

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

答：Model structure: 1. Latent dimension of user and movie: 7, 2. Bias of user and movie were added, 3. Training procedure: optimizer: adam, batch size: 256, 10% of training data was used for validation

Normalization的方式我用了min-max以及z-scores，min-max的公式為 $X' = \frac{X - X_{min}}{X_{max} - X_{min}}$ z-scores的公式為 $X' = \frac{X - \mu}{\sigma}$ ，要注意的是最後prediction時要反推回原本的值。發現沒有normalize的結果反而較好，推測是因為normalize後的rmse幾乎不會變動，因此有作normalize的loss function我改用mse，可能loss function與Kaggle上不同，造成normalize後的score反而較低。另外可以發現normalize後收斂的速度明顯比沒有normalize來得快。

	Private score	Public score	Epochs
w/o normalization	0.86497	0.87205	30
min-max	0.86900	0.87538	9
z-scores	0.86779	0.87347	15

2. (1%) 比較不同的embedding dimension的結果。
(Collaborators:)

答：Model structure: 1. Bias of user and movie were added, 2. Training procedure: optimizer: adam, batch size: 256, 10% of training data was used for validation

Embedding dimension越大則trainable parameters會越多，但也代表容易overfitting。從結果發現在dimension為20時，在Kaggle上的表現最好。

Embedding dimension	Private score	Public score
5	0.86586	0.87384
10	0.86009	0.87153
15	0.86378	0.87095
20	0.86053	0.86870
30	0.86231	0.87140

40	0.86178	0.87117
50	0.86459	0.86998

3. (1%) 比較有無bias的結果。
(Collaborators:)

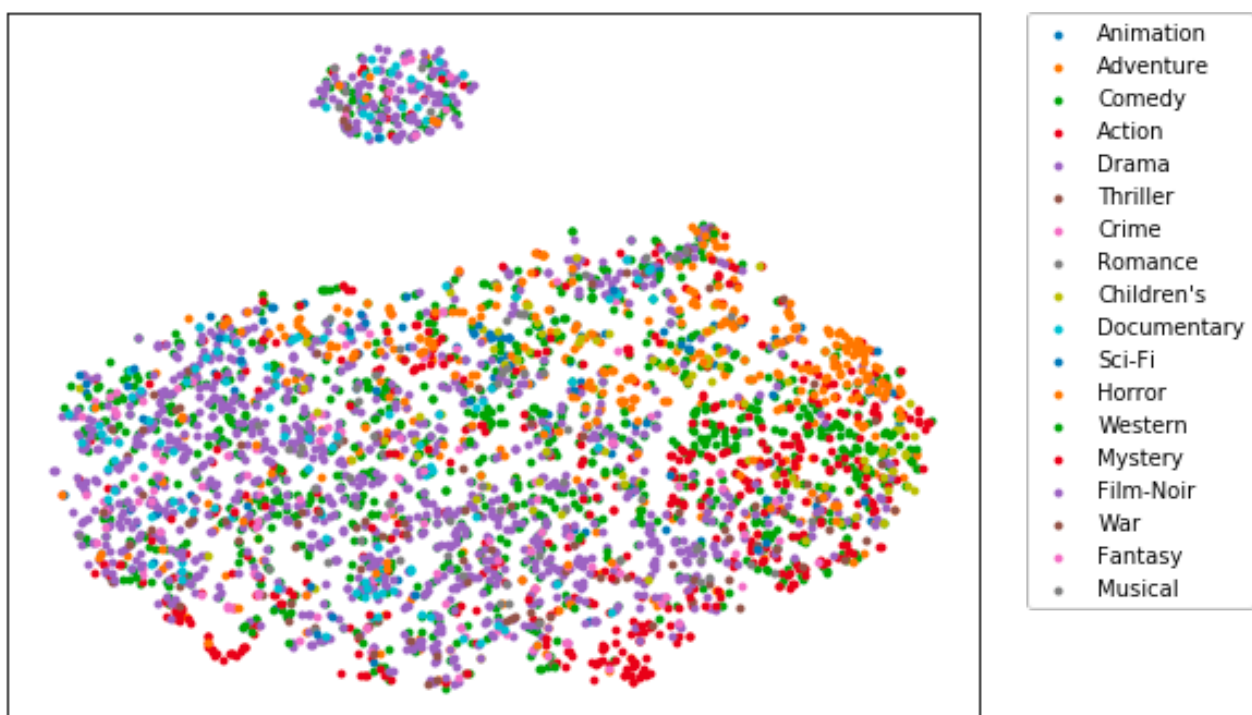
答：Model structure: 1. Latent dimension of user and movie: 7, 2. Training procedure: optimizer: adam, batch size: 256, 10% of training data was used for validation

在其他變數都固定下，有bias的model會有較好的結果，畢竟多了user bias跟movie bias兩個vectors的變數。

	Private score	Public score
w/ bias	0.86497	0.87205
w/o bias	0.86829	0.87541

4. (1%) 請試著將movie的embedding用tsne降維後，將movie category當作label來作圖。
(Collaborators:)

答：利用tsne將movie embedding的7維空間降至2維，若單一電影有多個類別，則取第一個類別當作其label。因為label的類別太多，導致類別的分佈雜亂，比較難看到分佈情形。最明顯的是上方有一塊分離的聚集，其主要是紫色label（drama）的電影。



5. (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。(Collaborators:)

答：利用users.csv內的資訊，除了User外，另外加上Gender、Age、Occupation、Zip-code四個embedding，因此User總共會有5個embedding。將5個embedding跟movie embedding concatenate成一個一維vector後，通過2層dense layer（分別有256、128個units），最後通過unit為1的output layer，在Kaggle上的表現如下表。

Private score	Public score
0.85600	0.86151