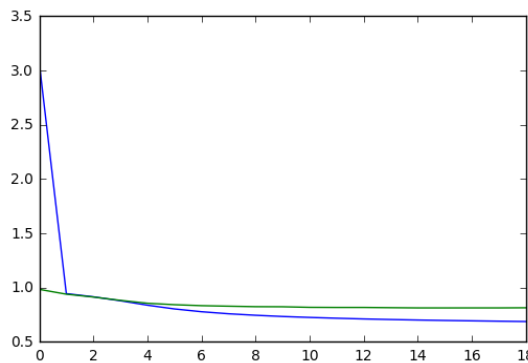
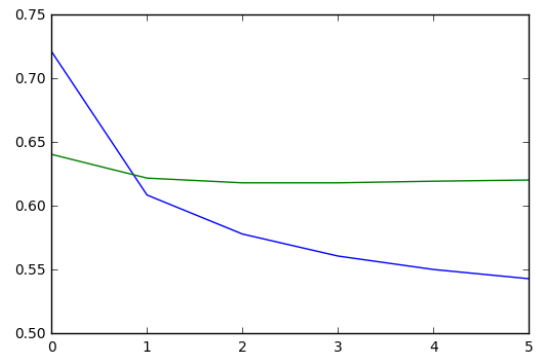


1. (1%)請比較有無 `normalize(rating)` 的差別。並說明如何 `normalize`。

(collaborator: 無)



(無 normalize)

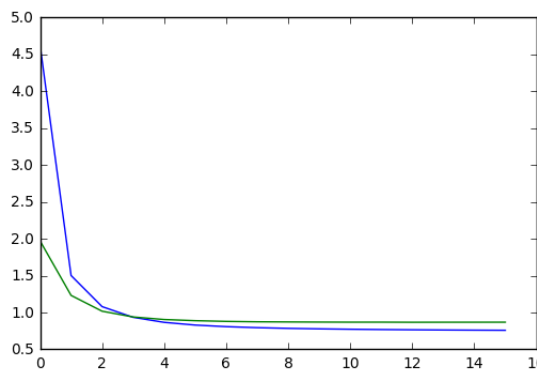


(有 normalize)

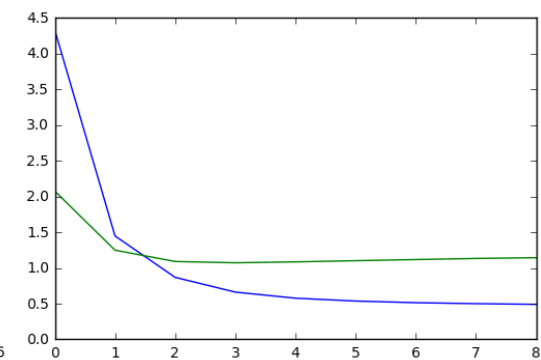
使用 `train` 的資料去減 `mean` 然後除以 `std` 的方式去 `normalize`，有 `normalize` 速度收斂的速度會加快，然後結果也較好。

2. (1%)比較不同的 `latent dimension` 的結果。

(collaborator: 無)



Latent dimension = 8

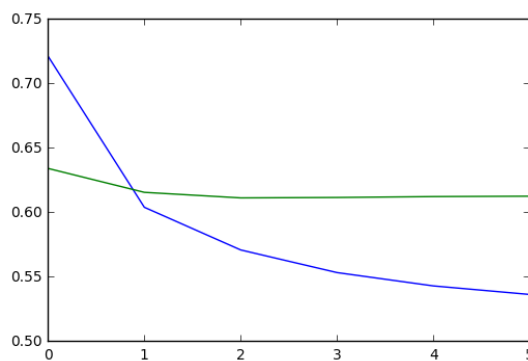


Latent dimension = 128

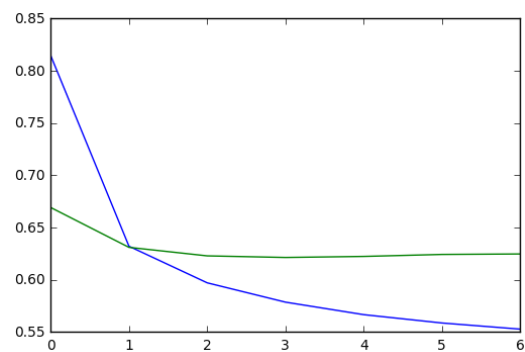
Latent dimension 並非越大越好，8 的 `val loss` 值明顯比 128 的好上許多

