

Machine Learning Project 6

Matrix Factorization

- b03902089 資工三 林良翰

Questions

1. 請比較對評分(Rating)有無標準化(Normalize)的差別, 並說明如何標準化.

- Embedding Features: user ID, movie ID.
- Embedding Dimensions: 512
- Validation Ratio: 0.05

RMSE	Not Normalized	Normalized
Training	0.6520	0.6382
Validation	0.8438	0.8568
Kaggle Public	0.8465	0.8561

- 標準化之後訓練成果會更好, 但是測試結果會差一點.
- Method of Normalization
 - 算出評分 R 的平均值 μ , 標準差 σ
 - 把每個評分做標準化:

$$R_n = \frac{(R - \mu)}{\sigma}$$

- 最後預測出來的結果 R_{n-p} 再反算回去原來的評分值 R_p

$$R_p = R_{n-p} \times \sigma + \mu$$

2. 比較不同的潛在維度(Latent Dimension)的結果.

- Embedding Features: user ID, movie ID.
- Validation Ratio: 0.05

Latent Dimension	Epochs	Validation RMSE
2	370	0.8803
4	211	0.8733
8	146	0.8636
16	96	0.8589
32	66	0.8535
64	34	0.8528
128	25	0.8470
256	16	0.8465
512	13	0.8438
1024	10	0.8470

- Latent Dimension = 512 的時候結果最佳.

3. 比較有無偏見(Bias)的結果.

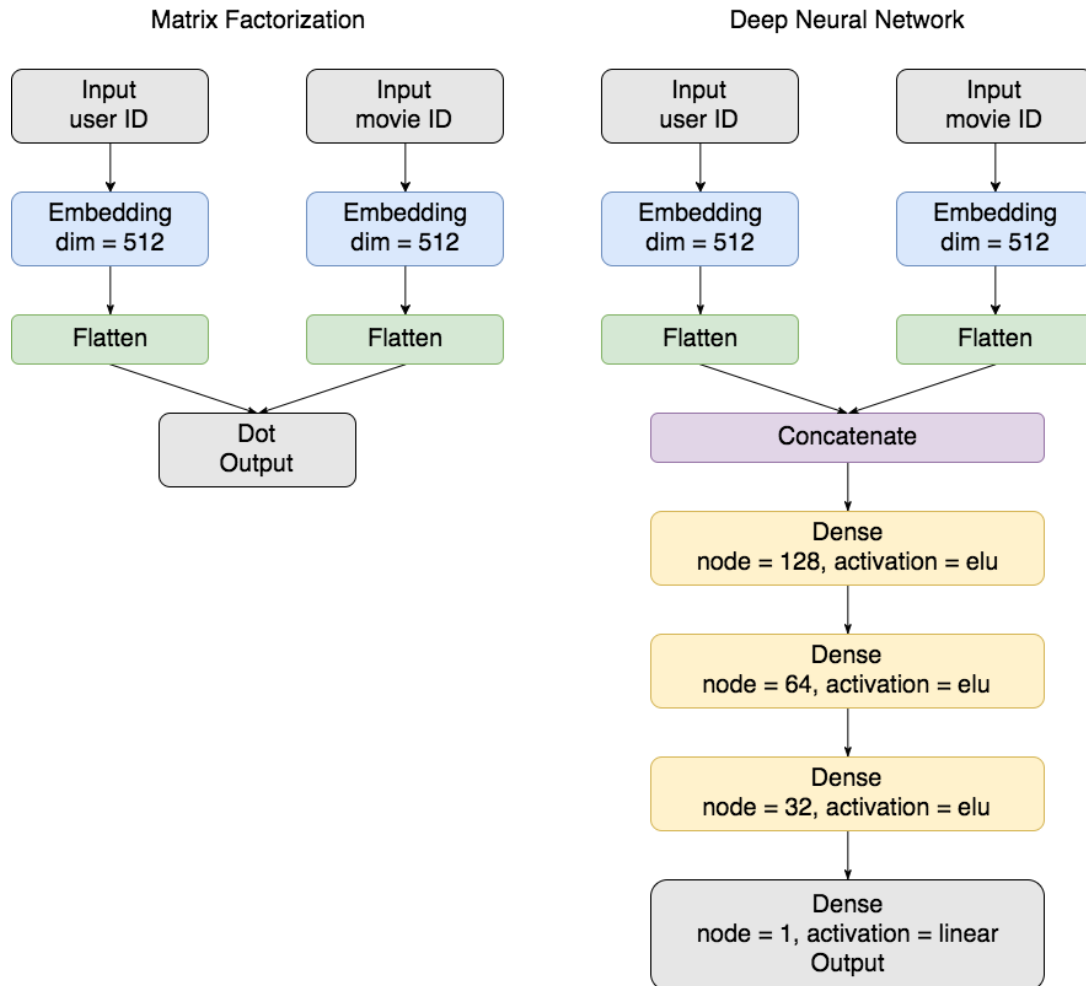
- Embedding Features: user ID, movie ID.
- Embedding Dimensions: 512
- Validation Ratio: 0.05

