

### 第三周作业 在 Rental Listing Inquiries 数据上练习 xgboost 参数调优

#### 数据说明：

Rental Listing Inquiries 数据集是 Kaggle 平台上的一个分类竞赛任务，需要根据公寓的特征来预测其受欢迎程度（用户感兴趣程度分为高、中、低三类）。其中房屋的特征  $x$  共有 14 维，响应值  $y$  为用户对该公寓的感兴趣程度。评价标准为 logloss。

**数据链接：**<https://www.kaggle.com/c/two-sigma-connect-rental-listing-inquiries>

为减轻大家对特征工程的入手难度，以及统一标准，数据请用课程网站提供的特征工程编码后的数据（RentListingInquiries\_FE\_train.csv）或稀疏编码的形式（RentListingInquiries\_FE\_train.bin）。xgboost 既可以单独调用，也可以在 sklearn 框架下调用。大家可以随意选择。若采用 xgboost 单独调用使用方式，建议读取稀疏格式文件。

关于特征工程的过程，可参看文件：FE\_RentListingInquiries.ipynb

#### 作业要求：

采用 xgboost 模型完成商品分类（需进行参数调优）。