基于 SVD 分析电影评分实践小项目报告

组员：梁展、刘姿廷

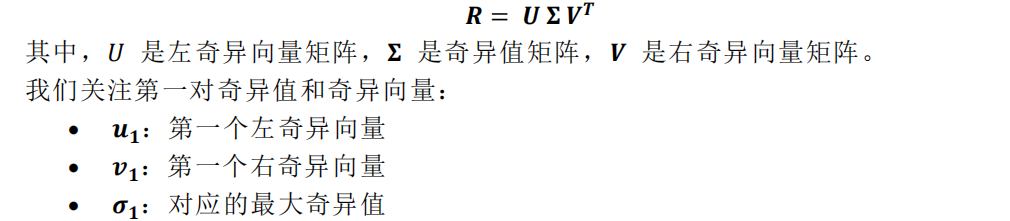
日期：2024年12月

学院：未来技术学院

## 简介

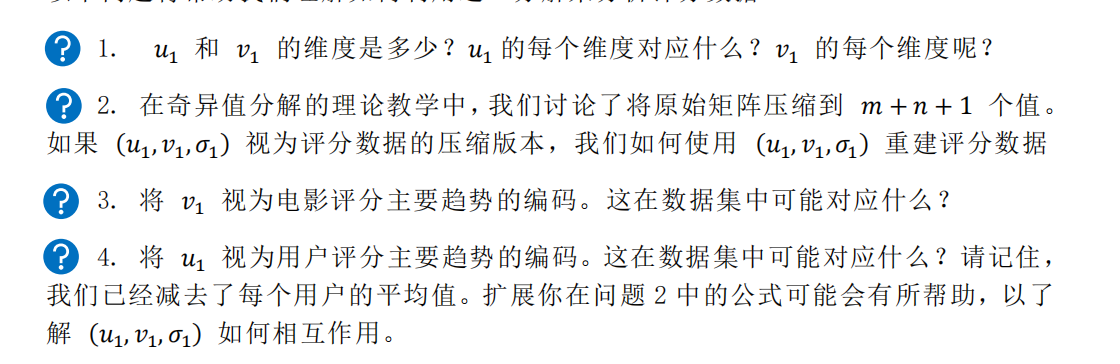
SVD是矩阵分解的一种方法，由𝑉𝐷𝑉𝑇 变换而来，完成行数和列数不相等情况下的特征值和特征矩阵的分解，从而提取多维度数据的特征。

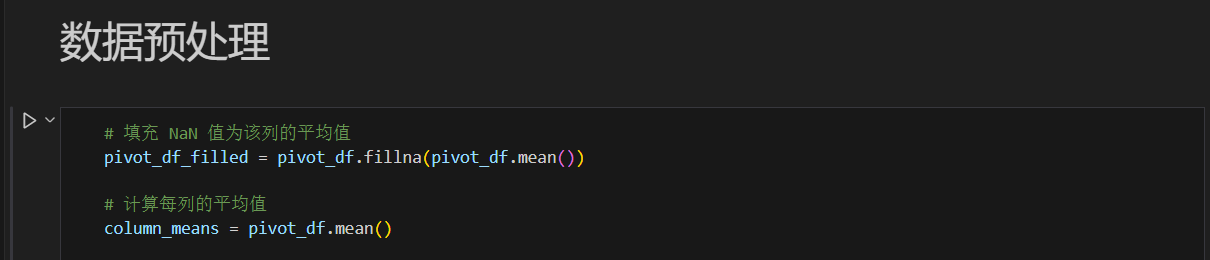
## 主要原理及其公式



## 项目问题回答

**一、使用 SVD 构建问题——SVD基本含义及其在代码中的对应**

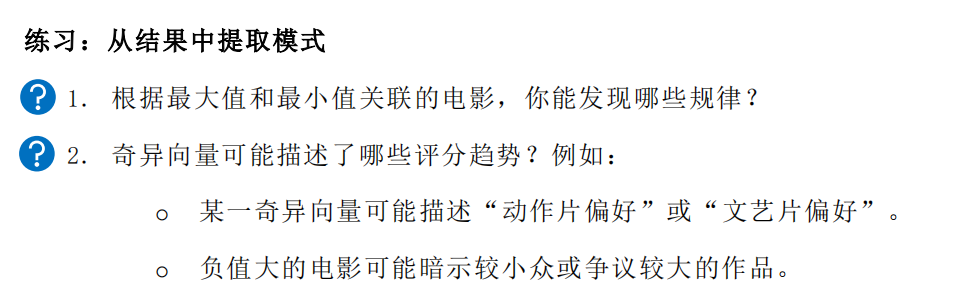


1. u为左奇异矩阵的列向量构成，v为右奇异矩阵的列向量构成， u的维度是9663\*1，v的维度是3790\*1。u的每个维度对应一个用户对一部电影的平均后的评分，v的每个维度对应一部电影里的一个用户的平均后的评分。
2. 可以通过重建评分公式其中bu为用户偏差，及数据处理中的对应的每位用户的平均值。