RMSE	Without Bias	With Bias
Training	0.6520	0.6388
Validation	0.8438	0.8438
Kaggle Public	0.8465	0.8466

- 結果差別不大.

4. 請試著用**DNN**來解決這個問題, 並且說明實作的方法. 比較**MF**和**DNN**的結果, 討論結果的差異.

- MF, DNN模型架構



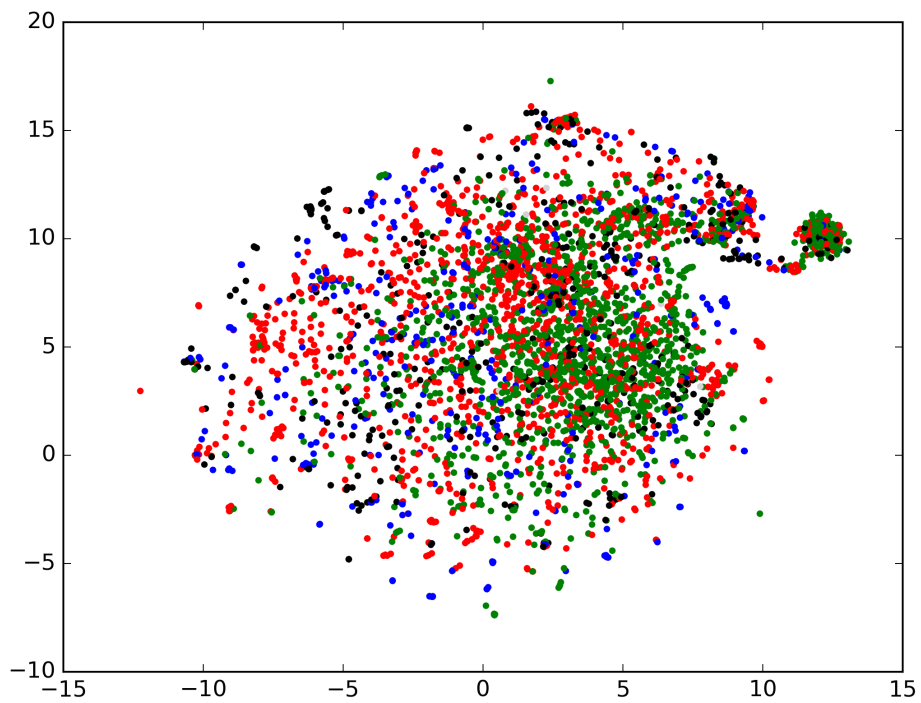
- 結果比較

RMSE	MF	DNN
Training	0.6520	0.8181
Validation	0.8438	0.8679
Kaggle Public	0.8465	0.8696

- DNN的結果會比較差一點, 而且DNN不能夠太多層, 否則容易over fitting.