3. (1%)比較有無 `bias` 的結果。

(collaborator: 無)



有 bias



無 bias

有 `bias` 效果稍微好一些

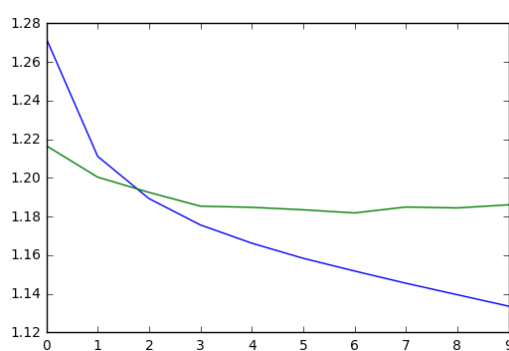
4. (1%)請試著用 DNN 來解決這個問題，並且說明實做的方法(方法不限)。並比較 MF 和 NN 的結果，討論結果的差異。

(collaborator:無)

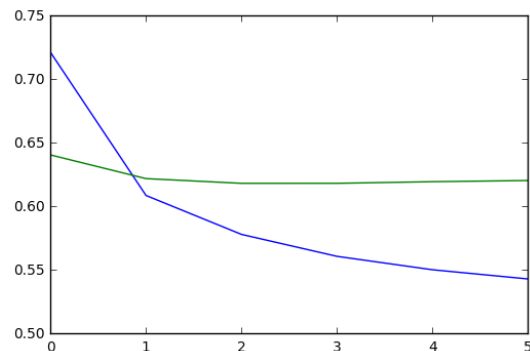
Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	(None, 1)	0	
input_1 (InputLayer)	(None, 1)	0	
embedding_2 (Embedding)	(None, 1, 128)	773248	input_2[0][0]
embedding_1 (Embedding)	(None, 1, 128)	497152	input_1[0][0]
flatten_2 (Flatten)	(None, 128)	0	embedding_2[0][0]
flatten_1 (Flatten)	(None, 128)	0	embedding_1[0][0]
concatenate_1 (Concatenate)	(None, 256)	0	flatten_2[0][0] flatten_1[0][0]
dense_1 (Dense)	(None, 128)	32896	concatenate_1[0][0]
dropout_1 (Dropout)	(None, 128)	0	dense_1[0][0]
batch_normalization_1 (Batch Normalization)	(None, 128)	512	dropout_1[0][0]
dense_2 (Dense)	(None, 256)	33024	batch_normalization_1[0][0]
dropout_2 (Dropout)	(None, 256)	0	dense_2[0][0]
batch_normalization_2 (Batch Normalization)	(None, 256)	1024	dropout_2[0][0]
dense_3 (Dense)	(None, 512)	131584	batch_normalization_2[0][0]
dense_4 (Dense)	(None, 5)	2565	dense_3[0][0]

