Machine Learning HW6 Report

學號：R07922134　系級：資工碩一　　姓名：陳紘豪

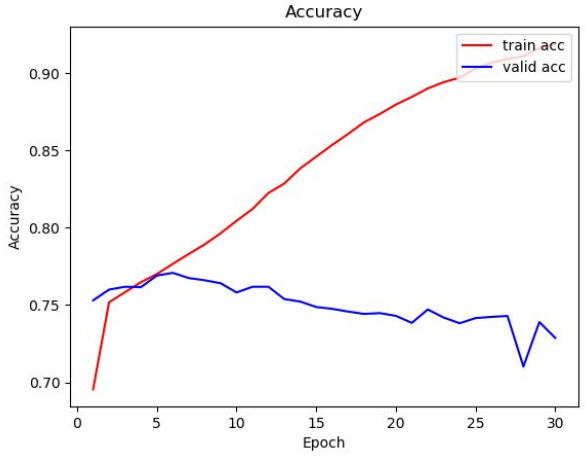
1. (1%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法，回報模型的正確率並繪出訓練曲線*

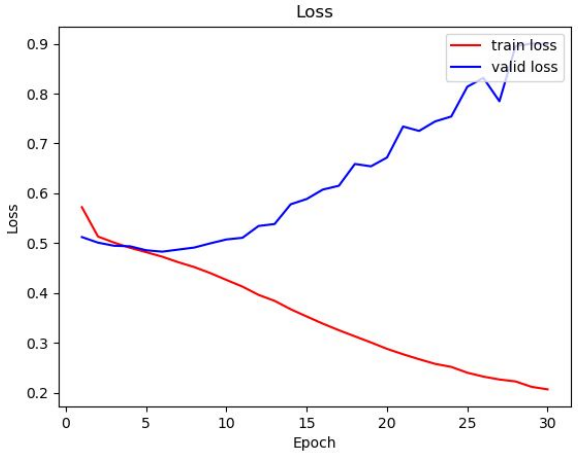
RNN架構：

RNN 的架構主要由兩層的雙向 GRU 所組成，hidden size 我設為 256，在一開始進 GRU 前，我有先讓 input tensor 過一層 layer normalization，經過GRU之後的 forward、backward 兩個方向的 outputs 我先個別做 average pooling，再把兩者 concatenate 起來丟入兩層的 Linear 中來預測是否為惡意言論，在 GRU、Linear 之中我都有加入 Dropout 來減輕 overfitting。

Word embedding方法：

我先用 jieba 對句子進行斷詞，然後透過 word2vec 套件把所有的詞轉成 128 維向量的型態，之後每則斷詞後的留言取前 100 個詞轉成向量以固定 RNN 的輸入長度。

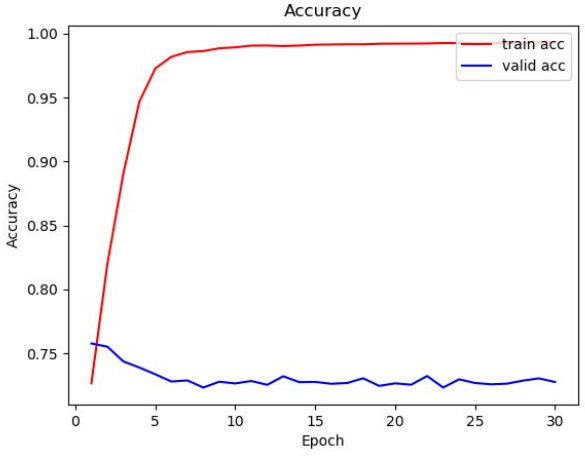


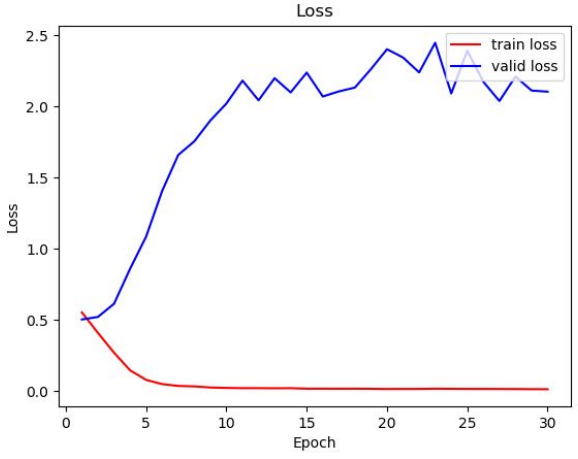


Public	Private
0.76260	0.76160

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*。

我先對斷完詞後的 dataset 建立一個 word2index 的字典，由於字典中的字有 127388 個，所以利用這字典去建立每個句子的 BOW 向量長度也會有 127388。DNN的架構由 5 層 Linear 組成，unit 數量為 127388 -> 1024 -> 512 -> 256 -> 128 -> 1，而且我有加入 Dropout 來減緩 overfitting，所以雖然在 training set 上 accuracy 趨近於 1 但在 validation set 上的 accuracy 不至於很爛，甚至跟 RNN 的 valiation accuracy 還差不多。



