學號：B04902044 系級： 資工二 姓名：朱柏澄

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

normalize 的方法：計算 training set rating 的平均和標準差，將 training set 減掉平均後除以標準差，用 normalize 過的 rating train。test 的時候算出的值要先乘以同樣的標準差再加上同樣的平均，才是預測的答案。

RMSE on public testing set:
沒有normalize: 0.85719
有normalize: 0.85981
無論有沒有加，得到的結果差不多。

training speed:
有normalize 的訓練速度會比沒有還要快，有的話只要 5 epochs 左右 validation rmse 就開始收斂了，而沒有的話至少要 20 epochs 以上。

2. (1%)比較不同的latent dimension的結果。

RMSE on validation set:
dimension = 5: 0.868
dimension = 15: 0.8662
dimension = 25: 0.8598
dimension = 100: 0.8562

latent dimension 不用取很大就有不錯的performance 了，增加 dimension 到 100 提升的幅度並不大，取約10~20 其實就可以了。

3. (1%)比較有無bias的結果。

RMSE on validation set:
有 bias : 0.8662
無 bias : 0.8681

bias 的影響並不大，我覺得是可有可無的一項。

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。

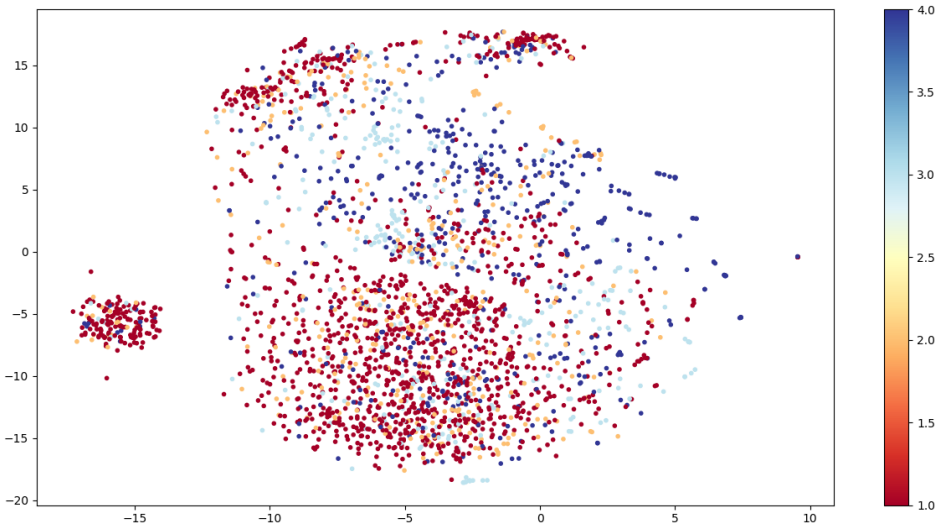
我在 embedding layer 和 output layer 中加了 10個 hidden layer, 每個 hidden layer 都是256個 units

RMSE on validation set:
MF: 0.8673
DNN: 0.8669

我做出來的結果兩者差異不大，因為 MF 的參數我調比較久，而 DNN 只有大概調一下而已，但我想如果多試一些 DNN 模型和參數，DNN 應該會有比較好的結果，因爲它的結構複雜許多。

```python
def get_model(n_users, n_movies, emb_dim):

    in_u = Input(shape=[1])
    in_m = Input(shape=[1])

    emb_u = Embedding(n_users, emb_dim)(in_u)
    emb_m = Embedding(n_movies, emb_dim)(in_m)

    emb_u = Flatten()(emb_u)
    emb_m = Flatten()(emb_m)

    concat = Concatenate()([emb_u, emb_m])

    hidden = Dense(256, activation='relu')(concat)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)
    hidden = Dense(256, activation='relu')(hidden)

    out = Dense(1)(hidden)

    model = Model(inputs=[in_u, in_m], outputs=out)

    return model
```

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



紅色是 Drama, Musical　黃色是 Horror, Crime, Thriller
淺藍是 Adventure, Animation, Children's　藍色是 Action, War