WDS

DataScience@Web

机器学习讨论班

2018年暑期

WDS
DataScience@Web

其他回归方法

王紫悦

线性回归

给定数据集D={(x_1,y_1) , (x_2,y_2) ,… , (x_m,y_m)}，其中x_i=(x_{i1}; x_{i2};… ; x_{id})
线性回归试图学得：

$$\mathrm{y}_{\mathrm{i}} = \hat{w}x_i + b$$

最小二乘法

$$X=\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} & 1 \\ x_{21} & x_{22} & \dots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{md} & 1 \end{bmatrix} \quad W=\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \\ b \end{bmatrix} \quad Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

线性模型 $y_i = \hat{w}x_i + b \Longrightarrow Y = XW$

均方误差 $e = \sum_{i=1}^{m}\left(y_i - \hat{w}x_i - b\right)^2 \Longrightarrow E = \left\|Y - XW\right\|^2$

$$= \left(Y - XW\right)\left(Y - XW\right)^T$$

最小二乘法

欲求均方误差最小，对W求导，得： $\frac{\partial \mathrm{E}}{\partial \mathrm{W}} = 2\mathrm{X}^{T}(XW - Y)$

当X^TX为满秩矩阵或正定矩阵，令上式为0，得：

$$\mathrm{W} = (\boxed{X^T X})^{-1} X^T Y$$

可逆？

现实任务中X^TX往往不是满秩矩阵。比如，许多任务中有大量的变量，其数目超过样例数，导致X的列数大于行数。此时可解出多个W，导致结果不稳定。

引 入 正 则 项

正则化

思路：在原先的W的最小二乘估计中加入正则项 $\boxed{\lambda I}$，使问题得以解决。

$$\mathrm{W}(\lambda) = (X^T X + \boxed{\lambda I})^{-1} X^T Y$$

$$(\hat{w}, b) = \underset{\hat{w}, b}{\operatorname{argmin}} \left\{ \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{d} x_{ij} w_j - b \right)^2 + \boxed{\lambda \sum_{j=1}^{d} \left| w_j \right|^q} \right\}$$

约束项

介绍内容

- 岭回归
- Lasso回归
- 弹性网络

岭回归(Ridge Regression,RR)

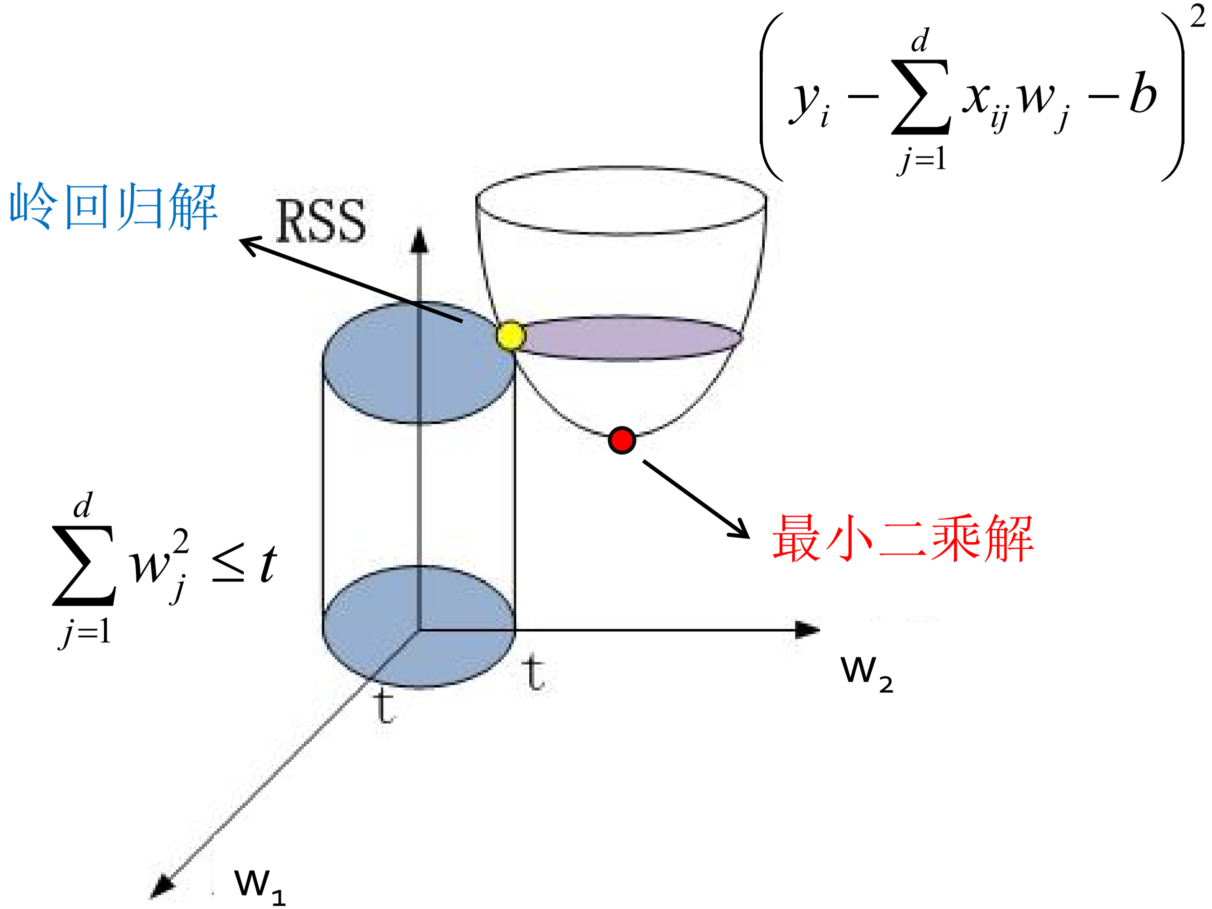
$$(\hat{w},b)=\underset{\hat{w},b}{\operatorname{argmin}}\left\{\sum_{i=1}^{m}\left(y_i-\sum_{j=1}^{d}x_{ij}w_j-b\right)^2+\boxed{\lambda}\sum_{j=1}^{d}w_j^2\right\}$$

