

智能系统原理与开发 Lab2

声明

- 出现抄袭现象，抄袭双方均按零分计。
- 请严格按照 deadline 提交，超出每天扣除总分的 20%。
- 更多问题 @四位助教咨询。

数据集说明

数据集将会发布在群内，`train.csv` 文件，共2967个user、4125个item。测试集文件将在最终测试准确率时公布。训练集和测试集结构如下（userID为0~2966，itemID为0~4124）：

	userID	itemID	rating	timestamp
0	2409	256	5.0	1114232409
1	2533	1116	4.0	1351541044
2	422	369	4.0	1068357977
3	840	53	4.0	1010512429
4	521	167	3.5	1107571443
...

测试集测试指标统一采用 $RMSE$ ：

$$RMSE = \sqrt{\frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} (r_{ui} - \hat{r}_{ui})^2}.$$

DEADLINE：2019/12/15 23:59:59

任务：

基于给定数据集完成协同过滤的相关算法。训练集含有大量用户对不同物品的评分，需要建立以下协同过滤算法推测用户对其他物品的评分：

1. 分别实现基于user_based、item_based的算法
2. 实现Probabilistic Matrix Factorization算法

3. 自由发挥提出改进的算法，提升精度

要求：

- （40分）使用 python 语言实现，可以使用 numpy 中基本的矩阵运算，不能使用已有的推荐系统的包，类似 surprise.SVD 等
- （20分）任务2中的PMF需使用 ALS 算法求解
- （20分）任务3中尝试改进算法提高精度，最终以测试集的班级精度排名为评分参考
- （20分）编写实验文档，可以包括但不限于：
 - 代码基本结构
 - 不同算法、实验参数的比较
 - 对算法的改进思路、实验等
- 面试时助教仅提供测试集文件，格式同训练集，建议事先写好 test.py 文件，调用自己的算法，输出 user_based 、 item_based 、 PMF 以及自己改进算法的 RMSE 结果。
- 文档要求工整、详实、美观，并导出为 PDF 格式。

提交

将所有代码、文档放在 学号-姓名 文件夹下，打包上传到 WORK_UPLOAD/LAB2 目录下，例如 12345054321-林则徐 。