

实验一 线性回归算法实验

一. 简单介绍 OLS,岭回归以及 Lars 三种线性回归算法的原理

二. 模拟数据实验

1. 数据生成：按照有无共线性，有无噪声生成四组模拟数据（数据维度自定义）。-----同时，模拟数据可参考：Sayan Mukherjee and Dingxuan Zhou. 《learning coordinate covariancs via gradients》一文中 5.1 节的模拟数据，或者其他自定义数据。

2. 分别利用三种算法对四组模拟数据建立线性回归模型，并利用 R^2 、MSE 和 MAE 等指标评价模型性能，从而对比三种算法的性能，分析三种算法在线性回归中的优缺点以及不同的作用。

3. 对 1 中同一数据进行多次采样，每次采样后使用 Lars 算法进行变量选择，进而探索 Lars 算法在变量选择中的稳定性---即每次所选变量的相同或者不同情况，并进行变量选择前后回归学习效果的对比。

三. Boston 数据实验

- 1.对 Boston 房价数据进行探索性分析；
- 2.数据的预处理：归一化，缺失值处理等；
- 3.数据的特征工程：特征选择、增加特征等；
- 4.对三种算法在该数据上的性能进行分析。