

线性回归分析

MATHWYZ

目录

1. 引论	1
1.1. 什么是回归	1
1.2. 统计软件 R	1
2. 回归的方法	1
2.1. 线性回归	1
2.2. 线性与非线性	1
2.3. 逻辑回归	2
2.4. 多项式回归	2
2.5. 岭回归	2
2.6. 最小角回归,Lasso 回归	2
2.7. 弹性回归	2
3. 其他线性回国问题	2
3.1. 异方差	2
3.2. 自相关	2
3.3. 过拟合	3
3.4. 多重共线性	3

1. 引论

1.1. **什么是回归.** 回归最早是由英国生物学家高尔顿在研究人类遗传问题时提出来的。他发现一种有趣的现象，

1.2. 统计软件 R.

1.2.1. *R* 语言的发展历史.

1.2.2. *R* 语言的优势和缺点. *R* 语言与 SAS,SPSS 1. 首先 *R* 是 GNU 计划的一部分, 是开源软件, 是免费软件. 选择了开源软件, 就选择了一套体系, 它不仅仅是免费。2.

R 语言与 Python

R 语言与 Hadoop 家族

1.2.3. *R* 语言的网站. 目前,CRAN 上面有 15573 个包。

2. 回归的方法

2.1. 线性回归.

2.2. **线性与非线性.** 数学家是这么一类人, 他们把所有的问题都转化成线性代数的问题。这是因为, 现实生活中的很多问题可以用线性关系来替代。比如动力系统, 我们用线性部分来近似的替代整个动力系统, 如果初始值在微小的范围内, 解任然在微小的范围内, 我们就说是稳定的。这个时候就可以近似的用线性部分来替代非线性部分。同样的, 我们在统计中, 经常也遇到这样的问题。我们的变量之间是线性的, 更准确的说是近似线性的。那么我们就可以假设它是最简单的线性的相关关系。我们就可以用线性的方法来近似。另一方面, 虽然很多问题并不是线性的, 但是我们把变量经过变换之后, 他们之间就会出现线性的相关关系, 我们仍然可以用线性的来近似。

定理 2.1 (Gauss-Markov 定理). 在线性无偏估计类中, 最小二乘估计是唯一的具有最小方差的估计。

2.3. 逻辑回归.

2.4. 多项式回归.

2.5. 岭回归.

2.5.1. 复共线性.

2.5.2. *MSE*.

定理 2.2. 在 *MSE* 的意义下存在岭估计, 它的 *MSE* 比线性回归要好.

2.5.3. 岭估计.

2.5.4. 计算岭回归的方法.

- Hoerl-Kennard 公式
- 岭迹法

岭回归在共线性数据分析中应用较多,也称为脊回归,它是一种有偏估计的回归方法,是在最小二乘估计法的基础上做了改进,通过舍弃最小二乘法的无偏性,使回归系数更加稳定和稳健。其中 R^2 值会稍低于普通回归分析方法,但回归系数更加显著,主要用于变量间存在共线性和数据点较少时。

2.6. 最小角回归,Lasso 回归. LASSO 回归的特点与岭回归类似,在拟合模型的同时进行变量筛选和复杂度调整。变量筛选是逐渐把变量放入模型从而得到更好的自变量组合。复杂度调整是通过参数调整来控制模型的复杂度,例如减少自变量的数量等,从而避免过拟合。LASSO 回归也是擅长处理多重共线性或存在一定噪声和冗余的数据,可以支持连续型因变量、二元、多元离散变量的分析。

2.6.1. 理论基础.

2.6.2. 例子.

2.6.3. 相关的 R 的包.

2.7. 弹性回归.

3. 其他线性回归问题

3.1. 异方差.

3.1.1. 具有异方差的线性回归模型.

3.2. 自相关.

3.2.1. 具有自回归的线性回归模型.

3.3. 过拟合.

3.4. 多重共线性.

3.4.1. 岭回归.

3.4.2. 主成分.

3.4.3. *LASSO*.