學號：b04901074 系級： 電機三 姓名：吳倉永

1. 請比較有無 normalize 的差別。並說明如何 normalize

   正常：kaggle→0.8689

   normalize：kaggle→0.8721

   我的方法是算出所有的 rating 的平均跟標準差，並把它記下來對所有 data normalized。我認為結果會比較差的原因在：因為不同的電影他的 rating 分布不一樣，不能把其他電影一起考慮近來，應該要對各自的電影的 rating 進行標準化才是比較好的

2. 比較不同的 embedding dimension 的結果

   18 維：kaggle→**0.8689**

   50 維：kaggle→0.8712

   100 維：kaggle→0.8738

   500 維：kaggle→0.8899

   結果來看維度越高效果越差，我觀察了影片的分類方式，總共就 18 類，其實 feature 不會太多，也解釋了維度對效果造成的影響

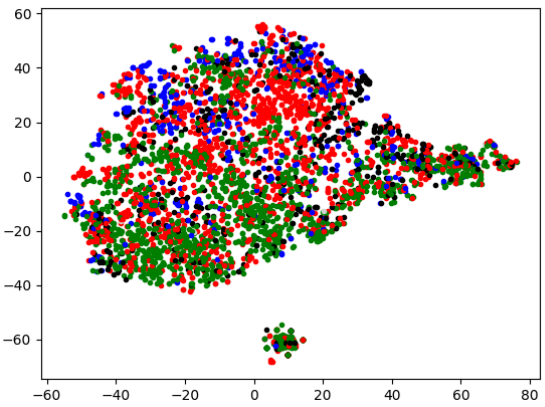
3. 比較有無 bias 的結果

   沒 bias: kaggle→**0.87283**

   有 bias: kaggle→0.86885

   沒有 bias 的結果差了一點，我認為是 bias 可能可以修正一些誤差，畢竟就像老師說的可能某個 movie 就是比較有名等等的

4. 請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當做 label 來作圖



這個 movie embedding 是利用 rating 來學習的，而我將影片大致分成 5 類，可以發現影片種類跟 rating 沒有太大的關係

5. 試著使用除了 rating 以外的 feature，並說明你的作法和結果，結果好壞不影響評分

   我利用 occupation、age(因分布較廣)納入 feature 中 train 出這些 attribute 對 rating 的 latent vector，並把他們與電影的 latent vector 各自做內積得出分數來。此方法得到的分數：kaggle→0.8699，比沒做稍微差了一點，但其實沒有差很多，所以這兩個 feature 跟 rating 沒什麼關係