

學號：B03902102 系級：資工三 姓名：廖廷浩

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，因為這次的作業是要 multiple label，而 softmax 會將機率的總和設為一，讓 threshold 難以調整，改為 sigmoid 會比較優良。

2. (1%)請設計實驗驗證上述推論。

分別測試 softmax 跟 sigmoid，在 threshold 為 0.3 的時候，可以發現 sigmoid 的結果會比 softmax 高很多，如果要得到一樣的結果，softmax 必須把 threshold 調得很小，很難找。

3. (1%)請試著分析 tags 的分布情況(數量)。

('HORROR'", 45)

('NOVEL', 124)

('ADVENTURE-NOVEL', 5)

('AUTOBIOGRAPHY', 6)

('FANTASY"', 84)

('HISTORICAL-NOVEL', 1)

('HORROR', 5)

('AUTOBIOGRAPHY"', 1)

('ROMANCE-NOVEL', 2)

('ALTERNATE-HISTORY"', 4)

('SUSPENSE', 192)

('WAR-NOVEL', 3)

('THRILLER', 56)

('CRIME-FICTION"', 183)

('ROMANCE-NOVEL"', 10)

('SCIENCE-FICTION', 251)

('NON-FICTION"', 27)

('FANTASY', 118)

('NOVEL"', 234)

('HISTORICAL-NOVEL'", 58)
('ADVENTURE-NOVEL'", 3)
('MYSTERY'", 17)
('MYSTERY', 242)
('DETECTIVE-FICTION'", 29)
('SPY-FICTION', 32)
('FICTION', 875)
('"', 25)
('SPECULATIVE-FICTION'", 100)
('CHILDREN\'S-LITERATURE'", 214)
('YOUNG-ADULT-LITERATURE'", 83)
('SCIENCE-FICTION'", 3)
('THRILLER'", 54)
("CHILDREN'S-LITERATURE", 128)
('FICTION'", 57)
('DETECTIVE-FICTION', 67)
('HISTORICAL-FICTION', 14)
('SPECULATIVE-FICTION', 355)
('MEMOIR', 2)
('BIOGRAPHY', 3)
('SUSPENSE'", 3)

從資料中可以看出 Crime-Fiction, Fiction, Children's-Literature, Speculative-Fiction 佔了大多數的 tags

4. (1%)本次作業中使用何種方式得到 word embedding?請簡單描述做法。
先上網下載 glove.100d.txt，並且在程式中，get_embedding_dict()會將 glove 檔的字切成許多的 dim vector，get_embedding_matrix()會將每個 feature 的句子轉換成向量，以便 model 的運行。
5. (1%)試比較 bag of word 和 RNN 何者在本次作業中效果較好。

RNN 的 model 會受前後文的影響（記錄歷史的特性）可能會影響到某些字對 tag 的關聯性，而 bag of word 會直接依照字出現在某個 tag 的頻率當作標準，因此在此次作業有可能 tag 與字的頻率相關性較重。