

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。

我覺得 softmax 不適合作為本次作業的 output layer，因為 softmax 會 normalize，所以原本機率較高的 tag 會使得其他 tag 的機率下降，而這次作業會有很多個 tag 在同一段文字的情況發生，因此不適用 softmax。

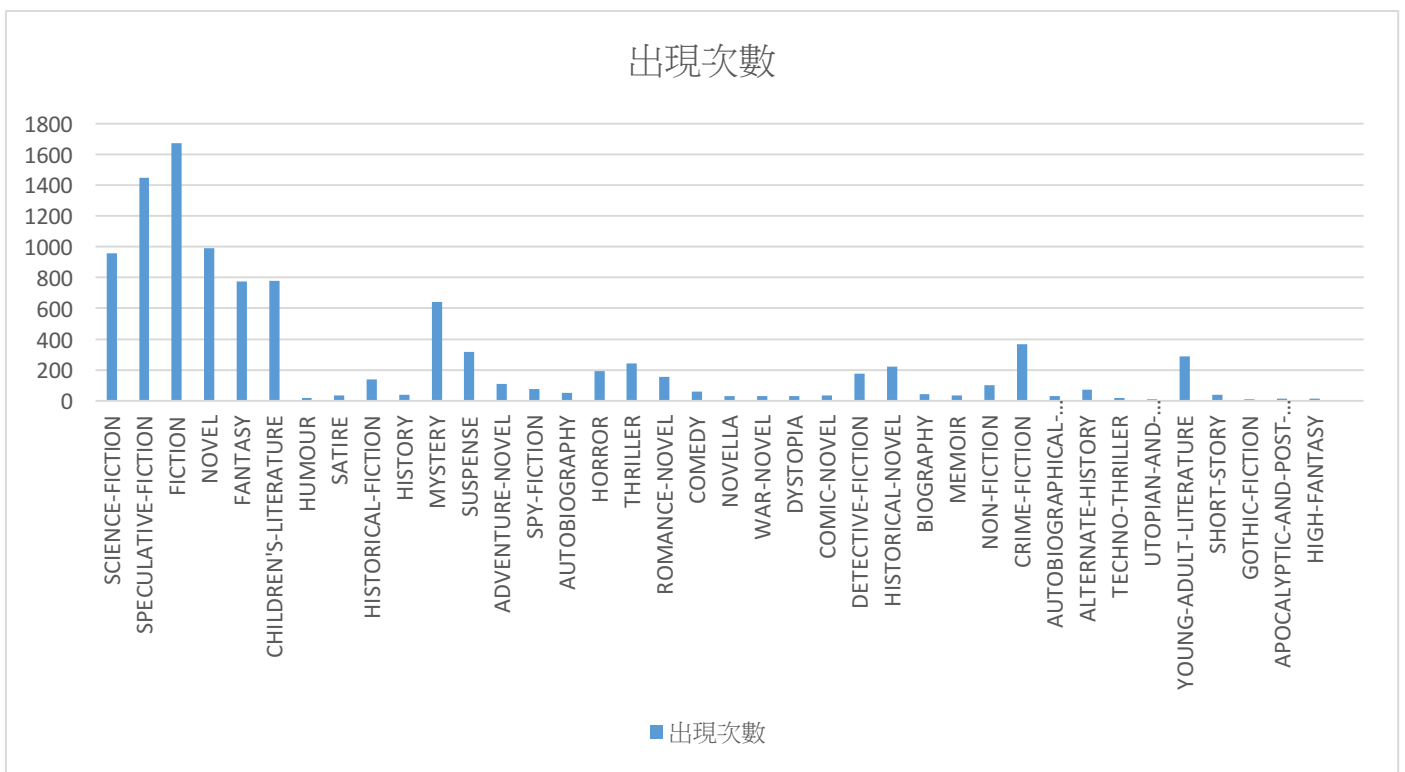
2. (1%)請設計實驗驗證上述推論。

由下表可以看出來，softmax 的效果很差，sigmoid 的效果就好很多！

次數	Softmax	Sigmoid
1	0.3012	0.4842
2	0.2783	0.4934
3	0.2854	0.4910
4	0.2615	0.4953
5	0.2711	0.5010

3. (1%)請試著分析 tags 的分布情況(數量)。

由下圖表可以看出來，fiction 的數量是最多的，大約在 1600 左右。



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

這次的作業我是用 glove，這個方法兼具了 prediction 跟 count 的優點，可以根據出現字的前後文，增加出現在同個位置字的同值性，藉著這點可以 train 出更好的 model。

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

在比較 bag of word 跟 RNN 的時候，我各分別 train 了五次，並看最後的準確率為多少。由下表可以發現，bag of word 的效果比 RNN 還要差，因為 bag of word 沒有考慮到 word 之間的順序性，所以會有比較差的結果也不是很令人意外。

次數	Bag of Word	RNN
1	0.4613	0.4842
2	0.4721	0.4934
3	0.4430	0.4910
4	0.4212	0.4953
5	0.4533	0.5010