

Homework 4 Report Problem Set

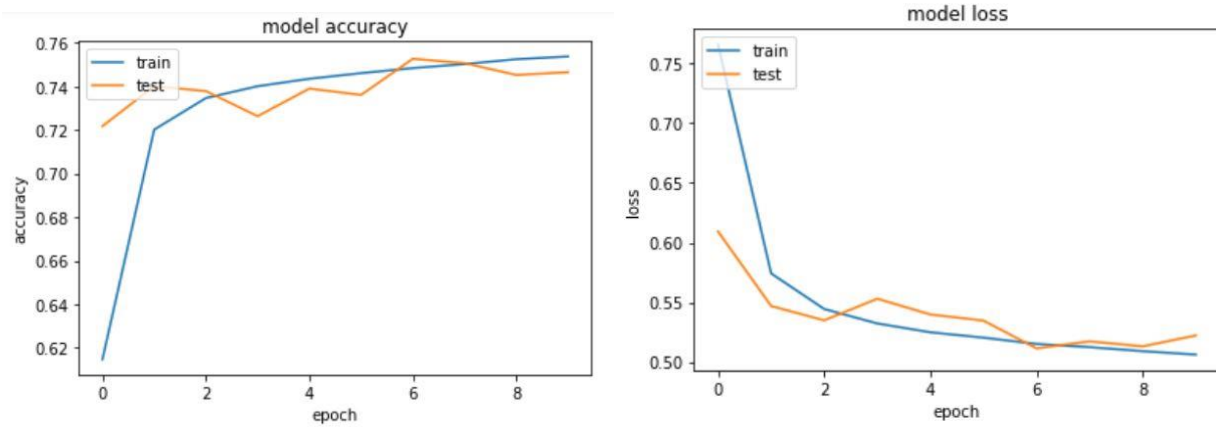
Professor Pei-Yuan Wu

EE5184 - Machine Learning

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。

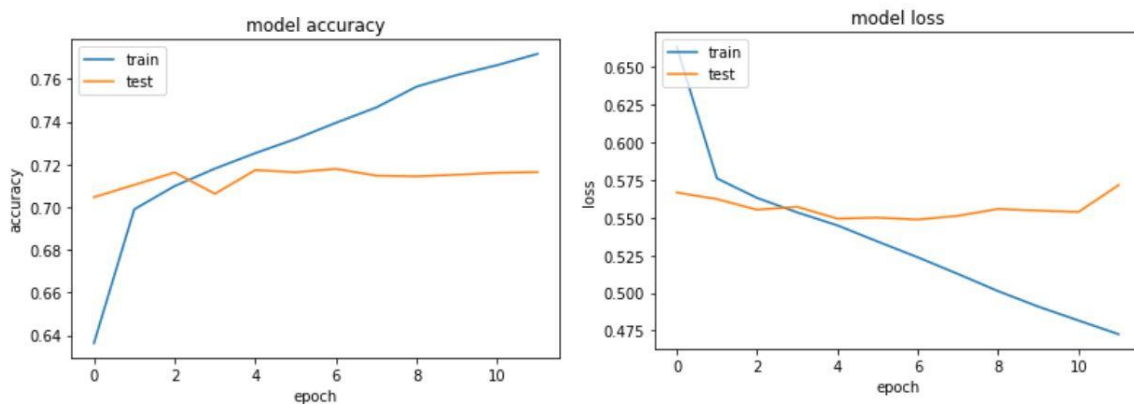
RNN:

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 200, 250)	9647500
cu_dnngru_3 (CuDNNGRU)	(None, 200, 20)	16320
dropout_6 (Dropout)	(None, 200, 20)	0
cu_dnngru_4 (CuDNNGRU)	(None, 200, 20)	2520
dropout_7 (Dropout)	(None, 200, 20)	0
flatten_2 (Flatten)	(None, 4000)	0
dense_5 (Dense)	(None, 256)	1024256
batch_normalization_4 (Batch Normalization)	(None, 256)	1024
dropout_8 (Dropout)	(None, 256)	0
dense_6 (Dense)	(None, 256)	65792
batch_normalization_5 (Batch Normalization)	(None, 256)	1024
dropout_9 (Dropout)	(None, 256)	0
dense_7 (Dense)	(None, 256)	65792
batch_normalization_6 (Batch Normalization)	(None, 256)	1024
dropout_10 (Dropout)	(None, 256)	0
dense_8 (Dense)	(None, 2)	514
Total params: 10,825,766		
Trainable params: 1,176,730		
Non-trainable params: 9,649,036		