3.将v1 视为电影评分主要趋势的编码。这在数据集中可能对应于电影的总体流行度或受欢迎程度等主要因素。

4.将 u1视为用户评分主要趋势的编码。这在数据集中可能对应于用户的平均评分倾向或其他个性化偏好。

1. **右奇异向量构建问题**

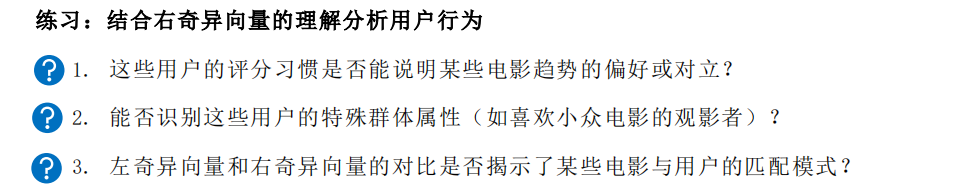


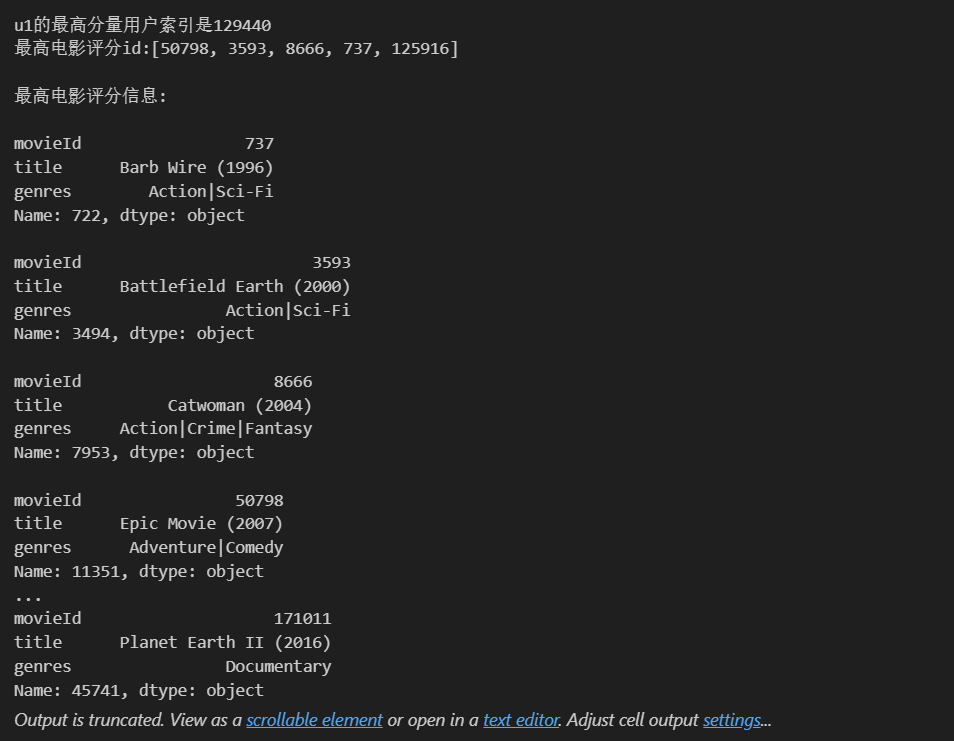
1. 通过表格得出与同一奇异值负相关最大和正相关最大两种情况下的影片多属于同一类型.

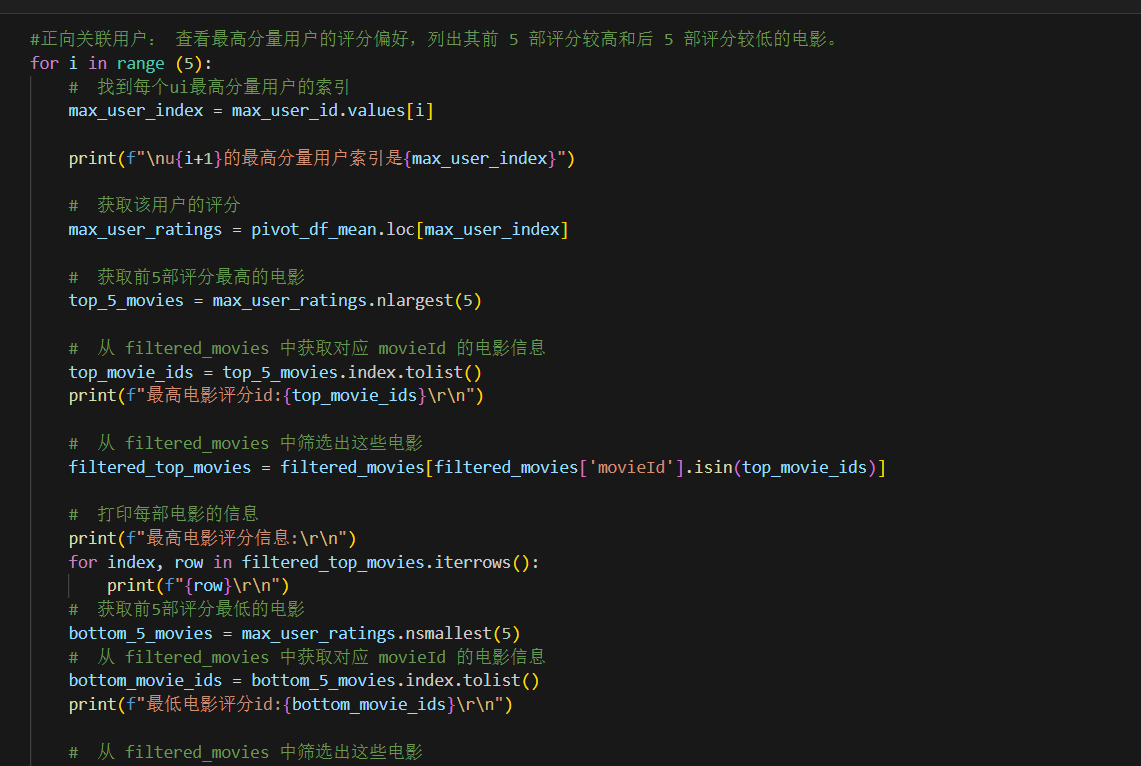
代码运行结果——与五个奇异值负相关最大的5部电影：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **向量** | **电影名1** | **电影名2** | **电影名3** | **电影名4** | **电影名5** |
| v1 | Children of Paradise (Les enfants du paradis)  天堂的孩子 | Widows’ Peak (1994)  寡妇岭 | Trip to Bountiful, The (1985)  布之旅 | Blue Angel, The (1930)  蓝天使 | 42 Up (1998)  42岁起 |
| v2 | Armageddon (1998)  世界末日 | Independence Day (a.k.a. ID4) (1996)  独立日 | Pearl Harbor (2001)  珍珠港 | Mummy Returns, The (2001)  木乃伊归来 | Wild Wild West (1999)  狂野西部 |
| v3 | Ghost (1990)  人鬼情未了 | Pretty Woman (1990)  风月俏佳人 | Top Gun (1986)  壮志凌云 | Crocodile Dundee (1986)  鳄鱼邓迪 | Home Alone (1990)  小鬼当家 |
| v4 | Fight Club (1999)  搏击俱乐部 | Pulp Fiction (1994)  低俗小说 | From Dusk Till Dawn (1996)  恶夜追杀令 | Starship Troopers (1997)  星河战队 | Natural Born Killers (1994)  天生杀人狂 |
| v5 | Fear and Loathing in Las Vegas (1998)恐惧与厌恶在拉斯维加斯 | Rocky Horror Picture Show, The (1975)  洛基恐怖秀 | Natural Born Killers (1994)  天生杀人狂 | Thin Red Line, The (1998)  细红线 | Mulholland Drive (2001)  穆赫兰道 |

