

第一周作业 利用线性回归技术实现 Ames 房价预测

数据说明： Ames 房价预测是 Kaggle 平台上的一个竞赛任务，需要根据房屋的特征来预测亚美尼亚州洛瓦市 (Ames , Iowa) 的房价。其中房屋的特征 x 共有 79 维，响应值 y 为每个房屋的销售价格 (SalePrice)。评价标准为预测值的对数和观测值的对数的 RMSE(Root-Mean-Squared-Error)。

数据链接 <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

训练数据 (Ames_House_train.csv) 和测试数据 (Ames_House_test.csv) 已经上传到课程网站 (<http://edu.csdn.net/course/detail/6368>)。

作业要求：

1. 对数据做数据探索分析
2. 适当的数据清洗 (异常值处理和缺失值处理)
3. 适当的特征工程
4. 用线性回归模型对房价进行预测 (最小二乘、岭回归、Lasso)，注意正则超参数的调优。

特征说明：

SalesPrice：房屋售价 (美元)，这是要预测的目标变量。

MSSubClass：房地产建筑类别

MSZoning：地区分类

LotFrontage：房屋到街道的直线距离 (英尺)

LotArea：土地的大小 (平方英尺)

Street：所在道路的类型

Alley：所在巷通道的类型

LotShape：房地产形状

LandContour：土地的平整性

Utilities：可用的工具

LotConfig：土地配置

LandSlope：财产的斜度

Neighborhood:市内的物理位置

Condition1: 靠近主干道或铁路

Condition2: 靠近主要道路或铁路 (如果还有第二个)

BldgType: 住宅类型

HouseStyle: 住宅风格

OverallQual: 整体材质和完成品质

OverallCond: 总体条件评级

YearBuilt: 最初建造日期

YearRemodAdd: 重建日期

RoofStyle: 屋顶类型

RoofMatl: 屋顶材料

Exterior1st: 房屋外墙

Exterior2nd: 房屋外墙 (如果多于一种材料)

MasVnrType: 表层砌体 (Masonry veneer) 类型

MasVnrArea: 表层砌体面积 (square feet)

ExterQual: 外部材料的质量

ExterCond: 外部材料的现状

Foundation: 地基类型

BsmtQual: 地下室高度

BsmtCond: 地下室的总体状况

BsmtExposure: Walkout 或花园层地下室的墙壁

BsmtFinType1: 完成的地下室的质量

BsmtFinSF1: 第一类型 (Type 1) 完成面积面积 (square feet)

BsmtFinType2: 第二类型完成面积(如果有)

BsmtFinSF2: Type 2 完成面积 (square feet)

BsmtUnfSF: 未完成的地下室总面积 (square feet)

TotalBsmtSF: 地下室总面积 (square feet)

Heating: 取暖类型

HeatingQC: 取暖质量和条件

CentralAir: 中央空调

Electrical: 电气系统

1stFlrSF: 第一层的面积 (square feet)

2ndFlrSF: 第二层的面积 (square feet)

LowQualFinSF: 低质量完成的面积 (square feet) (所有楼层)

GrLivArea: 地上居住面积 (square feet)

BsmtFullBath: 地下室全浴室数目

BsmtHalfBath: 地下室半浴室数目

FullBath: 地上全浴室数

HalfBath: 地上半浴室数目

Bedroom: 地下室之上的卧室数目

Kitchen: 厨房树木

KitchenQual: 厨房质量

TotRmsAbvGrd: 地上房间总数(不包括浴室)

Functional: 家庭功能评级

Fireplaces: 壁炉的数目

FireplaceQu: 壁炉质量

GarageType: 车库位置

GarageYrBlt: 车库建造年份

GarageFinish: I 车库内部装修

GarageCars: 车库大小 (能停的车辆数目)

GarageArea: 车库大小 (square feet)

GarageQual: 车库质量

GarageCond: 车库条件

PavedDrive: 铺设的车道

WoodDeckSF: 木头 deck 面积 (square feet)

OpenPorchSF: 开放门廊面积 (square feet)

EnclosedPorch: 封闭门廊面积 (square feet)

3SsnPorch: 三季门廊面积 (square feet)

ScreenPorch: 观景门廊 (Screen porch) 面积 (square feet)

PoolArea: 游泳池面积 (square feet)

PoolQC: 游泳池质量

Fence: 围栏质量

MiscFeature: 没被包含在其他类别的杂项功能

MiscVal: 杂项功能价值 (\$)

MoSold: 销售月份

YrSold: 销售年份

SaleType: 销售类型

SaleCondition: 销售条件

CSDN学院 edu.csdn.net