學號：B03901030 系級： 電機三 姓名：蕭晨豪

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。**

我認為 softmax 並不適合做為 output layer，softmax 以往作為 output layer 是因為我們認為這些預測出來的結果會根據某個機率分布，因此使用 softmax 可以使他們的和為 1，可是在 multi label problem 裡面並沒有這個性質，因此這些 label 都應該分開來看而非當成全體符合一個機率分布。最後我所使用的 output layer 為 sigmoid，因為我認為這個問題像是針對每一個 label 都去看他是否為 1 的 binary classification。

p.s.我的模型結構見第二題

2. **(1%)請設計實驗驗證上述推論。**

我的 model 架構如下:

```
model = Sequential()
model.add(Embedding(num_words, embedding_dim, weights=[embedding_matrix], input_length=max_article_length, trainable=False))
model.add(GRU(256,activation='tanh',dropout=0.5, return_sequences = True,recurrent_dropout = 0.5))
model.add(GRU(256,activation='tanh',dropout=0.5,recurrent_dropout = 0.5 ))
model.add(Dense(256,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(128,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(64,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(38,activation='sigmoid'))
```

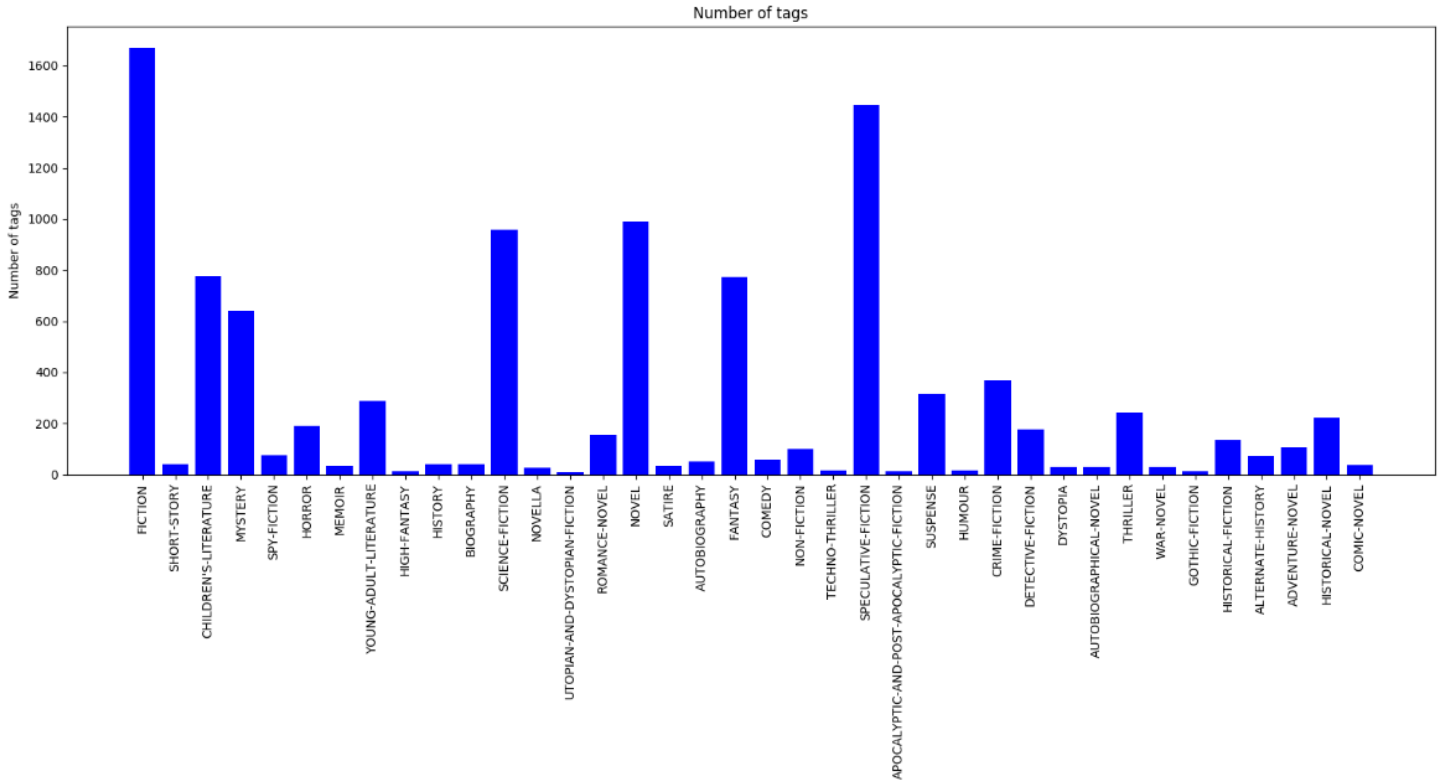
由於一層 GRU 效果還不是很理想，因次我多加了一層，另外 output layer 如第一題所述使用的是 sigmoid

