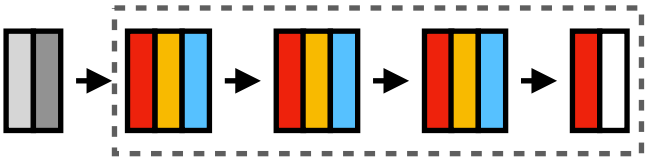
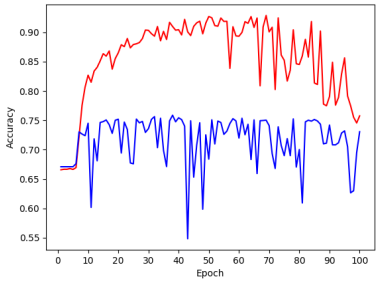
ML Homework #5 學號：B0902120 系級： 資工四 姓名：曾鈺婷

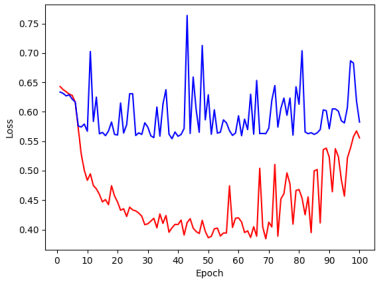
1. 請說明你實作之 RNN 模型架構及使用的 word embedding 方法，回報模型的正確率並繪出訓練曲線。

淺灰色：Embedding ( )
深灰色：LSTM (input_size = 100, hidden_size = 100, num_layers = 2, dropout = 0.3)
虛線　：Classifier ( )
紅色是 Linear　　黃色是 ReLU　　藍色是 Dropout (0.5)　　白色是 Softmax



word embedding 的部分，首先建立字典、編號、與 vector 的關係（只記錄有出現過的字，以減少儲存所需的空間），再將 comment 的內容從 words 轉換為『編號』表示。下圖分別是 accuracy 與 loss，藍色為 training 的結果，而紅色則是 validation 的結果。

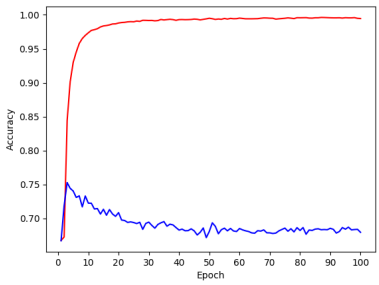


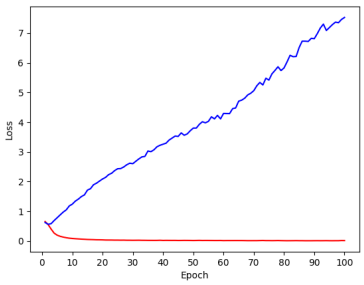


2. 請實作 BOW + DNN 模型，敘述你的模型架構，回報模型的正確率並繪出訓練曲線。

紅色是 Linear　　黃色是 ReLU

bag of words 的部分，首先建立字典、編號、與出現次數的關係，再將 comment 的內容用全部 words 的 vector 表示（出現的 word 為『出現次數』）。下圖分別是 accuracy 與 loss，藍色為 training 的結果，而紅色則是 validation 的結果 。





3. 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。

除了最一般的 tokenize，我首先進行的是 lemmatization，將不同型態的 words 還原成最基礎的型態，例如 running 或 runs 還原為 run，減少他被視為不同字或相差甚遠 vector 的機率，也可以減少 vocabulary 的數量。再來，由於這些是從 twitter 上的句子，會有 @user、URL 等等出現，因此我有去掉這些不必要的資訊，以免造成混淆。最後，我有做移除 stopwords 的動作，像是 the、in、a 等等詞語，其實對語句是沒有任何意義的，但他們出現的次數卻算多次，所以我把它們移除，避免影響模型的準確率。

