

1. (1%)請問 softmax 適不適合作為本次作業的 output layer? 寫出你最後選擇的 output layer 並說明理由。

不太適合，因為如果有一個 label 輸出很大的值，softmax 會造成其他 label 的數值被壓低，但我們是希望輸出多個值而不是一個最有可能的，所以每個 output 分別用 sigmoid 是比較合理的作法

2. (1%)請設計實驗驗證上述推論。

我有訓練起來的 model 都是使用 sigmoid，使用 softmax 的 f1 都卡在 0.006 左右升不起來

3. (1%)請試著分析 tags 的分布情況(數量)。

'ADVENTURE-NOVEL': 109,  
'ALTERNATE-HISTORY': 72,  
'APOCALYPTIC-AND-POST-APOCALYPTIC-FICTION': 14,  
'AUTOBIOGRAPHICAL-NOVEL': 31,  
'AUTOBIOGRAPHY': 51,  
'BIOGRAPHY': 42,  
'CHILDREN'S-LITERATURE': 777,  
'COMEDY': 59,  
'COMIC-NOVEL': 37,  
'CRIME-FICTION': 368,  
'DETECTIVE-FICTION': 178,  
'DYSTOPIA': 30,  
'FANTASY': 773,  
'FICTION': 1672,  
'GOTHIC-FICTION': 12,  
'HIGH-FANTASY': 15,  
'HISTORICAL-FICTION': 137,  
'HISTORICAL-NOVEL': 222,  
'HISTORY': 40,  
'HORROR': 192,  
'HUMOUR': 18,  
'MEMOIR': 35,  
'MYSTERY': 642,  
'NON-FICTION': 102,

'NOVEL': 992,  
'NOVELLA': 29,  
'ROMANCE-NOVEL': 157,  
'SATIRE': 35,  
'SCIENCE-FICTION': 959,  
'SHORT-STORY': 41,  
'SPECULATIVE-FICTION': 1448,  
'SPY-FICTION': 75,  
'SUSPENSE': 318,  
'TECHNO-THRILLER': 18,  
'THRILLER': 243,  
'UTOPIAN-AND-DYSTOPIAN-FICTION': 11,  
'WAR-NOVEL': 31,  
'YOUNG-ADULT-LITERATURE': 288

可以明顯看出數量分布非常不均

4. (1%) 本次作業中使用何種方式得到 **word embedding**? 請簡單描述做法。

我一開始試了類似 tfidf 的作法，分每個 tag 計算(In label) / (out of label) 的值。簡單的切開後 f1 大概在 0.49，最後使用 GloVe 作為 Embedding

5. (1%) 試比較 **bag of word** 和 **RNN** 何者在本次作業中效果較好。

**RNN** 應該要比較好，但是因為訓練資料量小，加上 **label** 分布非常不均勻，使得 **bag of word** 有著接近甚至超過的分數。**Bag of word** 用 6 層 DNN 配 elu，**RNN** 用 GRU+3 層 DNN，我自己訓練起來兩個 **validation f1** 大約都在 0.51~0.52。