

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

(collaborator:)

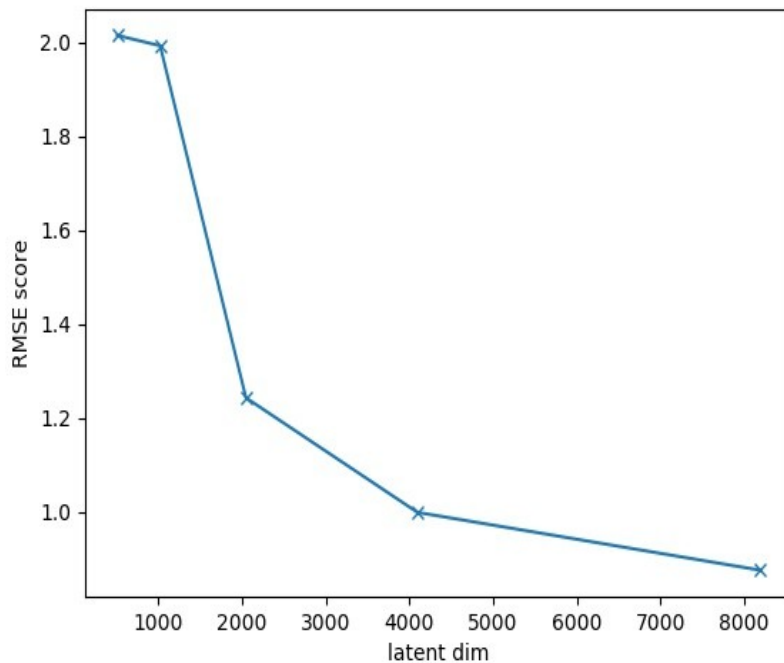
在這裏我做了兩種不同的方法對 data 進行處理, 並與沒有處理的 data 相比較. 因為 rate 只在範圍 1-5,

(a) 直接把 Data 壓到 0~1 之間: $x' = (x - 1)/4$,(b) 把 Data 取平均再除 standard deviation: $x' = (x - \langle x \rangle) / x_std$,**Kaggle score (Public) :**

Without Process	(a)	(b)
0.89512	0.88227	0.87621

2. (1%)比較不同的 **latent dimension** 的結果。

(collaborator:)



以上結果為 average over 15 次實驗, 並用 20% data 當 validation 的 RMSE score 為 y 軸. Data normalized 並有加 bias.

由以上實驗結果可以看到 latent dimension 在可以 simulate 的範圍內, 越大, performance 越好

3. (1%)比較有無 **bias** 的結果。

(collaborator:)

Validation score:

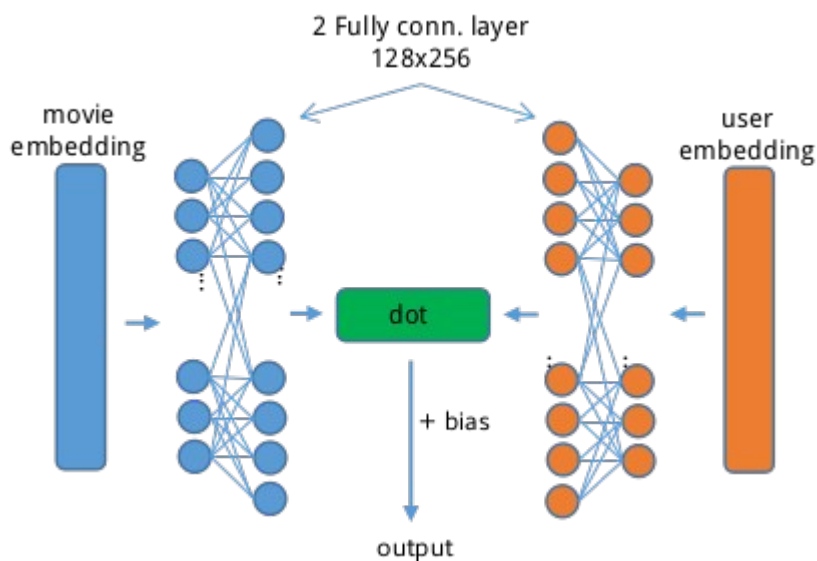
有 bias	無 bias
0.87754	0.89210

由以上結果我們發現有加 bias 顯著提升 performance .

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:)

DNN model 架構如下, 將兩個 embedding 分別過 DNN 後內積 再加上 bias 後 output.



Kaggle score (Public) :

MF	NN
0.88295	0.87621

*Data is normalized.

