

# Homework 6 Report - matrix factorization

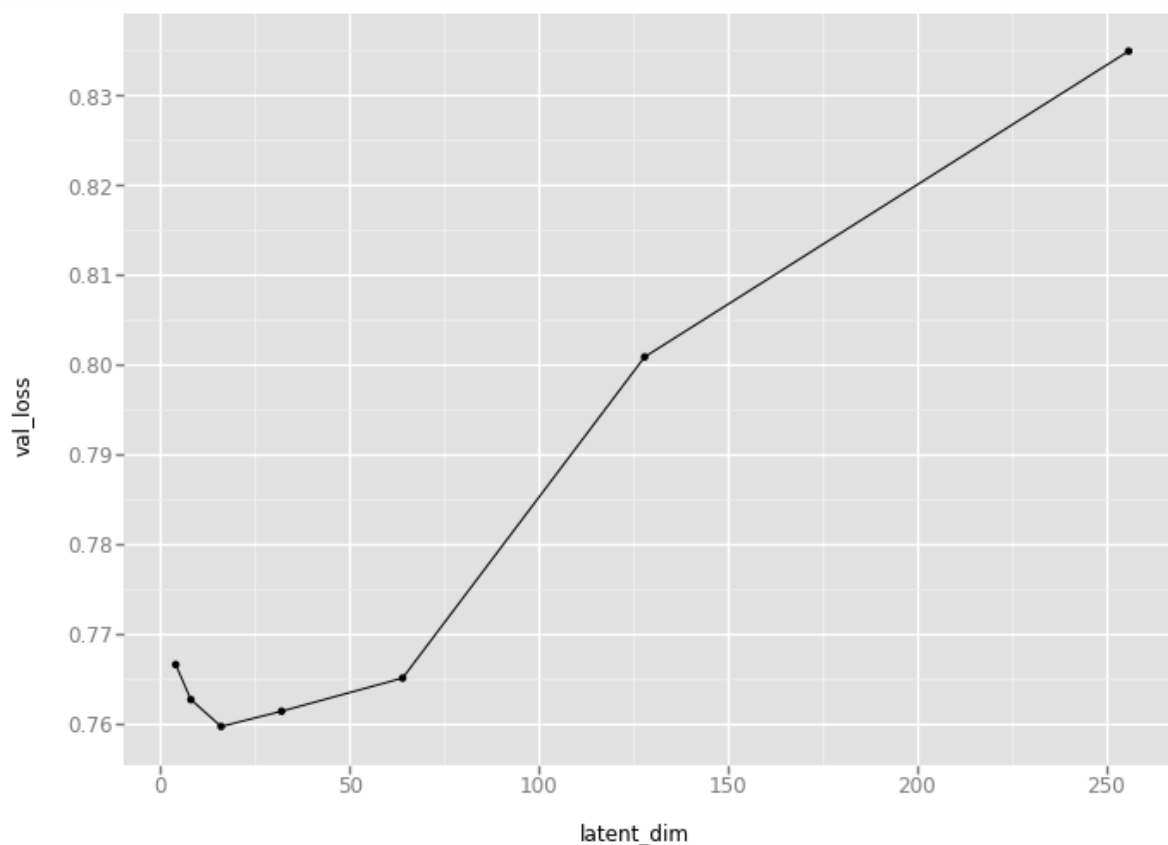
學號：B05902019 系級：資工二 姓名：蔡青邑

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

我 normalize 的方式跟一般標準化的方法一樣，對 training data 的 rating 取 mean 和 standard deviation，對其標準化。然後 testing data 也做一樣的事，不過用的是同一組平均和標準差。沒有標準化的 private score 是0.86316，有標準化的則是0.86444，感覺沒有差很多。

## 2. (1 %)比較不同的embedding dimension的結果。

以下是針對dimension為4, 8, 16, 32, 64, 128, 256，有做bias的training結果。可以發現從128開始結果開始明顯變差。



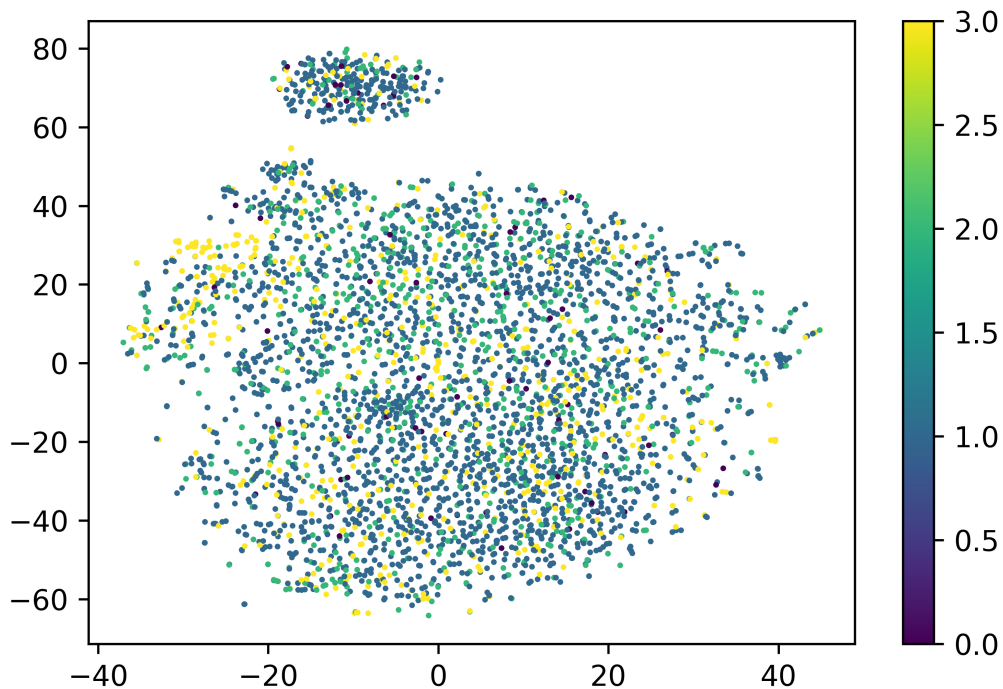
3. (1 %)比較有無bias的結果。

0.86984	0.87692	<input type="checkbox"/>
0.87536	0.88166	<input type="checkbox"/>

上面是有bias的，下面是沒有加入bias的，可以明顯發現有bias結果比較好。

4. (1 %)請試著將movie的embedding用tsne降維後，將movie category當作label來作圖。

```
dataset1: {"Comedy", "Romance", "Drama", "Musical", "Documentary", "Animation",  
"Children's"}  
dataset2: {"Fantasy", "Action", "Sci-Fi", "Adventure", "Western"}  
dataset3: {"Thriller", "Horror", "Crime", "Mystery", "War", "Film-Noir"}
```



## 5. (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。

我有試著將 user.csv 裡面的 gender, age, occupation 作為 feature 放到我的 model 裡面。主要就是經過 embedded layer後把這些 feature 的資訊放在 testing/training dataset 的 ID 之後一起實作。以下是結果，有比原本單純用 rating 來判別好一點。

yo.csv

11 days ago by B05902019蔡青邑

[add submission details](#)

0.85818

0.86691