1. 正值和负值的差距可以描述观众对电影评价的分歧程度
2. **左奇异向量构建问题——并结合右奇异向量分析客户行为**

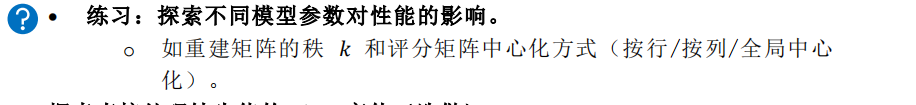


1．根据代码生成结果得出用户评分习惯会产生高评分电影趋势。

2. 能否识别这些用户的特殊群体属性

3.U和V作为相互的转置矩阵由有奇异向量得出的结果看来电影评分和用户种类相关联。

1. **SVD预测实践——评估推荐系统性能**



**暂无**

**五、SVD 算法的理解：**

**- 数学层面：SVD 的基本原理和在评分矩阵降维中的作用。**

SVD在行数和列数不相等情况下完成对矩阵的特征值和特征向量的分解，从而提取多维度数据的特征。从而完成评分矩阵降维

**- 应用层面：SVD 如何用于提取评分模式和预测缺失值。**

由分解后得到的特征值和特征矩阵可以总结用户打分习惯，并利用由得到的R矩阵反推用户的评分，从而预测缺失值。

**- 优缺点分析：SVD 应对数据稀疏性和维度问题的能力，及其局限性**

**优点：**

**提取重要特征：SVD能够从稀疏数据中提取出最重要的特征，通过保留最大的奇异值对应的奇异向量，可以有效地降低数据的稀疏性，提高后续数据分析的准确性。**

**降噪：SVD在分解矩阵时，可以去除噪声和冗余信息，只保留最重要的特征，这在处理稀疏数据时尤其有用，因为它可以减少噪声对模型的影响。**

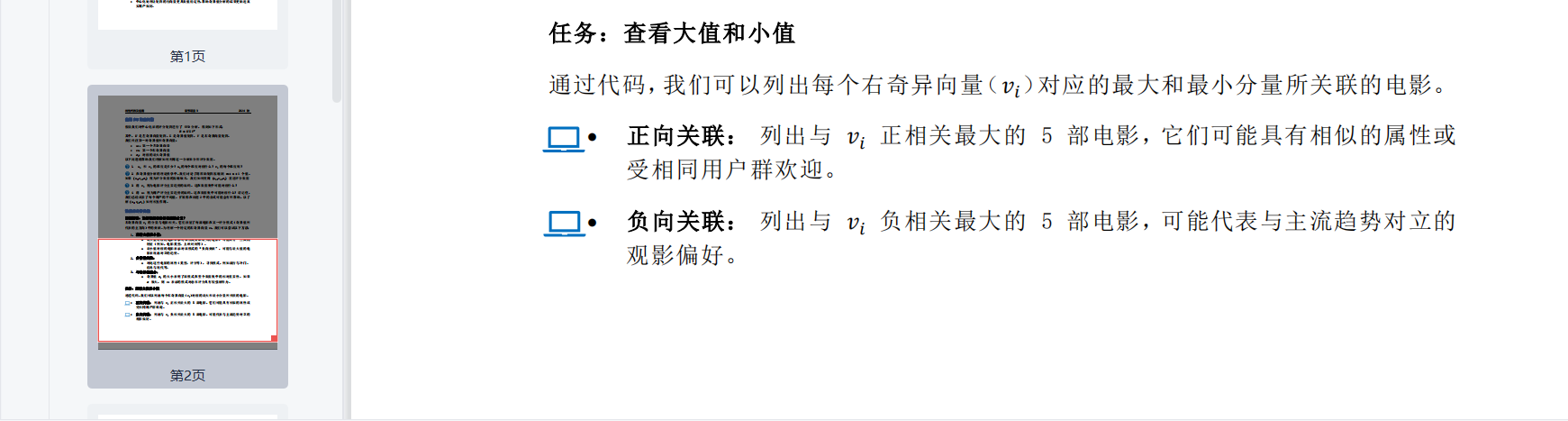
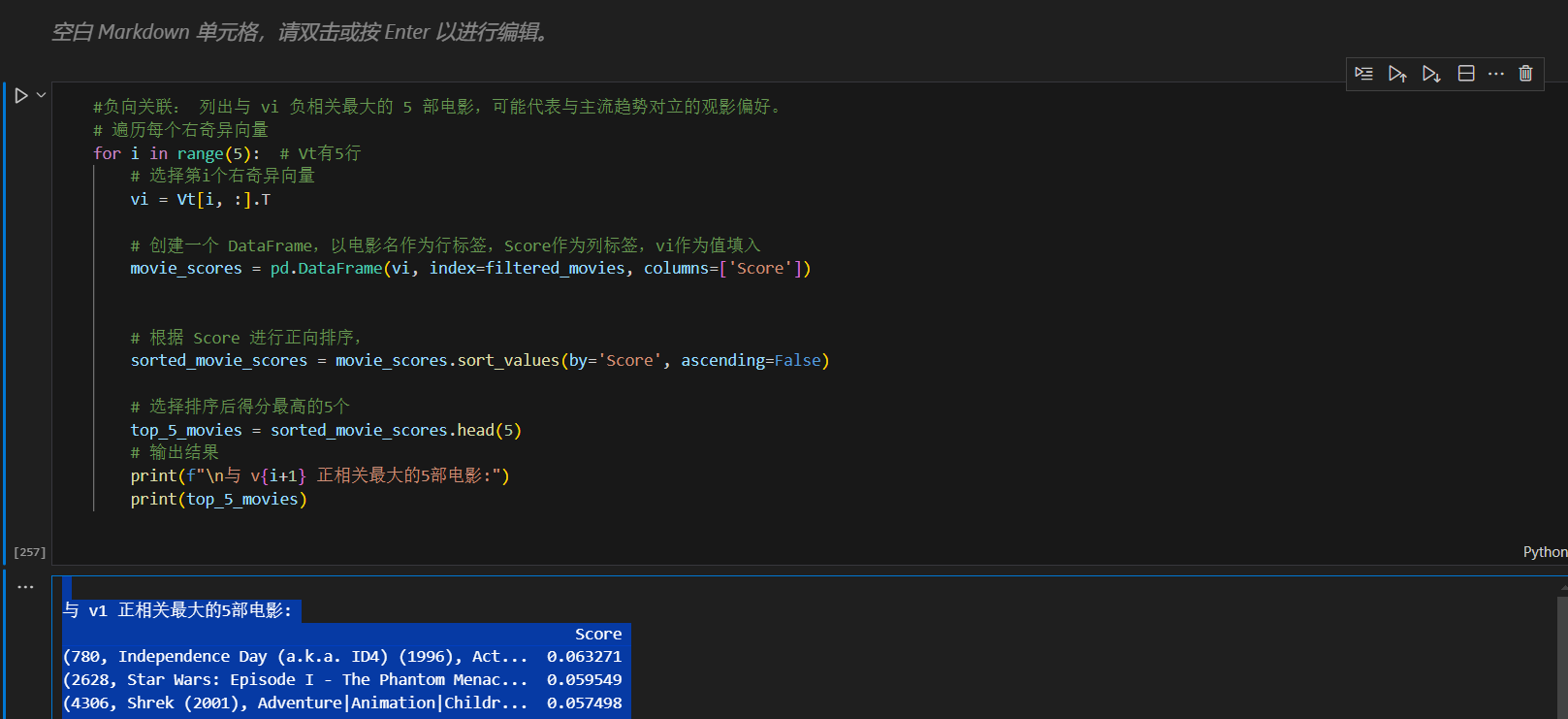
**缺点：**

