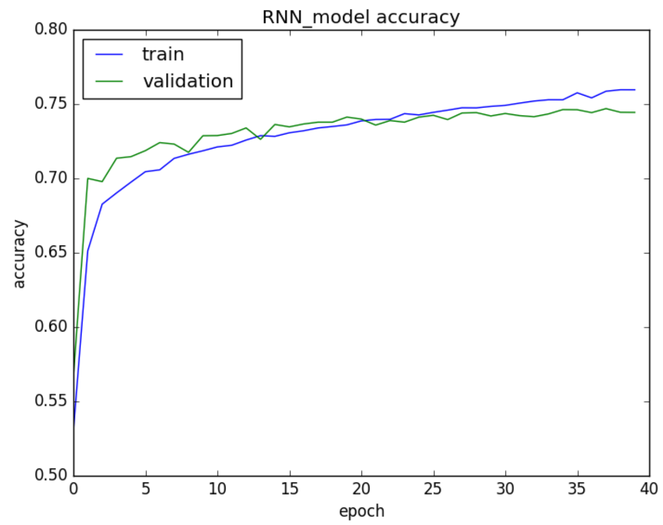
Homework 4 Report Problem Set

Professor Pei-Yuan Wu
EE5184 - Machine Learning b04504042 電機三 劉家豪

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。

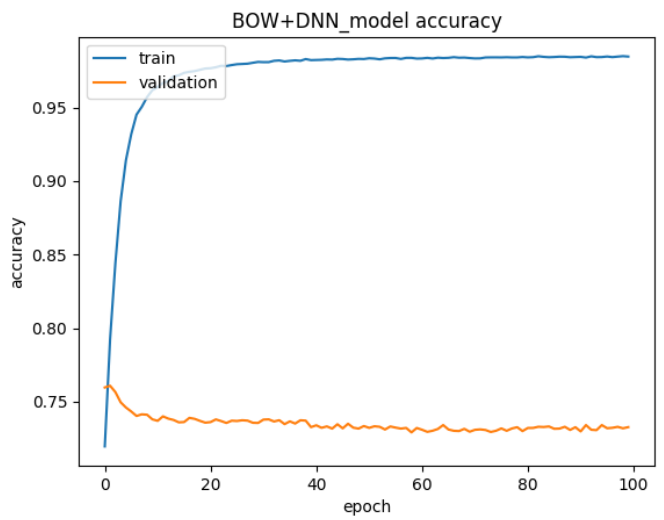
* 訓練曲線 (Training curve):顯示訓練過程的 loss 或 accuracy 變化。橫軸為 step 或 epoch,縱軸為 loss 或accuracy。



word2vec : size = 256 , min_count = 5 , iter =10

training data 先用jeiba切成詞並且轉換成word2vec 向量，我把每個句子都pad成(48,256)並且餵入以下model。

lstm : 兩層。第一層return sequence ＝ true ，所以輸出為(48,256) 。第二層return sequence ＝ false ，取最後一個輸出，再丟入兩層nn ，用sigmoid產生分數結果。



word2vec : size = 256 , min_count = 5 , iter =10

training data 維度(120000 , vocabulary size)

架構使用三層fully connected network ， output size 分別為 256 , 128 , 64 。 drop out rate 均為 0.5 ，activation function 均為relu。

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。

在做word2vec model時，發現iter對model影響很重要，太少的話train得不夠完整，於是我把iter設成10。min count 則是設成5 ， 代表一個字要出現五次才會出現在vocabulary裡面，因為有些字出現太少可能是錯字或是沒意義的字。training架構方面，我使用兩層lstm，表現比一層還好，並且發現加了recurrent_dropout這個參數，表現進步很多。另外一點是，一開始在training時發現很快就overfit了，因此加上dropout 可以避免太快overfit。

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。

實作後發現沒有斷詞的結果比較好，推測是因為這種網路留言判斷是否惡意，前後語序關係不是那麼重要，重要的是關鍵詞有沒有出現。而以字為單位去train，前後字與字的關係相對來講就是詞的形成。 經過斷詞的，輸出的是詞與詞之間關係，但這比判斷一個詞有沒有出現複雜很多，因此效果會略差。而沒斷詞的，假如有關鍵詞出現，就會保留組成該詞的字與字的關係。

