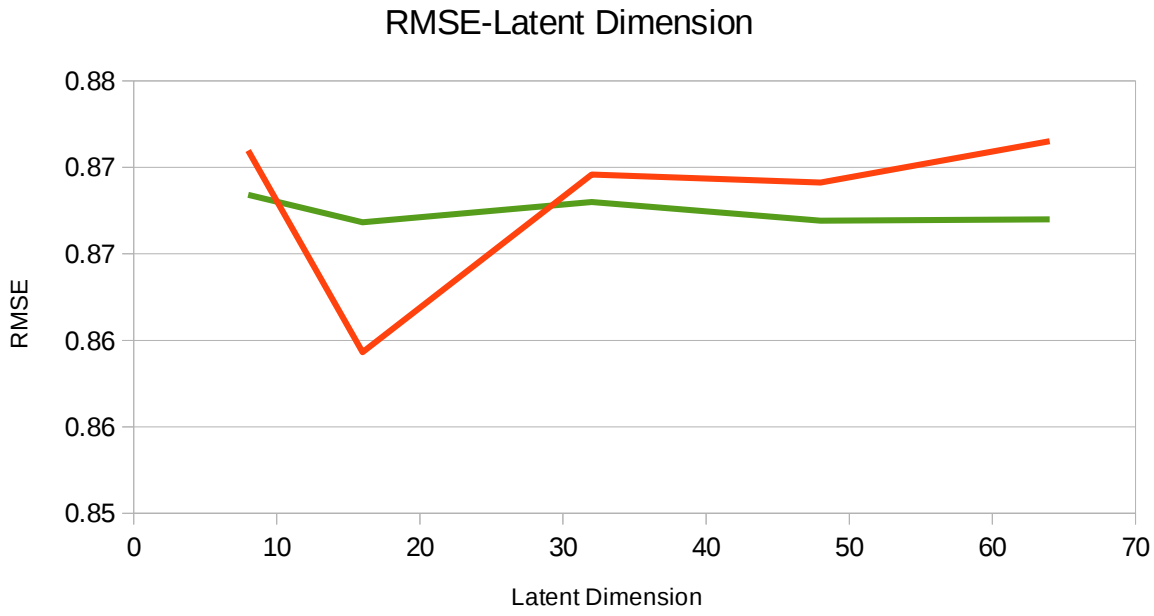
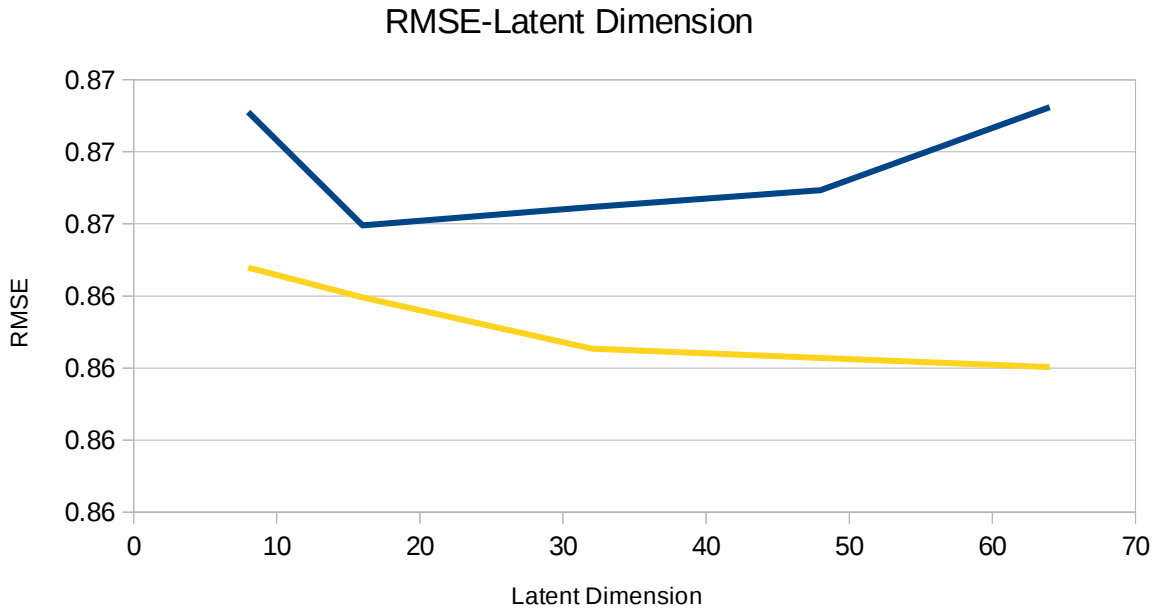
學號：B05901189 系級： 電機一 姓名：吳祥叡