Total params: 1,472,005
 Trainable params: 1,471,237
 Non-trainable params: 768

DNN model



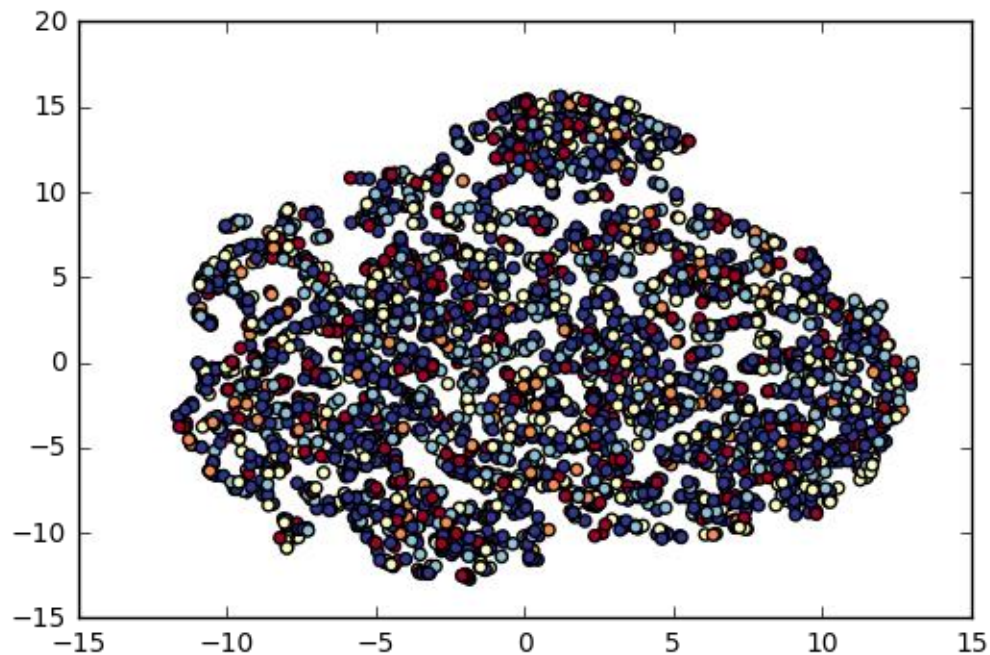
DNN 結果



MF 結果

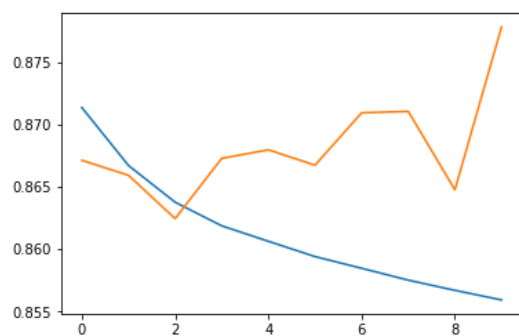
透過將 movie 和 user 兩個 embedding concatenate 在一起，然後經過 3 層的 DNN 將其分做 5 類做預測，而就 loss 的結果來說 MF 比 DNN 好像好一點，在預測上 DNN 分 5 類預測的結果是整數，而 MF 則是小數點，在 kaggle 上的 RMSE 評分中較為精確一些。

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

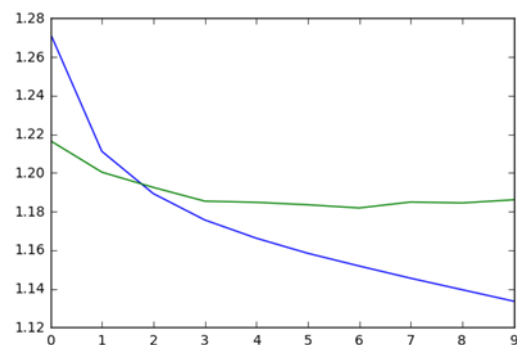


6. (BONUS)(1%)試著使用除了 rating 以外的 feature, 並說明你的作法和結果，結果好壞不會影響評分。
(collaborator:無)

我將原本 18 個類別簡化為性質較相近的 5 大類—Action、Adventure、Fantasy、Sci-Fi 和 War 為奇幻動作類；Crime、Film-Noir、Mystery 和 Documentary 為犯罪解謎類；Horror 和 Thriller 為驚悚恐怖類；Animation、Children's 和 Comedy 為喜劇動畫類；Drama、Musical 和 Romance 為浪漫藝術類。再把這 5 大類疊加到原來的 DNN 上做預測，發現比起原來的 DNN，其 loss 介在 1.18~1.20 間，新的 loss 可降到在 0.875~0.800 之間，效果變得比較好。



新的 DNN 結果



舊的 DNN 結果