

ML 2019 Hw5 Report

學號：b05705042 系級：資管四 姓名：皇甫立翔

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。

(1) 我的 RNN Model 使用的是 LSTM，其中的參數(以最好的結果來看)是：

```
self.lstm = nn.LSTM(256, 128, num_layers=1, dropout=0.37, batch_first=True)
```

再將 LSTM 的結果展開，利用 fully-connected neural network 計算出結果。

(2) word embedding 的部分我是混合同時使用 gensim 的 word2vec 以及 nn 裡面自帶的 Embedding 方法，使用的方法如下：

```
self.embedd = nn.Embedding(weigh.shape[0], 256)
self.embedd.weight.data.copy_(torch.from_numpy(weigh))
self.embedd.weight.requires_grad = False
```

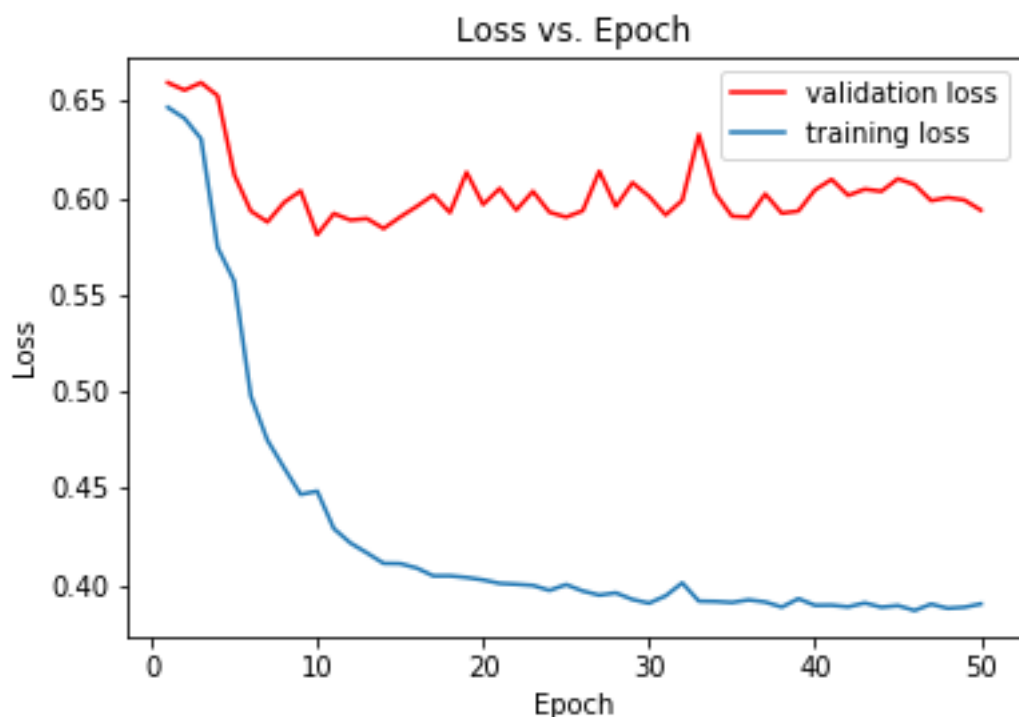
data.copy_後面使用的就是已經 train 好的 word2vec，兩者一起使用。

