

1. (1%)請比較有無normalize(rating)的差別。並說明如何normalize.
(collaborator: None)

我做 normalize 的方式採用 min-max scaling。Parse 出 rating 的值後分別計算 rating 的平均與標準差。接著每個 rating 值修改如下：

$$\text{rating} = (\text{rating} - \text{mean}) / (\text{std} + 1\text{e-}100)$$

將此處理過的 rating 拿去訓練。在 test 的時候，要記得把預測出來的結果轉換回來！（一開始忘記轉很慘）

$$\text{predict} = \text{predict} * \text{std} + \text{mean}$$

model	public score
無 normalize	0.87295
有 normalize	0.86582

2. (1%)比較不同的latent dimension的結果。
(collaborator: None)

這題我切 10% 測資作為 validation set。並使用 RMSE 來收斂，BATCH_SIZE 調 512，然後 EPOCH 設為 20。並調整不同的 embedding dimension。

EMBEDDING_DIM	training_loss	training_RMSE	valid_loss	valid_RMSE
64	0.7103	0.8091	1.0904	1.0077
200	0.6207	0.716	1.1571	1.0144
512	0.5262	0.5811	1.2436	1.0170

可以看到分數有點差異，但丟上 kaggle 卻都拿到差不多的分數。

3. (1%)比較有無bias的結果。
(collaborator: None)

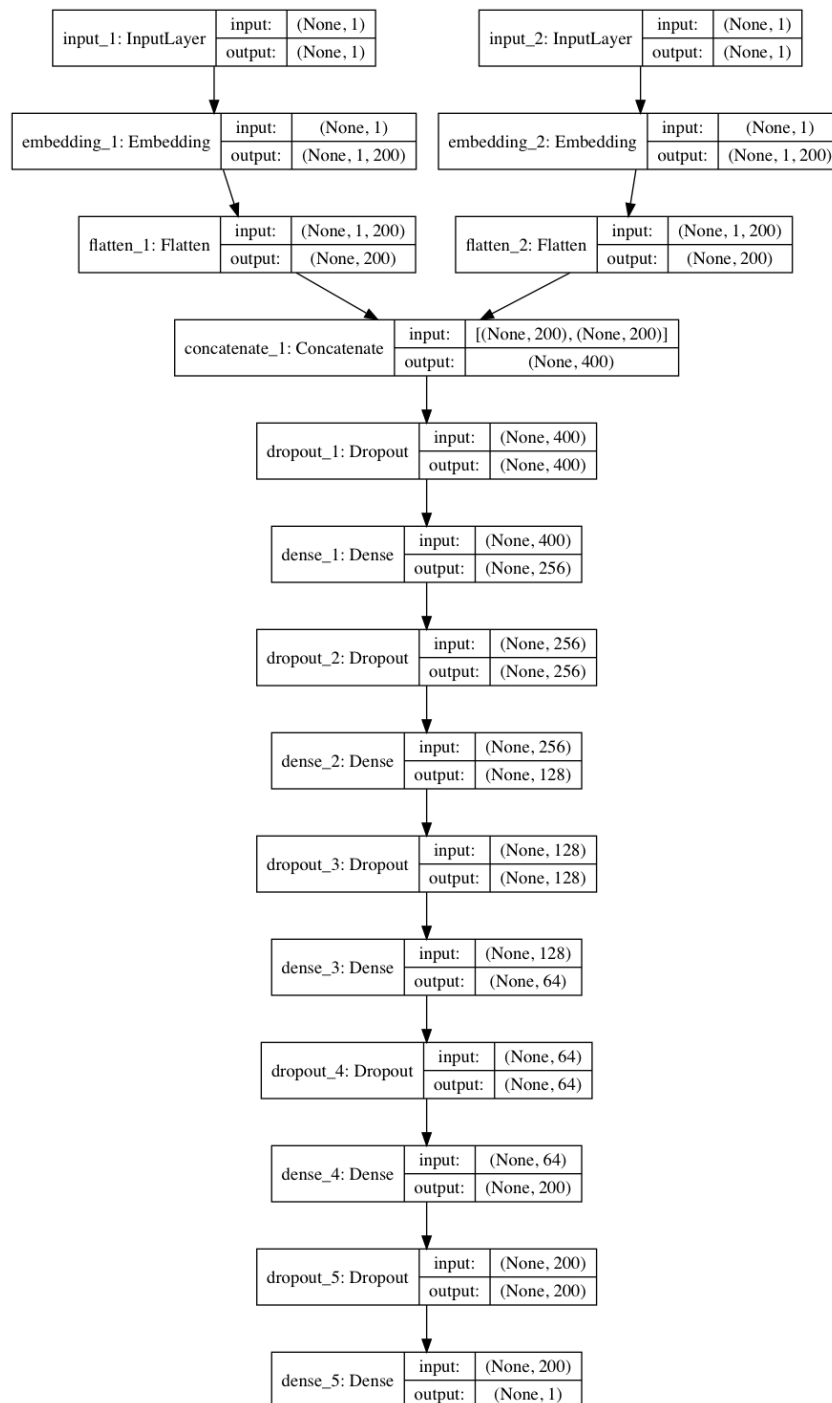
收先定義 bias，另外參數設定與上題一樣，EMBEDDING_DIM 採用 200。

```
U_bias = Embedding(MAX_USERID, 1, embeddings_regularizer=l2(0.00001))(U_input)
U_bias = Reshape((1,))(U_bias)

V_bias = Embedding(MAX_USERID, 1, embeddings_regularizer=l2(0.00001))(V_input)
V_bias = Reshape((1,))(V_bias)
```

	training_loss	training_RMSE	valid_loss	valid_RMSE
有 bias	0.6207	0.716	1.1571	1.0144
無bisa	0.6317	0.739	1.1592	1.0161

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: None)



我把 user 與 movie 的資料過 embedding layer 後各自做 flatten，之後再把他們接在一起，然後過三層 Dnn。

	RMSE	Public Score
Deep model	0.6202	0.89454
MF model	0.7009	0.87295

原本看到 DNN 訓練完後的分數想說不錯，結果丟上去分數反而有點差，我想可能是 DNN 在訓練時充分用到所有資訊，但是訓練結果是 overfit 的。

5. (1%)請試著將movie的embedding用tsne降維後，將movie category當作label來作圖。
(collaborator: None)

我的 model 在 Drama | Musical 蠻分散的，沒有規律，而 Fantasy 是集中在中下角。

