Machine Learning HW5 Report

學號：T08902109　系級：資工系　姓名：賈成銪

(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 模型主要使用兩層的雙向 LSTM+DNN 實現，最終輸入層的維度為 300，序列長度為 120，具體結構見下圖：

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 120, 300)	5249400
bidirectional_1 (Bidirection	(None, 120, 256)	439296
bidirectional_2 (Bidirection	(None, 128)	164352
batch_normalization_1 (Batch	(None, 128)	512
dense_1 (Dense)	(None, 64)	8256
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 32)	2080
dropout_2 (Dropout)	(None, 32)	0
dense_3 (Dense)	(None, 2)	66

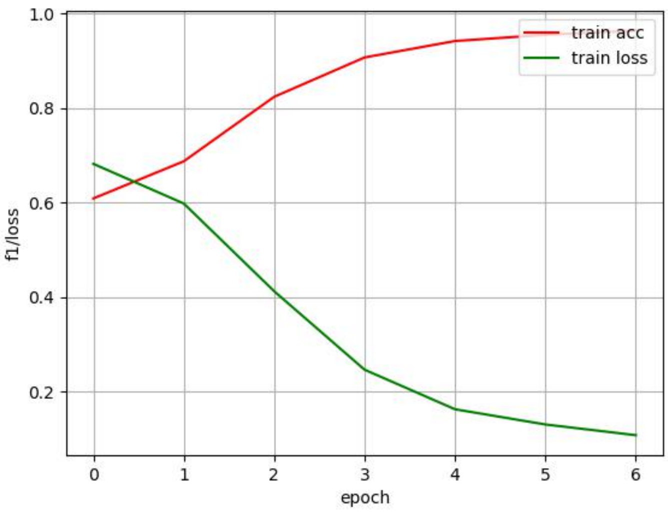
Word embedding 部分使用 gensim.models.Word2Vec，維度取 300，具體參數見下圖:

```
w2v_model = gensim.models.Word2Vec(data, size=300, window=5, min_count=0, workers=8)
w2v_model.save('./wordEmbed')
```

F1_score:

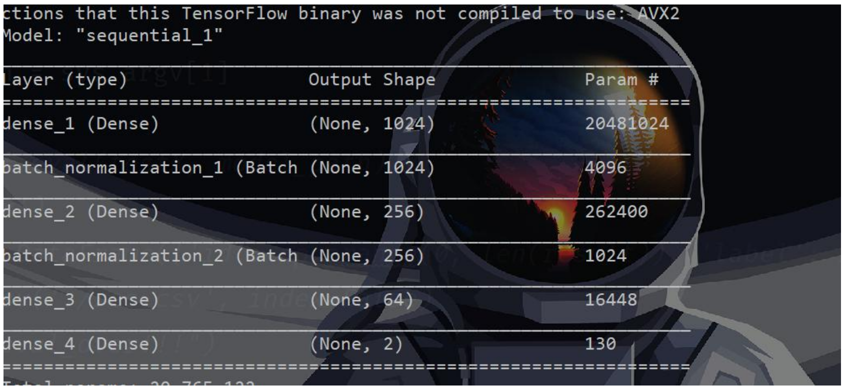
Private : 0.83488　Public: 0.77674

訓練曲線：



(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+DNN 模型我取了頻率最高的 20000 個詞進行 one_hot_encode，DNN 部分具體架構見下圖：

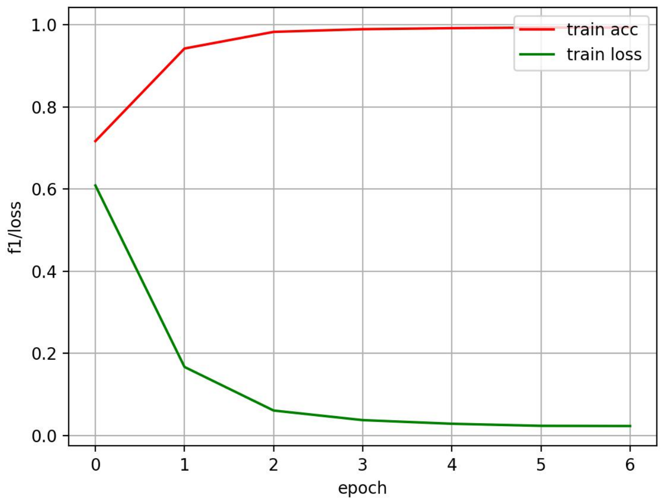


```
ctions that this TensorFlow binary was not compiled to use: AVX2
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 1024)              20481024
_________________________________________________________________
batch_normalization_1 (Batch (None, 1024)              4096
_________________________________________________________________
dense_2 (Dense)              (None, 256)               262400
_________________________________________________________________
batch_normalization_2 (Batch (None, 256)               1024
_________________________________________________________________
dense_3 (Dense)              (None, 64)                16448
_________________________________________________________________
dense_4 (Dense)              (None, 2)                 130
=================================================================
```