**计算成本高：对于稀疏矩阵，虽然SVD能够提取重要特征，但是直接计算SVD的时间复杂度和空间复杂度都较高，特别是当数据规模很大时，计算成本尤为显著。**

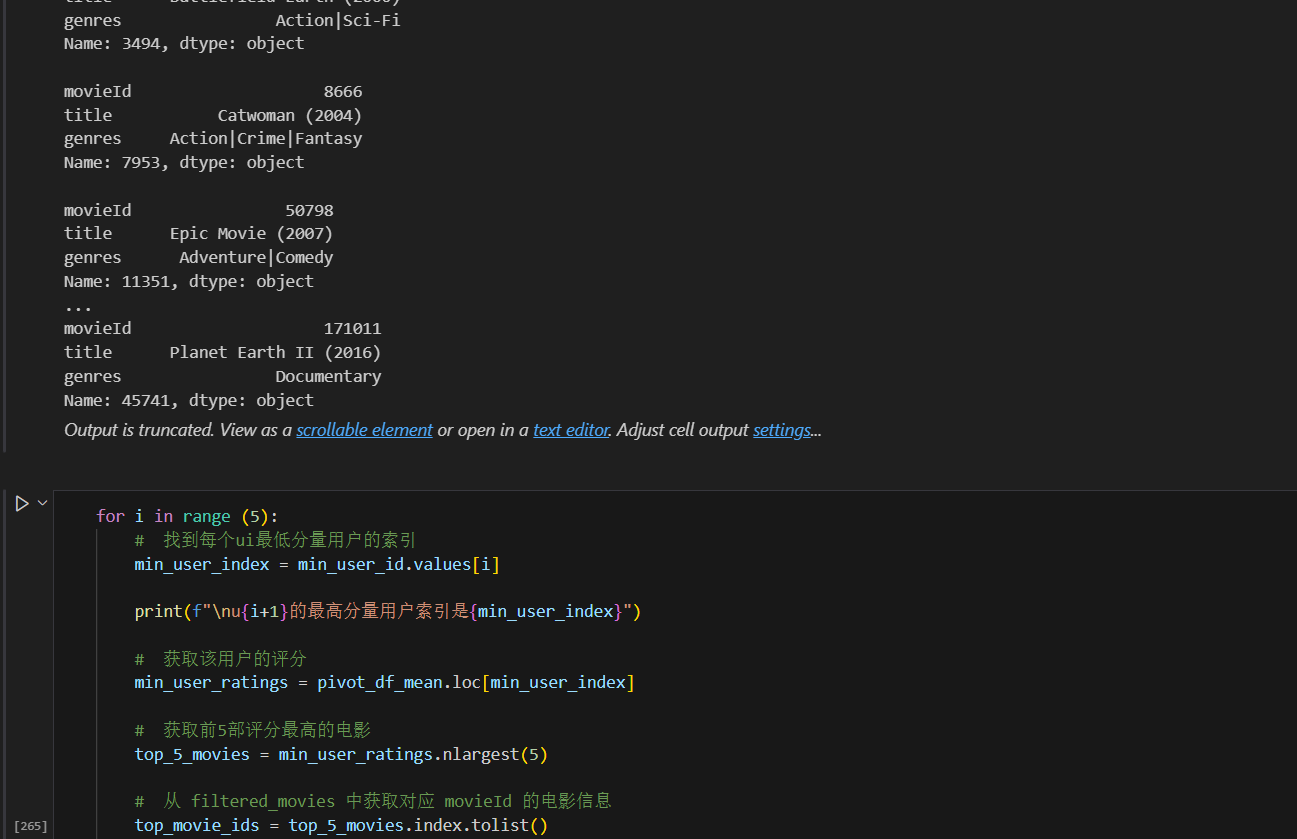
