學號：B04902045 系級： 資工二 姓名：孫凡耘

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。

   In softmax when increasing score for one label, all others are lowered (it's a probability distribution). You don't want that when you have multiple labels，因為他們並沒有機率合是一這個性質。最後我選擇sigmoid，因為保證output在0跟1之間，很直觀的可以理解為屬於這個class的機率。

2. (1%)請設計實驗驗證上述推論。

   固定其他參數(參數使用助教的sample code)的前提下，嘗試改變output layer的activation function。

   softmax, thresh = 0.4: 跟本train不起來，顯然因為thresh=0.4, cv < 0.1

   softmax, thresh = 0.1: cv = 0.41012

   sigmoid, thresh = 0.4: cv = 0.44803

   sigmoid, thresh = 0.5: cv = 0.44842

   tanh, thresh = 0: 也train不起來，估計應該是因為loss為crossentropy，而tanh的output可能會是負的 cv ~ 0.1

3. (1%)請試著分析tags的分布情況(數量)。

   FICTION 1672

   SPECULATIVE-FICTION 1448

   NOVEL 992

   SCIENCE-FICTION 959

   CHILDREN'S-LITERATURE 777

   FANTASY 773

   MYSTERY 642

   CRIME-FICTION 368

   SUSPENSE 318

   YOUNG-ADULT-LITERATURE 288

THRILLER 243

HISTORICAL-NOVEL 222

HORROR 192

DETECTIVE-FICTION 178

ROMANCE-NOVEL 157

HISTORICAL-FICTION 137

ADVENTURE-NOVEL 109

NON-FICTION 102

SPY-FICTION 75

ALTERNATE-HISTORY 72

COMEDY 59

AUTOBIOGRAPHY 51

BIOGRAPHY 42

SHORT-STORY 41

HISTORY 40

COMIC-NOVEL 37

SATIRE 35

MEMOIR 35

WAR-NOVEL 31

AUTOBIOGRAPHICAL-NOVEL 31

DYSTOPIA 30

NOVELLA 29

TECHNO-THRILLER 18

HUMOUR 18

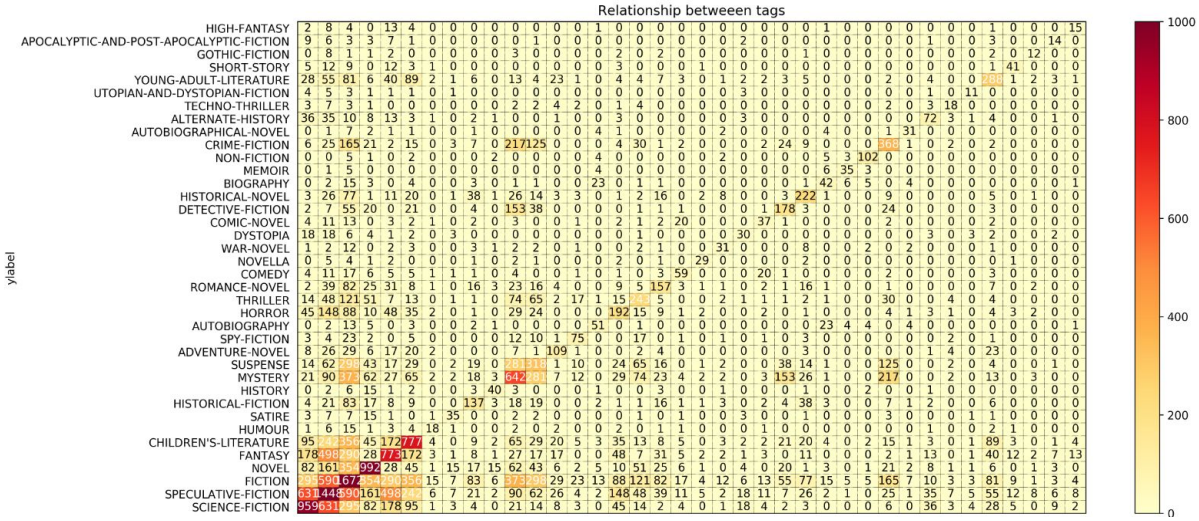
HIGH-FANTASY 15

APOCALYPTIC-AND-POST-APOCALYPTIC-FICTION 14

GOTHIC-FICTION 12

UTOPIAN-AND-DYSTOPIAN-FICTION 11

reference:
http://stackoverflow.com/questions/30222747/drawing-a-grid-in-python-with-colors-corresponding-to-different-values



為了近一步探索tags之間的關係，我用training data set 畫了這張圖。左邊為ylabel，下面為xlabel，則格子內的數字代表在所有屬於ylabel的sample中，同時也屬於xlabel的sample數量。

In other words, 把數字除以那列的最大值（即對角線上的值），即可obtain the probability (or ratio) of the text belongs to xlabel given that it belongs to ylabel.

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glove的word vector(Global Vector for Word Representation)

GloVe is essentially a log-bilinear model with a weighted least-squares objective.The training objective of GloVe is to learn word vectors such that their dot product equals the logarithm of the words' probability of co-occurrence. Owing to the fact that the logarithm of a ratio equals the difference of logarithms, this objective associates (the logarithm of) ratios of co-occurrence probabilities with vector differences in the word vector space. 簡單來說，就是以最小化兩個詞的向量積與它們共現次數的對數之間的差異。

$$J = \sum_{i,j=1}^{V} f(X_{ij}) \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2$$

當 wi 和 bi 分別作為詞語 i 的詞向量和偏差，w~j 和 bj 分別作為詞語 j 的文本詞向量和偏差，Xij 是在詞語 j 的文本中出現詞語 i 的次數，而 f 是將相對低的權重分配給稀有和頻繁共現的加權函數。

reference: https://kknews.cc/zh-tw/news/9evz2q.html

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

此處的數值皆為CV

If not fine-tuned:

兩者表現差不多，結果大約.46 ~ 0.49之間

If fine-tuned:

RNN 的表現稍微比bag of words好一些

RNN: 0.5159

Bag of Words: 0.50468