BOW+DNN:

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 256)	64256
batch_normalization_1 (Batch Normalization)	(None, 256)	1024
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 256)	65792
batch_normalization_2 (Batch Normalization)	(None, 256)	1024
dropout_2 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 256)	65792
batch_normalization_3 (Batch Normalization)	(None, 256)	1024
dropout_3 (Dropout)	(None, 256)	0
dense_4 (Dense)	(None, 256)	65792
batch_normalization_4 (Batch Normalization)	(None, 256)	1024
dropout_4 (Dropout)	(None, 256)	0
dense_5 (Dense)	(None, 2)	514
Total params: 266,242		
Trainable params: 264,194		
Non-trainable params: 2,048		



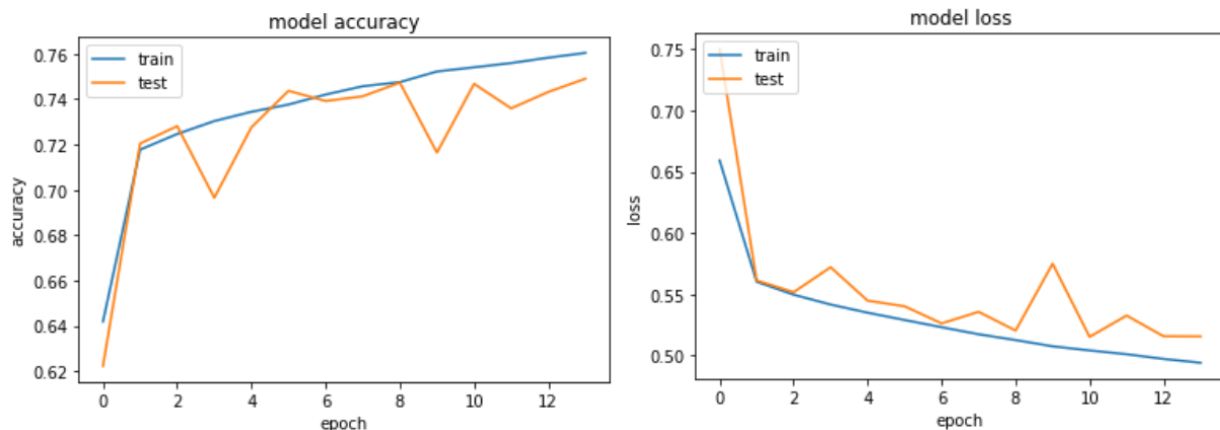
Problem 2. (1%) 請敘述你如何 improve performance (preprocess, embedding, 架構等)，並解釋為何這些做法可以使模型進步。

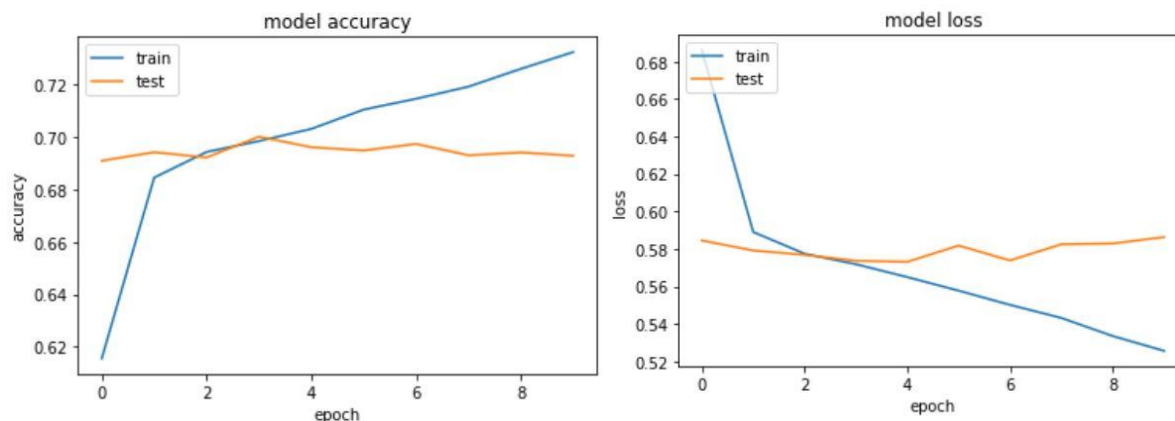
Word embedding 的 training process 有作參數調整。Size(詞向量維度大小)在 100~500 之間作測試，維度太小會無法有效表達詞與詞的關係，維度太大會使關係太稀疏而難以找出規則，最後使用 250 維度；wondows(看前後幾個字來預測)在 2~10 之間作測試，此參數可使算是 word embedding 的重點，最後實驗結果是大約 4~7 之間會有比較好的 performance。