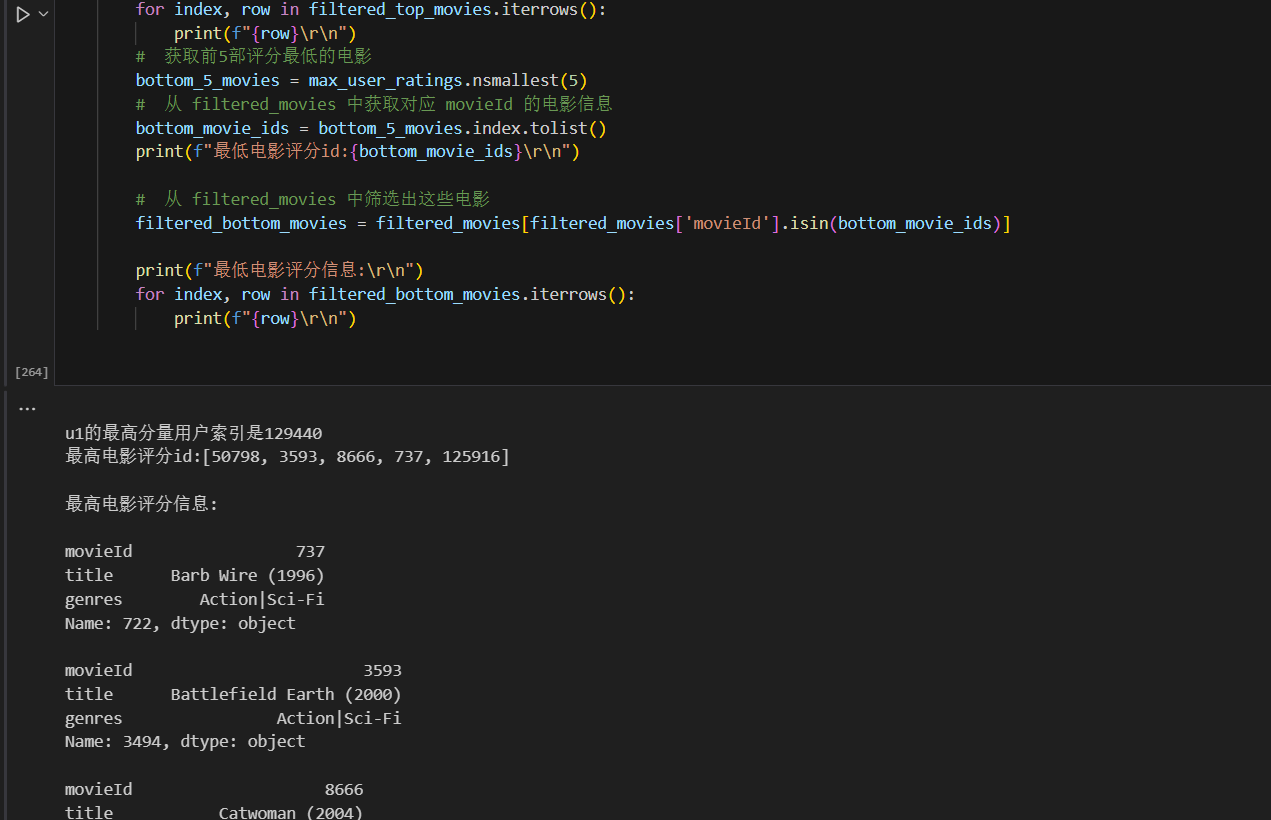
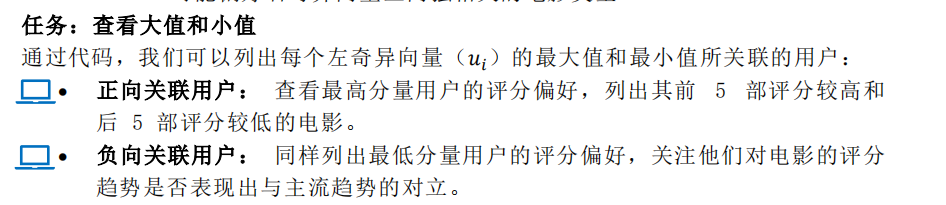
**稀疏性损失：在分解过程中，原始的稀疏矩阵可能会转化为密集矩阵，这可能导致稀疏性的损失，进而影响某些稀疏优化算法的性能**

