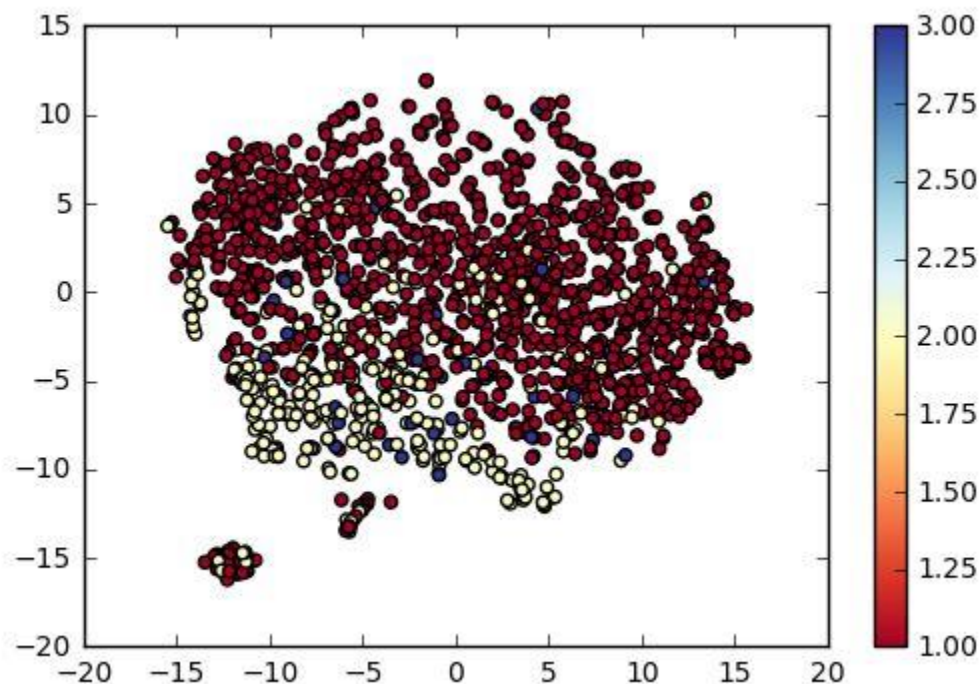


學號：R05945039 系級：生醫電資所碩一 姓名：張宇軒

1. (1%)請比較有無 **normalize(rating)**的差別。並說明如何 **normalize**.
用 train 的 rating 得到 mean 跟 standard_deviation，代入公式 $(x - \text{mean}) / (\text{standard_deviation})$ ，用這個值去做 MF 後再 kaggle 上的分數為 0.86025，與最好的結果 0.85117 相比結果是比較差的，可能是原本的 rating 的級距比較適合這個 model 的估計結果。
2. (1%)比較不同的 **latent dimension** 的結果。
最好的結果 latent dimension 為 400，kaggle 分數 0.85117，用 120 的 kaggle 分數為 0.85268，
3. (1%)比較有無 **bias** 的結果。
在同樣的 latent 跟 model 下，加了 bias 之後結果沒有變好，kaggle 分數 0.86758
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。
實際使用 DNN 可以過 simple baseline，但要往上提昇很困難，DNN 的結果是用 mse 取 loss 得到結果，val_loss 都會卡在 0.77 就會 overfitting，MF 的 val_loss 則可以到 0.73，
5. (1%)請試著將 **movie** 的 **embedding** 用 **tsne** 降維後，將 **movie category** 當作 **label** 來作圖。



上圖是用 comedy 跟 horror 做圖得到的結果，紅色是 comedy 黃色是 horror 藍色是兩者都有的電影，可以看到兩者的分佈有分開來，且藍色會出現在黃紅之間，能看出 comedy 與 horror 是可以用 embedding 的兩維來表現的。

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。