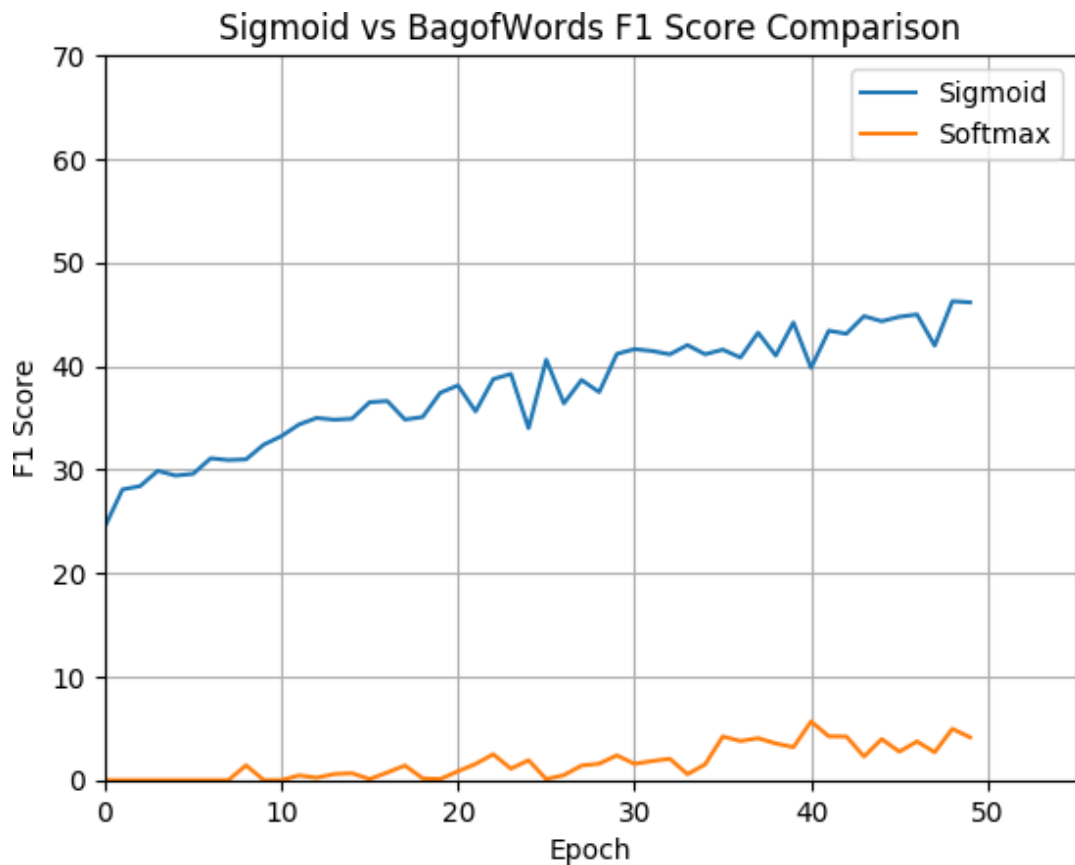


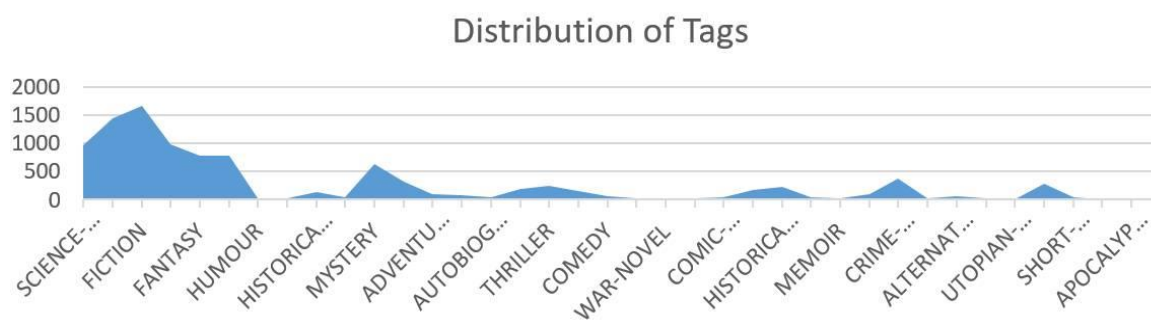
1. (1%)請問 softmax 適不適合作為本次作業的 output layer? 寫出你最後選擇的 output layer 並說明理由。

softmax 不適合為本次作業的 output layer。主要是因為 softmax 把輸出層的所有機率合起來為 1，意思是說要找出全部 label 中機率最高的一個答案，而不是多個答案。這次作業是針對 multilabel，即要每個輸出的最大值為 1，而非全部輸出合起來為 1。

2. (1%)請設計實驗驗證上述推論。



3. (1%)請試著分析 tags 的分布情況(數量)。



```
SCIENCE-FICTION:959
SPECULATIVE-FICTION:1448
FICTION:1672
NOVEL:992
FANTASY:773
CHILDREN'S - LITERATURE:777
HUMOUR:18
SATIRE:35
HISTORICAL - FICTION:137
HISTORY:40
MYSTERY:642
SUSPENSE:318
ADVENTURE - NOVEL:109
SPY - FICTION:75
AUTOBIOGRAPHY:51
HORROR:192
THRILLER:243
ROMANCE - NOVEL:157
COMEDY:59
NOVELLA:29
WAR - NOVEL:31
DYSTOPIA:30
COMIC - NOVEL:37
DETECTIVE - FICTION:178
HISTORICAL - NOVEL:222
BIOGRAPHY:42
MEMOIR:35
NON - FICTION:102
CRIME - FICTION:368
AUTOBIOGRAPHICAL - NOVEL:31
ALTERNATE - HISTORY:72
TECHNO - THRILLER:18
UTOPIAN - AND - DYSTOPIAN - FICTION:11
YOUNG - ADULT - LITERATURE:288
SHORT - STORY:41
GOthic - FICTION:12
APOCALYPTIC - AND - POST - APOCALYPTIC - FICTION:14
HIGH - FANTASY:15
```

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

這次的作業是用 GloVe 的 100 維度的 word embedding。

他們的演算法的步驟為，找出詞彙出現的頻率，然後如果有些詞彙的出現率低則相對應的給它比較低的權重。

$$decay = 1/offset$$

接下來用不同組合的詞彙找出其關

係。最後，用一個 cost function 來以防只學習出現頻率高的詞彙，而忽略了其他詞彙。

$$f(X_{ij}) = \begin{cases} (\frac{X_{ij}}{x_{max}})^\alpha & \text{if } X_{ij} < XMAX \\ 1 & \text{otherwise} \end{cases}$$

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

從中比較，雖然 bag of word 的上升很穩定，但是後來就很難再上升了。反過來，RNN 雖然訓練過程中精準度不是很穩定，但是如果調整好 batch size 和 learning rate，則比較容易得到更高的分數。

因此，在要求過 simple baseline，bag of word 比較容易；過 strong baseline 則用 RNN 比較好。