F1_score:

Private: 0.81162　Public:0.73953

訓練曲線：



(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:

數據清洗部分我先解析了 emoji，然後使用 spacy 的處理，去除標點符號，空格和停用詞（具體見下圖）。解析 emoji 可以保留跟多信息，去除標點符號，空格和停用詞，可以去

除文本的多餘無用信息。

```python
data = pd.read_csv(data_path)['comment'].values
data = [emoji.demojize(seg) for seg in data]
data = [seg.replace('@user', ' ') for seg in data]
nlp = spacy.load('en_core_web_sm')
```

```python
if (not token or not token.string.strip() or
    token.is_stop or token.is_punct):
        return False
```

Embedding：

embedding 部分使用 gensim.models.Word2Vec，主要通過使用不同的參數檢測結果的分數。最終使用了（1）中的參數。

架構：

RNN 部分我最早使用的是單向的 LSTM，後面改為雙向的 LSTM，獲得了較好的改進。

使用雙向的 LSTM 能使模型對前後語義理解更充分，結果也會獲得比較大的提升。

(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。

空格分開：public f1_score: 0.75418

Spacy('en_core_web_sm')斷詞： public f1_score: 0.77906

使用套件斷詞會使結果好一些，原因可能是套件在斷詞時會去掉一些無用的語氣詞，使數據的無用信息進一步減少。進而提升結果的分數。

(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。

RNN 分數：

```
[[0.8983901  0.10160995]
 [0.7675597  0.23244028]]
```

BOW 分數：

```
[[9.9828124e-01 1.7187380e-03]
 [9.9978095e-01 2.1904591e-04]]
```

可以看出當使用 BOW model 時，兩個句子的得分基本沒有差別。

但使用 RNN 時，這兩個不同句子就出現了差別。

說明 RNN 可以識別句子中單詞的順序，這個順序會對 model 的結果產生影響。但 BOW 則只能識別單詞，雖然單詞順序不同，但只要單詞相同，output 的結果就一樣。

