

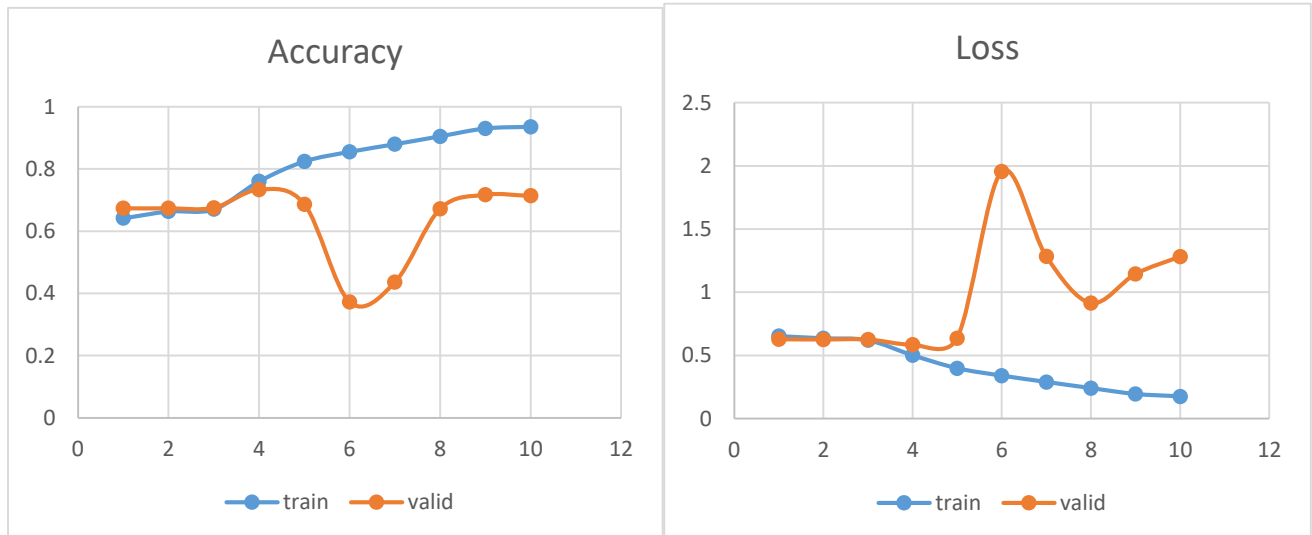
Machine Learning HW5 Report

學號：R08942025 系級：電信碩一 姓名：徐瑋辰

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*

本作業使用之架構如下，正確率為: public-0.77674/private-0.76976

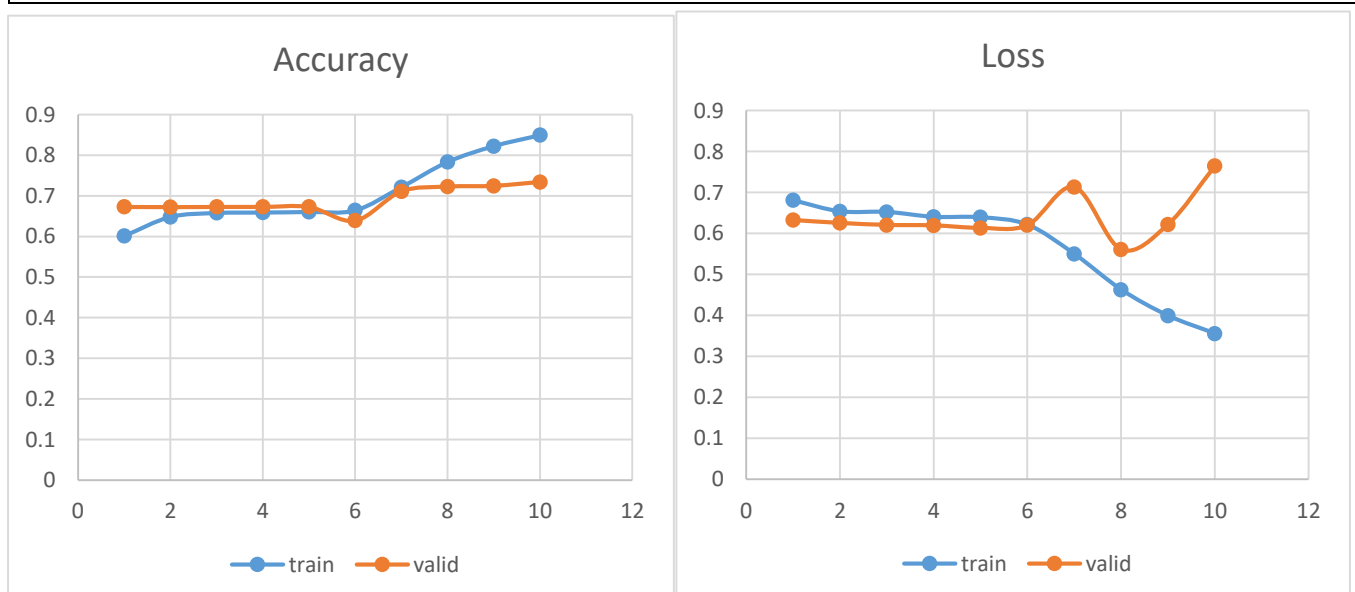
Word Embedding Layer
LSTM
4*FC Layers(linear -> BatchNorm1d -> ReLU -> Dropout)



