LASSO

中国人民大学《复杂数据统计方法-基于R的应用》人大经济论坛

岭回归的问题：

岭参数计算方法太多，差异太大

根据岭迹图进行变量筛选，随意性太大

岭回归返回的模型（如果没有经过变量筛选）包含所有的变量

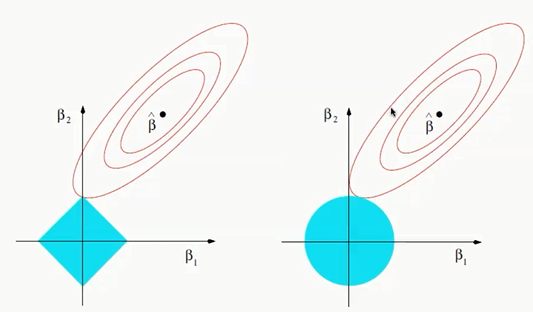
Lasso 1996由Tibshirani提出（The Least Absolute Shrinkage and Selectionator operator）算法

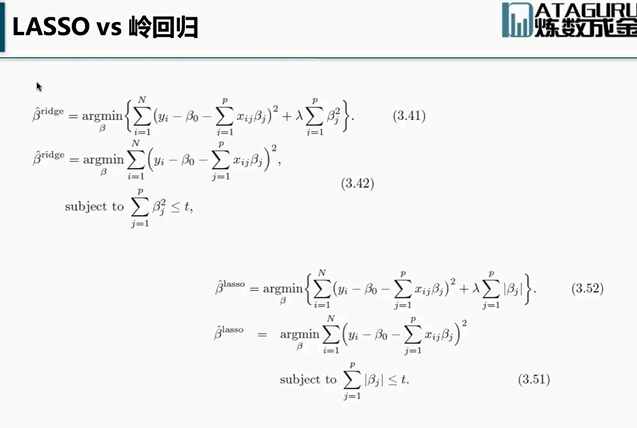
通过钩子哦一个一阶惩罚函数获得一个精炼的模型；通过最终确定一些指标（变量）的系数为零（岭回归估计系数等于0的机会微乎其微，造成筛选变量困难），解释力很强。

擅长处理具有多重共线性的数据，与岭回归一样是有偏估计。

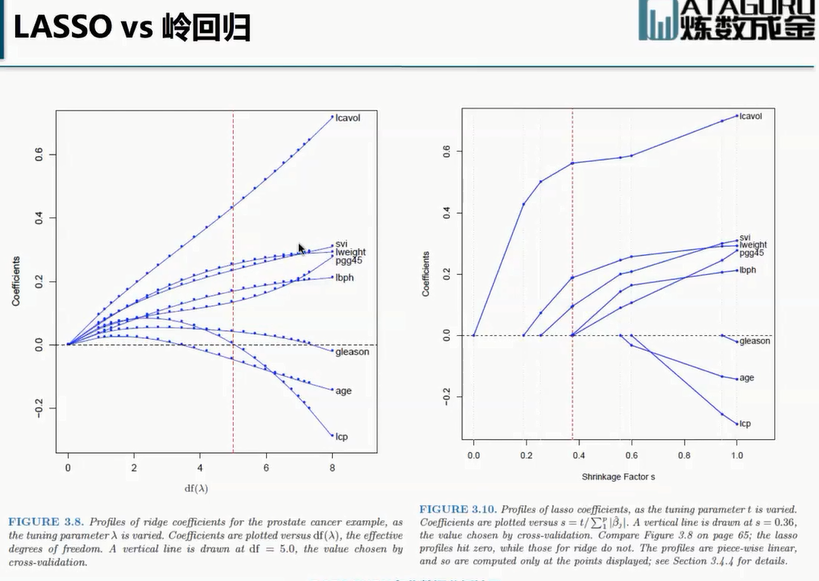
为什么Lasso能直接筛选变量？

右边的图形就是岭回归当中的。左边的图形是lasso的图形表示（菱形）、





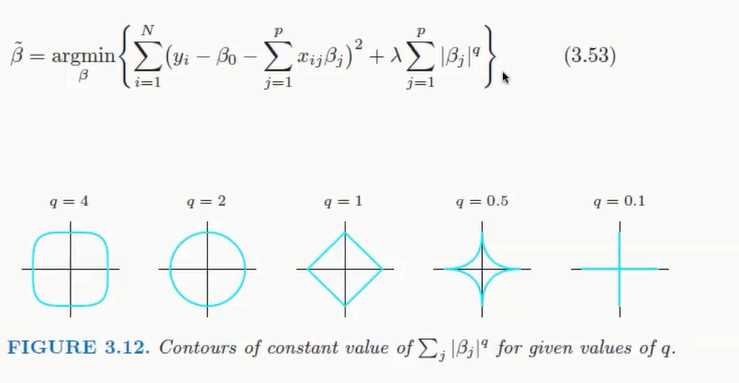
上图左上角岭回归，右下角lasso。



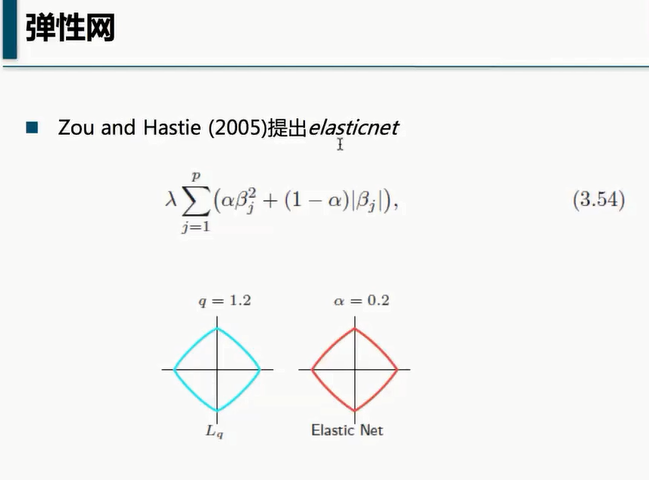
上图左边图形是岭回归的，横轴是自由度，和岭系数k有关。

右边图形描述lasso，从右到左是一个收缩的过程，可以想象成菱形再收缩。横轴是

========更一般化的模型==============



===========弹性网===========



既有岭回归的部分又有lasso的部分，画出来的形状是上图右侧。

筛选变量非常有效，而且损失的精度不会太多。

==========最小角回归====================