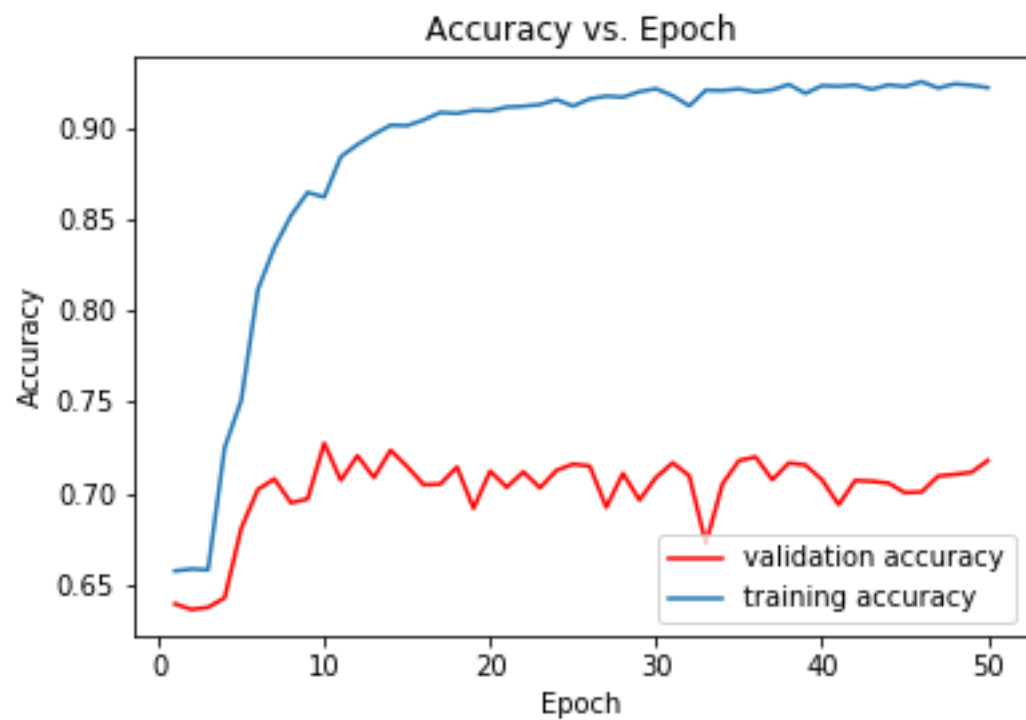
正確率：

Public Score = 0.77906

Private Score = 0.77441

訓練曲線：





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。

(1) BOW 將每個文本轉換成一串 vector，對應其在 dictionary 裡面出現過的次數。

(2) DNN 使用 fully-connected neural network，三層 linear，維度分別是：

`nn.Linear(input_dim, 1024)`

`nn.Linear(1024, 64)`

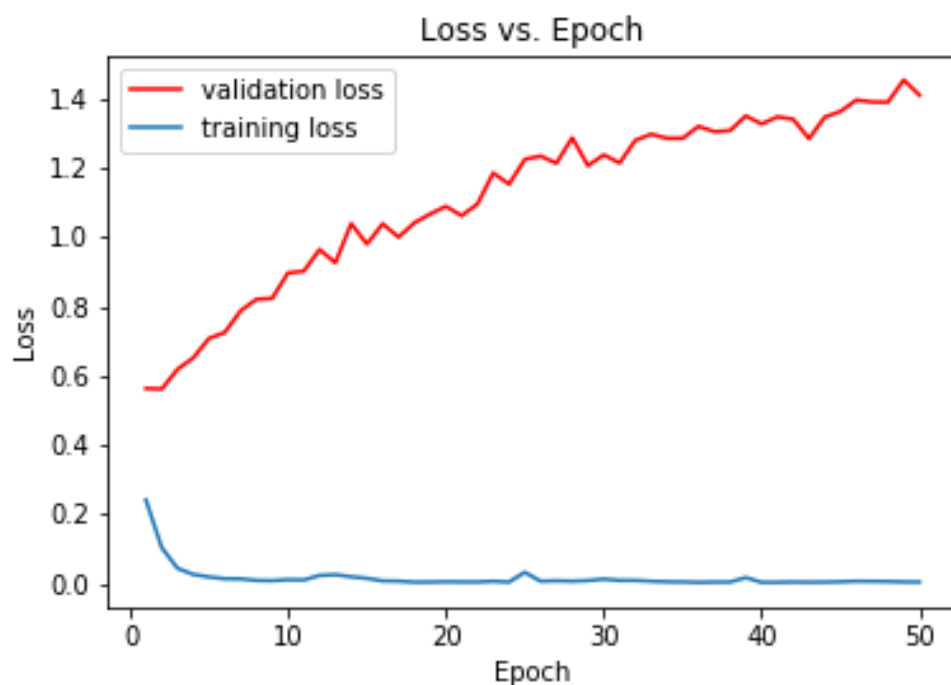
`nn.Linear(64, 2)`

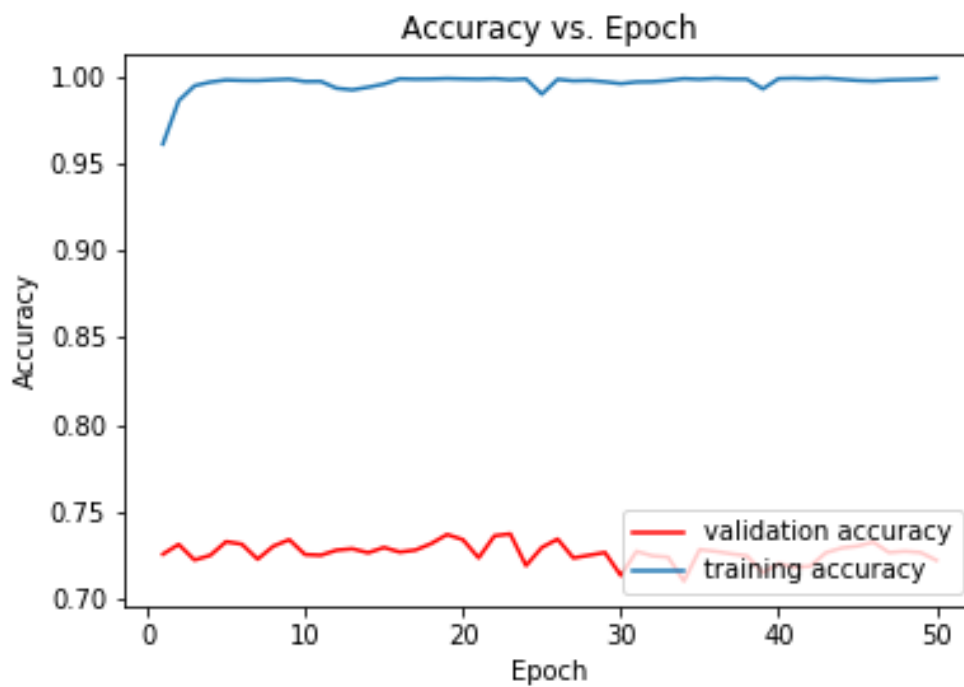
正確率：

Public Score = 0.77209

Private Score = 0.79534

訓練曲線：





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。

(1) Preprocessing

- a. 盡可能讓 stop words 齊全，避免大量重複的 token 出現，以減少 noise。
- b. 將表情符號去除或轉換成實質的文字意義。

(2) Word Embedding

- a. 利用 gensim train word2vec 提高 epoch 或是 vector 本身長度。
- b. 將 nn.Embedding 與 word2vec 混合使用。

(3) LSTM Model

- a. 調整 dropout 到適當的比率。
- b. 有時候使用 bidirection 會表現比較好。

4. 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。

	沒做斷詞	有做斷詞
Public Score	0.72325	0.77906
Private Score	0.74418	0.77441

如果沒有做斷詞的話，同樣的一個字可能會因為後面接的符號不同，而被當成不同的字來訓練，但本質上他們可能代表同一個意思，所以這種做法會讓整個詞庫變得相當 sparse，因此模型本身結果會比較差。

有做斷詞的話，比較能精準的將原始資料整理成良好的訓練資料，將本質相同的文字留下來並去除多餘的無意義符號與文字，因此模型的結果相對來說會比較好。

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。

第一句 "Today is hot, but I am happy."

RNN : [-0.0269, -0.0512]

BOW : [-0.4745, 0.1701]

第二句 "I am happy, but today is hot."

RNN : [-0.0353, -0.0541]

BOW : [-0.4745, 0.1701]

RNN Model 做出來的結果不一樣是因為，他們使用 embedding，所以不同順序的 input 就會是不同的 vector，但是 BOW 並不會因為順序上的差異而導致 input 不相同，所以結果計算出來會是一樣的。

6. Refer to math problem。

(1)

$$t = 1 \rightarrow z = 3, z_i = 90, z_f = 10, z_o = -10$$

$$\rightarrow c' = 3.0$$

$$\rightarrow y_1 = 0.0001361936$$