1. $t=1$: $z = wx+b = 3$ $\qquad g(z) = g(3) = 3$

$z^i = w_i x + b = 90 \qquad f(z^i) = f(90) = 1$

$z^f = w_f x + b = 10 \qquad f(z^f) = f(10) = 1$

$c = c\,f(z^f) + f(z^i) \cdot g(z)$

$= 0 \times 1 + 1 \times 3 = 3$

$z^o = w_o x + b = -10 \qquad f(z^o) = f(-10) = 0$

$h(c) = h(3) = 3$

$y = h(c) \cdot f(z^o) = 3 \times 0 = 0$

thus $t=1$ $\quad y=0$ $\quad$ now $c=3$

$t=2$: $z = wx+b = -2 \qquad g(z) = g(-2) = -2$

$z^i = w_i x + b = 90 \qquad f(z^i) = f(90) = 1$

$z^f = w_f x + b = 10 \qquad f(z^f) = f(10) = 1$

$c = c\,f(z^f) + f(z^i) \cdot g(z)$

$= 3 \times 1 - 2 \times 1 = 1$

$z^o = w_o x + b = 90 \qquad f(z^o) = f(90) = 1$

$h(c) = h(1) = 1$

$y = h(c) \cdot f(z^o) = 1$ $\quad$ now $c=1$

$t=3$: $z = wx+b = 4 \qquad g(z) = g(4) = 4$

$z^i = w_i x + b = 190 \qquad f(z^i) = f(190) = 1$

$z^f = w_f x + b = -90 \qquad f(z^f) = f(-90) = 0$

$c = c\,f(z^f) + f(z^i) \cdot g(z)$

$= 1 \times 0 + 4 \times 1 = 4$

$z^o = w_o x + b = 90 \qquad f(z^o) = f(90) = 1$

$h(c) = h(4) = 4$

$y = h(c) \cdot f(z^o) = 4$ $\quad$ now $c=4$

$t = 4:$ $z = wx + b = 0$ $g(z) = g(0) = 0$

$z^i = w_i x + b = 90$ $f(z^i) = f(90) = 1$

$z^f = w_f x + b = -10$ $f(z^f) = f(-10) = 0$

$c = c f(z^f) + f(z^i) \cdot g(z)$

$= 4 \times 0 + 0 \times 1 = 0$

$z^o = w_o x + b = 90$ $f(z^o) = f(90) = 1$

$h(c) = h(0) = 0$

$y = h(c) \cdot f(z^o) = 0$ now $c = 0$

$t = 5:$ $z = wx + b = 2$ $g(z) = g(2) = 2$

$z^i = w_i x + b = 90$ $f(z^i) = f(90) = 1$

$z^f = w_f x + b = 10$ $f(z^f) = f(10) = 1$

$c = c f(z^f) + f(z^i) \cdot g(z)$

$= 0 \times 1 + 1 \times 2 = 2$

$z^o = w_o x + b = -10$ $f(z^o) = f(90) = 0$

$h(c) = h(2) = 2$

$y = h(c) \cdot f(z^o) = 0$ now $c = 2$

$t = 6:$ $z = wx + b = -4$ $g(z) = g(-2) = -4$

$z^i = w_i x + b = -10$ $f(z^i) = f(90) = 0$

$z^f = w_f x + b = 110$ $f(z^f) = f(10) = 1$

$c = c f(z^f) + f(z^i) \cdot g(z)$

$= 2 \times 1 - 4 \times 0 = 2$

$z^o = w_o x + b = 90$ $f(z^o) = f(90) = 1$

$h(c) = h(1) = 1$

$y = h(c) \cdot f(z^o) = 2$ now $c = 2$

$t=7$: $\quad z = wx+b = 1 \qquad g(z) = g(1) = 1$

$z^i = w_i x + b = 90 \qquad f(z^i) = f(90) = 1$

$z^f = w_f x + b = -90 \qquad f(z^f) = f(-90) = 0$

$c = c\, f(z^f) + f(z^i) \cdot g(z)$

$= 2 \times 0 + 1 \times 1 = 1$

$z^o = w_o x + b = 90 \qquad f(z^o) = f(90) = 1$

$h(c) = h(1) = 1$

$y = h(c) \cdot f(z^o) = 1 \qquad$ now $c = 1$

$t=8$: $\quad z = wx+b = 2 \qquad g(z) = g(2) = 2$

$z^i = w_i x + b = 90 \qquad f(z^i) = f(90) = 1$

$z^f = w_f x + b = 10 \qquad f(z^f) = f(10) = 1$

$c = c\, f(z^f) + f(z^i) \cdot g(z)$

$= 1 \times 1 + 2 \times 1 = 3$

$z^o = w_o x + b = 90 \qquad f(z^o) = f(90) = 1$

$h(c) = h(1) = 1$

$y = h(c) \cdot f(z^o) = 3 \qquad$ now $c = 3$

thus: $\quad y: \ 0 \ 1 \ 4 \ 0 \ 0 \ 2 \ 1 \ 3$

2. Word Embedding

↓ next page:

$$L = -\log \prod_{c=1}^{C} \frac{\exp(u_{c,j^*})}{\sum_{j=1}^{V} \exp(u_{c,j})}$$

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

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

$$\frac{\partial L}{\partial W_{ij}} = \sum_{k=1}^{V} \sum_{c=1}^{C} \frac{\partial L}{\partial u_{c,k}} \frac{\partial u_{c,k}}{\partial W_{ij}}$$

$$\frac{\partial L}{\partial u_{c,j}} = -\delta_{jj_c^*} + y_{c,j}$$

thus: $$\frac{\partial L}{\partial W'_{ij}} = \sum_{k=1}^{V} \sum_{c=1}^{C} \frac{\partial L}{\partial u_{c,k}} \frac{\partial u_{c,k}}{\partial W'_{ij}} = \sum_{c=1}^{C} \frac{\partial L}{\partial u_{c,j}} \frac{\partial u_{c,j}}{\partial W'_{ij}}$$

$$\boxed{= \sum_{c=1}^{C} (-\delta_{jj_c^*} + y_{c,j}) \left(\sum_{k=1}^{V} w_{ki} x_k \right)}$$

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

$$= \boxed{\sum_{k=1}^{V} \sum_{c=1}^{C} (-\delta_{jj_c^*} + y_{c,j}) \left(\sum_{k=1}^{V} w_{ki} x_k \right)}$$