

构建并评估回归模型

使用sklearn估计其构建回归模型

回归分为学习和预测

学习：是通过训练样本数据来拟合回归方程

预测：是利用学习过程中拟合出来的方程，将测试数据放入方程中求出预测值。

常用回归模型

回归模型名称	通用条件	算法描述
线性回归	因变量与自变量是线性关系	对一个或多个自变量和因变量直接按的线性关系进行建模，可用最小二乘求解模型系数
非线性回归	因变量与自变量之间不是线性关系	对一个或多个自变量和因变量之间的非线性关系进行建模。如果非线性关系可以通过简单的函数变换转成线性关系，用线性回归的思想求解，如果不能转换，用非线性最小二乘方法求解。
Logistic回归	因变量一般有1和0（是与否）两种取值	是广义线性回归的特例，利用logistic函数将因变量的取值范围控制在0和1之间，表示取值为1的概率。
岭回归	参与建模的自变量之间具有多重共线性	是一种改进最小二乘估计的方法。
主成分回归	参与建模的自变量之间具有多重共线性	主成分回归是根据主成分分析的思想提出的，是对最小二乘法的一种改进，它是参数估计的一种有偏估计。可以消除自身之间的多重共线性。

Sklearn库常用回归算法函数

- Sklearn内部提供了不少回归算法，常用的函数如下
- 可以利用预测结果和真实结果画出折线图做对比，以便更直观看出线性回归模型效果。

模块名称	函数名称	算法名称
linear_model	LinearRegression	线性回归
svm	SVR	支持向量回归
neighbors	KNeighborsRegression	最近邻回归
tree	DecisionTressRegression	回归决策树
ensemble	RandomForestRegressor	随机森林回归
ensemble	GradientBoostingRegressor	梯度提升回归树

回归模型评价指标

- 回归模型的性能评估不同于分类模型，虽然都是对照真实值进行评估，但由于回归模型的预测结果和真实值都是连续的，所以不能够求取Precision、Recall和F1值评价指标。回归模型拥有一套独立的评价指标。
- 平均绝对误差、均方误差和中值绝对误差靠近0，模型性能越好。可解释方差值和R方值则越靠近1，模型性能越好。

方法名称	最优值	sklearn函数
平均绝对误差	0.0	metrics.mean_absolute_error
均方误差	0.0	metrics.mean_squared_error
中值绝对误差	0.0	metrics.median_absolute_error
可解释方差值	1.0	metrics.explained_variance_score
R方差	1.0	metrics.r2_score