# 資料探勘研究與實務 Data Mining Research & Practice HW4

# 0853426陳紹雲

# 步驟一：讀入資料並刪除換行符號

# 

# 步驟二：製作token函式，並去除stopword，改用SnowballStemmer

# 

# 步驟三：製作單詞文本以及給先設定model參數

# 

# 步驟四：Tokenizer

# 

# 步驟五：padding/truncate 到maxlength （dataframe.review文本

# 

# 步驟六：向量累加製作embedding\_matrix

# 

# 步驟七：建模

# 

# 

# 步驟八：Training

# 

# 結果：

# 

# Overfitting Result

# 

# LSTM result SimpleRNN

# 討論：

# 我覺得跑出來的結果有點驚訝，起初我只用epochs＝８，跑出來看不太出來，調成epochs=30後才發現大概在週期10以後開始收斂，但結果實在是太爛。把callbacks參數ReduceLR註解掉，validation loss 卻不會動，所以我想是過擬合了，之後改用輸入層Dropout(0.2)，隱藏層Dropout(0.5)，才正常收斂，大概在epochs=15的時候val\_loss才沒繼續下降，代表收斂。

# RNN vs. LSTM

# 因為繹安學長說將LSTM層直接改成SimpleRNN就可以跑了，但起初因為我用原先的LSTM(Dropout(0.2))跟recurrent\_dropout，發現plot不出任何值來，後來我將dropout拿掉後，才有值出來，所以我在猜可能是因為SimpleRNN的梯度遺失的關係，所以造成以上結果。另外效能方面，SimpleRNN 訓練的速度比LSTM快，但是在準確率和loss上，比LSTM差很多。

# 