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： RNN : 0.4596 , BOW : 0.6972
在說別人之前先想想自己,白痴： RNN : 0.6278 , BOW : 0.6972

BOW在兩個例子的分數一樣高，原因是因為BOW不在乎前後文，只在乎出現的詞，兩個句子的詞組成是一樣的，所以對BOW來說這兩個例子其實是一樣的，且因為存在白痴這個詞，所以判斷為惡意。而RNN可以保留句子前後文的關係，語序會影響結果，雖然都有白癡這個詞，但第一個例子他不會判斷為惡意。

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	+	-	+	+	+	-	-	+	-	-

collaborator：b04901065 游博翔

b04203058 蘇軒

5. $\underline{t=1}$ 切在 4,5 中間 $f_1(x)$ (左+右−)

error rate $\varepsilon_1 = 0.2$ $d_1 = \sqrt{\frac{(1-0.2)}{0.2}} = 2$

$u_1^n = 1$ $\alpha_1 = \ln d_1 = \ln 2 = 0.693$

${}^{+}u_2^0 = 0.5$ ${}^{-}u_2^1 = 2$ ${}^{+}u_2^2 = 0.5$ ${}^{+}u_2^3 = 0.5$ ${}^{+}u_2^4 = 0.5$ ${}^{-}u_2^5 = 0.5$

${}^{-}u_2^6 = 0.5$ ${}^{+}u_2^7 = 2$ ${}^{-}u_2^8 = 0.5$ ${}^{-}u_2^9 = 0.5$

$\underline{t=2}$ $Z_2 = 8$ 切在 1,2 中間 $f_2(x)$ (左−右+) error rate $\varepsilon_2 = \frac{1}{8}(0.5+0.5\times4)$

$= 0.3125$

$u_3^0 = 0.5\times1.48 = 0.74$ $u_3^1 = \frac{2}{1.48} = 1.35$

$u_3^2 = \frac{0.5}{1.48} = 0.338$ $u_3^3 = 0.338$

$u_3^4 = 0.338$ $u_3^5 = 0.74$

$u_3^6 = 0.74$ $u_3^7 = 1.35$ $u_3^8 = 0.74$ $u_3^9 = 0.74$

$d_2 = \sqrt{\frac{(1-0.3125)}{0.3125}} = 1.48$

$\alpha_2 = \ln d_2 = 0.394$

$\underline{t=3}$ $Z_3 = 7.414$ 切在 0,1 之間 $f_3(x)$ (左+右−) error rate $\varepsilon_3 = \frac{1}{7.414}(0.34\times3+1.35)$

$= 0.32$

$d_3 = \sqrt{\frac{(1-0.32)}{0.32}} = 1.45$

$\alpha_3 = \ln d_3 = 0.372$

Result:

0	1	2	3	4	5	6	7	8	9
+	−	+	+	+	−	−	+	−	−

$0.372 f_3$ $0.394 f_2$ $0.693 f_1$

$\Downarrow$

0	1	2	3	4	5	6	7	8	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
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

