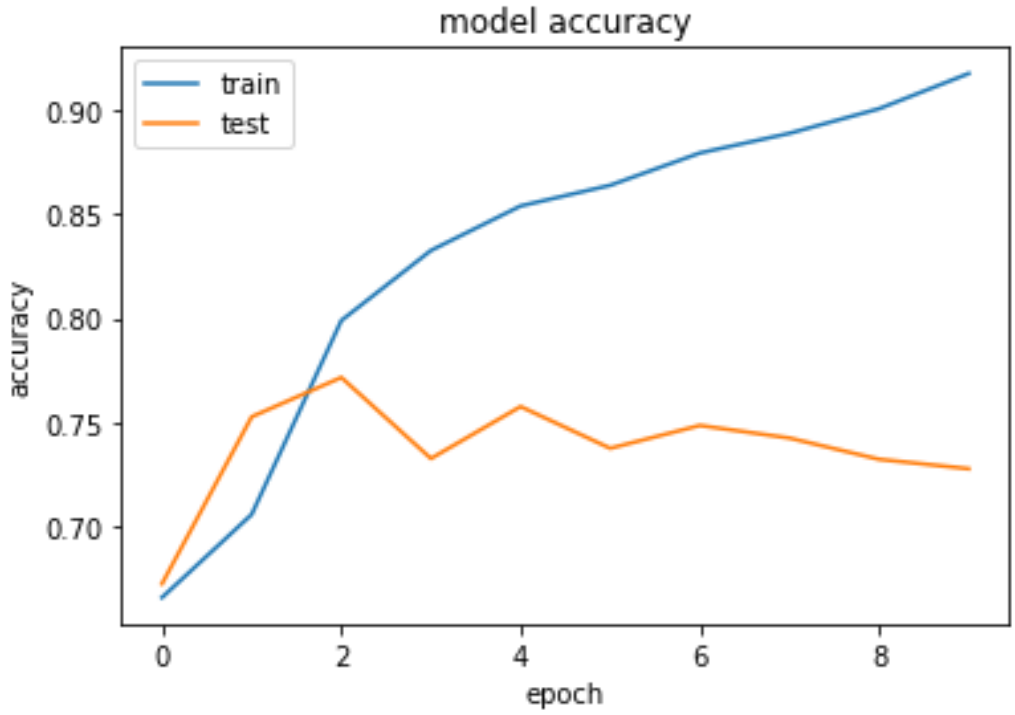
Machine Learning HW5 Report

學號：R06522709　系級：機械碩三　　姓名：鄭呈毅

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*

   我是使用 keras 的 embedding 層來實作 word embedding。因此在資料進入 model 前，會先使用 keras 的 tokenizer 來將每個詞轉換為一個對應的 token，而每個句子的長度最多為 50 個詞。模型如下:

   embedding layer->LSTM(512)->LSTM(256)->LSTM(256)->Dense(128)->Dense(64)->Dense(32)->Dense(1)



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*。

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。

   我一開始先用這個模型，然後 epoch 設定 10 次，可是成績不是很理想，僅約 0.75。所以我就試著先用 gensim 做 word embedding 之後再丟入 LSTM，但可能是 implementation 有誤，每個 epoch 的 train 跟 valid accuracy 都上升到某個特定值(取決於 hyper parameter)之後就不會再改變了。絕望之際我打開了沒有 gensim，僅採用 embedding layer 的模型，把 epoch 由 10 降為 4，結果就通過 strong baseline，真是可喜可賀。從結果來看，單獨採用 embedding layer 的效果並不會比 gensim

等套件來得差；而另一方面，epoch 降低所導致準確率的提升，也顯示了 early stopping 的重要性。

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。
   以我目前使用的這個 model 來說，如果我不做斷詞，tokenizer 會將整個句子視為一個單字，也就是說當我將資料轉為 sequence 時，每一筆 training data 會變成只有一個數字的 sequence，而由於 tokenizer 沒看過 testing data，因此會出現 testing data 沒有對應到的 sequence 的窘境，這會使準確度變得很低，只有約 0.3 而已。
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。

手寫題(助教抱歉，因光源的關係導致圖片有點模糊)

MLHW5 R06522709 鄭呈毅

1. $w=[0,0,0,1]$ $b=0$.
$w_1=[100,100,0,0]$, $b_1=-10$.
$w_f=[-100,-100,0,0]$, $b_f=110$.
$w_0=[0,0,100,0]$, $b_0=-10$.

