Homework 4 Report Problem Set

Professor Pei-Yuan Wu
EE5184 - Machine Learning

B05902074 資工三 魏佑珊

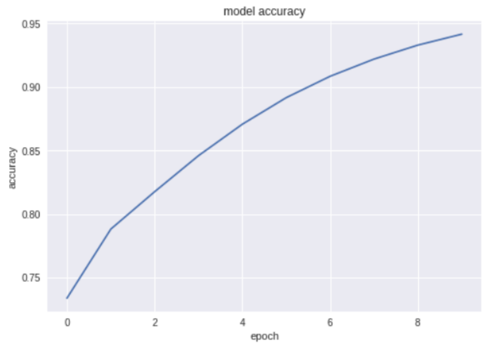
Note：以下數據由較新版本的 keras 跑出，純粹用作分析，在 github 上的模型版本是符合作業要求的。

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。

我以 gensim 的 word2Vec 做 word embedding。Model 則是 embedding layer → LSTM → Dense。Word dimension 及 LSTM unit 數目皆為 100。

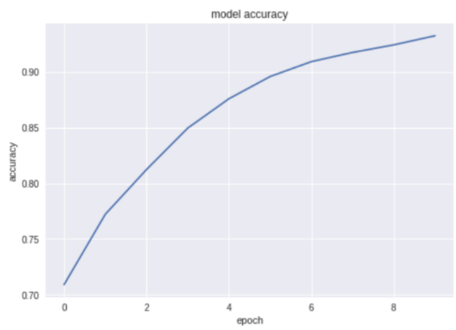
訓練曲線如下



BOW 則是以三層 Dense 實作，並且為了不讓字典中的字數太多，限制只有出現 10 次以上的字才放入字典

正確率：public 0.75657, private 0.75345

訓練曲線如下



正確率：public 0.74132, private 0.73655

我的訓練曲線中只有繪出 training data 的正確率，但其實觀察 validation data 會發現通常在 3~5 個 epochs 後 validation accuracy 就開始降低了，overfit 得很快。

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。

我使用了

(1) word embedding – 相較於 bag of word，word embedding 較能考慮字詞在句子中的前後順序，而這在判斷善意/惡意的留言時可能很有用，因為同樣的一句話換個順序說，意思可能會截然不同。

(2) word embedding 參數：在訓練 word embedding 的 model 時，我只取用出現過 5 次以上的字歸入字典，避免字典過於龐大或是一些不常出現的偏僻字詞也被納入。

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。

不做斷詞：public score 0.75020, private score 0.74655

做斷詞：public score 0.75657, private score 0.75345

中文字不同於英文字，英文字詞用空白分開就可以；中文的字有一些無法單獨表達意義，有一些則是和別的字組合後會產生不同意義。如果沒做好斷詞，model 就學不到考慮這件事，正確率會降低。

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。

RNN →

在說別人白痴之前，先想想自己—0.45625

在說別人之前先想想自己，白痴—0.61562

BOW→

在說別人白痴之前，先想想自己—0.32421

在說別人之前先想想自己，白痴—0.32421

RNN 有「順序性」，亦即能學到詞語出現在一句話的不同地方產生的結果；但 Bag of words 單純統計每個詞語在句子中的出現次數作為特徵。因此就上述兩句話而言，RNN 會認定他們是不同的，Bag of words 則會覺得都是相同的（因為每個詞出現的次

數都一樣。）

我覺得比較特別的是 BOW 並不認為這兩句話是惡意留言…我猜也有可能是我創建 BOW 是只取 training data 中有出現過 10 次以上的詞，可能”白痴”沒有出現 10 次以上吧…。但仍可看出在 BOW 判斷下，這兩句話都會判斷出同樣的分數。

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_t^n used for training, the weighted error rate ϵ_t, scaling coefficient α_t, and the classification function $f_t(x)$. The initial weights u_1^n 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	+	-	+	+	+	-	-	+	-	-

5. T=1:

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

(first classifier)

$$f_1(x) = \begin{cases} x \le 4 : + \\ x > 4 : - \end{cases}$$

misclassified: $x = 1$ and $x = 7$. $\varepsilon_1 = \frac{2}{10} = 0.2$, $d_1 = \sqrt{\frac{0.8}{0.2}} = 2$, $\alpha_1 = 0.69$

$\Rightarrow u_2^1 = u_2^7 = 1 \times 2 = 2$, $u_2^2, u_2^3, u_2^4, u_2^5, u_2^6, u_2^8, u_2^9 = \frac{1}{2} = 0.5$

