

线性回归, 岭回归以及 Lasso 回归

邵李翔 赵张弛 祖劭康

日期: April 12, 2021

1 数据集描述

本次实验采用的数据集为 Boston (波士顿房价) 数据集, 它记录了波士顿周围 506 个街区的 medv (房价中位数)。我们将设法用 13 个预测变量如 rm (每栋住宅的平均房间数), age (平均房龄), lstat (社会经济地位低的家庭所占比例) 等来预测 medv (房价中位数)。

表 1: 数据集中部分数据展示

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
1	0.01	18.00	2.31	0.00	0.54	6.58	65.20	4.09	1.00	296.00	15.30	396.90	4.98	24.00
2	0.03	0.00	7.07	0.00	0.47	6.42	78.90	4.97	2.00	242.00	17.80	396.90	9.14	21.60
3	0.03	0.00	7.07	0.00	0.47	7.18	61.10	4.97	2.00	242.00	17.80	392.83	4.03	34.70

2 线性回归

使用 `lm()` 函数对所有预测变量进行多元回归, 得到的结果如下

表 2: 多元线性回归拟合结果

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	36.4595	5.1035	7.14	0.0000
crim	-0.1080	0.0329	-3.29	0.0011
zn	0.0464	0.0137	3.38	0.0008
indus	0.0206	0.0615	0.33	0.7383
chas	2.6867	0.8616	3.12	0.0019
nox	-17.7666	3.8197	-4.65	0.0000
rm	3.8099	0.4179	9.12	0.0000
age	0.0007	0.0132	0.05	0.9582
dis	-1.4756	0.1995	-7.40	0.0000
rad	0.3060	0.0663	4.61	0.0000
tax	-0.0123	0.0038	-3.28	0.0011
ptratio	-0.9527	0.1308	-7.28	0.0000
black	0.0093	0.0027	3.47	0.0006
lstat	-0.5248	0.0507	-10.35	0.0000