4. 請比較不做斷詞（例如：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。

有做斷詞會忽略連續空格的情況，並且有能有型態還原、去除標點符號等工作，而無斷詞的話，則會因為有過多不重要資訊，進而造成準確率較低。

	Public Score	Private Score
不做斷詞 BOW + DNN	0.75348	0.79767
不做斷詞 RNN	0.75581	0.81395
有做斷詞 BOW + DNN	0.76511	0.80465
有做斷詞 RNN	0.78604	0.82093

5. 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。

在 RNN 的時候，有考慮 word 的順序，因此語序會影響他的輸出結果；然而 BOW + DNN 的部分只在乎 word 的出現與否，最終得到相同的結果。

BOW + DNN	Inoffensive	Offensive
Sentence 1	0.649308265271134	0.346785302032181
Sentence 2	0.649308265271134	0.346785302032181
RNN	**Inoffensive**	**Offensive**
Sentence 1	0.967632577250912	0.155093212956762
Sentence 2	0.880070716606048	0.297318453212775

6. 手寫題

1. LSTM Cell

分別計算 y 在不同 t 的情況。

$t = 1$

$$z = wx^1 + b = (0,\ 0,\ 0,\ 1)(0,\ 1,\ 0,\ 3) + 0 = 3$$
$$z_i = w_i x^1 + b_i = (100,\ 100,\ 0,\ 0)(0,\ 1,\ 0,\ 3) - 10 = 90$$
$$z_f = w_f x^1 + b_f = (-100,\ -100,\ 0,\ 0)(0,\ 1,\ 0,\ 3) + 110 = 10$$
$$z_o = w_o x^1 + b_f = (0,\ 0,\ 100,\ 0)(0,\ 1,\ 0,\ 3) - 10 = -10$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-90}} * 3 + 0 * \frac{1}{1+e^{-10}} = 3$$
$$y_1 = f(z_o)h(c') = \frac{1}{1+e^{-(-10)}} * 3 = 0$$

$t = 2$

$$z = wx^2 + b = (0,\ 0,\ 0,\ 1)(1,\ 0,\ 1,\ -2) + 0 = -2$$
$$z_i = w_i x^2 + b_i = (100,\ 100,\ 0,\ 0)(1,\ 0,\ 1,\ -2) - 10 = 90$$
$$z_f = w_f x^2 + b_f = (-100,\ -100,\ 0,\ 0)(1,\ 0,\ 1,\ -2) + 110 = 10$$
$$z_o = w_o x^2 + b_f = (0,\ 0,\ 100,\ 0)(1,\ 0,\ 1,\ -2) - 10 = 90$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-90}} * -2 + 3 * \frac{1}{1+e^{-10}} = 1$$
$$y_2 = f(z_o)h(c') = \frac{1}{1+e^{-90}} * 1 = 1$$

$t = 3$

$$z = wx^3 + b = (0,\ 0,\ 0,\ 1)(1,\ 1,\ 1,\ 4) + 0 = 4$$
$$z_i = w_i x^3 + b_i = (100,\ 100,\ 0,\ 0)(1,\ 1,\ 1,\ 4) - 10 = 190$$
$$z_f = w_f x^3 + b_f = (-100,\ -100,\ 0,\ 0)(1,\ 1,\ 1,\ 4) + 110 = -90$$
$$z_o = w_o x^3 + b_f = (0,\ 0,\ 100,\ 0)(1,\ 1,\ 1,\ 4) - 10 = 90$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-190}} * 4 + 1 * \frac{1}{1+e^{-(-90)}} = 4$$
$$y_3 = f(z_o)h(c') = \frac{1}{1+e^{-90}} * 4 = 4$$

