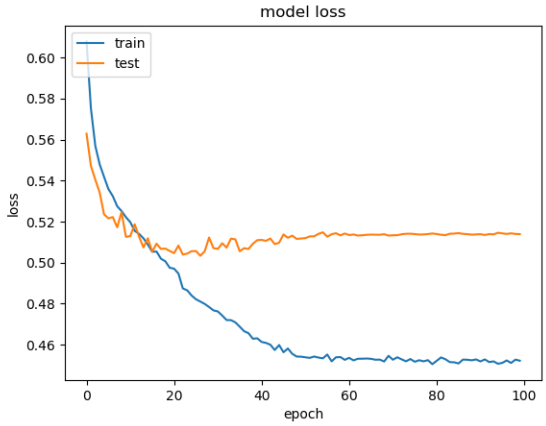
Problem 1. (0.5%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法,回 報模型的正確率並繪出訓練曲線 *。(0.5%) 請實作 BOW+DNN 模型,敘述你的模型架構, 回報正確率並繪出訓練曲線。

* 訓練曲線 (Training curve):顯示訓練過程的 loss 或 accuracy 變化。橫軸為 step 或 epoch,縱軸為 loss 或 accuracy。

我實作之 RNN, Word Embedding= word2vec

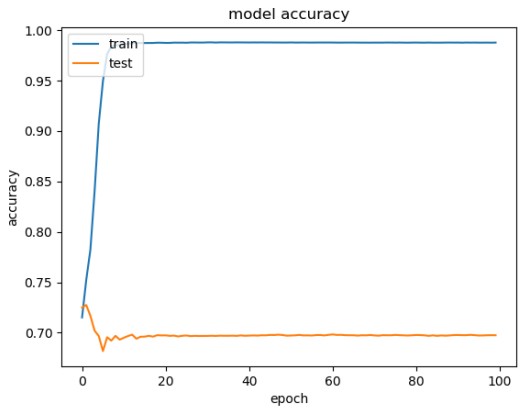
```
Layer (type)                 Output Shape              Param #
=================================================================
gru_1 (GRU)                  (None, 50, 512)           1181184
_________________________________________________________________
gru_2 (GRU)                  (None, 512)               1574400
_________________________________________________________________
dense_1 (Dense)              (None, 512)               262656
_________________________________________________________________
dense_2 (Dense)              (None, 512)               262656
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 513
=================================================================
Total params: 3,281,409
Trainable params: 3,281,409
Non-trainable params: 0
_________________________________________________________________
Train on 108000 samples, validate on 12000 samples
```

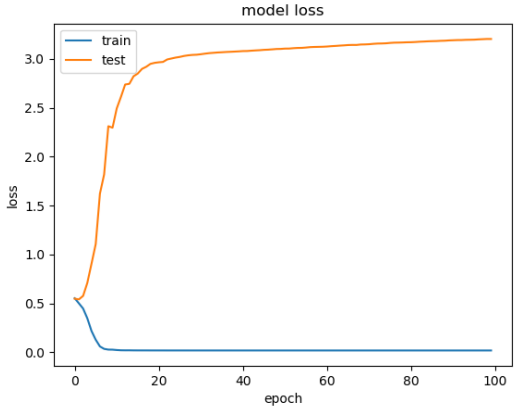




BOW+DNN

```
Layer (type)                 Output Shape              Param #
=================================================================
flatten_1 (Flatten)          (None, 12800)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              13108224
_________________________________________________________________
dense_2 (Dense)              (None, 512)               524800
_________________________________________________________________
dense_3 (Dense)              (None, 128)               65664
_________________________________________________________________
dense_4 (Dense)              (None, 64)                8256
_________________________________________________________________
dense_5 (Dense)              (None, 1)                 65
=================================================================
Total params: 13,707,009
Trainable params: 13,707,009
Non-trainable params: 0
_________________________________________________________________
Train on 108000 samples, validate on 12000 samples
```





	RNN	DNN+BOW
Accuracy	76.2%	約 70%

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。

主要對 performance 影響較大的是 preprocess，embedding 的過程。

Preprocess：首先我對 train_x.csv,test.csv 進行 training

1. 斷詞方面，原本使用 jieba 斷詞使用搜索引擎模式(cut_for_search)來增加詞量，不過改成精準模式後，Accuracy 上升了。我想搜索引擎模式應該製造太多冗詞造成

2. 表示詞的向量維度對 accuracy 也有一定的影響，太大、太小都不行，最後選在 256

3. 另外依據大部分留言的長度，我們把詞數限制在 50 以內，太多就砍掉，太少就做 padding。有嘗試調高或調低，但是 accuracy 都下降。我認為這會上升因為大部分留言經過斷詞後都會在 50 詞以內，不過當詞數太少時很可能會看不到某些詞而產生誤判。

在架構方面：

這裡沒有做太多的變化，把 GRU 換成 LSTM 或其他的 cell 都獲得差不多的結果

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。

Ans:

	做斷詞	不做斷詞
Training Accuracy	77.83	77.80
Testing Accuracy	76.075	75.775

可以看到做斷詞的效果會好，原因在於中文詞是有意義的，斷字會造成這個意義消失，所以 Accuracy 會變低。

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。

Ans:

	"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	0.5595	0.2732
BOW+DNN	0.7353	0.6778

結果如上圖所示造成這個差異的原因我認為是由於 RNN 把“白痴”和其他的詞的連接性 train 起來了。但對於這個結果我蠻意外的，以我們的直覺“在說別人之前先想想自己,白痴”應該・比較接近惡意留言。但這裡的結果缺不一樣。我想可能是因為逗點的關係，RNN 把他們之間的關聯性破壞了，造成“在說別人之前先想想自己,白痴”這句話的 output 值較小。

而 BOW 只是依據“白痴”這一個詞來判斷是否是而已留言。所以可以看到這兩個句子都被判斷成惡意留言。

Problem 5. (1%) In this exercise, we will train a binary classifier with AdaBoost algorithm on the data shown in the table. Please use decision stump as the base classifier. Perform AdaBoost algorithm for T = 3 iterations. For each iteration (t = 1, 2, 3), write down the weights u^n_t used for training, the weighted error rate ε_t, scaling coefficient α_t, and the classification function $f_t(x)$.

