學號：b03705013 系級： 資管三 姓名：徐逸然

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

我使用MIN\_MAX scaling的方式標準化rating分數，然而做出來效果很差，同樣的參數下與沒有做標準化的Kaggle分數落差達0.11 (0.97與0.86)，可見此題normalize之後並不會有幫助。由於rating都是固定在1-5分的區間，因此標準化資料以邏輯來說確實不需要。

1. 比較不同的latent dimension的結果。

經過實驗之後發現latent dimension 越大 ，收斂的速度越快，Error也能做到越小，然而由於CPU效能不太足夠，因此只將iteration設定200次並做到K = 5， 無法測試當latent dimension極大狀況如K = 100 時的情況。 我個人推測當latent dimension極大時會容易產生overfit。

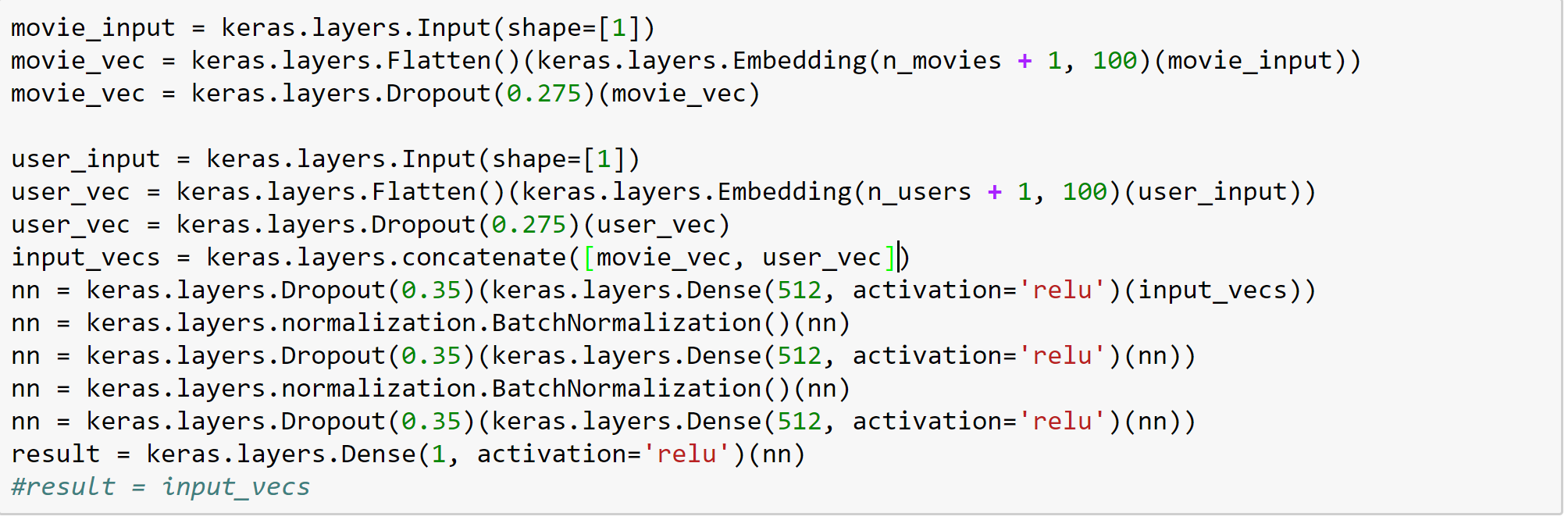
|  |  |
| --- | --- |
| Dimension | Error |
| K = 1 | 733866.12 |
| K = 2 | 673984.34 |
| K = 3 | 651721.59 |
| K = 4 | 635745.21 |
| K = 5 | 610880.03 |

1. 比較有無bias的結果。

在K = 5, iteration = 200 的情況下，有無bias的kaggle成績只差0.002 (0.866與0.864)在誤差的範圍內，因此bias並不會造成決定性的影響。

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。

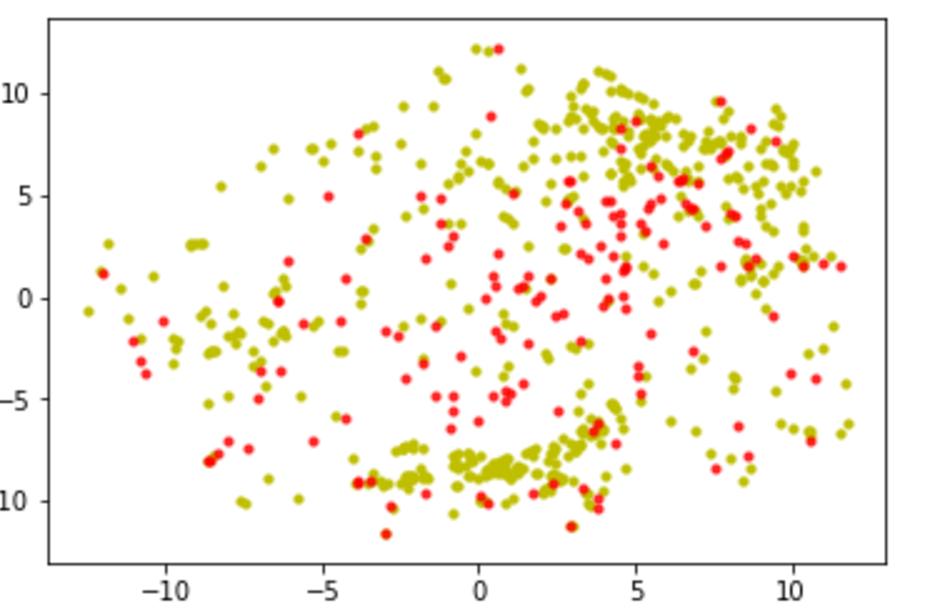
以下是我的DNN model，我採用三層hidden layer，每層512個neuron，根據幾次實驗後發現用regression準確率比分類高出許多，因此最後output採用RELU而非Softmax。由於能用Gpu 訓練，因此相比普通Matrix Factorization大幅縮訓練時間，一個epoch約2-3秒。最好的結果validation error (MSE)能做到0.697，kaggle則拿到0.846左右的分數



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

根據幾次作圖後，我發現幾組畫出來較相近的分類為 :

1. Action & Adventure : 黃色為Adventure，紅色為Action，可以看出兩者分布均靠右上方，左上方則較稀疏。



2. Horror & Thriller : 黃色為Horror，紅色為Thriller，可以看出兩者的分布較靠近正上方、右下方，左下方則較稀疏。