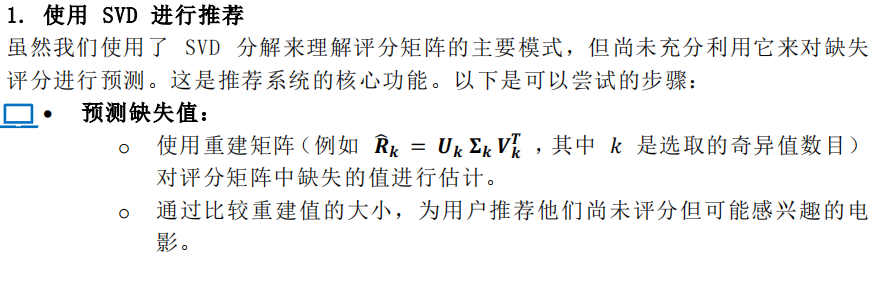
## 代码功能实现

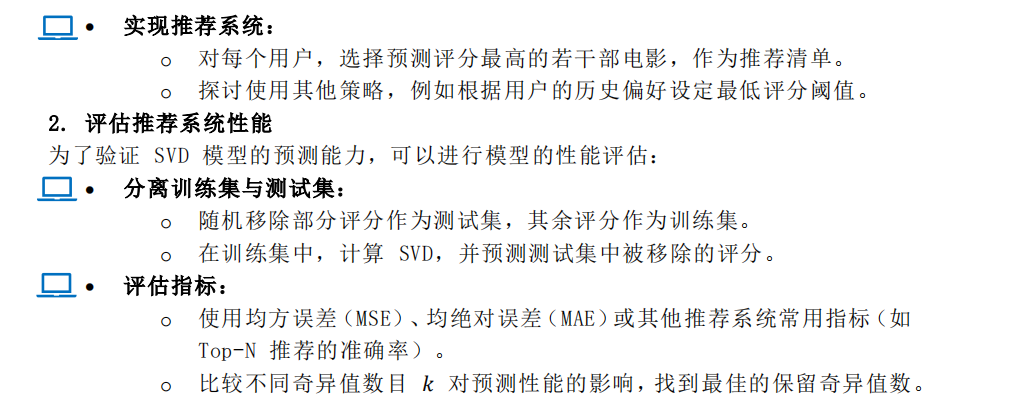
1. 右奇异向量构建——查看大值和奇异值

