學號：R06921037 系級： 電機所碩一 姓名：許哲瑋

1. **(1%)請比較有無 normalize(rating)的差別。並說明如何 normalize.**
   (collaborator: R06942010 蘇建翰)

	Kaggle Public
Without Normalized	**0.87367**
Normalized	**0.86060**

Normalize 作法:算出所有 Rating 的平均 μ 和標準差 σ 後，rating = (rating- μ)/ σ 。

Predict:將所有 predict 出來的值先乘上標準差 σ 後再加上平均 μ 才是最後的值。

加上 Normalize 效果變好，明顯有進步。

2. **(1%)比較不同的 latent dimension 的結果。**
   (collaborator: R06942010 蘇建翰)

Latent Dim	Kaggle Public
16	**0.86323**
32	**0.86471**
64	**0.86060**
128	**0.86178**

比較不同 Latent Dimension，可以發現其實分數差距不算太大，隨著 dimension 不同而有所起伏。而猜測可能最佳的 Latent Dimension 位於 64~128 之間。

3. **(1%)比較有無 bias 的結果。**
   (collaborator: R06942010 蘇建翰)

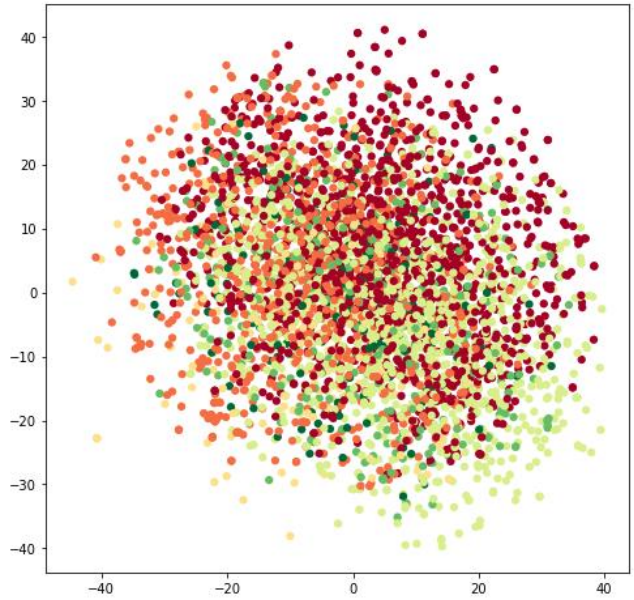
	Kaggle Public
Without bias	**0.86611**
With bias	**0.86060**

加入 bias 後預測結果稍微有改善，分數有變好一些。這樣的成果是可預期的，有些 User 本來就有自己的偏好(偏向把分數打得比較高或是很低)，加入 bias 可以把這樣的情況考慮進來，因此預測結果會比較好一些。

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: R06942010 蘇建翰)

DNN 實作方法:前面和 MF 一樣，都是先把 User 和 Movie 做 embedding(latent dim 這裡使用 64)，接著把 User_embedding 和 Movie_embedding concatenate 在一起 ，接到三層的 NN(128，64，32)，其中在每一層都加了 BatchNormalization 和 dropout，最後接到輸出層，activation 為 linear。但是最後的訓練結果反而有些退步，Kaggle Public 為 0.86492，相較於用 MF 的 0.86060 有點落差。猜測可能是疊了太深參數用了太多，導致讓每個 User 的多維向量都比較偏於擬合 train set，而在 test data 上面的表現較差一些。

5. **(1%)請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當作 label 來作圖。**
   (collaborator: R06942010 蘇建翰)



分類方法如下:

Movie Classes	Label
Animation，Children's，Comedy	0
Crime，Thriller，Horror，Film-Noir	1
Mystery，Fantasy，Sci-Fi	2
Musical，Drama	3
Adventure，Action，Western	4
War，Documentary，Romance	5

其中有一些沒有標註的電影，便將其放入 Label=0 的 class 當中，數量大約有一百多部。受限於分類類別影響(自己的分類方式)，因此點與點之間仍然不是分得很開，但是可以大概觀察出來橘色的點大約偏左上方，綠色的點比較偏右下方，紅色的點比較偏中上方。

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: R06942010 蘇建翰)

將每個 User 的性別和年齡都加入考慮(Occupation 和 Zip-Code 對於電影愛好影響應該較小)，賜予不同的 weight 後結合在 User 的 embedding layer 裡，接著將原來的 User embedding 和性別、年齡層作加總變成新的 User embedding layer，接上 NN 去訓練。出來的結果爛掉了，Kaggle public 為 0.91150，單純的把層與層作線性的加總似乎對於額外資訊的給予不太適合。