1. (1%)請比較有無 normalize(rating)的差別。並說明如何 normalize.
   如同 TA hour 投影片所述，將 training data 的 rating 做 normalize 使得標準差為 1 平均值為 0。testing 的時候將 predict_test=predict_test*train_std+train_mean。以下結果為有 bias 的 loss，改變不同 latent dimension 和調整有無 normalization 的比較。



綠色為無 normalize 的 private score，紅色為有 normalize 的。在大部分情況下沒有 normalize 反而表現較好。

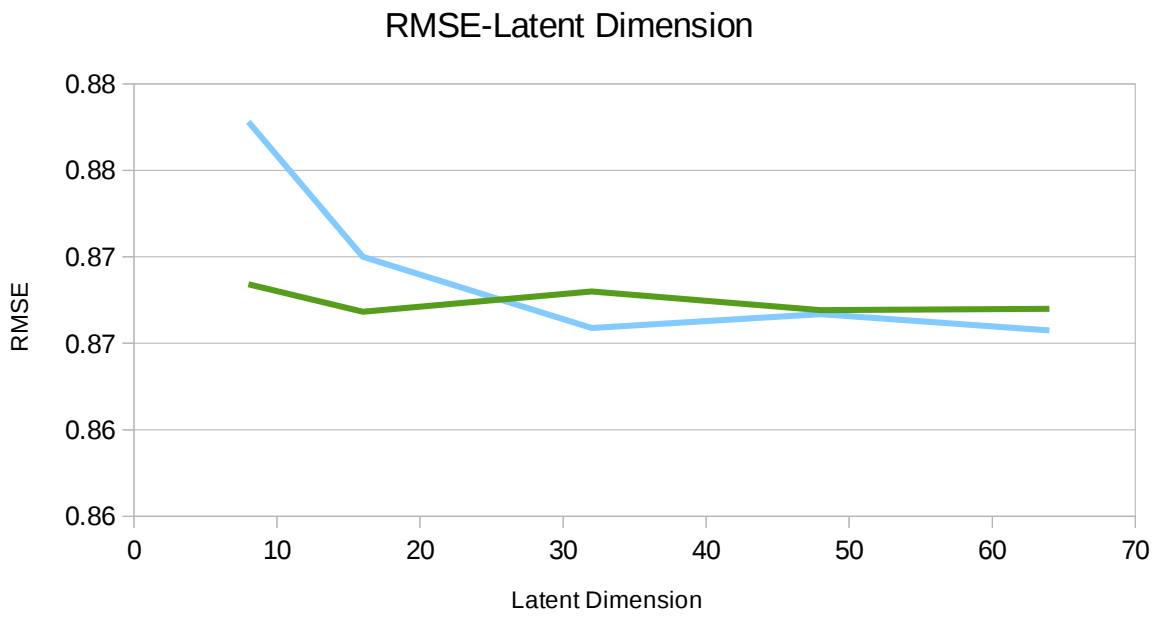


上圖為同樣方法做的 public score，藍色是有 normalize 的，黃色則沒有。

結果顯示沒有對 rating 做 normalize 結果反而較好。

2. (1%)比較不同的 latent dimension 的結果。
   固定做第一題的 normalize，且使用 bias 調整 latent dimension 訓練得到上圖的結果。由第一題的兩張圖顯示在 latent dimension 為 16 附近有較佳的表現。

