Machine Learning HW5 Report

學號：R08921A01　系級：電機碩一　姓名：陳允中

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*

   RNN 模型架構:

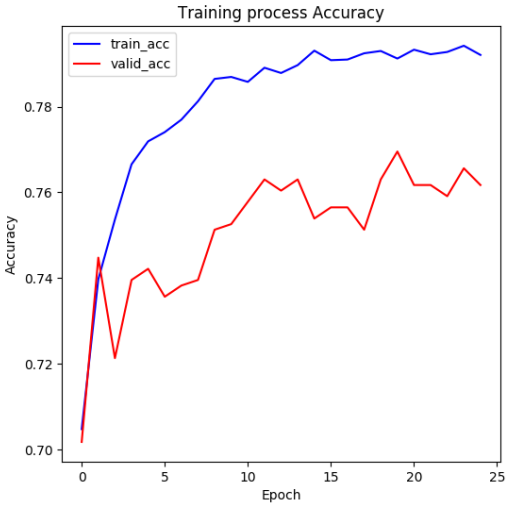
   一層 LSTM (參數: hidden_size=vector_space, n_layers=6, bidirectional=True, dropout=0.5)

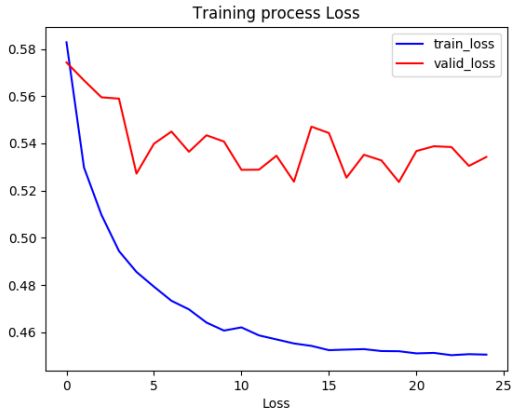
   一層 Linear 降到 2 維後過 LogSoftmax

   word embedding 方法:

   使用 gensim.models.Word2Vec (參數:size=25, window=5, min_count=1)

   Public score: 0.78604





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*。

   BOW 總共有 5205 個 words，使用 tfidf

   nn.Linear(5205,2048),

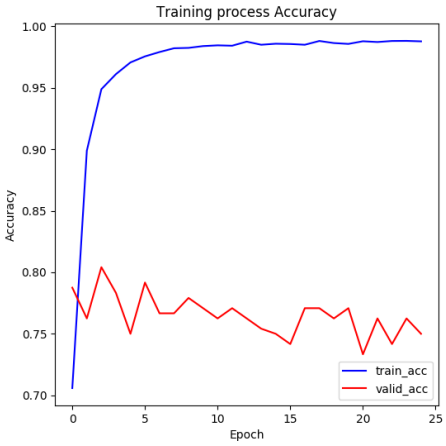
   nn.Linear(2048,1024),

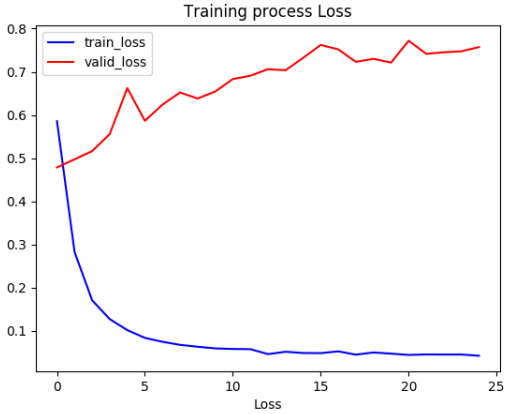
   nn.Linear(1024,512),

   nn.Linear(512,2),

   nn.LogSoftmax(dim=1)

   Public score: 0.77209





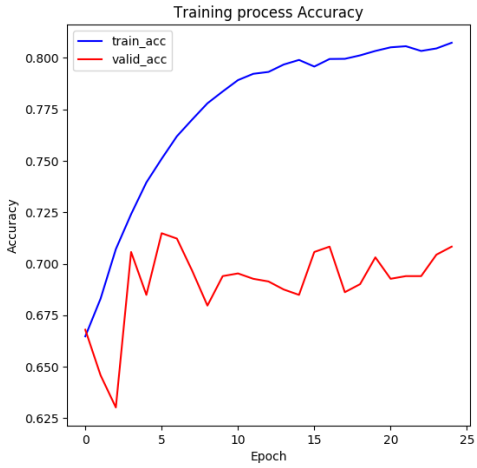
3. (1%) 請敘述你如何 improve performance（preprocess, embedding, 架構等），並解釋為何這些做法可以使模型進步。

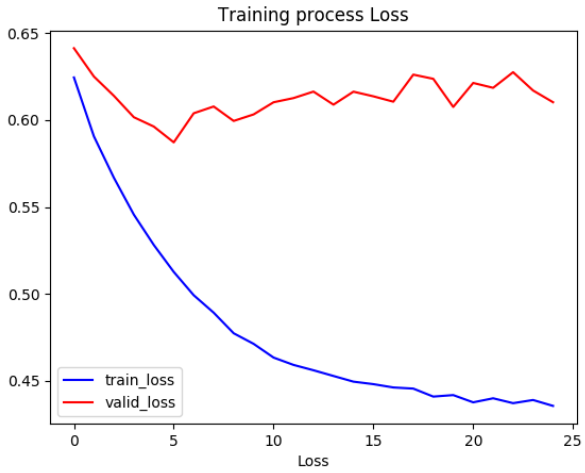
   LSTM 的 n_layers 太大會 train 不起來，最後選擇為 6

   embedding 的維度不是越大越好，太高會讓 validation score 下降，最後選擇為 25，原因是維度太大無法確實表示單字之間的關係，一堆沒意義的數字反而會導致 overfit

