

學號：r06942095 系級：電信碩一 姓名：劉翔瑜

1. (1%)請比較有無 normalize(rating)的差別。並說明如何 normalize.

(collaborator:)

以 kaggle 上的分數比較

normailze 前 : 0.856 → normalize 後 : 1.09

準確度變低了!

Normailze 之方法為 : 取出 rate 之  $\text{var}(\text{sigma}^2)$  和 mean

$\text{rate} = (\text{rate} - \text{mean}) / \text{sigma} + \text{mean}$

predict 後的處理 :

$\text{test} = (\text{predict} - \text{mean}) * \text{sigma} + \text{mean}$

2. (1%)比較不同的 latent dimension 的結果。

(collaborator:)

\*測試時 user 及 movie 的 embedding dim 相同

dim=20: public=0.86277

dim=30: public=0.86079

dim=40: public=0.85742

dim=50: public= 0.85693

dim=70: public= 0.85408

由上可知 embedding dim 越高，所得出的資料會越準確

3. (1%)比較有無 bias 的結果。

(collaborator:)

加上 bias : 0.856

無 bias : 0.870

由於某些人會偏好以較高的分數評每部電影，有些則相反，加上 bias 可以更好的區別出這個差異，因此較準確

4. (1%)請試著用 DNN 來解決這個問題，並且說明實做的方法(方法不限)。並比較 MF 和 NN 的結果，討論結果的差異。

(collaborator:)

Layer (type)	Output Shape	Param #	Connected to
movie (InputLayer)	(None, 1)	0	
user (InputLayer)	(None, 1)	0	
movie_bias (Embedding)	(None, 1, 3)	11859	movie[0][0]
user_em (Embedding)	(None, 1, 40)	241640	user[0][0]
movie_em (Embedding)	(None, 1, 40)	158120	movie[0][0]
user_bias (Embedding)	(None, 1, 3)	18123	user[0][0]
flatten_4 (Flatten)	(None, 3)	0	movie_bias[0][0]
flatten_1 (Flatten)	(None, 40)	0	user_em[0][0]
flatten_2 (Flatten)	(None, 40)	0	movie_em[0][0]
flatten_3 (Flatten)	(None, 3)	0	user_bias[0][0]
dense_2 (Dense)	(None, 1)	4	flatten_4[0][0]
dot (Dot)	(None, 1)	0	flatten_1[0][0] flatten_2[0][0]
dense_1 (Dense)	(None, 1)	4	flatten_3[0][0]
add_1 (Add)	(None, 1)	0	dense_2[0][0] dot[0][0] dense_1[0][0]
Total params: 429,750			

此為將原用做 bias 的 embedding layer 調整，變為 dim=3，各輸入一個個數=1 的 Dense layer 最後與 dot 相加，最後 loss=0.854

Layer (type)	Output Shape	Param #	Connected to
movie (InputLayer)	(None, 1)	0	
user (InputLayer)	(None, 1)	0	
user_em (Embedding)	(None, 1, 40)	241640	user[0][0]
movie_em (Embedding)	(None, 1, 40)	158120	movie[0][0]
movie_bias (Embedding)	(None, 1, 1)	3953	movie[0][0]
flatten_1 (Flatten)	(None, 40)	0	user_em[0][0]
flatten_2 (Flatten)	(None, 40)	0	movie_em[0][0]
user_bias (Embedding)	(None, 1, 1)	6041	user[0][0]
flatten_4 (Flatten)	(None, 1)	0	movie_bias[0][0]
dot (Dot)	(None, 1)	0	flatten_1[0][0] flatten_2[0][0]
flatten_3 (Flatten)	(None, 1)	0	user_bias[0][0]
add_1 (Add)	(None, 1)	0	flatten_4[0][0] dot[0][0] flatten_3[0][0]
Total params: 409,754			
Trainable params: 409,754			
Non-trainable params: 0			

而原本直接將 embedding layer (dim=1)輸出作為 bias，loss=0.860

因此加入了 DNN，有可能使結果變好

5. (1%)請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當作 label 來作圖。

(collaborator:)

6. (BONUS)(1%)試著使用除了 rating 以外的 feature, 並說明你的作法和結果，結果好壞不會影響評分。

(collaborator:)

\*由於 kaggle 次數用光了，以 training data 中的 valid set 來做比較

加入 users.csv 中性別的 data 做測試，M->[1 0],F->[0,1]

將此資料一起套入 model，另外接一個神經個數=1 之 dense layer，最後與 dot,bias 相加即為 output，將此 model 做訓練

原本加上 bias 之 loss=0.853

再加上性別之  $\text{loss}=0.855$

結果些許變差，原本預計會變更好