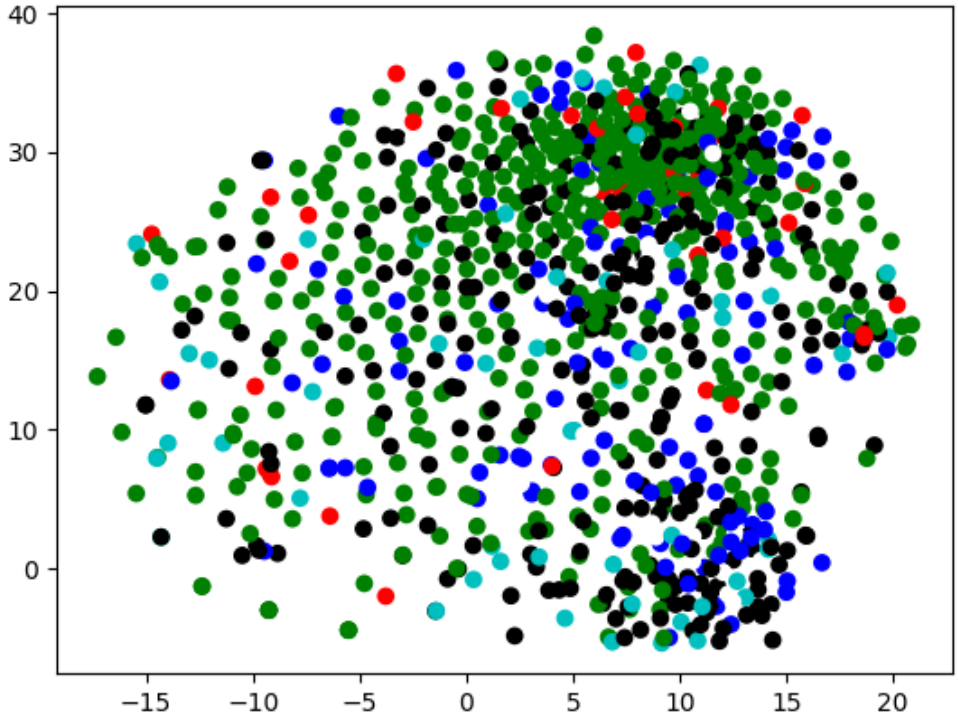
3. (1%)比較有無 bias 的結果。
   不將 rating 做 normalization，調整 latent dimension，比較有無 bias 的 private score，作圖如下：



   藍色為無 bias，綠色為有 bias。如圖可見在 dimension 太小的時候無 bias 會有較大的 error，推測因為 bias 顯示了不同 user 和不同 movie 的偏好，等於是 user-wise 和 movie-wise 的一種 normalization。

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。
   實做方法:
   將 movie_id 和 user_id 做 embedding 變成 16 維。
   將兩個 embedding vector concatenate
   通過 x 層 y 個 neuron 的 DNN。
   實做 x=2 y=500，public score 為 0.87113
   x=2 y=200，public score 為 0.87111 ，表現均不如 MF。

5. (1%)請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當



作 label 來作圖。

圖中綠色為 Drama 或 Musical
黑色為 Comedy 或 Romance
深藍色為 Thriller 或 Horror 或 Crime
紅色為 Documentary 或 Film-Noir 或 War
淺藍色為 Animation 或 Children's 或 Adventure 或 Fantasy.

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。

先將 user_id 和 movie_id 分別做 32 維的 embedding，最後將兩個 embedding 和 users 的其他 feature 全部 concatenate 並通過 3 層 500 個 neuron 的 DNN。但最好的結果 rmse 也只能到 0.88，和 MF 相差甚遠。
對 user feature 做 normalization 也沒有進步。