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+DNN 模型之架構如下，正確率為: public-0.75813/private-0.75116

FC Layers(linear -> BatchNorm1d -> ReLU -> Dropout(0.5))
FC Layers(linear -> BatchNorm1d -> ReLU -> Dropout(0.5))
FC Layers(linear -> BatchNorm1d -> ReLU -> Dropout(0.25))
FC Layers(linear -> BatchNorm1d -> ReLU -> Dropout(0.25))
Softmax



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。

在資料前處理時，將表情符號等刪除，避免干擾判斷

Embedding 的過程則是由 gensim 產生後跟 model 一起訓練，可以更快收斂

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。

下表為斷詞與否的正確率比較

	Public	Private
有斷詞	0.77674	0.76976
無斷詞	0.76046	0.75581

以詞為單位作 embedding，可以使訓練之模型較能掌握詞與詞之間之關係，比起整句一起處理有著更好的效果

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。

兩個句子在不同模型之分數如下：

	Today is hot, but I am happy.	I am happy, but today is hot.
RNN	[1.3885164 -1.7341924]	[1.2636071 -1.5947468]
BOW	[1.0746359 -1.2542158]	[1.0156772 -1.2955631]

BOW 不考慮詞的順序，故分數之差距會比 RNN 來的小。RNN 則考慮語意的轉折等因素，所以會有著不同的分數。

6. (2%) Math Problems

LSTM Cell

	1	2	3	4	5	6	7	8
z	3	-2	-4	0	2	-4	1	2
z_i	90	90	190	90	90	-10	190	90
z_f	10	10	-90	10	10	110	-90	10
z_o	-10	90	90	90	-10	90	90	90
c	0	3	1	4	4	6	6	1
c'	3	1	4	4	6	6	1	3
y	$1.36 \cdot 10^4$	1	4	4	$2.71 \cdot 10^4$	6	1	3

Word Embedding

$$h = W^T x$$

$$u = W'^T h$$

$$y = \text{Softmax}(u)$$

$$= \text{Softmax}(W'^T W^T x)$$

$$L = -\log \prod_{c \in C} \frac{\exp(u_c)}{\sum_{i \in V} \exp(u_i)}$$

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

$$\frac{\partial L}{\partial u_{c,i}} = y_{c,i} - t_{c,i} = e_{c,i}$$

$$\begin{aligned} \frac{\partial L}{\partial W_{ij}^{T'}} &= \sum_{c=1}^C \frac{\partial L}{\partial u_{c,i}} \frac{\partial u_{c,i}}{\partial W_{ij}^{T'}} \\ &= \sum_{c=1}^C (e_{c,i}) \cdot \left(\sum_{k=1}^V W_{kj}^T x_k \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial W_{ij}^{T'}} &= \sum_{k=1}^V \sum_{c=1}^C \frac{\partial L}{\partial u_{c,k}} \frac{\partial u_{c,k}}{\partial W_{ij}^{T'}} \\ &= \sum_{k=1}^V \sum_{c=1}^C \frac{\partial L}{\partial u_{c,k}} + \frac{\sum_{k=1}^V \sum_{\ell=1}^V W_{hk}^{T'} W_{\ell n}^T x_\ell}{\partial W_{ij}^{T'}} \\ &= \sum_{k=1}^V \sum_{c=1}^C e_{c,k} W_{kj}^{T'} x_i \end{aligned}$$