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。

   不做斷詞會讓 word 的數量增加，且有些會是沒有意義的，那些無意義的單字會導致訓練學到的東西只是單純湊出來的，對 testing 沒有幫助





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."	"I am happy, but today is hot."
RNN	(-0.062, -2.810)	(-0.066, -2.750)
BOW	(-0.009, -4.687)	(-0.009, -4.687)

   文字順序不影響 BOW 產生的 input，這 2 個輸入對於 BOW 是一樣的

1. LSTM Cell

$x_1 = (0,1,0,3)$, $C_1 = 0$, $g(z_1) = 3$, $f(z_1^i) = f(90)$, $f(z_1^o) = f(-10)$

$C_1' = f(90) \times 3 + 0 \times f(z_1^f) = 3$, $y_1 = f(-10) \times 3 = 0$ #

$x_2 = (1,0,1,-2)$, $C_2 = 3$, $g(z_2) = -2$, $f(z_2^i) = f(90)$, $f(z_2^o) = f(90)$

$C_2' = f(90) \times -2 + 3 \times f(\underset{10}{z_2^f}) = -2+3 = 1$, $y_2 = f(90) \times 1 = 1$ #

$x_3 = (1,1,1,4)$, $C_3 = 1$, $g(z_3) = 4$, $f(z_3^i) = f(190)$, $f(z_3^o) = f(90)$

$C_3' = f(190) \times 4 + 1 \times f(\underset{-90}{z_3^f}) = 4+0 = 4$, $y_3 = f(90) \times 4 = 4$ #

$x_4 = (0,1,1,0)$, $C_4 = 4$, $g(z_4) = 0$, $f(z_4^i) = f(90)$, $f(z_4^o) = f(90)$

$C_4' = f(90) \times 0 + 4 \times f(\underset{10}{z_4^f}) = 0+4 = 4$, $y_4 = f(90) \times 4 = 4$ #

$x_5 = (0,1,0,2)$, $C_5 = 4$, $g(z_5) = 2$, $f(z_5^i) = f(90)$, $f(z_5^o) = f(-10)$

$C_5' = f(90) \times 2 + 4 \times f(\underset{10}{z_5^f}) = 2+4 = 6$, $y_5 = f(-10) \times 6 = 0$ #

$x_6 = (0,0,1,-4)$, $C_6 = 6$, $g(z_6) = -4$, $f(z_6^i) = f(-10)$, $f(z_6^o) = f(90)$

$C_6' = f(-10) \times -4 + 6 \times f(\underset{110}{z_6^f}) = 0+6 = 6$, $y_6 = f(90) \times 6 = 6$ #

$x_7 = (1,1,1,1)$, $C_7 = 6$, $g(z_7) = 1$, $f(z_7^i) = f(190)$, $f(z_7^o) = f(90)$

$C_7' = f(190) \times 1 + 6 \times f(\underset{-90}{z_7^f}) = 1+0 = 1$, $y_7 = f(90) \times 1 = 1$ #

$x_8 = (1,0,1,2)$, $C_8 = 1$, $g(z_8) = 2$, $f(z_8^i) = f(90)$, $f(z_8^o) = f(90)$

$C_8' = f(90) \times 2 + 1 \times f(\underset{10}{z_8^f}) = 2+1 = 3$, $y(8) = f(90) \times 3 = 3$ #

2. Word Embedding

$$\frac{\partial L}{\partial w'_{ij}} = \sum_{k=1}^{V}\sum_{c=1}^{C} \frac{\partial L}{\partial u_{c,k}} \cdot \frac{\partial u_{c,k}}{\partial w'^{T}_{ij}}$$

$$\frac{\partial L}{\partial u_{c,k}} = -\delta_{k k_c^*} + y_{c,k} \quad \cdot \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

k_c^* 是 w_c one-hot encoding 為 1 的位置

$$\frac{\partial u_{c,k}}{\partial w'^{T}_{ij}} = h_i = \sum_{k=1}^{V} w_{ik} x_k$$

$$\frac{\partial L}{\partial w'^{T}_{ij}} = \sum_{c=1}^{C} \left(-\delta_{k k_c^*} + y_{c,k}\right) \cdot \sum_{k=1} w_{ki} x_k \quad ※$$

$$\frac{\partial L}{\partial w^{T}_{ij}} = \sum_{k=1}^{V}\sum_{c=1}^{C} \frac{\partial L}{\partial u_{c,k}} \cdot \frac{\partial u_{c,k}}{\partial w^{T}_{ij}}$$

$$u_{c,k} = \sum_{m=1}^{N}\sum_{l=1}^{V} w'_{mk} w_{lm} x_l$$

$$\frac{\partial u_{c,k}}{\partial w^{T}_{ij}} = w'_{jk} x_i$$

$$\frac{\partial L}{\partial w^{T}_{ij}} = \sum_{k=1}^{V}\sum_{c=1}^{C} \left(-\delta_{k k_c^*} + y_{c,k}\right) w'_{jk} x_i \quad ※$$