由於有些資料自述太多(甚至還有把十二個月的星座運勢分析)，因此勢必要作縮短大小的動作。觀察許多特別長的 training data 例如「台大>>>>>台科大>>>>>忘記哪一間了...」、「一月運勢...二月運勢...~十二月運勢...」之類的。原本想說要刪除句子中重複的字(變成台大>台科大...)，但有許多重複的字是有意義的(ex. 好棒 v.s 好棒棒)，這樣做會對意義有很大的影響。後來發現，其實大部分很長的 data 都有重複的 pattern。因此比較好的做法可能直接打過長的部分剪掉，以上面的運勢分析為例，應該前兩三個月的句子就看得出 pattern 了不需要完整的十二個月。因此最後選用這種作法。

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。

RNN:





RNN 因為有考慮詞與詞之間的關係，因此 performance 還是較 BOW 好。但也有不小的下降。可以發現到 validation 的 accuracy 比較浮動，猜測可能是有些詞被拆成字之後，意義有了巨大的轉變，但為了使 training loss 下降，RNN 可能會用句子中其他的部分來作彌補與修正。而由於判斷句子是正面還負面的關鍵字通常在切斷或顛倒過後會有巨大的意義轉變(ex. 好棒 v.s 好棒棒)，因此這種上述這種現象可能導致 model 抓不到關鍵字而 overfiit training data。

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	[0.3778, 0.6221]	[0.3043, 0.6956]
BOW	[0.3273, 0.6726]	[0.3273, 0.6726]

BOW model 只考慮句子中出現過的詞，因此兩個句子者的 training input 是一樣的，而結果也當然一樣。而 RNN + word embedding 有考慮詞的前後關係因子會有不一樣的結果。可以由上表發現，雖然在我的 RNN model 下，兩者最後的 prediction 會是一樣的(負面)，但細看兩個 model 的 output 後可以發現，RNN 在處理這種相同字但順序不同的句子效果是比較好的(第一具有稍微正面一點)。也與我們的預期相符。

Problem 5. (1%)

Hw4 - Problem 6

NO.

DATE

X	0	1	2	3	4	5	6	7	8	9
Y	+	-	+	+	+	-	-	+	-	-
W	1	1	1	1	1	1	1	1	1	1
			+						-	

$$-[u(x-45) - u(45-x)]$$

$$\xi_1 = 0.2, \alpha_1 = 0.69, u_2^n = \begin{cases} 0.5 \cdot u_1^n, & \hat{y}^n = f_1(x^n) \\ 1.99 \cdot u_1^n, & \hat{y}^n \neq f_1(x^n) \end{cases}$$

X	0	1	2	3	4	5	6	7	8	9
Y	+	-	+	+	+	-	-	+	-	-
W	0.5	1.99	0.5	0.5	0.5	0.5	0.5	1.99	0.5	0.5
						+				

$$[u(x-1.5) - u(1.5-x)]$$

$$\xi_2 = 0.31, \alpha_2 = 0.4, u_3^n = \begin{cases} 0.69 \cdot u_2^n, & \hat{y}^n = f_2(x^n) \\ 1.49 \cdot u_2^n, & \hat{y}^n \neq f_2(x^n) \end{cases}$$

X	0	1	2	3	4	5	6	7	8	9
Y	+	-	+	+	+	-	-	+	-	-
W	0.75	1.33	0.37	0.37	0.37	0.75	0.75	1.33	0.75	0.75
			+							

$$-[u(x-0.5) - u(0.5-x)]$$

$$\therefore H(x) = \begin{cases} -0.69[u(x-45) - u(45-x)] \\ + 0.4[u(x-1.5) - u(1.5-x)] \\ - 0.37[u(x-0.5) - u(0.5-x)] \end{cases}$$

$$\xi_3 = 0.32, \alpha_3 = 0.37$$

$$\#$$

Problem 6. (1%)

Hw 4 Problem 7

NO.

