

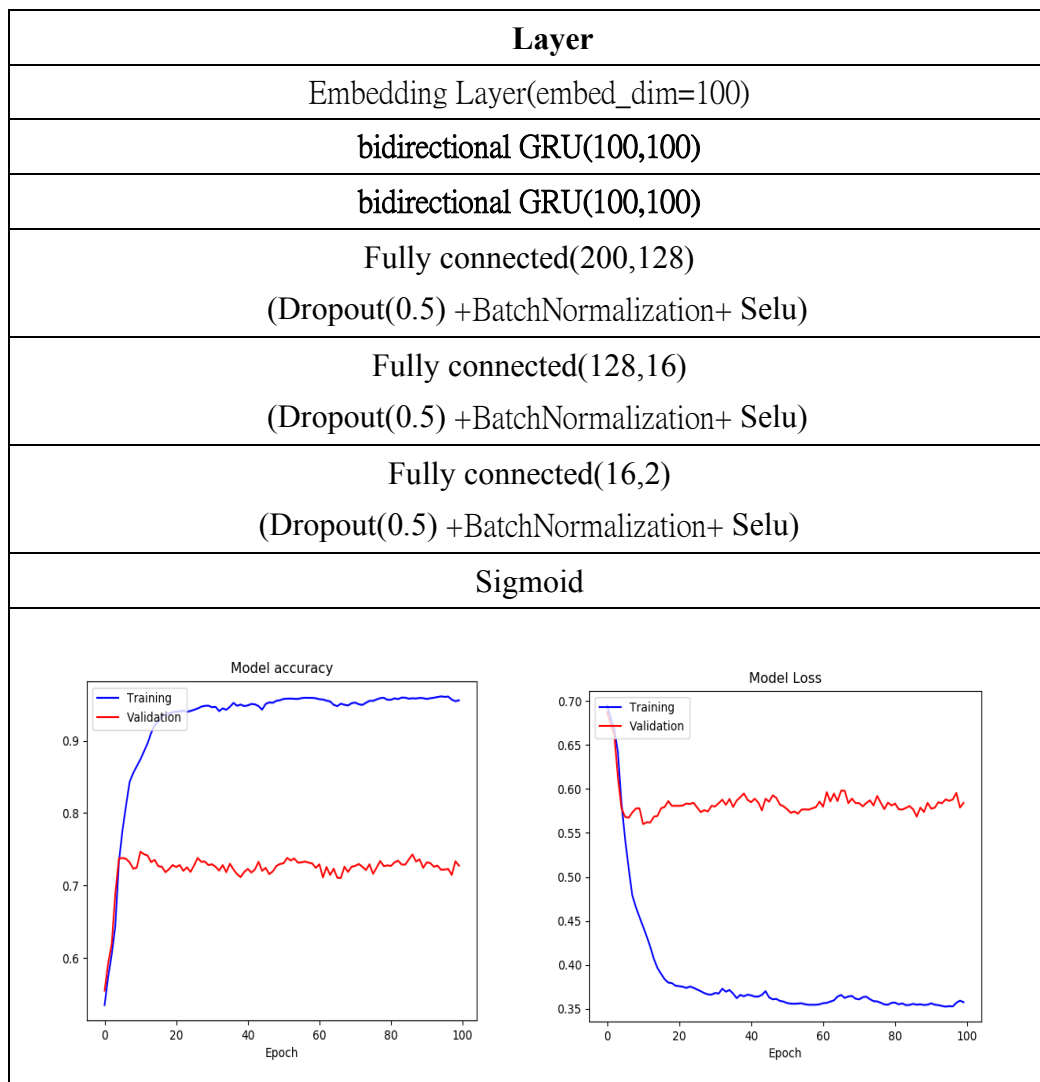
1. (1%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法，回報模型的正确率並繪出訓練曲線*

Word Embedding :

Embedding 採用 word2vector，套件使用 gnism，使用 Skip gram(如同手寫作業的第二題)。其中 embed_dim 為 100，sentence sequence 的最大長度為 30

