

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

(collaborator:)

使用 Keras 的 Lambda Layer 做 normalize，training set 的平均值大約為 3.5，標準差為 1.1，因此在實作上我只有扣除平均值來做 normalize。可以發現有 normalize 比沒有 normalize 的 RMSE 分數還要好。收斂速度方面有 normalize 的 RMSE 從約 0.98 開始收斂，沒有 normalize 的 RMSE 從約 3.2 開始收斂，有 normalize 的收斂速度也比沒有 normalize 的還要快。

	有 normalize	沒有 normalize
RMSE	0.85992	0.86993

2. (1%)比較不同的 latent dimension 的結果。

(collaborator:)

我實作了 16、32、64、128 四種不同 latent dimension，可以發現 dimension = 64 為最好的 model，因此並非 latent dimension 越大 RMSE 就越小。

Dimension	16	32	64	128
RMSE	0.86747	0.85992	0.85709	0.86061

3. (1%)比較有無 bias 的結果。

(collaborator:)

有 bias 的 RMSE 比沒有 bias 的 RMSE 還要小了快要 0.02，因此 bias 對於 training 的幫助很大。

	有 bias	沒有 bias
RMSE	0.85992	0.87780

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:)

DNN 是將兩個 Embedding 用 Concatenate 接起來後，通過兩層 Dense(128)，Dropout 皆為 0.5，activation 為 relu，最後再一層 Dense(1)，activation 為 linear 作為 output。可以顯示出 DNN 的 RMSE 比 MF 的 RMSE 還要差很多。

	MF	DNN
RMSE	0.85992	0.90773

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

(collaborator:)

將一些類別合併處理後總共分為 5 大類為下：

Class1: Action, Adventure, War, Documentary, Western

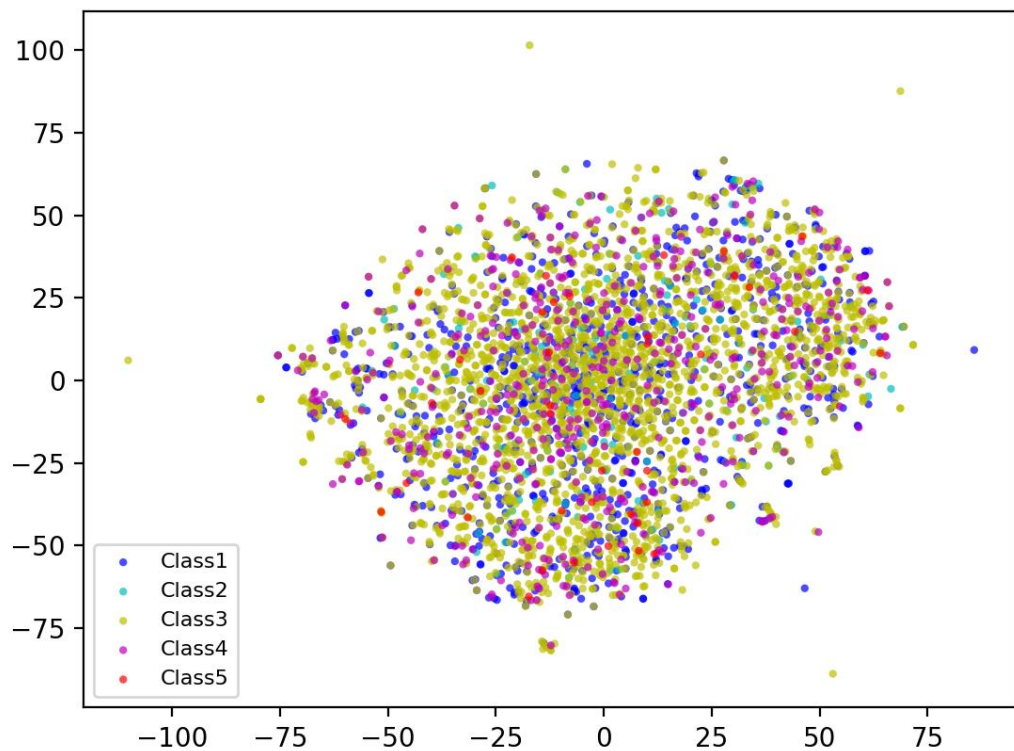
Class2: Crime, Film-Noir, Horror, Thriller

Class3: Comedy, Drama, Musical, Romance

Class4: Fantasy, Mystery, Sci-Fi

Class5: Animation, Children's

結果為下圖：



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:)