Report

電機三 b04502139 戴瑋辰

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

   Normalize 方法：
   針對 rating 進行 normalize，將 training data 的 rating 進行標準化，並把平均 (rating_mean) 與標準差 (rating_std) 記起來，要取得 predict data 時，把 MF 算出來的值*rating_std+rating_mean 即可。

   MSE

	有 normalize	無 normalize
Kaggle score	0.87099	0.87236

   由表中可知經過 normalize 過後的 rating 訓練出來的 model 稍微好一點點。在做 matrix factorization 時，有無 normalize 並不會有很大的區別。

2. 比較不同的 embedding dimension 的結果

Embedding dimension	5	10	50	100	666
Val loss	0.8662	0.864	0.8627	0.8579	0.8661

   從表中可以看出，embedding dimension 對於結果沒有太大的影響，要要超過五，幾乎都可以 train 得起來。在經過測試之後發現將其設為五可以 train 比較快且 kaggle 上分數也較好。

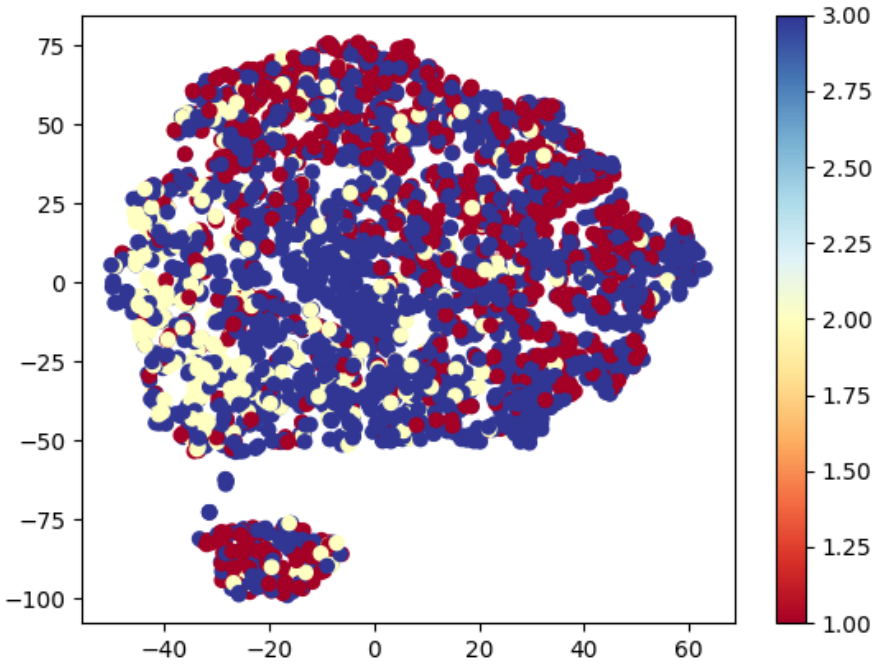
3. 比較有無 bias 的結果

	有 bias	無 bias
Val loss	0.8662	0.8912

從表中可以看出 bias 對整體的影響很大，如果有把每一個對應到的人的 bias 考慮進去，train 出來的東西會好很多。

4. 請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當作 label 來做圖。

   Drama, Musical 歸為一類，以紅點表示
   Thriller, Horror, Crime 歸為一類，以白點表示
   其餘歸為一類，以藍點表示



5. 試著使用除了 rating 以外的 feature，並說明你的做法和結果。

   除了 rating 以外，我另外加入了 movie category 以及 age 來 train。
   作法如下，先將 18 個 movie category 用 BOW 表示，接這將 participant ID, Movie ID, movie category, age 個字接到一層 size 為 128 的 Dense，再將這四個 128 維的東西 concatenate 起來，接到兩層 Dense，其結果比 MF 好了一些

	M F	Improved model
Kaggle score	0.87236	0.86515