DATE

$$\text{let } \mathbf{z}_A = \begin{bmatrix} z \\ z_1 \\ z_f \\ z_0 \end{bmatrix} = \begin{bmatrix} w^T \\ w_1^T \\ w_f^T \\ w_0^T \end{bmatrix} \cdot \mathbf{x}^t + \begin{bmatrix} b \\ b_1 \\ b_f \\ b_0 \end{bmatrix} = \mathbf{W}_A \cdot \mathbf{x}^t + \mathbf{b}_A$$

$$\Rightarrow \mathbf{W}_A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 100 & 100 & 0 & 0 \\ -100 & -100 & 0 & 0 \\ 0 & 0 & 100 & 0 \end{bmatrix}, \mathbf{b}_A = \begin{bmatrix} 0 \\ -10 \\ 110 \\ -10 \end{bmatrix}$$

$$\boxed{t=1} \quad \mathbf{z}_A = \begin{bmatrix} 0 \\ 90 \\ 0 \\ -10 \end{bmatrix} \Rightarrow \begin{cases} g^1 = 0 \\ c^1 = f(90) \cdot 0 + 0 = 0 \\ y^1 = f(-10) \cdot 0 = 0 \end{cases}$$

$$\boxed{t=2} \quad \mathbf{z}_A = \begin{bmatrix} 2 \\ 90 \\ 0 \\ 90 \end{bmatrix} \Rightarrow \begin{cases} g^2 = 2 \\ c^2 = f(90) \cdot 2 + 0 = 2 \\ y^2 = f(90) \cdot 2 = 2 \end{cases}$$

$$\boxed{t=3} \quad \mathbf{z}_A = \begin{bmatrix} 4 \\ 190 \\ -100 \\ 90 \end{bmatrix} \Rightarrow \begin{cases} g^3 = 4 \\ c^3 = f(190) \cdot 4 + f(-100) \cdot 2 = 4 \\ y^3 = f(90) \cdot 4 = 4 \end{cases}$$

$$\boxed{t=4} \quad \mathbf{z}_A = \begin{bmatrix} 0 \\ 90 \\ 0 \\ 90 \end{bmatrix} \Rightarrow \begin{cases} g^4 = 0 \\ c^4 = f(90) \cdot 0 + f(-10) \cdot 4 = 2 \\ y^4 = f(90) \cdot 2 = 2 \end{cases}$$

$$\boxed{t=5} \quad \mathbf{z}_A = \begin{bmatrix} 2 \\ 90 \\ 0 \\ -10 \end{bmatrix} \Rightarrow \begin{cases} g^5 = 2 \\ c^5 = f(90) \cdot 2 + f(-10) \cdot 2 = 1 \\ y^5 = f(-10) \cdot 1 \approx 0 \end{cases}$$

NO.

DATE

$$\boxed{t=6} \quad z_A = \begin{bmatrix} -4 \\ -10 \\ 100 \\ 90 \end{bmatrix} \Rightarrow \begin{cases} g^6 = -4 \\ C^6 = f(-10) \cdot (-4) + f(100) \cdot 1 = 1 \\ y^6 = f(90) \cdot 1 = 1 \end{cases}$$

$$\boxed{t=7} \quad z_A = \begin{bmatrix} 1 \\ 190 \\ -100 \\ 90 \end{bmatrix} \Rightarrow \begin{cases} g^7 = 1 \\ C^7 = f(190) \cdot 1 + f(-100) \cdot 1 = 1 \\ y^7 = f(90) \cdot 1 = 1 \end{cases}$$

$$\boxed{t=8} \quad z_A = \begin{bmatrix} 2 \\ 90 \\ 0 \\ 90 \end{bmatrix} \Rightarrow \begin{cases} g^8 = 2 \\ C^8 = f(90) \cdot 2 + f(0) \cdot 1 = 3.5 \\ y^8 = f(90) \cdot 3.5 = 3.5 \end{cases}$$

t	1	2	3	4	5	6	7	8
g	0	2	4	0	2	-4	1	2
C	0	2	4	2	1	1	1	3.5
y	0	2	4	2	0	1	1	3.5

#