

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。

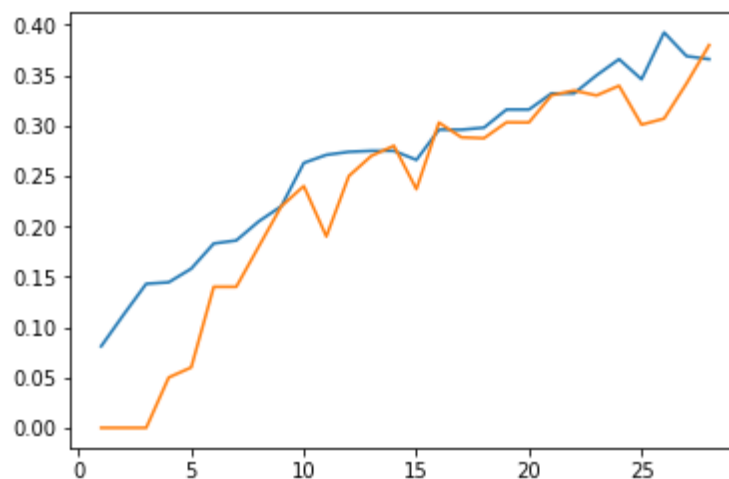
Softmax 不適合作為本次作業的輸出層，因為這次作業是多標籤問題，而當標籤很多時會變成每個標籤機率都很低(因為softmax讓輸出中所有標籤機率和為1)。

而這會造成thresh設高一點會抓不到半個標籤，很多筆測資的輸出變成沒東西，thresh設低一點會抓到一堆標籤，使正確率下降。

因此我最後選擇sigmoid當作輸出層，原因有兩個，其一是因為sigmoid本身介在0~1之間，適合當作機率看待，其二是sigmoid在 $x=0$ 附近很敏感(x 變動一點 y 就會有很大變動)，較不會有thresh設不好的問題。

2. (1%)請設計實驗驗證上述推論。

導致softmax會變差的原因是因為多標籤，因此我們只要試試看只有一個標籤的訓練資料。



上圖為只有一個label的訓練資料的情況，x軸為epoch，y軸為f1_score，藍線是sigmoid，黃線是softmax，可以看出來兩個差不多。

3. (1%)請試著分析tags的分布情況(數量)。

['SCIENCE-FICTION', 'SPECULATIVE-FICTION', 'FICTION', 'NOVEL', 'FANTASY',
"CHILDREN'S-LITERATURE", 'HUMOUR', 'SATIRE', 'HISTORICAL-FICTION',
'HISTORY', 'MYSTERY', 'SUSPENSE', 'ADVENTURE-NOVEL', 'SPY-FICTION',
'AUTOBIOGRAPHY', 'HORROR', 'THRILLER', 'ROMANCE-NOVEL', 'COMEDY',
'NOVELLA', 'WAR-NOVEL', 'DYSTOPIA', 'COMIC-NOVEL',
'DETECTIVE-FICTION', 'HISTORICAL-NOVEL', 'BIOGRAPHY', 'MEMOIR',
'NON-FICTION', 'CRIME-FICTION', 'AUTOBIOGRAPHICAL-NOVEL',
'ALTERNATE-HISTORY', 'TECHNO-THRILLER',
'UTOPIAN-AND-DYSTOPIAN-FICTION', 'YOUNG-ADULT-LITERATURE',
'SHORT-STORY', 'GOTHIC-FICTION',
'APOCALYPTIC-AND-POST-APOCALYPTIC-FICTION', 'HIGH-FANTASY']
[959. 1448. 1672. 992. 773. 777. 18. 35. 137. 40.
642. 318. 109. 75. 51. 192. 243. 157. 59. 29.
31. 30. 37. 178. 222. 42. 35. 102. 368. 31.
72. 18. 11. 288. 41. 12. 14. 15.]

上面是tag名稱，下面是對應出現次數，FICTION 和SPECULATIVE-FICTION佔大多數。

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

首先使用keras的tokenizer把單字對應一個數字(1維)，之後用現成的embedding，glove把單字轉成100維的向量，再把一個句子裡的所有字都轉成向量。

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

我bag of word 的使用方式是先把逗點和句點去掉，之後用utlk把一些stop word去掉，剩下的字取前5000個較常出現的來count，最後對每個label做normalize。

bag of word 的模型用4層dense+relu來實作，輸出用sigmoid，試出來的結果和rnn差不多，不過bag of word training 速度很快。