5. 請試著將電影的潛在維度(Embedding)用TSNE降維後，將電影類別當作標記(Label)來作圖

- 使用TSNE降維結果

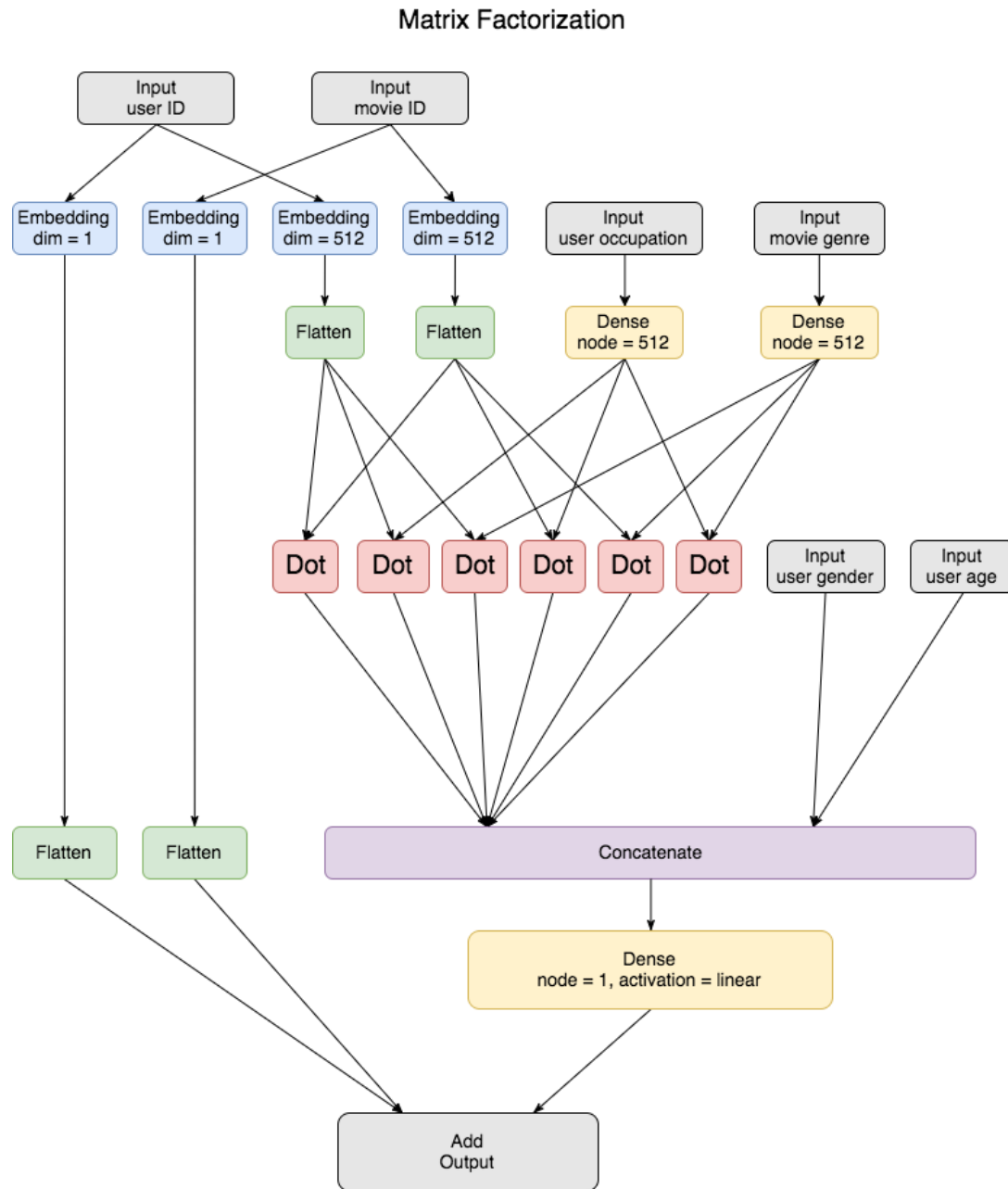


- Legends

Color	Genre
Red	Animation, Children's, Comedy, Adventure
Green	Romance, Drama, Documentary, Musical
Blue	Fantasy, Action, Sci-Fi, War, Western
Black	Crime, Thriller, Horror, Film-Noir
Gray	Other