用此 model，只更改 output layer，sigmoid 的結果在 Kaggle 上的分數為 0.512

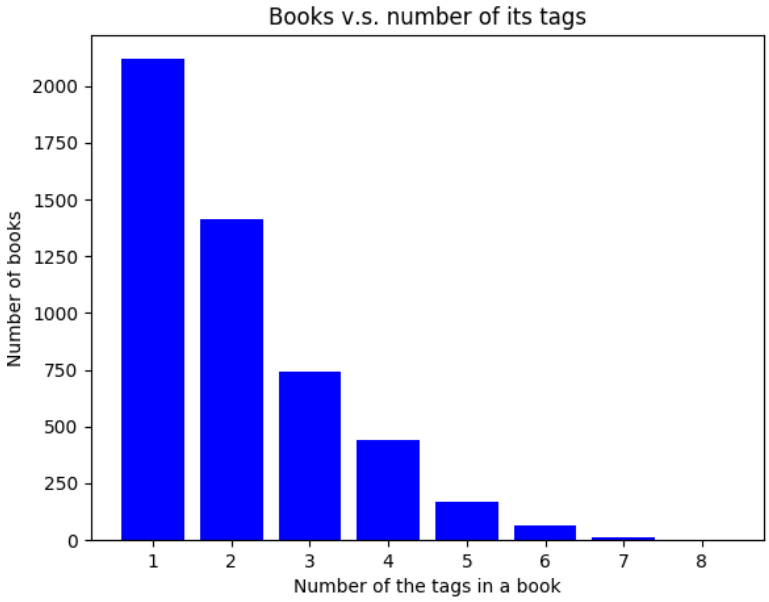
而 softmax 的結果在 Kaggle 上的分數則為 0.4613

此差距已超過誤差範圍，因此可認定 sigmoid 作為 output layer 比 softmax 好

3. (1%)請試著分析 tags 的分布情況(數量)。



我將 training data 中的 tag 分布取出來後用 matplotlib 標示出此柱狀圖，可以看到 FICTION 和 SPECULATIVE-FICTION 最多，SCIENCE-FICTION 和 NOVEL 也很多，可看出此 training set 主要是由許多虛構小說取出。



可以看到大部分的書都只有 1~3 個 tag

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

我使用的是第三方 train 好的 glove 200 維，以下簡介 glove

Glove 為 Global Vectors for Word Representation，是 Stanford 發表的一種 word embedding 方法，與 skip-gram 和 language model 有相似之處，但主要是基於要是語意相近，則在同一區域出現的條件機率會提高，因此透過類似 language model 的方式算條件機率，之後求兩個條件機率的比值。對這個比值取一些處理之後當作 objective function 來訓練。

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

我使用 bag of word 的方法只能到 5.1163，但使用 glove+兩層 GRU 可以到達 0.512，在後來更改 threshold 後更可以到達 5.1926，因此 RNN 的方法看起來比較好