T=2:

u	0.5	2	0.5	0.5	0.5	0.5	0.5	2	0.5	0.5
x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-

(second classifier)

$$f_2(x) = \begin{cases} x \le 1 : - \\ x > 1 : + \end{cases}$$

misclassified: $x = 0, 5, 6, 8, 9$ $\varepsilon_2 = \frac{2.5}{8} = 0.3125$, $d_2 = \sqrt{\frac{0.6875}{0.3125}} \doteqdot 1.48$

$\alpha_2 = 0.39$

$\Rightarrow u_3^0, u_3^5, u_3^6, u_3^8, u_3^9 = 0.5 \times 1.48 = 0.74$, $u_3^1 = u_3^7 = 2/1.48 = 1.35$,

$u_3^2 = u_3^3 = u_3^4 = \frac{0.5}{1.48} = 0.34$

T=3:

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

(third classifier)

$$f_3(x) = \begin{cases} x \le 0 : + \\ x > 0 : - \end{cases}$$

misclassified: $x = 2, 3, 4, 7$. $\varepsilon_3 = \frac{2.37}{7.42} = 0.3194$, $d_3 = \sqrt{\frac{0.6806}{0.3194}} \doteqdot 1.46$

$\alpha_3 = 0.38$

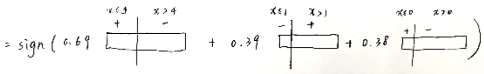
$\Rightarrow u_4^0 = 0.74/1.46 \doteqdot 0.51$, $u_4^1 = 1.35/1.46 \doteqdot 0.92$, $u_4^2 = u_4^3 = u_4^4 = 0.34 \times 1.46 \doteqdot 0.50$

$u_4^5 = 0.74/1.46 \doteqdot 0.51$, $u_4^6 = 0.74/1.46 \doteqdot 0.51$, $u_4^7 = 1.35 \times 1.46 \doteqdot 1.97$,

$u_4^8 = 0.74/1.46 \doteqdot 0.51$, $u_4^9 = 0.74/1.46 \doteqdot 0.51$

Final Classifier $H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t f_t(x)\right)$

$= \text{sign}\left(0.69 \cdot f_1(x) + 0.39 \cdot f_2(x) + 0.38 \cdot f_3(x)\right)$



⇒ Result:

$x=0$: +, $x=1$: −, $x=2$: +, $x=3$: +, $x=4$: +, $x=5 \sim 9$: −.

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.

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

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

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

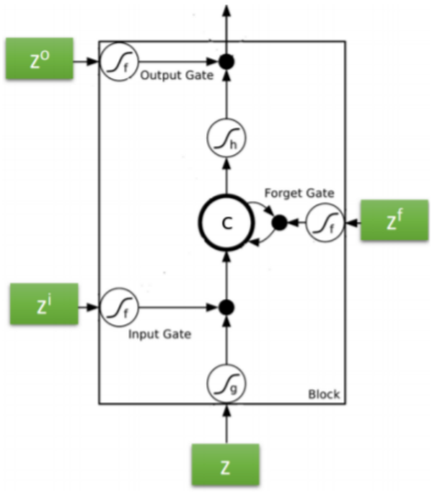


Figure 1: The LSTM cell

T = 1:

Z = W x + b = 3, Zi = Wi x + b = 90, Zf = Wf x + b = 10, Zo = Wo x + b = -10, c’ = f(Zi)g(Z) + c*f(Zf) = 3, y = f(Zo)h(c’) = 0

T = 2:

Z = -2 , Zi = 90, Zf = 10, Zo = 90, c’ = 1, y = 1

T = 3:

Z = 4, Zi = 190, Zf = -90, Zo = 90, c’ = 4, y = 4

T = 4:

Z = 0, Zi = 90, Zf = 10, Zo = 90, c’ = 4, y = 4

T = 5:

Z = 2, Zi = 90, Zf = 10 , Zo =-10, c’ = 6, y = 0

T = 6:

Z = -4, Zi = -10, Zf = 110, Zo = 90, c’ = 6, y = 6

T = 7:

Z = 1, Zi = 190, Zf = -90, Zo = 90, c’ = 1, y = 1

$T = 8$:

$Z = 2$, $Zi = 90$, $Zf = 10$, $Zo = 90$, $c' = 3$, $y = 3$