

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。

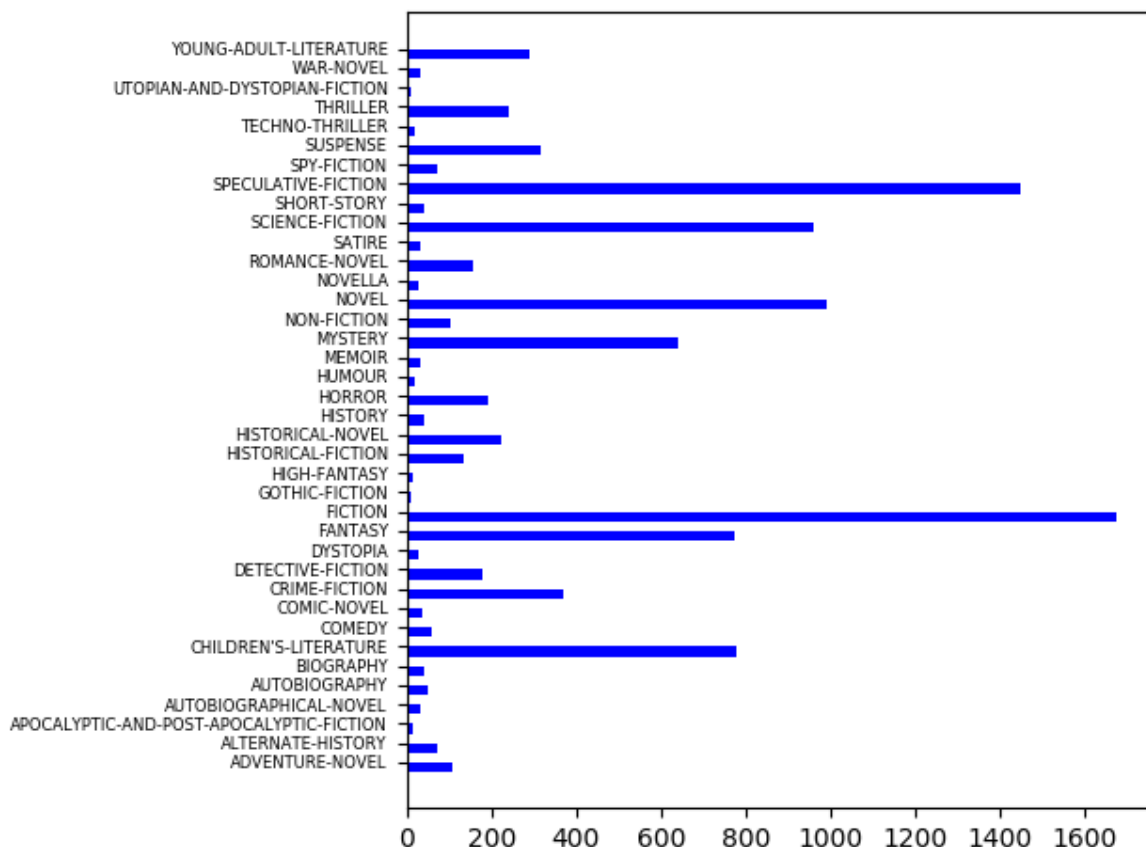
不適合。因為 softmax 會使得 output layer 的和為 1，所以當某一 label 的 output 上升會使得其他 label 下降，又每筆資料的 label 數不一樣，threshold 不好取，故不適合拿來做 multi-label 的 output。

最後選擇用 sigmoid，這樣可以讓每個 label 之間的 output 是獨立的，並且設一個 threshold，超過 threshold 代表有那個 label。

2. (1%)請設計實驗驗證上述推論。

同樣的 model 只改最後一層的 output layer，softmax 的 f1 score 為 0，完全 train 不起來，而 sigmoid 的 f1 score 可以到 0.5。

3. (1%)請試著分析 tags 的分布情況(數量)。



Label 的分部很不平均，因此準確預測出現頻率很低的 label 會有難度。

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

先用 keras 的 tokenizer 將全部的 data tokenize 之後，便可用來做 bag of word。

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

Bag of word 在本次作業中效果比較好，因為 data 量較少，導致 RNN train 不太起來。