學號：R06922075 系級： 資工碩一 姓名：翁瑋

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

Normalize方法為對rating標準化，將每個值減掉平均後除掉標準差，使得整體的標準差為1，平均為0，以下表格為比較有無標準差後的訓練表現與kaggle結果：

	沒有標準化	有標準化
Local train val_loss	0.8841	0.7651
kaggle public score	0.88685	0.85535
kaggle private score	0.88516	0.85455

loss採用自訂的rmse做標準，可以看到有標準化後的表現比較好，而且在訓練時有經過標準化的資料收斂速度也較快。

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

latent dimension	kaggle public score	kaggle private score
16	0.92331	0.92050
32	0.89668	0.89652
64	0.87339	0.87481
128	0.87098	0.86927
256	0.86449	0.86428

可以看到，隨著latent dimension的上升，在kaggle上的rmse也跟著減少，我認為是因為較多的維度可以做出來的組合越多,在預測方面也能更精準。

3. (1%)比較有無bias的結果。
   (collaborator:)

	有bias	無bias
Local train val_loss	0.8712	0.8779
kaggle public score	0.87098	0.88000
kaggle private score	0.86927	0.88152

沒有bias的模型在表現上略遜有bias的模型一籌，大概是因為對於特定的User跟Movie真的有特定的評分偏差。

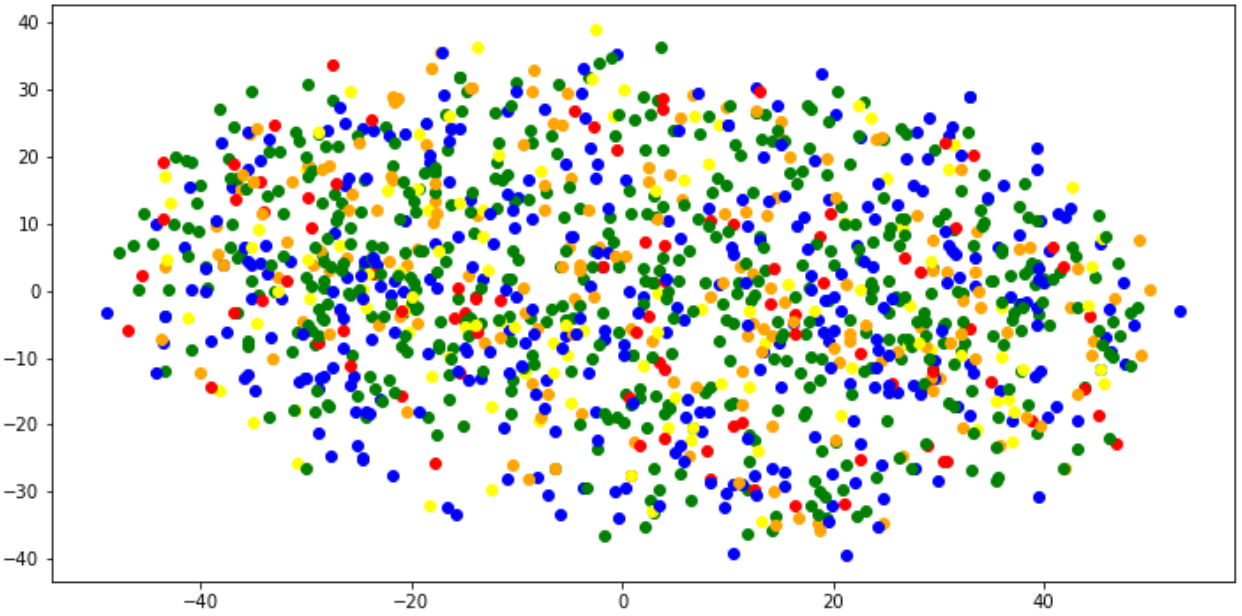
4. (1%)請試著用DNN來解決這個問題，並且說明實做的方法(方法不限)。並比較MF和NN的結果，討論結果的差異。
(collaborator:)

一樣將User跟Movie先embedding成128的vector，再使用concatenate層將兩個向量合成一個256維的向量，後接3層DNN，分別為256,128,64，用relu當activation，最後一層輸出一維，activation使用linear，以下為比較：

	MF	DNN
kaggle public	0.85535	0.87857
kaggle private	0.85455	0.87808

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

試著用sklearn的TSNE將原本128維的向量降為2維的向量，並用以下的分類方式將每部電影標上顏色：Children's, Animation, Comedy, Musical, Action, Adventure, Fantasy, Sci-Fi, Drama, Romance, Documentary, Western, Thriller, War, Mystery, Horror, Crime, Film-Noir



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:)

使用User的Gender作為bias，觀察是否不同的性別對於電影評分會有特定高分或低分的偏差，將每個使用者的性別依照男女編為0或1，並將0/1 embedding成一個特定的值當作bias，最後跟原本dot的結果還有movie，user的bias加起來當作output，以下為kaggle結果:

	有使用Gender	沒有使用Gender
kaggle public	0.87462	0.86093
kaggle private	0.87906	0.86415