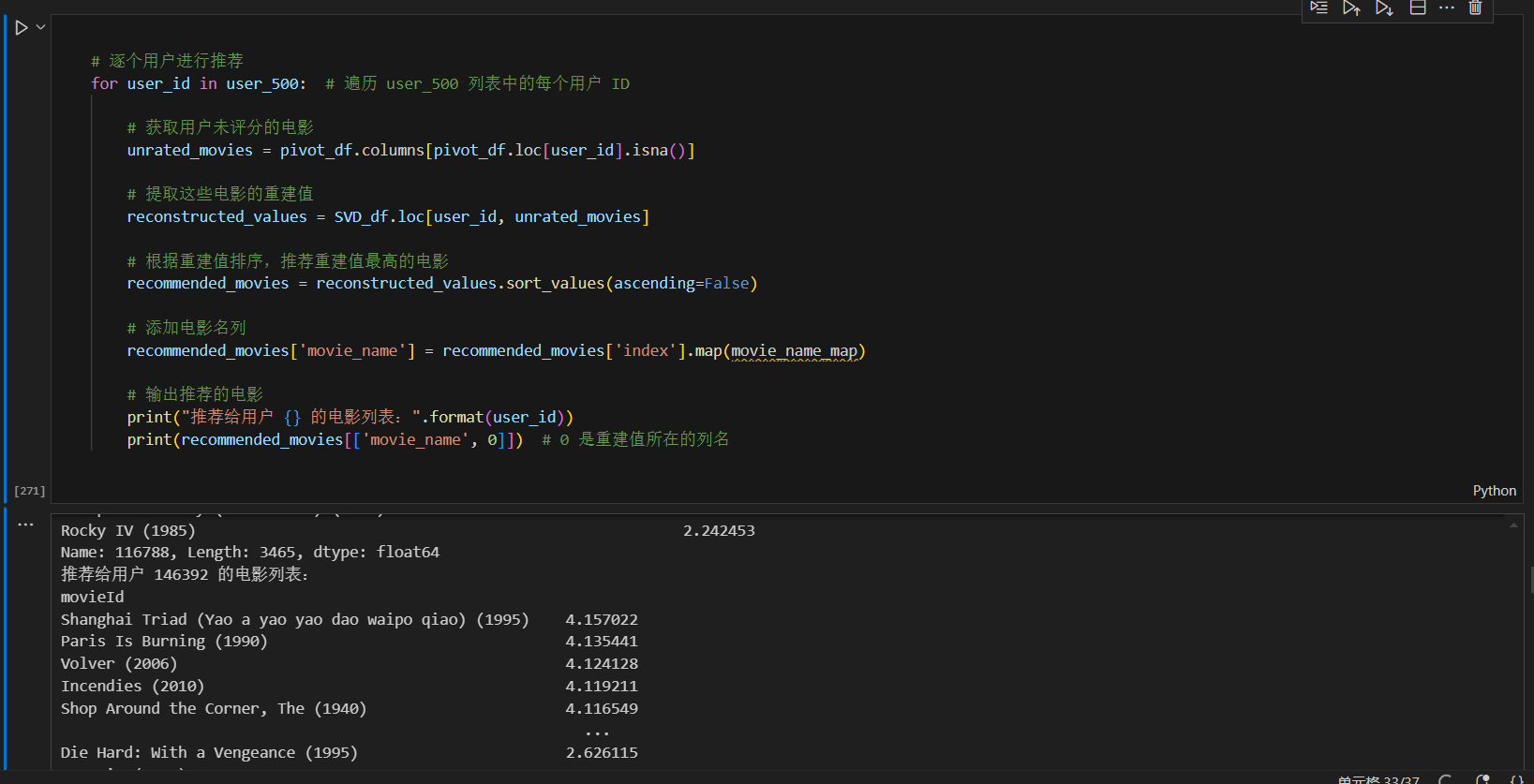
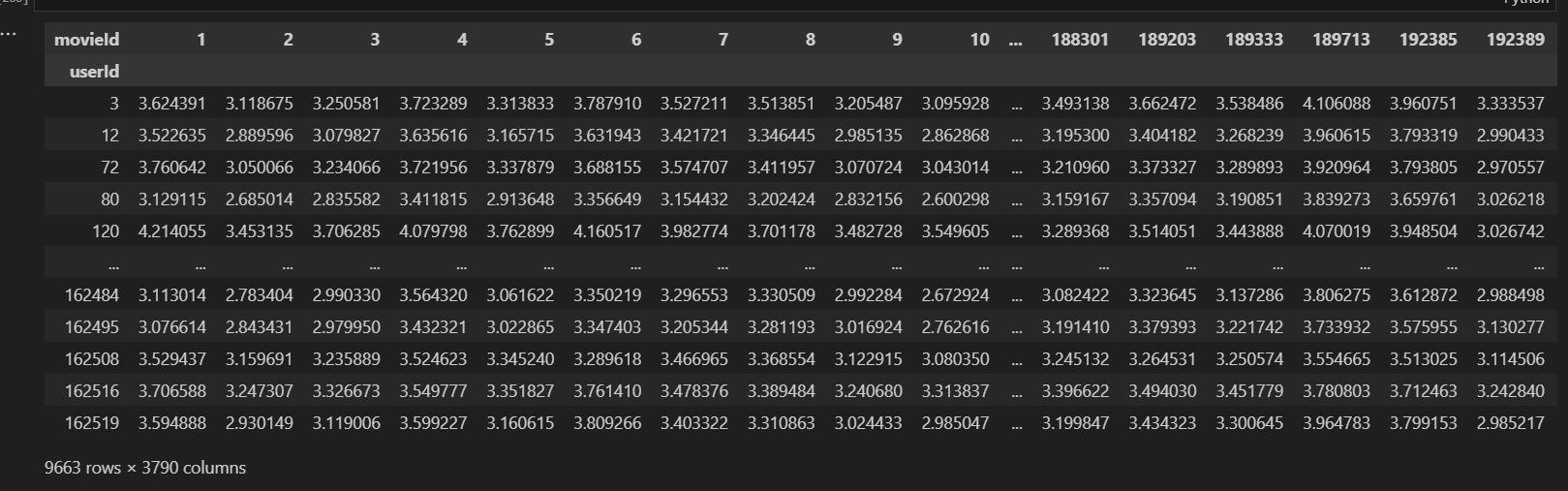
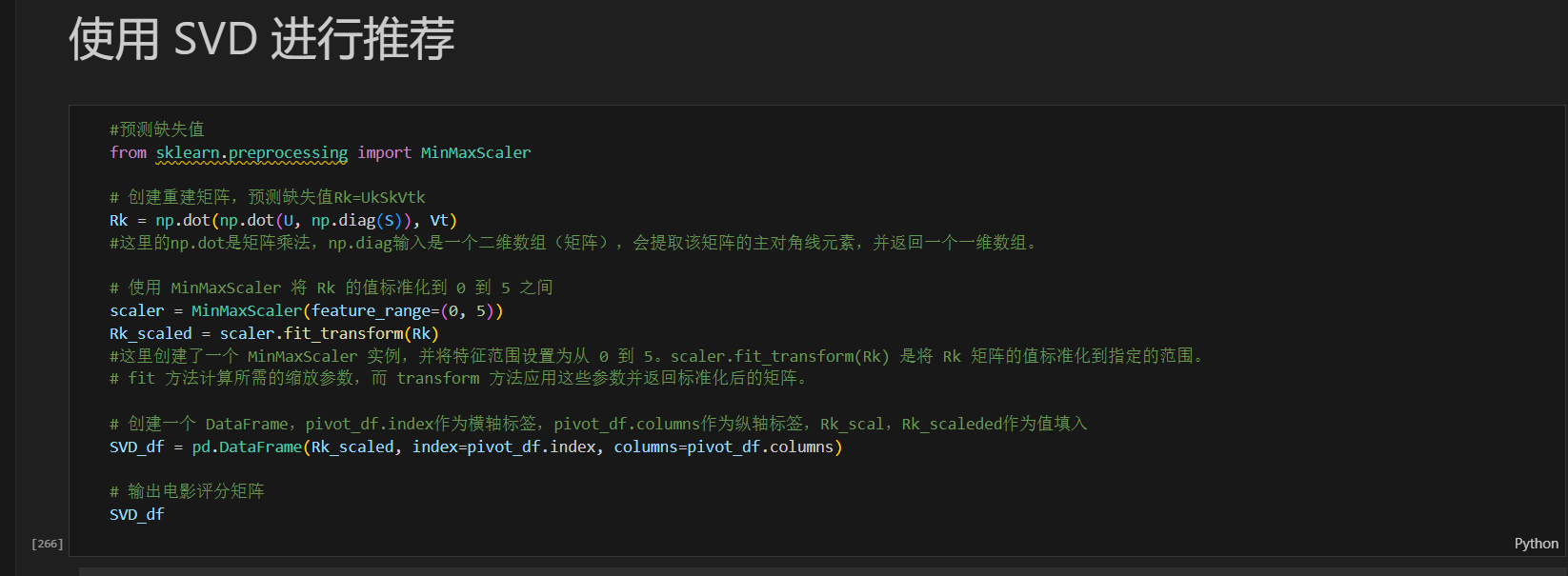
