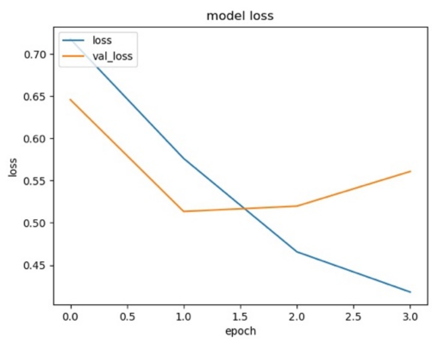
Machine Learning HW5 Report

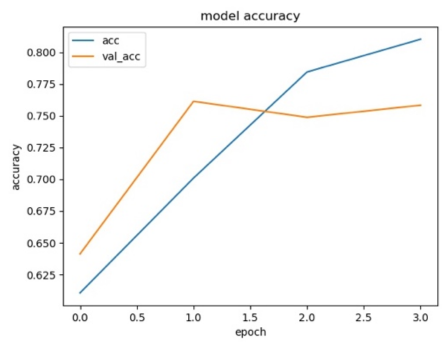
學號：r08942085 系級：電信一 姓名：陳芃彣

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

我的 RNN 模型我用了一層 hidden size 等於 256 維的 embedding，其中使用 gensim 的 Word2Vec 作為 embedding 的初始參數，然後使用 bidirectional 的 GRU，頂且使用 nn.initial weight 給他出使參數，在最後將兩個方向的 hidden state concat 起來丟進一個 fully connected layer 內。在訓練時，我發現 epoch 等於 4 或是 5 的時候會有較高的 validation accuracy，因此以下兩張圖為只有 4 個 epoch 的結果。

最好的成績為：private: 0.82325, public: 0.78139

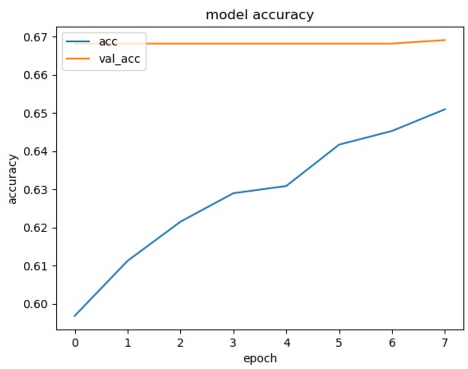


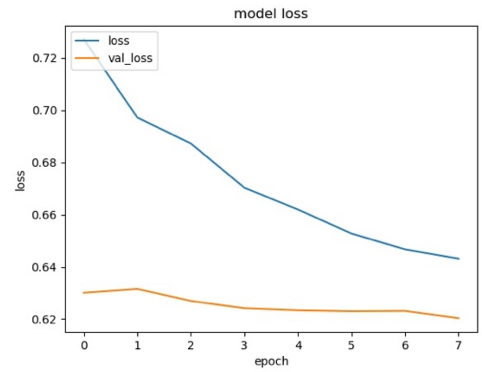


2. (1%) 請實作 BOW+DNN 模型，敘述你的模型架構，回報模型的正確率並繪出訓練曲線*。

我使用 gensim 的 corpora 來做 BOW，首先先使用 gensim.utils 的 simple_preprocess 做 tokenize，然後取出每個字在 dictionary 的 index，再用 gensim corpora 的 doc2bow 換算每一句話字的出現頻率，然後我使用 for loop 建立 bag of word，最後丟到一層 fully connected layer，activation 使用 ReLu。

private: 0.73255, public: 0.70930





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。

   在前處理的部份，起初我只做了單一個 ”@user” 的 replace，算出最長句子的長度，再將每一句話都 padding 成該長度（每個 padding 我都是加 ”<pad>”），然後我使用 gensim 的 Word2Vec 換算出 training data + testing data 中每個字的向量，可以做為之後 train embedding 的初始值，其中我在字典最後加了一個屬於 “other word“ 的字，當 data 中出現字典裡沒有的字時 (ex 被前處理掉的語助詞等等)，就會使用 other word 的向量表示。
   我的改善方式是用 spacy 的 ”en_core_web_sm” 做為 tokenizer（本來僅用 split），判斷了 stopping words，以及使用 lemma。
   模型進步的部分我加了 initial weight，將 learning 調大成 0.001，epoch 只取 4，hidden size = 256，以及 Word2Vec 的字詞維度調成 300。

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。

   兩者皆是有做斷詞的效果好一點點，但沒有很顯著的差別，也許是因為我前處理做得不夠好，可能包含一些不意義的標點符號或是 hashtag 或是韓文，因此就算做斷詞也會學到一些不必要的文字。

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。

"Today is hot, but I am happy."

RNN: 0.4578　　　　　　BOW: 0.5023

"I am happy, but today is hot."