$$t = 2 \rightarrow z = -2, z_i = 90, z_f = 10, z_o = 90$$

$$\rightarrow c' = 0.9998638$$

$$\rightarrow y_2 = 0.999863806$$

$$t = 3 \rightarrow z = 4, z_i = 190, z_f = -90, z_o = 90$$

$$\rightarrow c' = 4.0$$

$$\rightarrow y_3 = 4.0$$

$$t = 4 \rightarrow z = 0, z_i = 90, z_f = 10, z_o = 90$$

$$\rightarrow c' = 3.99981840$$

$$\rightarrow y_4 = 3.9998184085$$

$$t = 5 \rightarrow z = 2, z_i = 90, z_f = 10, z_o = -10$$

$$\rightarrow c' = 5.999636$$

$$\rightarrow y_5 = 0.0002723707$$

$$t = 6 \rightarrow z = -4, z_i = -10, z_f = 110, z_o = 90$$

$$\rightarrow c' = 5.9994552$$

$$\rightarrow y_6 = 5.9994552338$$

$$t = 7 \rightarrow z = 1, z_i = 190, z_f = -90, z_o = 90$$

$$\rightarrow c' = 1.0$$

$$\rightarrow y_7 = 1.0$$

$$t = 8 \rightarrow z = 2, z_i = 90, z_f = 10, z_o = 90$$

$$\rightarrow c' = 2.999954$$

$$\rightarrow y_8 = 2.99995460213$$

(2)

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

$$\frac{\partial L}{\partial u_{c,j}} = (-\delta_{jjc} * + y_{c,j})$$

$$\Rightarrow \frac{\partial L}{\partial W_{ij}^T}$$

$$= \sum_{k=1}^V \sum_{c=1}^C \frac{\partial L}{\partial u_{c,k}} \frac{\partial}{\partial W_{ij}^T} \left(\sum_{m=1}^N \sum_{l=1}^V W'_{mk} W_{lm} x_l \right)$$

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

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

$$\frac{\partial L}{\partial u_{c,j}} = (-\delta_{jjc} * + y_{c,j})$$

$$\Rightarrow \frac{\partial L}{\partial W'_{ij}}$$

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

only the one that $k = j$ would be nonzero

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

$$= \sum_{c=1}^C (-\delta_{jjc} * + y_{c,i}) \left(\sum_{k=1}^V W_{ki} x_k \right)$$