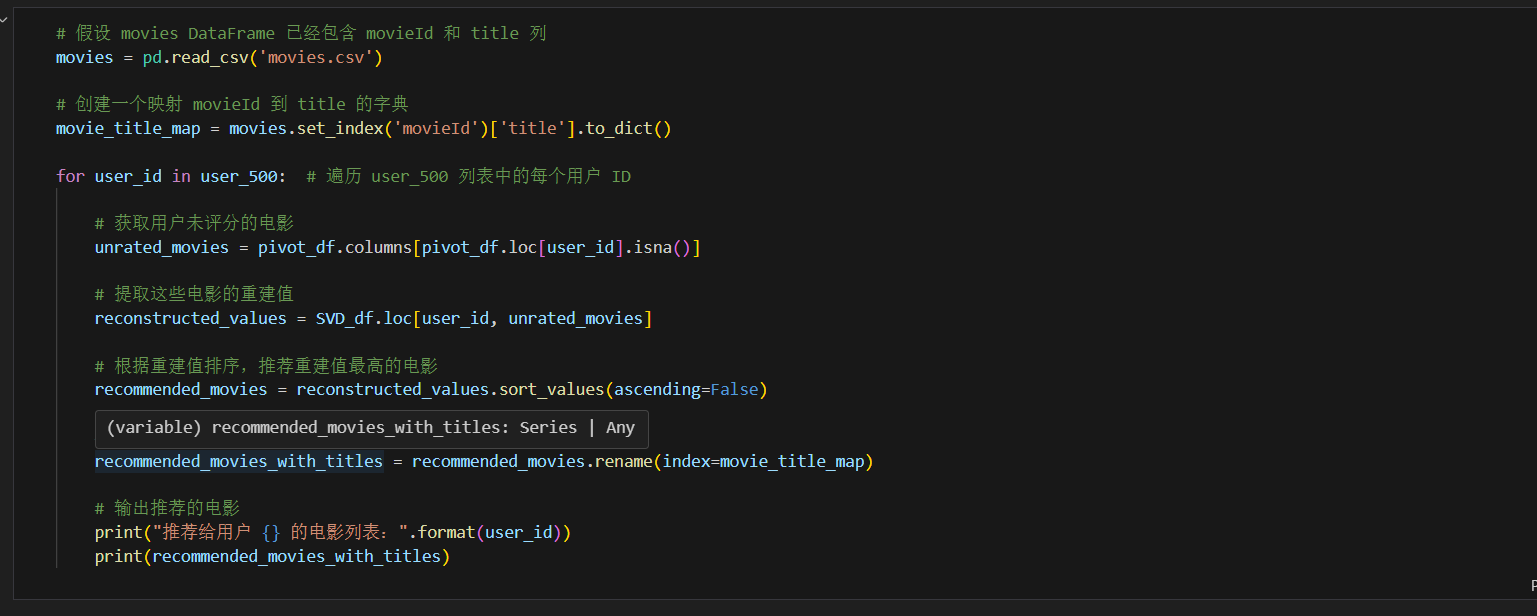
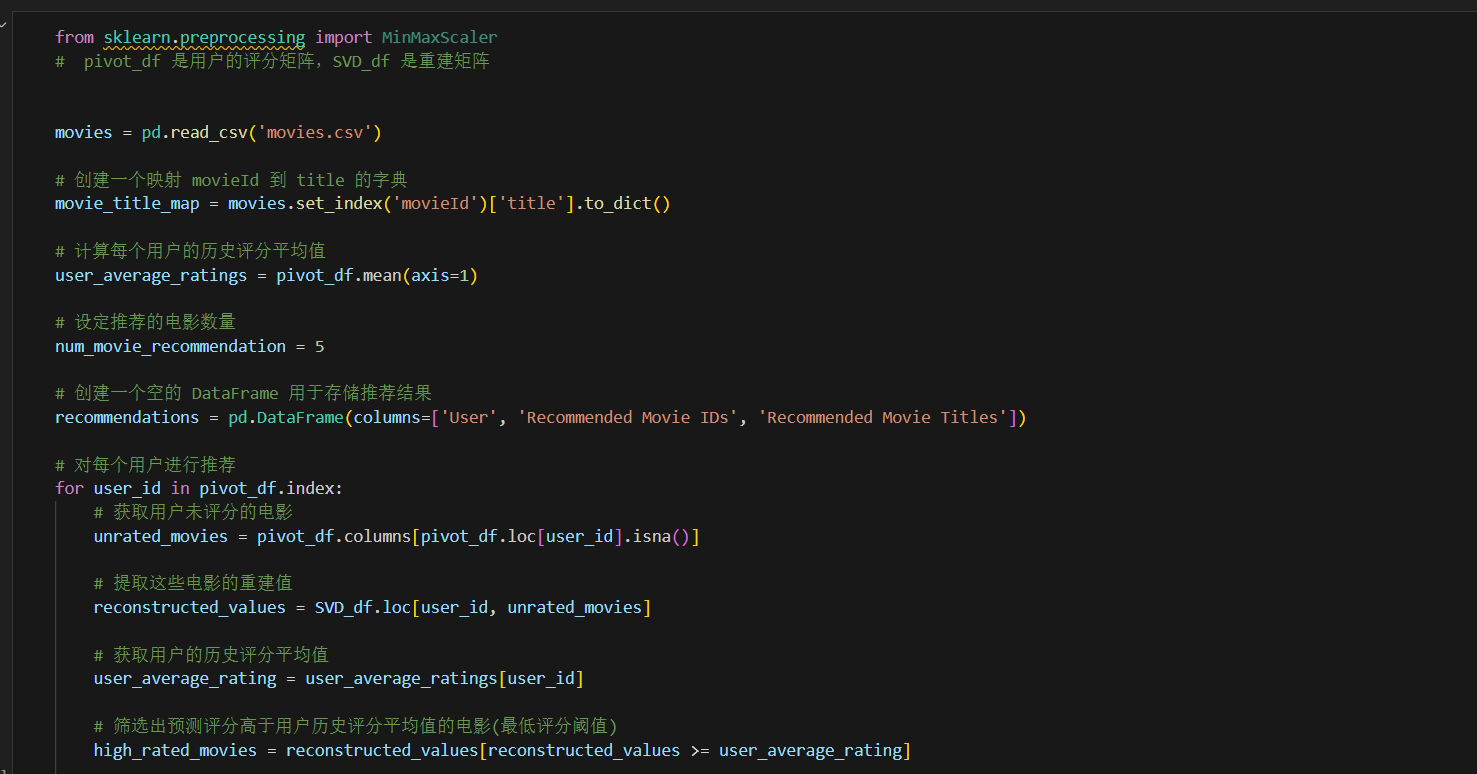
 

1. 左奇异向量构建—— 查看大值和奇异值

三、SVD运用





对应代码：  

四、可视化模块

暂无