**《信息检索》课程作业**

**题目：基于矩阵分解算法的评分预测实现**

**主要任务：**

1. 查阅相关资料，了解矩阵分解算法的基本概念、应用场景及其难点。重点了解SVD（Singular Value Decomposition，奇异值分解）系列方法。
2. 掌握Python语言的基本使用。
3. 了解梯度下降算法概念，熟悉并复现矩阵分解算法。
4. 在标准评测数据集MovieLens上验证矩阵分解算法。

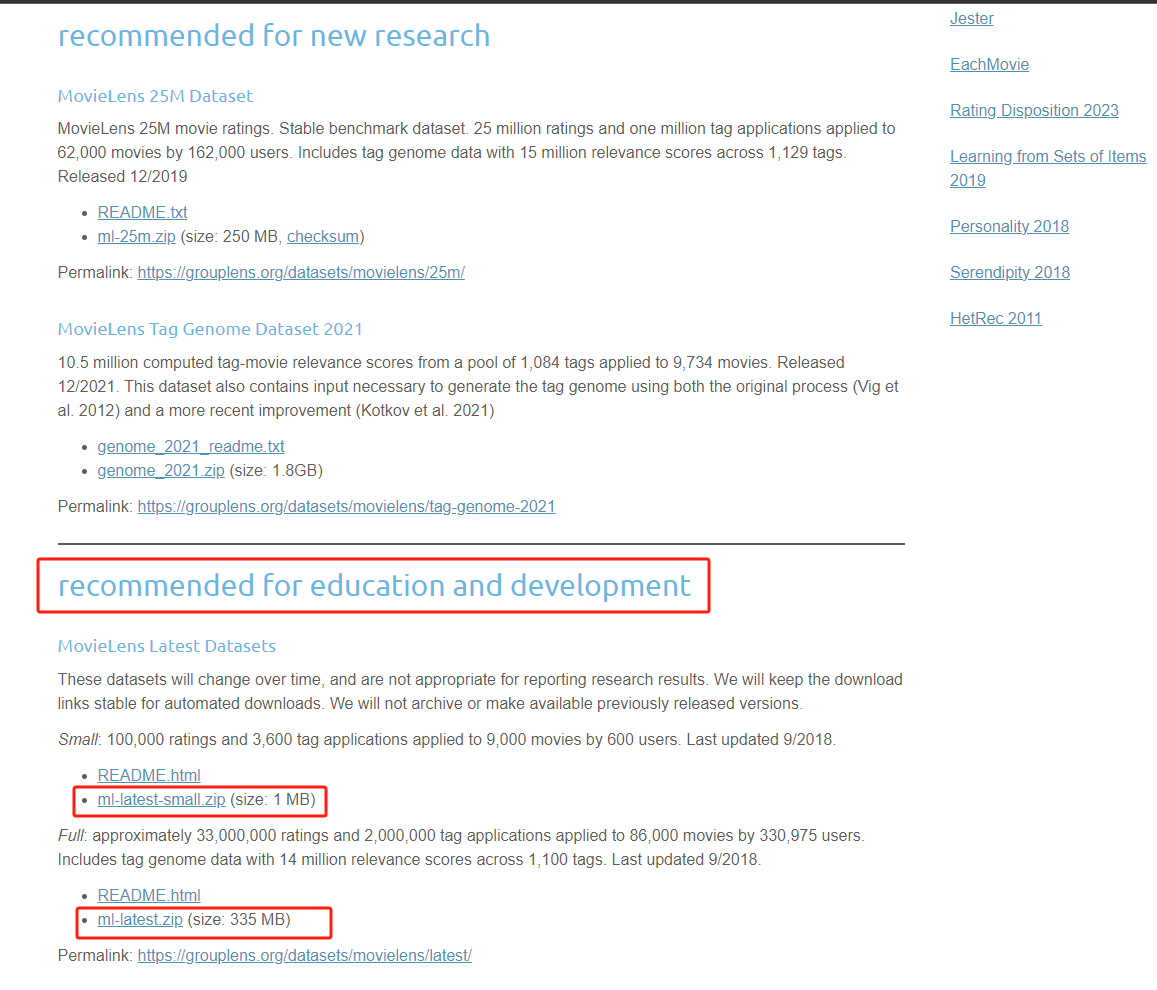
**实验环境：**

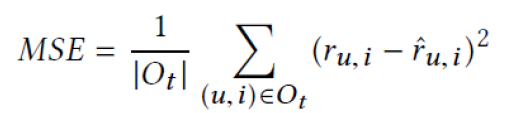
Windows或Linux, Python 3及以上

**数据集简介：**

MovieLen数据集包含多个用户对多部电影的评分数据，也包括电影元数据信息和用户属性信息。这个数据集经常用来做推荐系统，机器学习算法的测试数据集。本次作业主要利用MovieLens Latest Datasets (ml-latest.zip)解压后的ratings.csv和movies.csv文件，其中包含用户ID、电影ID、评分、电影名称、电影类型。若硬件条件不足，也可以使用ml-latest-small.zip数据集。

数据集下载地址为：<https://grouplens.org/datasets/movielens/>。



**评测标准：**

均方误差(Mean Squared Error, MSE):

**前期准备工作：**

1. 学习numpy或pandas的基本使用方法，能够对ratings.csv和movies.csv文件中的数据进行提取。
2. 使用numpy或sklearn中的知识，随机划分训练集、验证集、测试集，比例为8:1:1。不要把训练集作为测试集。

**预期成果或目标：**

使用Python语言或利用PyTorch、TensorFlow等深度学习库，复现矩阵分解算法，在标准评测数据集MovieLens上验证该算法，并且能够取得较低的均方误差(MSE不能高于1.5)。

**参考书目：**

《推荐系统实践》；

《Python编程从入门到实践》；

《动手学深度学习》。

**参考教程：**

1. [知乎-推荐基础算法之矩阵分解MF](https://zhuanlan.zhihu.com/p/268079100)，全文可阅读；
2. [推荐系统实战之评分预测问题](https://zhuanlan.zhihu.com/p/241968278)，重点阅读第2.3节；
3. [Netflix Prize 矩阵分解(Matrix factorization)预测用户评分](https://blog.csdn.net/SJTUzhou/article/details/106596803#:~:text=%E7%AC%94%E8%80%85Github%E9%93%BE%E6%8E%A5%EF%BC%9A,https%3A%2F%2Fgithub.com%2FSJTUzhou%2FNetflixPrizeMatrixFactorization)，可参考。

**助教：**

邓世龙 [465193106@qq.com](mailto:sldeng@std.uestc.edu.cn)

**作业提交：**

将作业过程和作业结果整合为实验报告，发送到助教邮箱[465193106@qq.com](mailto:sldeng@std.uestc.edu.cn)。

格式（邮件名和文件名）：姓名-学号-信息检索编程作业（例如：张三-20XX-信息检索编程作业）

截止日期：2023年11月20日