

# Machine Learning HW5

學號：R05943005 系級：電子所碩一 姓名：呂丞勛

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 並不適合使用於這次的作業當中，由於 softmax 會一起考慮所有的 classes，並轉換成機率，但是在 multilabel 的 task 當中，每個 class 最好都分開考慮，因此用 sigmoid，這樣出來的 output 等同於該 class 為 1 的機率，不受到其他 classes 的影響。

2. (1%)請設計實驗驗證上述推論。

以下是利用 kaggle 上傳的分數來作佐證，model 架構為 GRU (800), Dense (600, 250, 38)並設定 dropout rate = 0.5。

Output Layer	Kaggle Score
Softmax	0.45553
Sigmoid	0.50181

3. (1%)請試著分析 tags 的分布情況(數量)。

SATIRE	35
HORROR	192
DETECTIVE-FICTION	178
ALTERNATE-HISTORY	72
UTOPIAN-AND-DYSTOPIAN-FICTION	11
THRILLER	243
COMIC-NOVEL	37
MEMOIR	35
FANTASY	773
HISTORICAL-NOVEL	222
SUSPENSE	318
COMEDY	59
TECHNO-THRILLER	18
HISTORICAL-FICTION	137
CHILDREN'S-LITERATURE	777
ROMANCE-NOVEL	157
SPECULATIVE-FICTION	1448
AUTOBIOGRAPHY	51
WAR-NOVEL	31
SPY-FICTION	75
DYSTOPIA	30
HIGH-FANTASY	15
SHORT-STORY	41
NOVEL	992

CRIME-FICTION	368
BIOGRAPHY	42
FICTION	1672
HISTORY	40
SCIENCE-FICTION	959
NOVELLA	29
APOCALYPTIC-AND-POST-APOCALYPTIC-FICTION	14
NON-FICTION	102
GOTHIC-FICTION	12
ADVENTURE-NOVEL	109
HUMOUR	18
MYSTERY	642
YOUNG-ADULT-LITERATURE	288
AUTOBIOGRAPHICAL-NOVEL	30

可以看出，不同種類之間的差異其實算滿大的，這對 training 的過程會有些影響，對於這種 unbalanced data，keras 中可以定義 class\_weights 來送進 fit 中，讓訓練時對於較少 sample 的種類比較重視，否則的話 model 很容易將 data 分類到較多 sample 的種類。

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

本次使用的 word embedding 的方式主要參考 keras 以及助教的範例。

Glove 主要是利用 unsupervised 的方式來訓練，Model 為 log-bilinear model 並利用機率模型為基礎。

- 先利用 keras 內部的 tokenizer fit 所有 text (train + test)
- 在將 texts 轉成 sequence 後利用 pad\_sequence 轉換
- 最後利用 glove 6B 100d 的資料，轉換成 embedding matrix，查不到的字就給與全零的向量
- 將 embedding matrix 當成 layer 的 weights 即可

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

在本次的作業中，分別實作的 bag of word 以及 RNN，拿 bag of word + 3xDense，與 word embedding + 1xGRU + 3xDense，以 Kaggle 的 public testing score 來比較，bag of word 可以得到較好的結果，猜測是因為 bag of word 會把所有單字都考慮進來，不會像是 word embedding 那樣把不認識的字拿掉，可能會因此 loss 掉許多有用的參數。