

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

Normalize 的作法為

$$\text{Rates}[i] = (\text{Rates}[i] - \text{mean}(\text{Rates})) / \text{std}(\text{Rates});$$

結果如下，其實沒有什麼差別，normalize 只有幫助收斂更快而已。

	testing private score	testing public score
w/o normalization	0.88810	0.88642
with normalization	0.88525	0.88512

2. (1%)比較不同的latent dimension的結果。

取 latent dimension = 7777, 777, 77 做測試，結果如下，最好的latent dimension 大約落在777。

dim	testing private score	testing public score
7777	0.97256	0.97236
777	0.88810	0.88642
77	0.88925	0.89118

3. (1%)比較有無bias的結果。

測試使用latent dimension = 777. 結果如下:

可以看出雖然只有一點，但加入bias 的ERROR 是比較低的。

	testing private score	testing public score
w/o bias	0.88921	0.88863
with bias	0.88810	0.88642

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。

NN model 如下: (將兩個embedding concatenate 起來，再接fully connected layers)

Layer (type)	Output Shape	Param #	Connected to
user_input (InputLayer)	(None, 1)	0	
item_input (InputLayer)	(None, 1)	0	
embedding_1 (Embedding)	(None, 1, 777)	4693080	user_input[0][0]
embedding_2 (Embedding)	(None, 1, 777)	3070704	item_input[0][0]
flatten_1 (Flatten)	(None, 777)	0	embedding_1[0][0]
flatten_2 (Flatten)	(None, 777)	0	embedding_2[0][0]
concatenate_1 (Concatenate)	(None, 1554)	0	flatten_1[0][0] flatten_2[0][0]
dense_3 (Dense)	(None, 168)	261240	concatenate_1[0][0]
dropout_1 (Dropout)	(None, 168)	0	dense_3[0][0]
dense_4 (Dense)	(None, 64)	10816	dropout_1[0][0]
dropout_2 (Dropout)	(None, 64)	0	dense_4[0][0]
dense_5 (Dense)	(None, 32)	2080	dropout_2[0][0]
dropout_3 (Dropout)	(None, 32)	0	dense_5[0][0]
dense_6 (Dense)	(None, 1)	33	dropout_3[0][0]
Total params: 8,037,953			
Trainable params: 8,037,953			
Non-trainable params: 0			

	testing private score	testing public score
NN	0.89549	0.89762
MF	0.88810	0.88642

結果上來說NN 的效果並沒有比較好，但由於這次的validation loss 和testing loss 實在是相差太多，因此很難調NN 參數，否則應該NN 會有更好的表現。

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

將電影種類編號如下：

{'War': 10, 'Crime': 6, 'Sci-Fi': 12, 'Documentary': 5, 'Animation': 13, 'Thriller': 3, 'Film-Noir': 14, 'Comedy': 0, 'Horror': 7, 'Action': 1, 'Musical': 17, 'Romance': 4, 'Children's': 9, 'Mystery': 11, 'Western': 8, 'Fantasy': 16, 'Drama': 2, 'Adventure': 15}

使用 10000 筆資料作圖，結果如下：