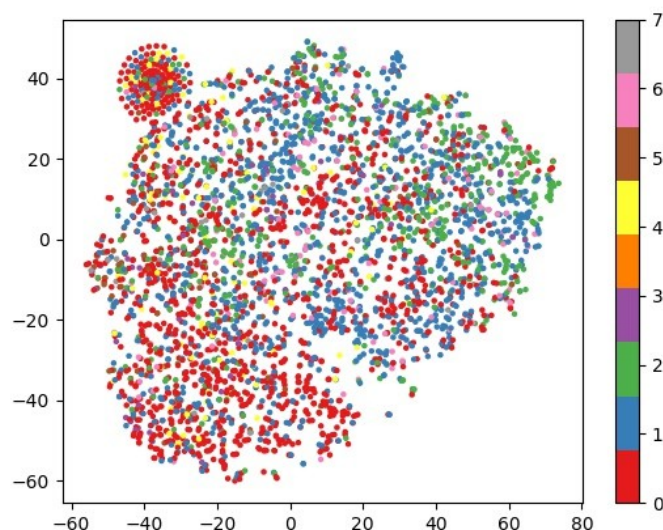
For NN, 2 Fully conn. Layers are with dropout 0.2 in training. learning rate 0.001, L2 regularization $10e-6$, using Adam optimizer training with 12 epoch. For MF, learning rate 0.005, L2 regularization $10e-7$, using Adam optimizer training with 15 epoch.

由上述結果發現 NN 比 MF 效果來得好，而由之前作業我們知道, multi-layer 比 single layer 描述一個 model 來得有效率, performance 較好. 因此上述結果是符合預期的.

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

(collaborator:)

0	Romance, Drama, Musical
1	Comedy, Animation, Children's, Action, Adventure
2	Crime, Thriller, Horror
3	War
4	Documentary
5	Mystery, Film-Noir
6	Fantasy, Sci-Fi
7	Western

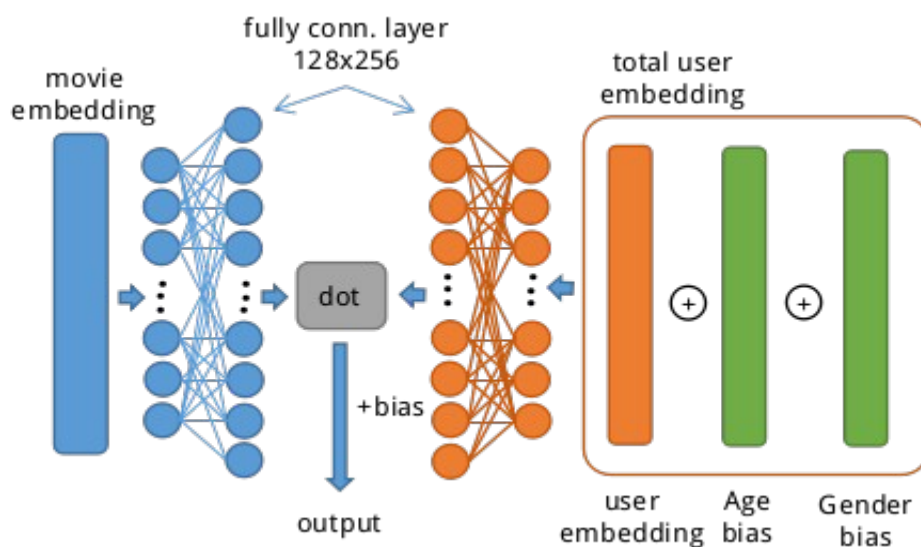


6. (BONUS)(1%)試著使用除了 **rating** 以外的 **feature**, 並說明你的作法和結果，結果好壞不會影響

評分

(collaborator:)

考慮進其他的 feature, 我修改 NN model , 另外對每一個 feature 作一層 bias 並把這些 bias vector 與 user embedding 相加後當作 user 的 total embedding vector,再與 movie 的 embedding vector 作內積. 架構如下:



*Data is normalized.

For NN, learning rate 0.001, L2 regularization 10e-6, using Adam optimizer training with 12 epoch. For NN+extra feature model : Batch size =256, embedding dimension = 128. learning rate = 0.002, L2 regularization= 10e-7. using Adam optimizer with 14 epoch. The fully conn layers in both model are with dropout 0.1 in training.

validation score :

NN	NN+ extra feature
0.87812	0.86912

由上述實驗結果可以發現, 在考慮進 user 的 extra feature 後, performance 提升.