大数据分析第六次作业

1. 使用原理：

根据用户与其他用户的相似度，根据观看电影的类型以及观看电影后的打分来进行相似度计算，从而寻找到对应的相似人群，再将相似人群对于某个电影的打分作为预测评分的依据推测出该用户对于这个电影可能的打分，再按打分由高到低排序，从而来对电影进行推荐。

1. 处理过程
2. 数据集划分，使用随机切分来对数据集进行划分，在代码中， 80%是训练集，20%是数据集。
3. 进行相似度计算，使用数据透视表生成userId\movieId\rating的表格矩阵，然后利用这个生成的矩阵求出相似矩阵，方便下一步处理
4. 进行数据预测：使用两个用户的相似度\*用户的评分求和来作为预测的评分
5. 寻找推荐：将评分预测最高的几个电影作为推荐推荐给用户
6. 进行模型准确率计算：我使用了标准误差作为模型准确率的计算，即均方根误差。

三．使用算法：协同过滤