

## HW6 Report

系級：電機三 學號：b04901009 姓名：林孟瑾

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

	Private score(rmse)	Public score(rmse)
有 normalize	0.85519	0.86237
無 normalize	0.85650	0.86283

將 rating 扣掉 rating 的平均，再除以 rating 的標準差，得到的就是 normalize 後的 rating。而在做 testing 時，將 predict 出來的結果，乘以 rating 的標準差，再加回 rating 的平均。由結果可以看出，經過 normalize 之後的 rmse 較小，效果較好。

2. (1 %)比較不同的 embedding dimension 的結果。  
(train.csv 中的 0.1 比例作為 validation data)

	Validation rmse
latent_dim = 20	0.85731
latent_dim = 50	0.85578
latent_dim = 100	0.85132
latent_dim = 150	0.84948
latent_dim = 200	0.84829
latent_dim = 500	0.84683

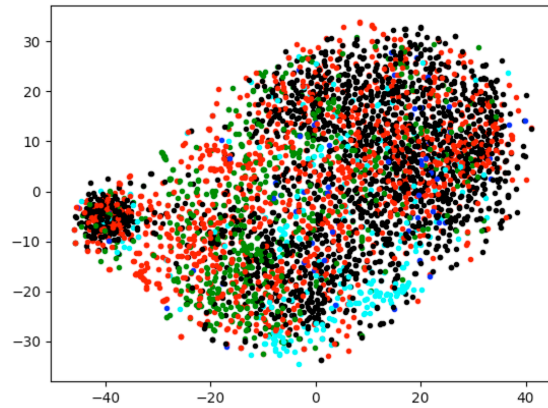
由結果可以看出，” latent\_dim = 500” 時 rmse 最小，效果最好。

3. (1 %)比較有無 bias 的結果。

	Private score(rmse)	Public score(rmse)
有 bias	0.85650	0.86283
無 bias	0.86732	0.87316

由結果可以看出，有加上 bias 的訓練出來的 model 預測較為準確，rmse 較小，效果較好。

4. (1 %)請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當作 label 來作圖。



點的顏色	電影分類
淺藍色	Animation, Children' s
黑色	Romance, Drama, Musical
綠色	Adventure, Fantasy, Action, Sci-Fi, Film-Noir
藍色	War, Western
紅色	Crime, Thriller, Horror

可以看出，綠色類的較為集中於左下角，淺藍色較為集中於下方，而紅色呈現散落的分佈，黑色集中於左方一小球和右側。

5. (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。

我使用了電影的種類還有使用者的年齡、性別、職業，作為額外的 feature。電影的種類和使用者的職業分別都處理成 one hot encoding。因此，model 的 input layer 多了四層，再將此四種 feature 經過 Dense 之後，和原本的 userID 和 movieID 的 input 兩兩做 Dot 層（包含 userID 和 movieID 之間的 Dot，總共取 5 種組合），再將這 5 個 Dot 層 Concatenate 起來，然後此結果再和原本的 bias 層 Add 起來。結果為 private score 為 0.86563，public score 為 0.86851，沒有比原本的 model 進步。

（原本的 model，private score 為 0.85650，public score 為 0.86283）