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

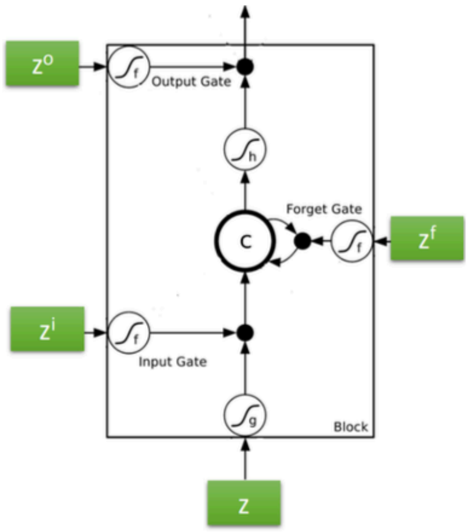


Figure 1: The LSTM cell

6. $Z = Wx + b$

$Z_i = W_i x + b_i$

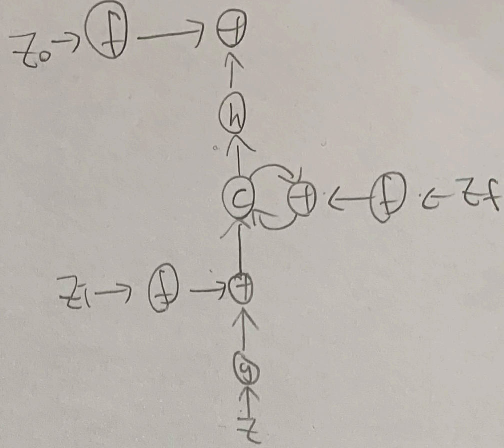
$Z_f = W_f x + b_f$

$Z_o = W_o x + b_o$

$y = f(Z_o)\,h(c')$

$c' = f(Z_i)\,g(Z) + c\,f(Z_f)$

$C_0 = 0$



$t=1$ $\quad Z = 3+0 = 3 \quad Z_i = 100 - 10 = 90 \quad Z_f = -100 + 110 = 10 \quad Z_o = 0 - 10 = -10$

$c' = f(90)g(3) + 0f(10) = \frac{1}{1+e^{-90}} \times 3 = 3 = C_1$

$y_1 = f(Z_o)\,h(c') = \frac{1}{1+e^{110}} \times 3 = 0$

$t=2$ $\quad Z = -2+0 = -2 \quad Z_i = 100 - 10 = 90 \quad Z_f = -100 + 110 = 10 \quad Z_o = 100 - 10 = 90$

$c' = f(90)g(-2) + C_1 f(10) = -2 + 3 \times \frac{1}{1+e^{-10}} = 1 = C_2$

$y_2 = f(Z_o)h(c') = \frac{1}{1+e^{-90}} \times 1 = 1$

$t=3$ $\quad Z = 4 \quad Z_i = 200 - 10 = 190 \quad Z_f = -200 + 110 = -90 \quad Z_o = 100 - 10 = 90 \quad c' = f(190)g(4) + C_2 f(-90) = 4 = C_3$

$y_3 = f(Z_o)h(c') = 1 \times 4 = 4$

$t=4$ $\quad Z = 0 \quad Z_i = 100 - 10 = 90 \quad Z_f = -100 + 110 = 10 \quad Z_o = 100 - 10 = 90 \quad c' = f(90)g(0) + 4f(10) = 4 = C_4$

$y_4 = f(Z_o)h(c') = 1 \times 4 = 4$

$t=5$ $\quad Z = 2 \quad Z_i = 100 - 10 = 90 \quad Z_f = -100 + 110 = 10 \quad Z_o = 0 - 10 = -10 \quad c' = f(90)g(2) + 4f(10) = 6 = C_5$

$y_5 = f(Z_o)h(c') = 0$

$t=6$ $\quad Z = -4 \quad Z_i = -10 \quad Z_f = 110 \quad Z_o = 100 - 10 = 90 \quad c' = f(-10)g(-4) + C_5 f(110) = 6 = C_6$

$y_6 = f(Z_o)\,h(c') = 1 \times 6 = 6$

$t=7$ $\quad Z = 1 \quad Z_i = 100 + 100 - 10 = 190 \quad Z_f = -100 - 100 + 110 = -90 \quad Z_o = 100 - 10 = 90$

$c' = f(90)g(1) + C_6 f(-90) = 1 = C_7 \qquad y_7 = f(Z_o)h(c') = 1 \times 1 = 1$

$t=8$ $\quad Z = 2 \quad Z_i = 100 - 10 = 90 \quad Z_f = -100 + 110 = 10 \quad Z_o = 100 - 10 = 90$

$c' = f(90)g(2) + C_7 f(10) = 3 = C_8$

$y_8 = f(Z_o)\,h(c') = 1 \times 3 = 3$