

學號：r05525096 系級：工海碩二 姓名：郭捷

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

(collaborator:)

(1)實驗的設定為有 bias，latent dimension 為 10，Batch size 為 128

(2)normalize 的做法為取 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 $\text{pred}_{\text{test}} = \text{pred}^* * \text{train}_{\text{std}} + \text{train}_{\text{mean}}$

(3)實驗結果為

	Public Score
Non-normalized	0.85883
Normalized	0.86106

2. (1%)比較不同的 latent dimension 的結果。

(collaborator:)

(1)實驗的設定為有 bias，沒有 normalization，Batch size 為 128

(2)實驗結果為

Latent dimension	Public Score
10	0.85883
15	0.85536
20	0.86274
50	0.86912

(3)最後選擇 15 作為 latent dimension。由 training 的收斂速度來說，latent dimension 越小，收斂的越快。Latent dimension 越大，越容易 overfitting。

3. (1%)比較有無 bias 的結果。

(collaborator:)

(1)實驗的設定為 latent dimension 為 15，沒有 normalization，Batch size 為 128

(2)實驗結果為

	Public Score
有 bias	0.85536
無 bias	0.85876

(3) 每個 user 可能都會有自己評分傾向，像是傾向於把每部電影都評得很高分或者很低分；同樣的電影也會有這樣的趨勢。所以有 bias 會比沒有 bias 更好。

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。

(collaborator:)

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, 1)	0	
input_2 (InputLayer)	(None, 1)	0	
embedding_1 (Embedding)	(None, 1, 15)	90615	input_1[0][0]
embedding_2 (Embedding)	(None, 1, 15)	59295	input_2[0][0]
dropout_1 (Dropout)	(None, 1, 15)	0	embedding_1[0][0]
dropout_2 (Dropout)	(None, 1, 15)	0	embedding_2[0][0]
flatten_1 (Flatten)	(None, 15)	0	dropout_1[0][0]
flatten_2 (Flatten)	(None, 15)	0	dropout_2[0][0]
concatenate_1 (Concatenate)	(None, 30)	0	flatten_1[0][0] flatten_2[0][0]
dense_1 (Dense)	(None, 150)	4650	concatenate_1[0][0]
dense_2 (Dense)	(None, 50)	7550	dense_1[0][0]
dense_3 (Dense)	(None, 1)	51	dense_2[0][0]
Total params: 162,161			
Trainable params: 162,161			
Non-trainable params: 0			

(1)實驗的設定為 latent dimension 為 15，沒有 normalization，Batch size 為 128，將 user embedding 以及 movie embedding concatenate 在一起再過 DNN 得出 rating。

(2)實驗結果：DNN 在 public score 為 0.87886。比 MF 差的原因可能為我沒有仔細調整 DNN 的參數，以及 Dense layer 的設定過大等。

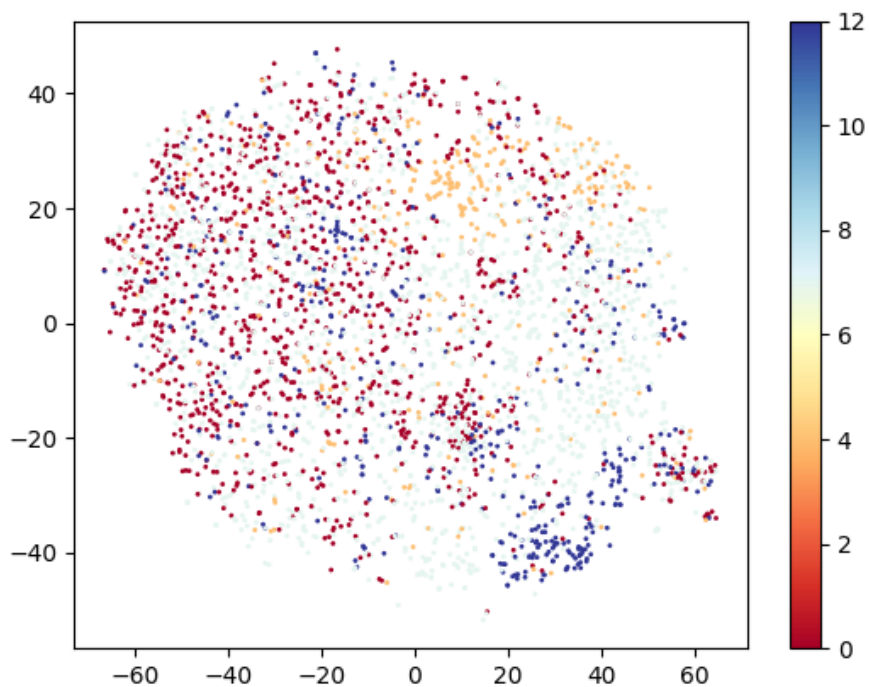
5. (1%)請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當作 label 來作圖。

(collaborator:黃雅博)

‘Drama’，‘Musical’ :紅色

‘Crime’，‘Thriller’，‘Horror’ :深藍

‘Animation’，‘Childrens’，‘Adventure’ :藍色



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。

(collaborator: 謝朋諺)

有無外部資料	Public Score
無	0.85536
有	0.87423

我的作法是加上外部 users.csv 的性別資料, 我認為性別會是影響到評分依據的, 男性可能比較喜歡刺激、恐怖、戰爭、冒險的影片, 女性可能喜歡音樂、浪漫、幻想、喜劇的影片。我的 Model 先將 user id、user age、movie id 這三組都先經過 embedding layer, 再將 user id 與 user age 做 Dot, movie id 與 user age 做 Dot, 再將這兩個 Dot 的結果再 Dot 一次, 效果看起來沒有比較好。