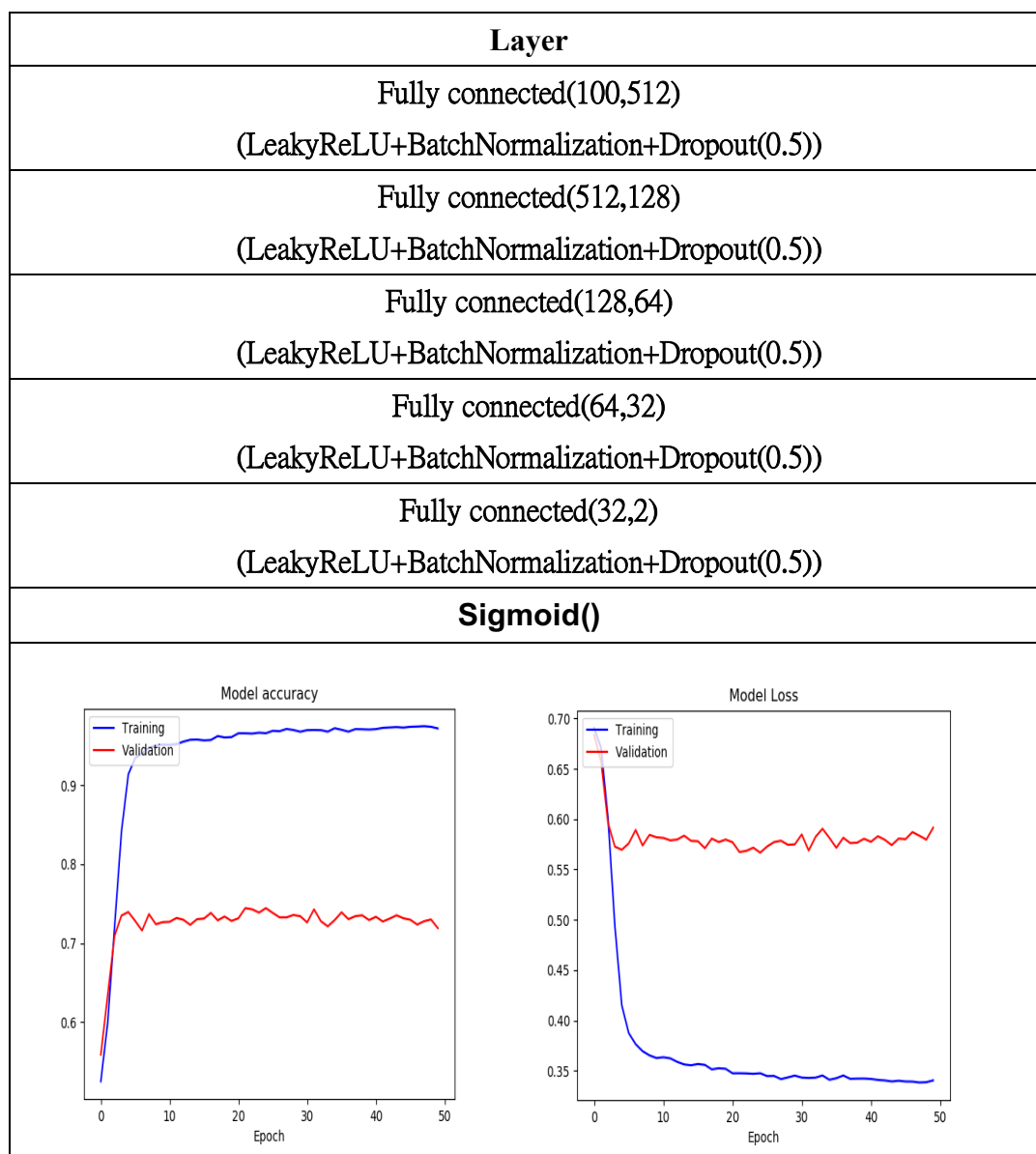
NN 架構：

最前面先用兩層 bi-directional 的 GRU，之後在接上三層的 fully-connected layer。



	Test validation	Public	Private
Accuracy	0.74	0.78604	0.78837

2. (1%) 請實作 BOW+DNN 模型，敘述你的模型架構，回報模型的正確率並繪出訓練曲線*。



Accuracy	Test validation	Public	Private
	0.72	0.74418	0.77906

3. (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。

Preprocess:

一開始在看 data 時，發現每一筆資料一開始都有"@user"，這一個不會影響判斷的資料，所以在 tokenize 的時候，如果是"@user"會選擇直接把他捨棄掉。可以看出如果把沒意義的字串拿掉，可以提高最後判斷的正確率。

	保留"@user"	捨棄"@user"
public	0.78604	0.77209
private	0.78827	0.76744

Word Embedding :

Embedding 採用 word2vector，套件使用 gnism，使用 Skip gram(如同手寫作業的第二題)。其中 embed_dim 為 100，sentence sequence 的最大長度為 30，若太長會把多餘的 word 截取掉，若太短會將其 padding 到最大的長度，也就是文句最後補上“padding”。可以發現有做 Word2vector 的方式會比 Bag of Word 來的好很多。

	Word to Vector	Bag of Word
public	0.78604	0.74418
private	0.78837	0.77906

4. (1%) 請比較不做斷詞 (e.g.,用空白分開) 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。

	斷詞	split
public	0.78604	0.74883
private	0.78837	0.75813

不做斷詞只用單個字當作輸入的 model 正確率比較低，有使用斷詞的 model 正確率較高，會產生這樣子的原因。我推測可能原因是因為在英文中會有很多詞性變化的情形，所以我在使用 spacy 的套件做斷詞時，我有使用 lemmatization，進行詞性的變化，例如:go、going、gone 都是走的意思，所以最後的結果都會被轉換成 go;而直接用 space 做分詞，go、going、gone 雖然都是走的意思，但卻會被對應到三種不同的詞彙。

5. (1%) 請比較 RNN 與 BOW 兩種不同 model 對於 "Today is hot, but I am happy."與"I am happy, but today is hot." 這兩句話的分數（model output），並討論造成差異的原因。

Output label:

	RNN	BOW
"Today is hot, but I am happy."	0	0
I am happy, but today is hot."	0	0

Probability:

	RNN		BOW	
label	0	1	0	1
"Today is hot, but I am happy."	1	5.48e-07	9.99e-01	1.5171e-04
I am happy, but today is hot."	1	1.741e-07	9.99e-01	1.5171e-04

(Label 0:non-malicious, Label 1:malicious)

雖然不論是 RNN 還是 BOW，他們 output 出來的 label 都會相同。但可以注意到的是，對於 BOW 而言他們 output 出來的機率是相同的，因為 BOW 沒有考慮一句話之中，文字前後順序的差別。

然而 RNN 之中，雖然 output 出來的 label 是一樣的，其中的差別主要來自於 RNN 因為有考慮到語句順序，所以機率是有差別的 5.48e-07 對 1.741e-07。

手寫作業

6.

Handwritten mathematical work on a spiral notebook, showing a table of values, matrix operations, and a final output.

Table:

$t \backslash p$	1	2	3	4	5	6	7	8
z	3	-2	4	0	2	-4	1	2
z_i	90	90	190	90	90	-10	190	90
z_f	110	10	-90	10	10	110	-90	10
z_0	-10	90	90	90	-10	90	90	90
$f(z_i)$	1	1	1	1	1	4.5	1	1
$g(z_i)$	3	-2	4	0	2	-4	1	2
$c=0$	0	3	1	4	4	6	6	1
$f(z_f)$	1	1	0	1	1	1	0	1
c'	3	1	4	4	6	6	1	3
$h(c)$	3	1	4	4	6	6	1	3
$f(z_0)$	1	1	1	1	1	1	1	1
output	2	0	4	4	4	0	6	3

Matrix Operations:

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 3 & -2 & 4 & 0 & 2 & -4 & 1 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 100 & -100 & 0 \\ 0 & 100 & -100 & 0 \\ 0 & 0 & 0 & 100 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 100 & -100 & 0 \\ -2 & 100 & 100 & 100 \\ 4 & 200 & -200 & 100 \\ 0 & 100 & -100 & 100 \\ 2 & 100 & -100 & 0 \\ -4 & 0 & 0 & 100 \\ 1 & 200 & -200 & 100 \\ 2 & 100 & -100 & 100 \end{pmatrix} + \begin{pmatrix} 0 & -10 & 110 & -10 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 90 & -10 & 90 \\ -2 & 90 & 110 & 90 \\ 4 & 140 & -90 & 90 \\ 0 & 90 & 10 & 90 \end{pmatrix} + \begin{pmatrix} 2 & 90 & 10 & -10 \\ -4 & -10 & 110 & 90 \\ 1 & 190 & -90 & 90 \\ 2 & 90 & 10 & 90 \end{pmatrix}$$

7.

$$i \in \mathcal{V} \\ j \in \mathcal{R}^N$$

$$u_c = w^T w^T x = h^T x \\ y_c = \text{softmax}(u)$$

No.
Date

$$2, \text{ Loss} = -\log \pi \frac{\exp(u_c)}{\sum_{i \in \mathcal{V}} \exp(u_i)}$$

for: $t_j = \begin{cases} 1 & j=c \\ 0 & j \neq c \end{cases}$
 t_j : the word
 1: the target word
 0: other words

$$= -\sum_{c \in \mathcal{C}} \log \frac{\exp(u_c)}{\sum_{i \in \mathcal{V}} \exp(u_i)}$$

$$= -\sum_{c \in \mathcal{C}} \log(\exp(u_c) + \sum_{i \in \mathcal{V}} \log \sum \exp(u_i))$$

$$= -\sum_{c \in \mathcal{C}} u_c + \left(\log \sum_{i \in \mathcal{V}} \exp(u_i) \right)$$

$$= \sum_{c=1}^C \left(\log \sum_{i \in \mathcal{V}} \exp(u_i) - u_c \right)$$

$$\frac{\partial \text{Loss}}{\partial u_{c,j}} = \sum_{c=1}^C \left(\frac{\exp(u_{c,j})}{\sum_{i \in \mathcal{V}} \exp(u_i)} - t_{c,j} \right)$$

$$= \sum_{c=1}^C (y_{c,j} - t_{c,j}) = - \quad (1)$$

$$u_{c,j} = \sum_{i=1}^V w_{i,j}^T h_i \quad \frac{\partial u_{c,j}}{\partial w_{i,j}^T} = h_i \quad (2)$$

$$\frac{\partial \text{Loss}}{\partial w_{i,j}^T} = \frac{\partial \text{Loss}}{\partial u_{c,j}} \frac{\partial u_{c,j}}{\partial w_{i,j}^T} = h_i \sum_{c=1}^C (y_{c,j} - t_{c,j})$$

$$h_i = \sum_{k=1}^V x_k w_{i,k}^T = \sum_{c=1}^C (y_{c,j} - t_{c,j}) \left(\sum_{k=1}^V w_{k,i} x_k \right)$$

$$\frac{\partial u_{c,j}}{\partial h_i} = \sum_{k=1}^V w_{k,i} x_k$$

Q.E.D.

No. _____
Date _____

$$U_{cjk} = \sum_{m=1}^N \sum_{l=1}^V W_{mk}^T W_{lm} X_l$$

$$\frac{\partial U_{cjk}}{\partial W_{ij}^T} = \sum_{k=1}^V \frac{\partial}{\partial W_{ij}^T} \left(\sum_{m=1}^N (W_{mk}^T W_{lm}) X_l \right)$$

$$= \sum_{k=1}^V W_{jk}^T X_i + \textcircled{5} \quad \begin{matrix} l=1 \\ m=1 \end{matrix}$$

$$\frac{\partial \text{Loss}}{\partial W_{ij}^T} = \frac{\partial \text{Loss}}{\partial U_{cjk}} \frac{\partial U_{cjk}}{\partial W_{ij}^T} \quad (\textcircled{1}, \textcircled{3}) \text{ (7)}$$

$$= \sum_{k=1}^V \sum_{c=1}^C (y_{cjk} - t_{cjk}) \cdot W_{jk}^T X_i \quad \text{Q.E.D.}$$

$$t_{cjk} = \begin{cases} 1 & k=c \\ 0 & k \neq c \end{cases}$$