學號:R06922152 系級: 資工碩一 姓名:袁晟峻

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

(collaborator: 呂承洋、邵志宇)

200 dimension 無 normalize: kaggle public 0.84536、kaggle private 0.84737 200 dimension 有 normalize: kaggle public 0.90510、kaggle private 0.90662 結果顯示做了 normalize 之後比較差,使用的 normalize 方式是減去平均之後

除標準差。

2. (1%)比較不同的 latent dimension 的結果。

(collaborator: 邵志宇)

200 dimension: kaggle public 0.84536 \ kaggle private 0.84737

128 dimension: kaggle public 0.84814 \ kaggle private 0.84945

100 dimension: kaggle public 0.84837 \ kaggle private 0.84968

3. (1%)比較有無 bias 的結果。

(collaborator: 邵志宇)

200 dimension 有 bias: kaggle public 0.84536、kaggle private 0.84737

200 dimension 無 bias: kaggle public 0.85055、kaggle private 0.85092

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:)

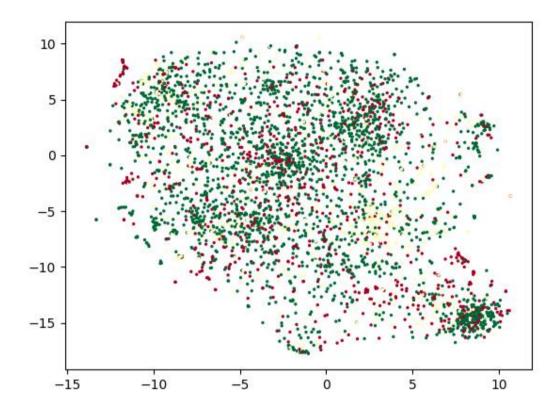
MF: kaggle public 0.84536 \ kaggle private 0.84737

NN: kaggle public 1.12365 \ kaggle private 1.12197

照著投影片講解的將每一筆的 user_id 和 movie_id embedding 後接起來當作一筆訓練資料,然後用 regression 方式丟進 DNN train,loss function 用 mse,train 到 loss 1.1 多的時候就下不去了。

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

(collaborator: 陳柏堯)



紅色: Action, Drama, Documentary, Romance, Mystery, Sci-fi

綠色: Musical, Animation, Children, Adventure, Fantasy, Comedy

黄色: Thriller, Crime, Horror, War, Film-Noir, Western

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:)