HW6 REPORT

B04705026 林彥廷

1. 請比較有無normalize的差別。並說明如何normalize
   - **架構-Matrix Factorization with bias**
   - **標準化方法**
     - mean = training data的平均
     - std= training data的標準差
     - Training data = (原資料-mean)/std
     - Prediction = 原預測*std+mean
   - **模型表現**

	No-norm	Normalization
public score	0.87167	0.87057
private score	0.86393	0.86038

   - **討論**
     - 可以由模型表現得出標準化可以降低錯誤率，但是其實進步幅度不大
     - 可能是因為現在只有1~5的評分其實分佈還沒有很廣很散，如果評分分數式0~100可能就會有很大的影響。
2. (1%) 比較不同的embedding dimension的結果。
   - **架構-Matrix Factorization with bias**
   - **模型表現**

Emb. Dim.	50	100
public score	0.87167	0.87107
private score	0.86393	0.86350

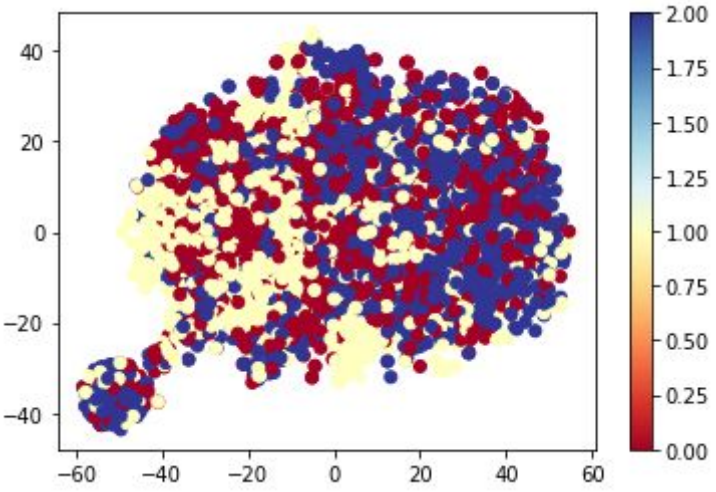
   - 討論
     - Embedding Dimension是超參數，如果資料數不夠多，卻使用太多dimension的話，會有太多變數其實是沒有意義。
     - 我使用Validation set來做選擇，當時表現最好的是50和100兩個候選人
     - 結果來說，100個latent variables可能才"夠"描述電影評分之問題，也許資料更多時，可以訓練出更多有意義的latent variables。

3. 比較有無bias的結果。
   a. **架構-Matrix Factorization with bias**
   b. **模型表現**

	NO-bias	bias
public score	0.87552	0.87107
private score	0.86371	0.86350

- 討論
  - **加上電影與個人的bias，可以抓住個人偏好與電影好壞的變異，這個作法有很直觀的意義。**

4. 請試著將movie的embedding用tsne降維後，將movie category當作label來作圖。
   - **圖片**
     - 電影分類：
     - 紅：'Adventure'Children's'Comedy''Fantasy''Mystery''Romance'
     - 黃：'Action''Crime''Film-Noir''Horror''Sci-Fi''Thriller''War''Western'
     - 藍：'Animation''Documentary''Drama''Musical'



   - **討論**
     - 將所有電影類型依照文藝氣息與動作感分成三大類，最後降維的結果可看出黃色類的電影都集中在左下角，藍色的電影偏右上角。紅色類型電影並沒有被分的很開，可能是因為這類型的電影相較於另外兩類電影並沒有特別差異，也可能是因為是2D無法呈獻3個類型之差異。

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

- **模型架構-DNN**
  - 使用movie和user id通過embedding再和額外feature結合，並通過FC層，得到rating。
- **模型表現**

	User_age	Only rating
public score	0.89347	0.91338
private score	0.89938	0.90637

- **討論**
  - 使用**Age**當作而外的feature
  - 加入額外的特徵模型表現有比較好，但是幅度並不大。
  - 加入全部feature可能又太龐大，需要再使用feature selection來篩選。