RNN: 0.4096　　　　　　BOW: 0.5112

其實這兩句話好像並沒有一個很正面一個很負面，因此 RNN 與 BOW 出來的結果都偏中立，或者是我的兩個 model 沒有 train 好，對 BOW 來說分數比較接近，可能是因為 BOW 不考慮詞的順序，因此兩句話對 BOW 來說應該很相近。

1. LSTM

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

$z=wx+b$ $z_f=w_f x+b_f$ $c'=f(z_i)g(z)+cf(z_f)$

$z_i=w_i x+b_i$ $z_o=w_o x+b_o$

$w=[0,0,0,1]$ $b=0$

$w_i=[100,100,0,0]$ $b_i=-10$

$w_f=[-100,-100,0,0]$ $b_f=110$

$w_o=[0,0,100,0]$ $b_o=-10$

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

t=1 : $z=3+0=3$

$z_i=100-10=90$

$z_f=-100+110=10$

$z_o=0-10=-10$

$c'=\frac{1}{1+e^{-90}}\cdot 3+c\cdot\frac{1}{1+e^{-10}}=3$ (C initial = 0)

$y_1=f(z_o)h(c')=f(-10)h(3)=0$

t=2 : $z=-2+0=-2$

$z_i=100-10=90$

$z_f=-100+110=10$

$z_o=100-10=90$

$c'=f(90)g(-2)+3f(10)=1$

$y_2=f(90)h(1)=1$

t=3 : $z=4$

$z_i=100+100-10=190$

$z_f=-200+110=-90$

$z_o=100-10=90$

$c'=f(190)g(4)+1\cdot f(-90)=4$

$y_3=f(90)h(4)=4$

t=4 : $z=0$

$z_i=100-10=90$

$z_f=-100+110=10$

$z_o=100-10=90$

$c'=f(90)g(0)+4\cdot f(10)=4$

$y_4=f(90)\cdot h(4)=4$

t=5 : $z=2$

$z_i=100-10=90$

$z_f=-100+110=10$

$z_o=-10$

$c'=f(90)g(2)+4\cdot f(10)=6$

$y_5=f(-10)\cdot h(6)=0$

t=6 : $z=-4$

$z_i=0-10=-10$

$z_f=0+100=110$

$z_o=100-10=90$

$c'=f(-10)g(-4)+6\cdot f(110)=6$

$y_6=f(90)h(6)=6$

t=7 : $z=1$

$z_i=200-10=190$

$z_f=-200+110=-90$

$z_o=100-10=90$

$c'=f(190)g(1)+6\cdot f(-90)=1$

$y_7=f(90)h(1)=1$

t=8 : $z=2$

$z_i=90$

$z_f=10$

$z_o=90$

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

$y_8=f(90)h(3)=3$

$\therefore$ 0, 1, 4, 4, 0, 6, 1, 3 #

2. $h = W^T x$

$u = W'^T h$

$y = \text{Softmax}(u) = \text{Softmax}(W'^T W^T x)$

$\text{Loss} = L = -\log \prod_{c \in C} P(w_{output,c} \,|\, w_{input}) = -\log \prod_{c \in C} \frac{\exp(u_c)}{\sum_{i \in V} \exp(u_i)}$

calculate $\frac{\partial L}{\partial W^T_{ij}}$ $\frac{\partial L}{\partial W^{T'}_{ij}}$

<sol> $\text{Loss} = -\log \prod_{c=1}^{C} \frac{\exp(u_{c,j^*})}{\sum_{j=1}^{V} \exp(u_{c,j})} = -\sum_{c=1}^{C} u_{c,j^*} + \sum_{c=1}^{C} \log \sum_{j=1}^{V} \exp(u_{c,j})$

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial W'_{ij}} = \sum_{k=1}^{V} \sum_{c=1}^{C} \frac{\partial \mathcal{L}}{\partial u_{c,k}} \frac{\partial u_{c,k}}{\partial W'_{ij}} \\ \frac{\partial \mathcal{L}}{\partial W_{ij}} = \sum_{k=1}^{V} \sum_{c=1}^{C} \frac{\partial \mathcal{L}}{\partial u_{c,k}} \frac{\partial u_{c,k}}{\partial W_{ij}} \end{cases}$$

$$\frac{\partial \mathcal{L}}{\partial u_{c,j}} = -\delta_{jj^*_c} + y_{c,j} \doteq e_{c,j}$$

$$\therefore \Rightarrow \frac{\partial \mathcal{L}}{\partial W'_{ij}} = \sum_{c=1}^{C} (-\delta_{jj^*_c} + y_{c,j}) \left(\sum_{k=1}^{V} W_{ki} x_k\right) \quad \#$$

$$\frac{\partial \mathcal{L}}{\partial W_{ij}} = \sum_{k=1}^{V} \sum_{c=1}^{C} (-\delta_{kk^*_c} + y_{c,k}) W'_{jk} x_i \quad \#$$