

學號：R06945003 系級：生醫電資碩一 姓名：林鈺盛

我使用的 MF model 以及 training 過程:

Layer (type)	Output Shape	Param #	Connected to
input_52 (InputLayer)	(None, 1)	0	
input_53 (InputLayer)	(None, 1)	0	
embedding_153 (Embedding)	(None, 1, 256)	1546240	input_52[0][0]
embedding_154 (Embedding)	(None, 1, 256)	1011712	input_53[0][0]
reshape_100 (Reshape)	(None, 256)	0	embedding_153[0][0]
reshape_101 (Reshape)	(None, 256)	0	embedding_154[0][0]
embedding_155 (Embedding)	(None, 1, 1)	6040	input_52[0][0]
embedding_156 (Embedding)	(None, 1, 1)	3952	input_53[0][0]
dropout_5 (Dropout)	(None, 256)	0	reshape_100[0][0]
dropout_6 (Dropout)	(None, 256)	0	reshape_101[0][0]
reshape_102 (Reshape)	(None, 1)	0	embedding_155[0][0]
reshape_103 (Reshape)	(None, 1)	0	embedding_156[0][0]
dot_23 (Dot)	(None, 1)	0	dropout_5[0][0] dropout_6[0][0]
add_41 (Add)	(None, 1)	0	reshape_102[0][0] reshape_103[0][0]
add_42 (Add)	(None, 1)	0	dot_23[0][0] add_41[0][0]

latent dim = 256

batch size = 10000

validation = 0.05 of training data

epoch = 78

optimizer = Adam

1. (1%)請比較有無 **normalize(rating)**的差別。並說明如何 **normalize**.
(collaborator:NO)

Normalize 方法：(rating-means)/std

本題結果原先在 kaggle public 為 0.84586

normalize 之後得到的結果為 0.85454，沒有預想中有好的結果，我想原因可能是經過 normalize 後會產生負數的 rating，造成 training 過程中在計算 loss(rmse)有所影響。

2. (1%)比較不同的 **latent dimension** 的結果。
(collaborator:NO)

嘗試了幾種的 latent dimension

64 : 0.85092

128 : 0.84713

256 : 0.84563

512 : 0.84745

以上皆為 public 的分數，可以發現 256 有較好的結果

3. (1%)比較有無 **bias** 的結果。

(collaborator:NO)

無 bias = **0.84745**

有 bias = **0.84563**

可以發現加了 bias(movie 及 user 都有加)後果然有較好的結果

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:NO)

batch size = 512

optimizer = Adam

loss = rmse

epoch 在 5 之後 val_loss 就沒有在進步了

訓練方法為將 user、movie 兩個 embedding concatenate 之後，也將 user

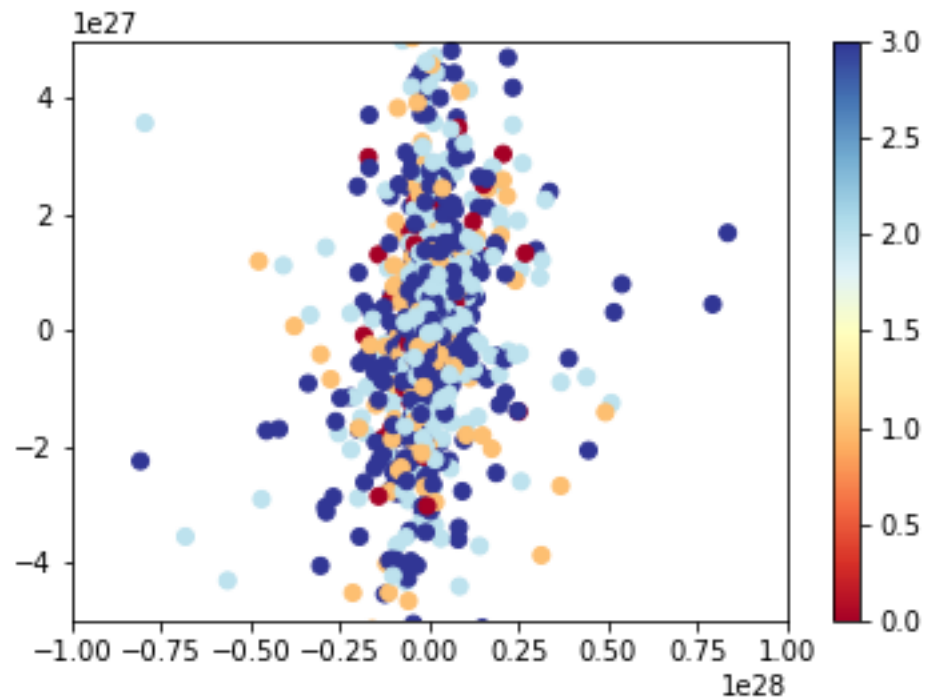
bias、movie bias 的 embedding concatenate，之後兩者相加丟進 DNN。

kaggle public : **0.86656**

結果還滿不理想的，我想原因是因為我沒有仔細的去調整參數，而且很容易就 overfit 了，或許疊的層數不能太多。

=====			
input_17 (InputLayer)	(None, 1)	0	
input_18 (InputLayer)	(None, 1)	0	
embedding_25 (Embedding)	(None, 1, 256)	1546240	input_17[0][0]
embedding_26 (Embedding)	(None, 1, 256)	1011712	input_18[0][0]
reshape_29 (Reshape)	(None, 256)	0	embedding_25[0][0]
reshape_30 (Reshape)	(None, 256)	0	embedding_26[0][0]
embedding_27 (Embedding)	(None, 1, 256)	1546240	input_17[0][0]
embedding_28 (Embedding)	(None, 1, 256)	1011712	input_18[0][0]
dropout_17 (Dropout)	(None, 256)	0	reshape_29[0][0]
dropout_18 (Dropout)	(None, 256)	0	reshape_30[0][0]
reshape_31 (Reshape)	(None, 256)	0	embedding_27[0][0]
reshape_32 (Reshape)	(None, 256)	0	embedding_28[0][0]
concatenate_15 (Concatenate)	(None, 512)	0	dropout_17[0][0] dropout_18[0][0]
concatenate_16 (Concatenate)	(None, 512)	0	reshape_31[0][0] reshape_32[0][0]
add_8 (Add)	(None, 512)	0	concatenate_15[0][0] concatenate_16[0][0]
dense_19 (Dense)	(None, 256)	131328	add_8[0][0]
dense_20 (Dense)	(None, 128)	32896	dense_19[0][0]
dense_21 (Dense)	(None, 1)	129	dense_20[0][0]
=====			

5. (1%)請試著將 **movie** 的 **embedding** 用 **tsne** 降維後，將 **movie category** 當作 **label** 來作圖。
(collaborator:NO)



紅:Animation or Children' s

黃:Action or Adventure

淺藍 : Drama

深藍:others

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:NO)

我加進了 movie 的 genres 當作 bias, 作法就跟先前的 MF 一樣, 只是除了先前的兩個 bias 之外, 多加了 genres 這個 embedding

得到的結果是 0.84573