

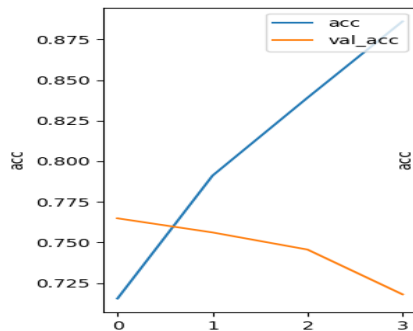
Machine Learning HW6 Report

學號：b05902018 系級：資工三 姓名：張凱程

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*

embedding_1 (Embedding)	(None, None, 250)	7872500
bidirectional_1 (Bidirectional)	(None, None, 512)	778752
bidirectional_2 (Bidirectional)	(None, None, 512)	1181184
time_distributed_1 (TimeDistributed)	(None, None, 256)	131328
dense_2 (Dense)	(None, None, 128)	32896
dropout_1 (Dropout)	(None, None, 128)	0
dense_3 (Dense)	(None, None, 64)	8256
dropout_2 (Dropout)	(None, None, 64)	0
dense_4 (Dense)	(None, None, 2)	130

其中 Embedding 設 weight = embedding matrix，embedding matrix 是該單字對應到的 id 經過 word2vec 訓練後的 vector；bidirectional 都是 GRU(unit = 256)，return sequence 皆為 true，time distribute 是 Dense 256。正確率：public: 0.76150,private: 0.76050

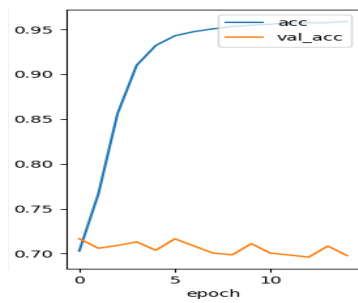


只有第一層過 0.76，所以只存第一層的 model。

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*。

Layer (type)	Output Shape	Param #
dense_7 (Dense)	(None, 512)	2560512
dropout_6 (Dropout)	(None, 512)	0
dense_8 (Dense)	(None, 512)	262656
dropout_7 (Dropout)	(None, 512)	0
dense_9 (Dense)	(None, 256)	131328
dropout_8 (Dropout)	(None, 256)	0
dense_10 (Dense)	(None, 128)	32896
dropout_9 (Dropout)	(None, 128)	0
dense_11 (Dense)	(None, 64)	8256
dropout_10 (Dropout)	(None, 64)	0
dense_12 (Dense)	(None, 2)	130

Dropout 分別為 0.1 0.1 0.2 0.2 0.3，正確率：public: 0.71640, private: 0.71740



epoch=5 時表現最佳。

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。

把 word2vec 的 vector 長度加長(250)，vector 長度加長可以增加每個詞的變因，也就是可以用更多維度全是一個詞，training 可以參考的值也會較多，刪除標點符號也可以使用，因為有些標點符號比較像是來亂的，刪除它可以減少 padding sequence 的差異。

4. (1%) 請比較不做斷詞 (e.g., 以字為單位) 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。

效果：有做斷詞稍微好一些，原因很可能就是兩個字變成一個詞可能造成意思上的變化，可能會有差的包含：愛死、不愧、自以為是、沒有醜女人（可能被認為「沒、有醜、女」）。

5. (1%) 請比較 RNN 與 BOW 兩種不同 model 對於 "在說別人白痴之前，先想想自己"與"在說別人之前先想想自己，白痴" 這兩句話的分數 (model output)，並討論造成差異的原因。

BOW + DNN: 前者是 1，後者是 1

Word2vec + RNN: 前者是 0，後者是 1

原因是因為詞的重複性很高，如果不考慮順序，則會被視為都是攻擊，因為裡面都有「白痴」兩個字，BOW 就是如此，如果考慮順序的話則未必，我的 model 測試結果是第一句幾乎不被視為攻擊，第二句才會，可見順序的重要性。