

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

我是先取 training data rating 的 mean 以及 standard deviation，再把 training data 的 rating 做 normalization 再用 normalized 過後的 rating 做為 training 的 target。在 testing 的時候則使用一下公式

$$pred_{test} = pred_{test}^* \times train_{std} + train_{mean}$$

然後用一樣的架構把結果上傳，可以看到沒有 normalize 的比較好，而有經過 normalize 的卻比較差，不知道為甚麼會這樣，我想可能是不管是 user 或是 movie，它的範圍都蠻大的，如果硬是把它 normalize 到 0~1 的話，可能沒辦法表現出原有的樣子。

normalize	Non normalize
1.0146	0.85630

2. (1%)比較不同的 latent dimension 的結果。

我在使用相同架構下，使用 k=16，上傳到 kaggle 的結果是 0.85630，使用 k=15，上傳到 kaggle 的結果是 0.85609，如果 k=12 的話，丟上去是 0.85801，如果是 k = 2 的話，丟上 kaggle 是 0.865，可以看到雖然 15 的結果比 16 還高，但我認為應該只是 random 的結果，latent dimension 應該越低，出來的結果會越高。

16	15	12	2
0.85630	0.85609	0.85801	0.8650

3. (1%)比較有無 bias 的結果。

有 bias	無 bias
0.85630	0.9367

我發現有無 bias 會有蠻大的差別的，之前一開始我都補 bias 調成 0，雖然 validation predict 都可以到達 0.6 多，但我發現我上傳後卻只有 0.9 多，但我把 bias 改成不是 0 後，就可以上升到 0.8 多，我想應該是 bias 是改變最後

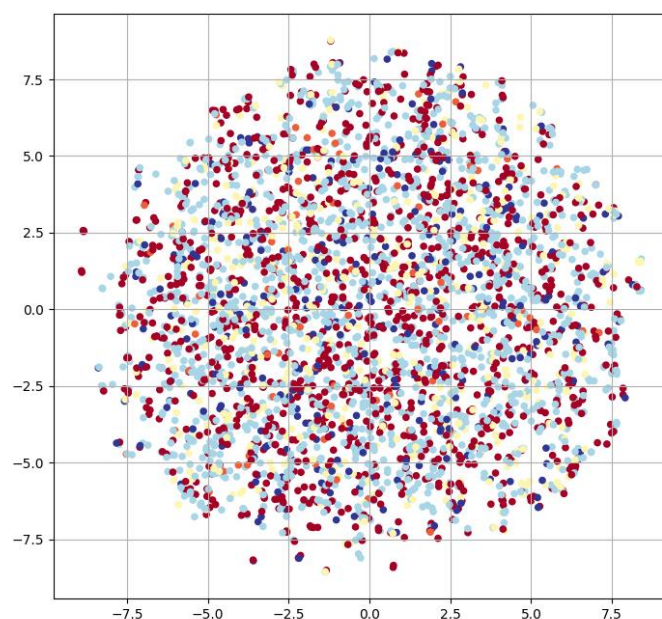
偏移的方法，如果你沒加 bias，可能容易 overfitting，所以雖然我跑出來的解果很好，但上傳的結果很爛，但如果有加 bias 的話，預測出來的和上傳結果就沒有相差太遠了。

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。

我是用 embedding 把 user_id 和 movie_id 投影 N 維和 M 維，然後再把 M 和 N 兩個接在一起，進行 normalize，然後丟進 DNN 的 model，我使用 softmax 當作輸出的最後一層，因為有 5 個 case，所以最後輸出是五維，使用 NN 的話，最後輸出只能是 1~5，不像 MF 可以到小數點下好幾位，所以 predict 出來的解果都大於 1，因為代表一錯都是相差 1 以上，我認為 NN 並不適合這次的功課。

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

Animation Adventure Children's	藍色
Musical Drama	紅色
Fantasy Mystery Romance	橘
Thriller Crime Horror	黃
Action Comedy	淺藍



這是我經過將 movie 的 embedding 用 tsne 降維後所畫得圖，我認為這圖並不太能分辨出來不同類型的電影，我想應該是我 embedding 出來的並不明顯，如果 kaggle 上的成績能更高，embedding 出來的效果會更好些，雖然這張圖不太能分辨，但有些地方還是值得觀察，例如藍色的點在左下比其他地方還要集中，感覺有區分出來，還有黃色點比較偏向左半部等等。

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。

我嘗試只使用 UserID 一個 feature 來做, 原本我 training 的時候是直接使用 UserID, 然後我嘗試直接使用 UserID 裡面的 age 來做, 我發現結果沒有我一開始想像的會非常差, 我想是雖然每個人的 age 都不一樣, 但相似度還是蠻高, 所以會有這種結果是可以預期的, 但我用另外一個 Gender 來做, 我發現使用這會非常不好, 我想是因為性別只有男生和女生, 單靠性別來區分是非常困難的, 看起來幾乎像亂猜, 所以我認為用 age 是非常不恰當的。以下是我用 10% 的 validation 所跑出的結果。

age	Gender
2.6481	1.6941