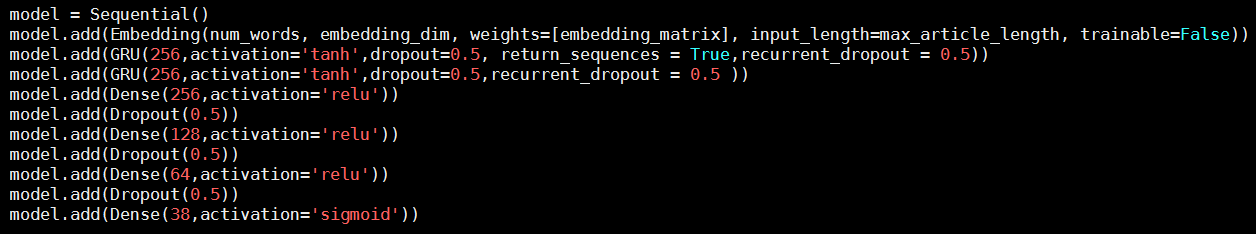
學號：B03901030 系級： 電機三 姓名：蕭晨豪

* 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

* 1. (1%)請設計實驗驗證上述推論。

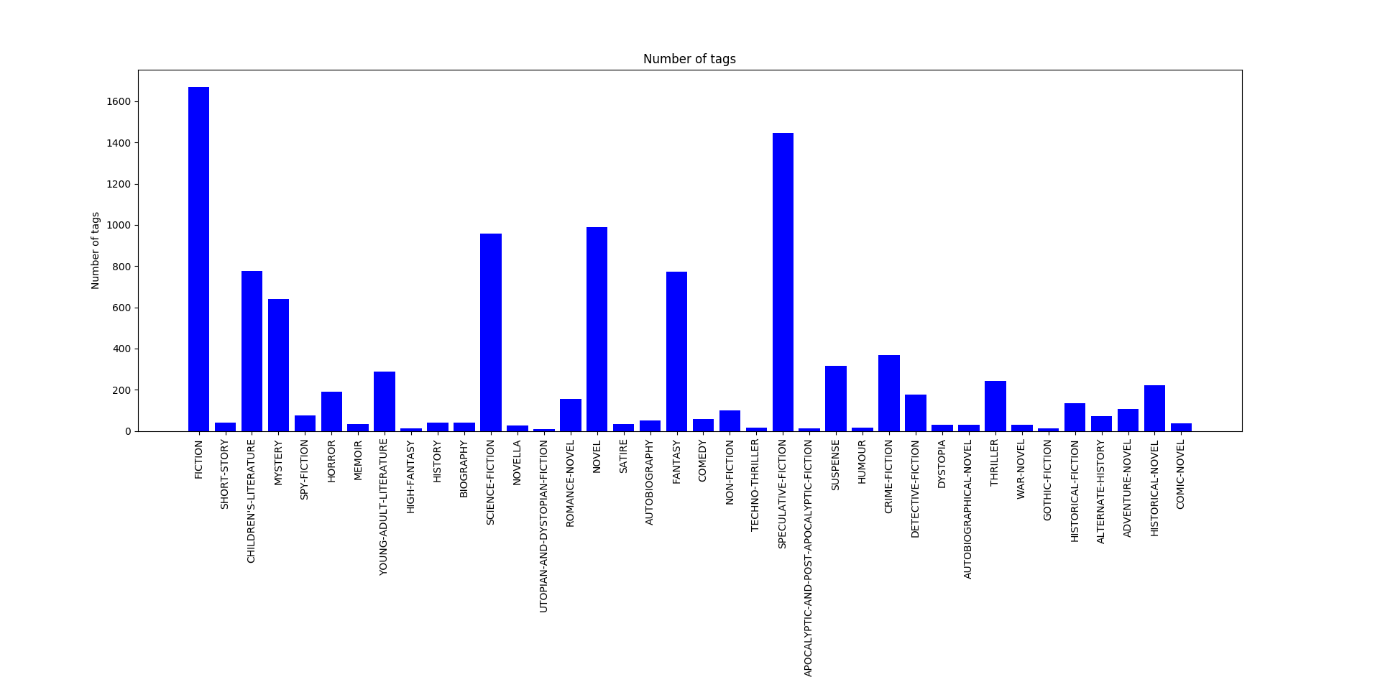
我的model架構如下:

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

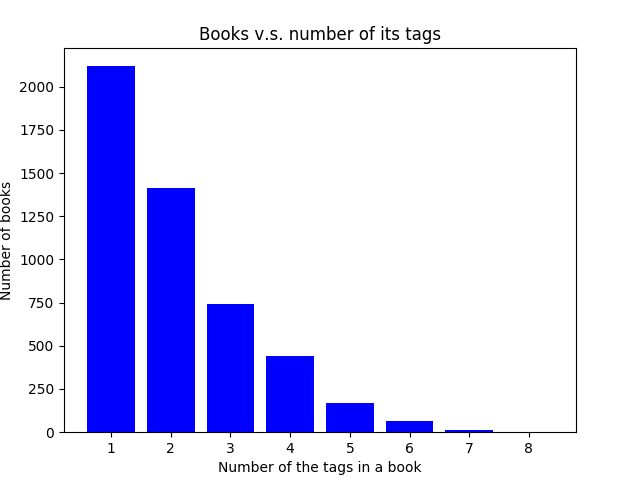
用此model，只更改output layer，sigmoid的結果在Kaggle上的分數為0.512

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

* 1. (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

* 1. (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。

* 1. (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