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

t=1
$z=3+0=3$
$z_1=100-10=90$
$z_f=-100+110=10$
$z_0=-10$

$c'=f(90)g(3)+0\cdot f(10)$
$=1\times 3+0=3$
$y_1=f(-10)h(3)=0$.

t=2.
$z=-2+0=-2$
$z_1=100-10=90$
$z_f=-100+110=10$
$z_0=100-10=90$.

$c'=f(90)g(-2)+3f(10)$
$=-2+3=1$
$y_2=f(90)h(1)=1$.

t=3.
$z=4+0=4$.
$z_1=100+100-10=190$.
$z_f=-200+110=-90$.
$z_0=100-10=90$

$c'=f(190)g(4)+1\cdot f(-90)$
$=4$.
$y_3=f(90)h(4)=4$ #

t=4
$z=0$
$z_1=100-10=90$
$z_f=-100+110=10$
$z_0=100-10=90$

$c'=f(90)g(0)+4\times f(10)$
$=4$.
$y_4=f(90)\cdot h(4)=4$.

$t=5$ $\begin{cases} z=2 \\ z_1 = 100-10 = 90 \\ z_f = -100+110 = 10 \\ z_0 = -10 \end{cases}$

$c' = f(90)\,g(2) + 4 \times f(10)$

$= 2+4 = 6$

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

$t=6$ $\begin{cases} z=-4 \\ z_1 = 0-10 = -10 \\ z_f = 0+110 = 110 \\ z_0 = 100-10 = 90 \end{cases}$

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

$= 0+6 = 6$

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

$t=7$ $\begin{cases} z=1 \\ z_1 = 200-10 = 190 \\ z_f = -100+110 = 90 \\ z_0 = 100-10 = 90 \end{cases}$

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

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

$t=8$ $\begin{cases} z=2 \\ z_1 = 100-10 = 90 \\ z_f = -100+110 = 10 \\ z_0 = 100-10 = 90 \end{cases}$

$c' = f(90)\,g(2) + 1 \times f(10) = 2+1 = 3$

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

$y_t = 0, 1, 4, 4, 0, 6, 1, 3$

2.

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

$$= -\log\left[\frac{\exp(u_1+u_2+u_3+\dots)}{\left(\sum_{i \in V} \exp(u_i)\right) \times \dots}\right]$$

$$= -\sum_{c=1}^{C} u_{cj^*} + \sum_{c=1}^{C} \log \sum_{i=1}^{V} \exp(u_{c,i})$$

kroneck delta
$= 1$, if $i = j^*$
$= 0$, otherwise

$$\frac{\partial L}{\partial w'_{ij}} = \sum_{k=1}^{V} \sum_{c=1}^{C} \frac{\partial L}{\partial u_{ck}} \frac{\partial u_{ck}}{\partial w'_{ij}} = \sum_{c=1}^{C} (-\delta_{ij_c^*} + y_{ij}) \left(\sum_{k=1}^{V} w_{ki} x_k\right)$$

$$\frac{\partial L}{\partial w_{ij}} = \sum_{k=1}^{V} \sum_{c=1}^{C} \frac{\partial L}{\partial u_{c,k}} \frac{\partial}{\partial w_{ij}} \left(\sum_{m=1}^{N} \sum_{l=1}^{V} w'_{mk} w_{lm} x_l\right)$$

$$= \sum_{k=1}^{V} \sum_{c=1}^{C} (-\delta_{kk_c^*} + y_{c,k}) w'_{jk} x_i \quad *$$