

**《机器学习》课程论文**

**[](http://bysj.zjgsu.edu.cn/)**

题目：XXXXXXX（自拟）

（看后请删除：课程论文总体要求：使用一个或者多个课堂实验未使用过的数据，进行多种算法的综合性分析。全文14页左右，格式见本模板，注意图、表的标题及位置。可参考数据下载网站：1. UCI数据库：

http://archive.ics.uci.edu/datasets（无需注册）2. Kaggle数据库：

https://www.kaggle.com/datasets（需要注册））

**学 院：统计与数学学院**

**专 业：**

**学 号：**

**学生姓名：**

**2024年 5 月**

摘要

最小角回归(Least Angle Regression)是一种模型选择算法，。。。。。。。

本文主要简述。。。。的步骤，对比。。。。。的结果，并分析了。。。。

**关键词：**A；B；C；D

(注1：上述只参考格式，不参考内容和字体颜色；

注2：课程论文涉及的数据请勿选择课程以及实验实践中使用过的数据；

注3：正文中使用的图，表的格式请参考模板，特别是图表的标题；

注4：要求总页数（含封面、目录）14页及以上。)

**目录**

[一、 LARS算法简述 １](#_Toc11860954)

[二、 其他算法 １](#_Toc11860955)

[三、 模拟实验 ２](#_Toc11860956)

[四、 总结 ９](#_Toc11860957)

[参考文献 １０](#_Toc11860958)

1. LARS算法简述

在高维的情况下，如何得到稳定准确的模型是一大难题。一方面，使用普通模型拟合高维数据常常具有很大的方差，另一方面，我们通常希望使用较少的特征进行预测。因此有必要使用算法对特征进行选择，以保证模型的稳定性和简捷性。常用的算法有岭回归(Ridge Regression),逐步回归(Forward Stepwise Regression)等。

1. 其他算法

为了方便比较各算法在不同情况下的表现，首先对各算法进行简单介绍。

1. 模拟实验

为了更直观地了解算法的表现，参考Lasso(Tibshirani 1996)，弹性网络(Hui Zou and Trevor Hastie 2003)中的模拟研究，对LARS，Lasso，岭回归,弹性

表 1 各例下100次重复实验MSE中位数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | Ex.1 | Ex.2 | Ex.3 | Ex.4 |
| LARS | 1.89 | 2.56 | 45.23 | 249.46 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

可见Lasso的MSE最小，其余三者在Ex.1的数据中的表现相近。

可见LARS和Lasso的变量选择效果较好，第1,2,5个特征被选择的频率明显高于其他特征。岭回归的变量选择效果最差，这是由于岭回归使用L2正则化，只会缩小不重要特征的系数，但不会将其设置为0。

对岭回归结果做额外处理，只有系数的特征才被认为是重要的，重新作图，得到岭回归的特征出现频率如下

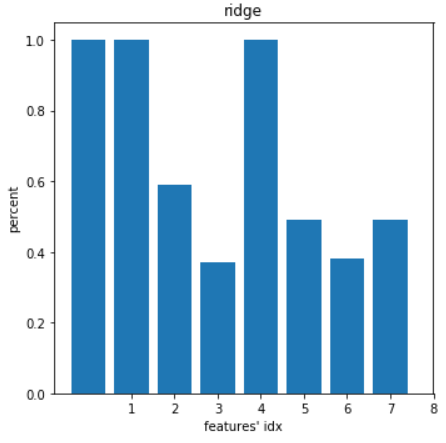


图 3 修正后的岭回归变量选择结果

**Example 4:** 数据集包含40个特征，共500个样本，训练集、验证集、测试集比例为，，样本按如下方式生成：

1. 总结

在少量特征起作用而且相关性不强的情况下(如example1)，Lasso和LARS的表现良好，而且筛选的特征有较高的准确率。

在特征权重相近的时(如example2)，

参考文献

[1] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1): 267-288.

[2] Zou H, Hastie T. Regularization and variable selection via the elastic net[J]. Journal of the royal statistical society: series B (statistical methodology), 2005, 67(2): 301-320.

[3] Park C, Yoon Y J. Bridge regression: adaptivity and group selection[J]. Journal of Statistical Planning and Inference, 2011, 141(11): 3506-3519.

[4] Efron B, Hastie T, Johnstone I, et al. Least angle regression[J]. The Annals of statistics, 2004, 32(2): 407-499.

[5] Efron B, Hastie T, Johnstone I, et al Least Angle Regression, Forward Stagewise and the Lasso[DB/OL] https://web.stanford.edu/~hastie/TALKS/larstalk.pdf