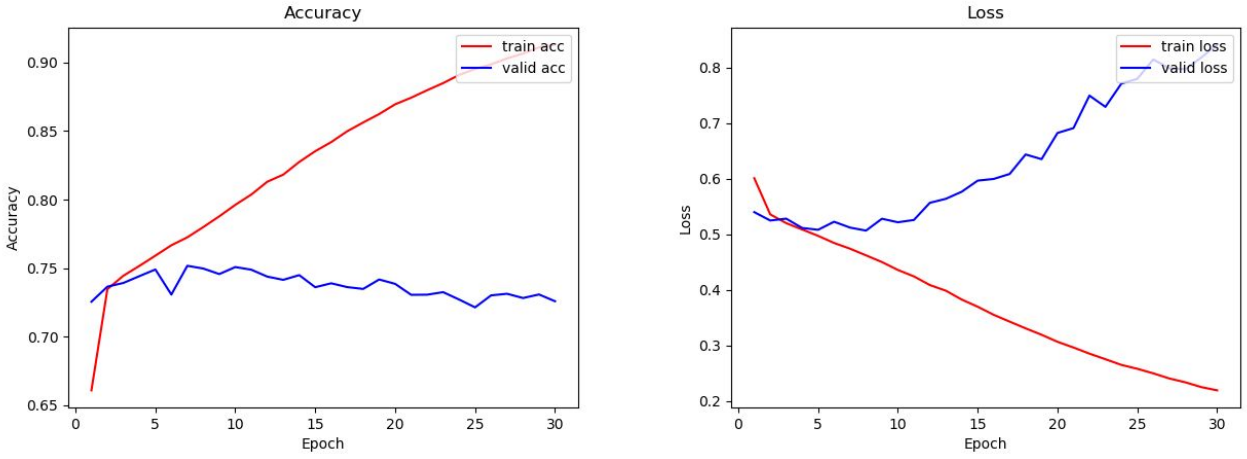


Public	Private
0.75800	0.75650

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。

   我一開始在做 word embedding 的時候，min count 是設 1，也就是我會把出現過的每一個的 word 都丟下去 train 一個 word embedding model，這樣的結果做出來的始終過不了 strong baseline，後來把 min count 設 5 就過了，我想原因應該是那些出現次數太少的對 model 來說是雜訊，可能會造成誤判。至於 preprocessing 的部分我沒有多做什麼，只是把句子做斷詞而已，若能夠在斷詞後先濾掉一些出現頻率比較高的字（例如：的、啦、吧之類的）再丟下去做 word embedding 可能可以再提高一點準確率。

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。



   上圖為沒做斷詞的 input 丟給 RNN 後的訓練曲線，大致上看起來和有做斷詞的 RNN 差不多，只是在 validation set 上的 accuracy 稍微低一點，沒有達到 0.76，所以從這曲線看起來可以推測出沒有做斷詞的表現會比有做斷詞還要來得差。有做斷詞的話，斷出來的詞是會有意義的，比較能完整地表達出跟句子相關的語意，例如「開心」、「去死」，沒有做斷詞而是把每個字當成一個詞的話，那麼每次 input 都是一堆零碎的字「開」、「心」、「去」、「死」，只能讓 model 強行去記住那些字跟 label 之間可能的關係，所以 performance 自然而然會比較差。

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。

	在說別人白痴之前，先想想自己	在說別人之前先想想自己，白痴
RNN	0.497946355	0.54770787
BOW+DNN	0.6893385	0.6893395

   這兩句話在使用 jieba 斷詞後的詞都一樣，只是順序不同。因為 BOW 只考慮詞出現的次數而不考慮其先後順序，所以吐出來的機率都一樣，但是在 RNN 中會去考慮詞和詞之間出現的順序，所以 RNN 吐出來的機率是不一樣的，在第一句中，雖然有一個「白痴」出現，但是在考慮前後文後，RNN會判斷成不是惡意言論，相較於 BOW 的判斷，RNN的對於此種情況的分類是較為正確的。