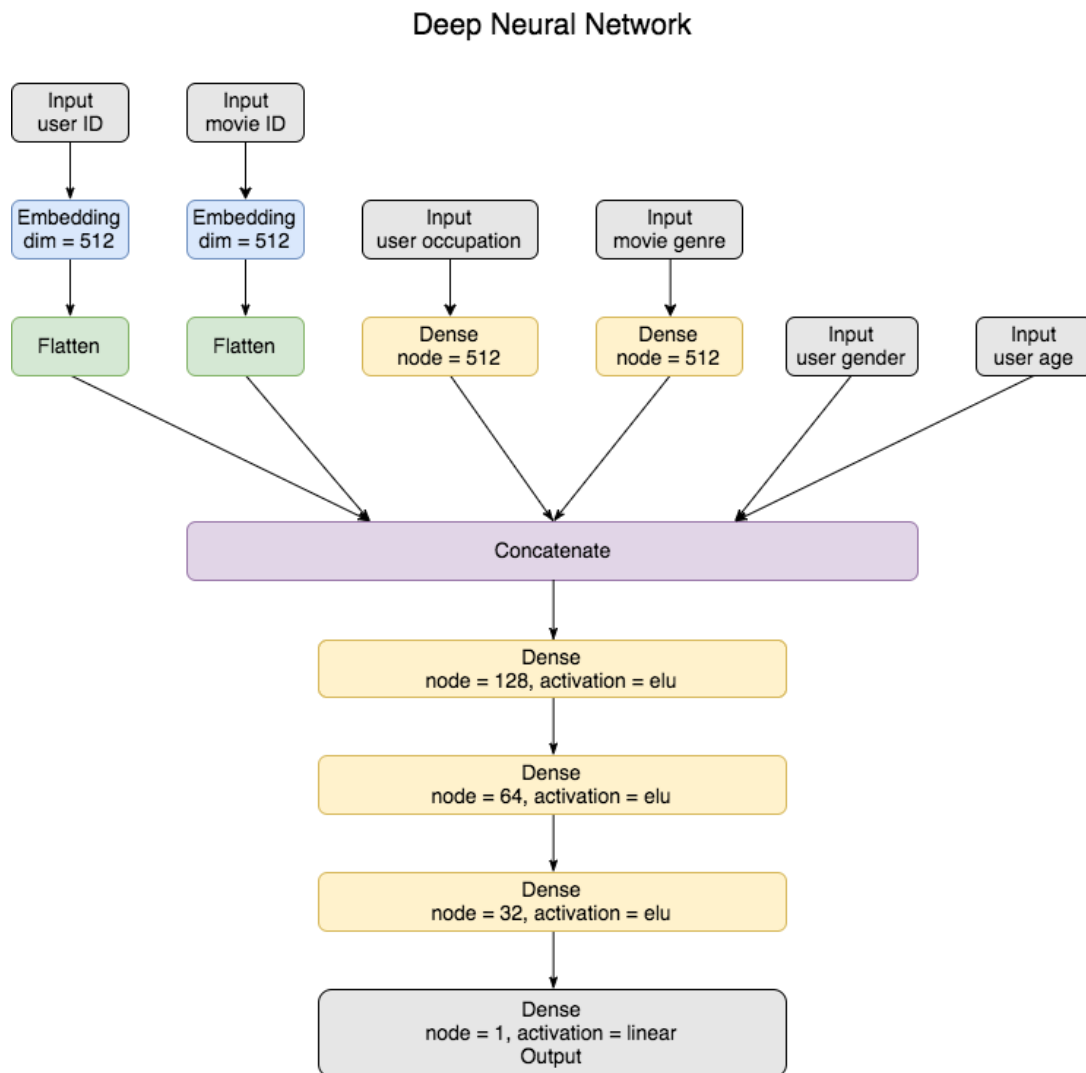
6. 試著使用除了評分以外的特徵(**Feature**), 並說明你的作法和結果, 結果好壞不會影響評分.

- 把user occupations和movie genres變成one-hot matrices
 - User Occupations: dim = 21
 - Movie Genres: dim = 18
- Validation Ratio: 0.05
- 使用複雜版本的MF模型



- Training RMSE: 0.6276
- Validation RMSE: 0.8697
- Kaggle Public RMSE: 0.8703

- 使用複雜版本的DNN模型



- Training RMSE: 0.8165
- Validation RMSE: 0.8574
- Kaggle Public RMSE: 0.8554