

第二周作业 在 Rental Listing Inquiries 数据上练习分类方法

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>

训练数据(RentListingInquiries_train.json)和测试数据(RentListingInquiries_test.json)已经上传到课程网站。

作业要求：

利用 Logistic 回归和 SVM 分类计数实现 Rental Listing Inquiries 数据的感兴趣程度预测。

1. 对数据做数据探索分析
2. 适当的数据清洗（异常值处理）
3. 适当的特征工程
4. 用 Logistic 回归模型对公寓感兴趣程度进行预测，注意正则超参数的调优。
5. 用 RBF 的 SVM 对公寓感兴趣程度进行预测，注意正则超参数和核函数参数的调优。