The initial weights u^n_1 are set to 1 (n = 0, 1, ..., 9). Please refer to the course slides for the definitions of the above notations. Finally, combine the three classifiers and write down the final classifier.

x 0 1 2 3 4 5 6 7 8 9 y + - + + + - - + - -

Ans:

When iteration t = 1,

$$f_1(x) = \begin{cases} -1, & x \geq 5 \\ 1, & x < 5 \end{cases} \quad \varepsilon_1 = 0.2, \ \alpha_1 = \ln 2 = 0.693, d_1 = 2$$

x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-
u^n_1	1	1	1	1	1	1	1	1	1	1

When iteration t = 2,

$$f_2(x) = \begin{cases} -1, & x \le 1 \\ 1, & x > 1 \end{cases} \quad \varepsilon_2 = 0.3125,\ \alpha_2 = 0.394,\ d_2 = 1.48$$

x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-
u_2^n	0.5	2	0.5	0.5	0.5	0.5	0.5	2	0.5	0.5

When iteration t = 3,

$$f_3(x) = \begin{cases} -1, & x \ge 1 \\ 1, & x < 1 \end{cases} \quad \varepsilon_3 = 0.319,\ \alpha_3 = 0.378,\ d_3 = 1.46$$

x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-
u_1^n	0.74	1.35	0.34	0.34	0.34	0.74	0.74	1.35	0.74	0.74

Finally, The final decision will be on :$f(x) = \alpha_1 * f_1(x) + \alpha_2 * f_2(x) + \alpha_3 * f_3(x)$

$$f(x) = 0.693 * f_1(x) + 0.394 * f_2(x) + 0.378 * f_3(x)$$

x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-

Problem 6. (1%) In this exercise, we will simulate the forward pass of a simple LSTM cell. Figure.1 shows a single LSTM cell, where z is the cell input, z_i, z_f, z_o are the control inputs of the gates, c is the cell memory, and f, g, h are activation functions. Given an input x, the cell input and the control inputs can be calculated by

$$z = w \cdot x + b$$

$$z_i = w_i \cdot x + b_i$$

$$z_f = w_f \cdot x + b_f$$

$$z_o = w_o \cdot x + b_o$$

Where w, w_i, w_f, w_o are weights and b, b_i, b_f, b_o are biases. The final output can be calculated by

$$y = f(z_o)h(c')$$

where the value stored in cell memory is updated by

$$c' = f(z_i)g(z) + cf(z_f)$$

Given an input sequence x^t (t = 1, 2, ..., 8), please derive the output sequence y^t. The input sequence, the weights, and the activation functions are provided below. The initial value in cell memory is 0. Please note that your calculation process is required to receive full credit.

Figure 1: The LSTM cell 2

$w = [0, 0, 0, 1]$, $b = 0$
$w_i = [100, 100, 0, 0]$, $b_i = -10$
$w_f = [-100, -100, 0, 0]$, $b_f = 110$
$w_o = [0, 0, 100, 0]$, $b_o = -10$

t	1	2	3	4	5	6	7	8
x^t	0	1	1	0	0	0	1	1
	1	0	1	1	1	0	1	0
	0	1	1	1	0	1	1	1
	3	-2	4	0	2	-4	1	2

$$f(z) = \frac{1}{1+e^{-z}} \qquad g(z) = z \qquad h(z) = z$$

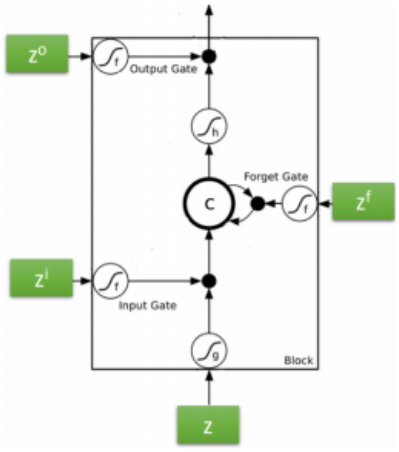


Figure 1: The LSTM cell

Ans:

t = 1:

	z	z_i	z_f	z_o	y	c'
value	3	90	10	-10	0	3

t = 2:

	z	z_i	z_f	z_o	y	c'
value	-2	90	10	90	1	1

t = 3:

	z	z_i	z_f	z_o	y	c'
value	4	190	-90	90	4	4

t = 4:

	z	z_i	z_f	z_o	y	c'
value	0	90	10	90	4	4

t = 5:

	z	z_i	z_f	z_o	y	c'
value	2	90	10	-10	0	6

t = 6:

	z	z_i	z_f	z_o	y	c'
value	-4	-10	110	90	6	6

t = 7:

	z	z_i	z_f	z_o	y	c'
value	1	190	-90	90	1	1

t = 8:

	z	z_i	z_f	z_o	y	c'
value	2	90	10	90	3	3

Output Sequence: $\boldsymbol{y^t = \{0, 1, 4, 4, 0, 6, 1, 3\}}$