$t = 4$

$$z = wx^4 + b = (0,\ 0,\ 0,\ 1)(0,\ 1,\ 1,\ 0) + 0 = 0$$
$$z_i = w_i x^4 + b_i = (100,\ 100,\ 0,\ 0)(0,\ 1,\ 1,\ 0) - 10 = 90$$
$$z_f = w_f x^4 + b_f = (-100,\ -100,\ 0,\ 0)(0,\ 1,\ 1,\ 0) + 110 = 10$$
$$z_o = w_o x^4 + b_f = (0,\ 0,\ 100,\ 0)(0,\ 1,\ 1,\ 0) - 10 = 90$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-90}} * 0 + 4 * \frac{1}{1+e^{-10}} = 4$$
$$y_4 = f(z_o)h(c') = \frac{1}{1+e^{-90}} * 4 = 4$$

$t = 5$

$$z = wx^5 + b = (0,\ 0,\ 0,\ 1)(0,\ 1,\ 0,\ 2) + 0 = 2$$
$$z_i = w_i x^5 + b_i = (100,\ 100,\ 0,\ 0)(0,\ 1,\ 0,\ 2) - 10 = 90$$

$$z_f = w_f x^5 + b_f = (-100,\ -100,\ 0,\ 0)(0,\ 1,\ 0,\ 2) + 110 = 10$$
$$z_o = w_o x^5 + b_f = (0,\ 0,\ 100,\ 0)(0,\ 1,\ 0,\ 2) - 10 = -10$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-90}} * 2 + 4 * \frac{1}{1+e^{-10}} = 6$$
$$y_5 = f(z_o)h(c') = \frac{1}{1+e^{-(-10)}} * 6 = 0$$

$t = 6$

$$z = wx^6 + b = (0,\ 0,\ 0,\ 1)(0,\ 0,\ 1,\ -4) + 0 = -4$$
$$z_i = w_i x^6 + b_i = (100,\ 100,\ 0,\ 0)(0,\ 0,\ 1,\ -4) - 10 = -10$$
$$z_f = w_f x^6 + b_f = (-100,\ -100,\ 0,\ 0)(0,\ 0,\ 1,\ -4) + 110 = 110$$
$$z_o = w_o x^6 + b_f = (0,\ 0,\ 100,\ 0)(0,\ 0,\ 1,\ -4) - 10 = 90$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-(-10)}} * (-4) + 6 * \frac{1}{1+e^{110}} = 6$$
$$y_6 = f(z_o)h(c') = \frac{1}{1+e^{-90}} * 6 = 6$$

$t = 7$

$$z = wx^7 + b = (0,\ 0,\ 0,\ 1)(1,\ 1,\ 1,\ 1) + 0 = 1$$
$$z_i = w_i x^7 + b_i = (100,\ 100,\ 0,\ 0)(1,\ 1,\ 1,\ 1) - 10 = 190$$
$$z_f = w_f x^7 + b_f = (-100,\ -100,\ 0,\ 0)(1,\ 1,\ 1,\ 1) + 110 = -90$$
$$z_o = w_o x^7 + b_f = (0,\ 0,\ 100,\ 0)(1,\ 1,\ 1,\ 1) - 10 = 90$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-190}} * 1 + 6 * \frac{1}{1+e^{-(-90)}} = 1$$
$$y_7 = f(z_o)h(c') = \frac{1}{1+e^{-90}} * 1 = 1$$

$t = 8$

$$z = wx^8 + b = (0,\ 0,\ 0,\ 1)(1,\ 0,\ 1,\ 2) + 0 = 2$$
$$z_i = w_i x^8 + b_i = (100,\ 100,\ 0,\ 0)(1,\ 0,\ 1,\ 2) - 10 = 90$$
$$z_f = w_f x^8 + b_f = (-100,\ -100,\ 0,\ 0)(1,\ 0,\ 1,\ 2) + 110 = 10$$
$$z_o = w_o x^8 + b_f = (0,\ 0,\ 100,\ 0)(1,\ 0,\ 1,\ 2) - 10 = 90$$
$$c' = f(z_i)g(z) + cf(z_f) = \frac{1}{1+e^{-90}} * 2 + 1 * \frac{1}{1+e^{-10}} = 3$$
$$y_8 = f(z_o)h(c') = \frac{1}{1+e^{-90}} * 3 = 3$$

2. Word Embedding