學號：R06943153 系級： 電子碩一 姓名：蘇旻彥

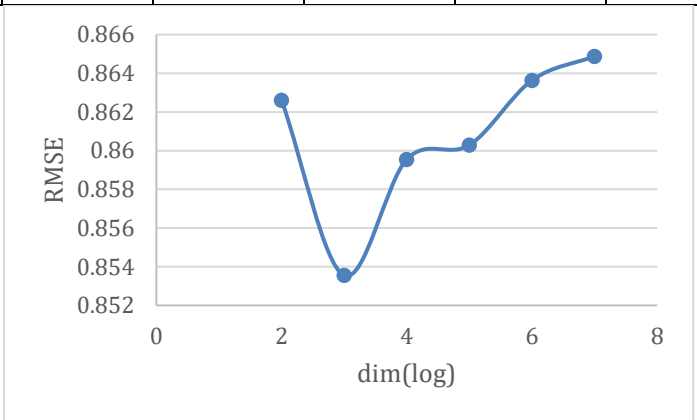
1. (1%)請比較有無 normalize(rating)的差別。並說明如何 normalize.
   (Collaborators: R06943147 李明庭、R06943084 游彥勝)
   以相同的 MF model 和相同條件做實驗(dimendion=8, batch_size=128)結果如下：

	有 normalize	無 normalize
Kaggle Public Score	0.85355	0.86056

   我做 normalize 的方法為找出 training data 的 rating(Y)的 mean 以及 std，並把每一筆資料做以下處理：y = (y - y_mean)/y_std。在 predict 的時候再反過來 predict = (predict * y_std + y_mean)。由實驗結果得知有 normalize 之後比較好。

2. (1%)比較不同的 latent dimension 的結果。
   (Collaborators: R06943147 李明庭、R06943084 游彥勝)
   從 2^2 做到 2^7 的結果如下：

latent dimension	4	8	16	32	64	128
Kaggle Public Score	0.86261	0.85355	0.85955	0.86029	0.86364	0.86488



   從以上數據可以觀察出 dim=8 的時候表現最好。

3. (1%)比較有無 bias 的結果。
   (Collaborators: R06943147 李明庭、R06943084 游彥勝)
   以相同的 MF model 和相同條件做實驗(dimendion=8, batch_size=128)結果如下：

	有 bias	無 bias
Kaggle Public Score	0.85355	0.86199

   由實驗結果可知有 bias 的分數比較好，推測是因為模型的複雜度比較高。

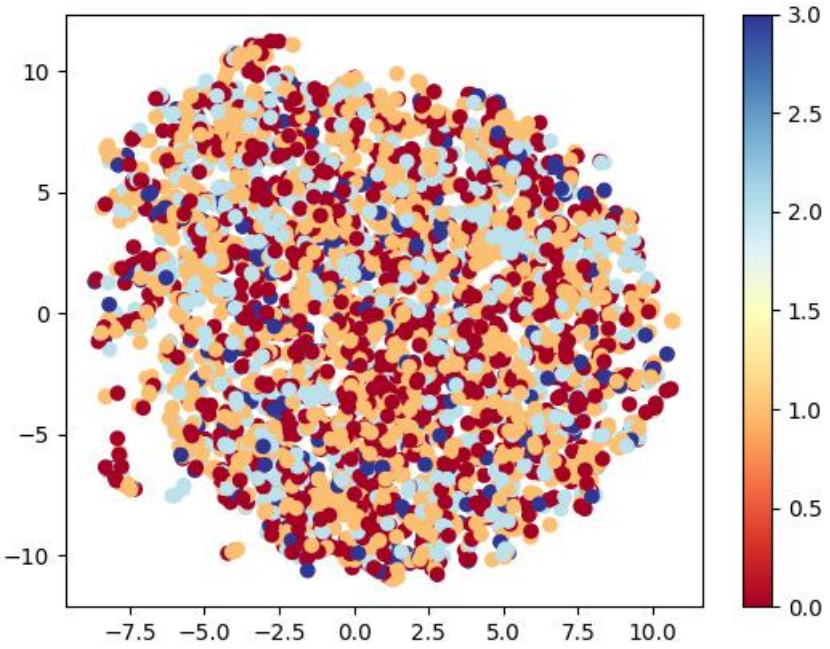
4. (1%)請試著用 DNN 來解決這個問題，並且說明實做的方法(方法不限)。並比較 MF 和 NN 的結果，討論結果的差異。
   (Collaborators: R06943147 李明庭、R06943084 游彥勝)
   我實作 DNN 的方式為：在 users 和 movies 進入 dot 層之前先通過四層 dense

layer，並在最後通過 softmax 選擇是哪一個 rating。實驗結果如下：

	MF	NN(加入 dense)
Kaggle Public Score	0.85355	0.85938

由以上數據可以看出以這個方法實作 DNN 的效果沒有原本的 MF 來的好，僅僅是輸了一點點，可能是有些細部的 parameters 需要 tune。

5. (1%)請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當作 label 來作圖。
(Collaborators: R06943147 李明庭、R06943084 游彥勝)



嘗試了很多種不同的組合都沒辦法看見很明顯的區分，此結果是將 Drama、Musical 為一類；Crime、Thriller、Horror 一類；Adventure、Animation、Children's 一類，剩餘其他為一類所做的圖。

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。
(Collaborators: R06943147 李明庭、R06943084 游彥勝)
我的作法是把所有的資料抽出 UserID、Gender、Age、Occupation、movieID、Genres 這六種 feature 當作 input 然後輸出是 rating，使用 classify 的方法把所有結果分成五種分數(1~5)，使用的是基本的 DenseNet，使用了六層 Dense Layer，除了 output 層使用 softmax 以外其餘皆使用 relu，最後在 Kaggle 上的分數是 0.85938(public score)。