

學號：R06725041 系級：資管碩一 姓名：彭証鴻

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

(Collaborators: sample code、keras official document)

答：

有 Normalize：

Public rmse: 0.87570

Private rmse: 0.86810

沒有 Normalize：

Public rmse: 0.88712

Private rmse: 0.88049

(rmse = root-mean-square-error)

Normalize 實作方法：

先算出 **train.csv** 中的 **rating** 那欄位的平均值與標準差後，將每個 **rating** 值減掉平均再除以標準差。

2. (1%) 比較不同的 **embedding dimension** 的結果。?

(Collaborators: sample code、keras official document)

答：

Embedding dimension=32：

Public rmse: 0.87172

Private rmse: 0.86339

Embedding dimension=128：

Public rmse: 0.85559

Private rmse: 0.85033

(rmse = root-mean-square-error)

3. (1%) 比較有無 **bias** 的結果。

(Collaborators: sample code)

答：

有加入 bias：

Public rmse: 0.85559

Private rmse: 0.85033

沒有加入 bias：

Public rmse:0.85588

Private rmse:0.85028

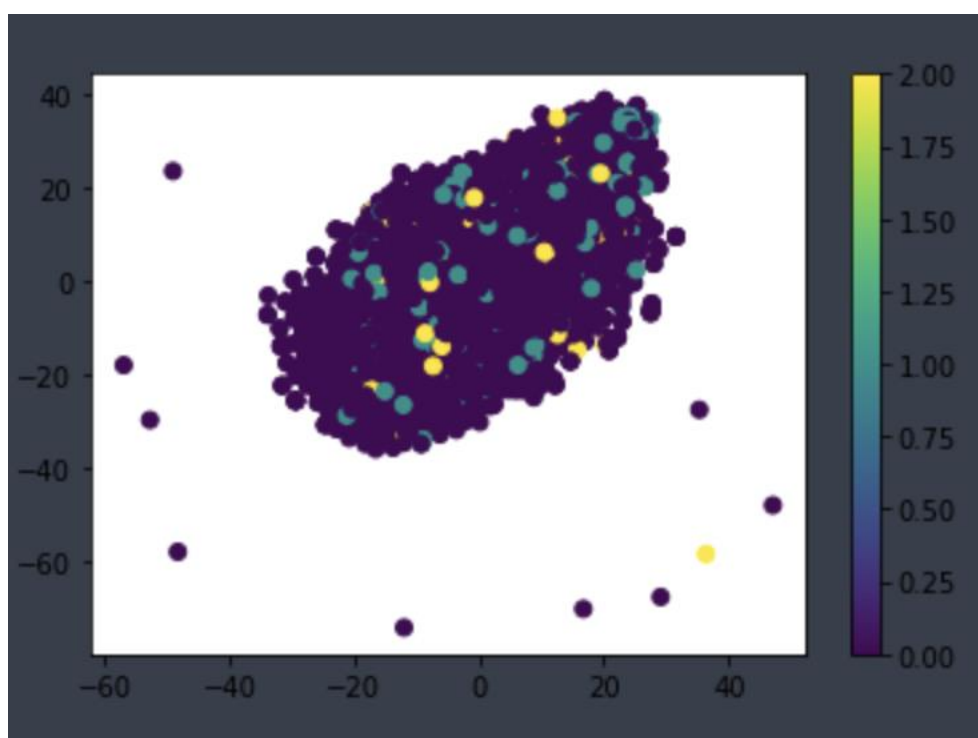
由上面可看出在我的 **model** 中，有沒有加入 **bias** 對 **rmse** 的差異並不是太大，無論是在 **public**、**private** 中差距都是小數點後 3 位而已。

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

(Collaborators: sample code、keras official document)

答：

我挑了 **Romance**、**Horror** 和 **Thriller** 三個電影類別，並將他分別 **lable** 為 **0**(紫色)、**1**(祖母綠)和 **2**(黃色)，可以看到在 **tsne** 降維後，雖然中間有一大團分不太開，不過可以看到 **Horror** 和 **Thriller** 確實較為幾乎完全重疊(較關聯)，且外圍的點也僅有一個是 **Horror**。



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: sample code)

答：

我將 **users** 中 **Gender** 和 **Age** 這兩個 **feature** 加入我原先的 **model**，在經

過 **Embedding layer** 後與原有的 **userID, movieID, rating** concatenate 起來，結果如下：

加入前：

**Public rmse:0.85559**

**Private rmse:0.85033**

加入後：

**Public rmse:0.85882**

**Private rmse:0.85005**

從上面的結果看來，其實加入 **age**、**gender** 這兩個 **feature** 後，並沒有有效的降低 **rmse**。