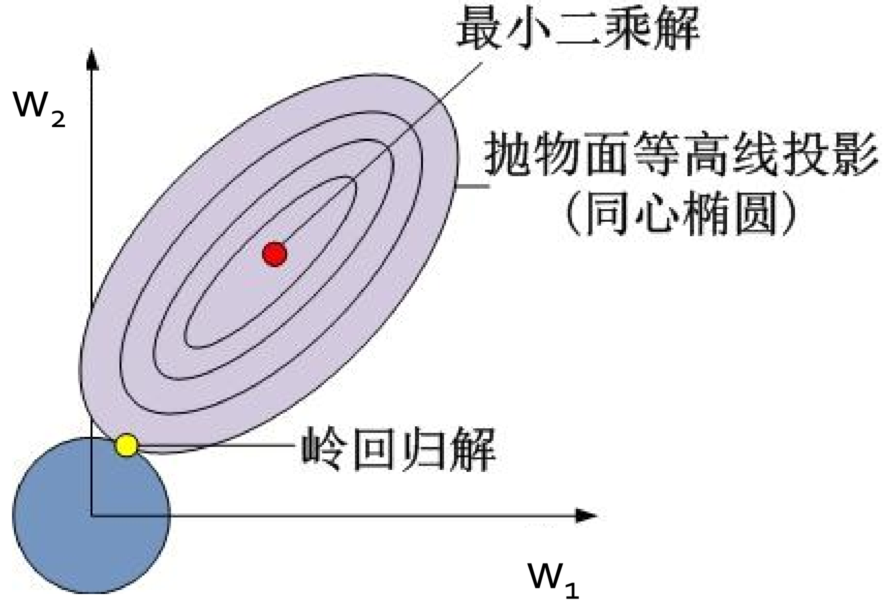
$$\Updownarrow$$

$$(\hat{w},b)=\underset{\hat{w},b}{\arg\min}\sum_{i=1}^{m}\left(y_i-\sum_{j=1}^{d}x_{ij}w_j-b\right)^2$$

$$subject\quad to\quad \sum_{j=1}^{d}w_j^2\leq\boxed{t}$$

RR几何意义


$\left(y_i - \sum_{j=1}^{d} x_{ij} w_j - b\right)^2$
岭回归解
RSS
最小二乘解
$\sum_{j=1}^{d} w_j^2 \le t$
t
t
w₂
w₁

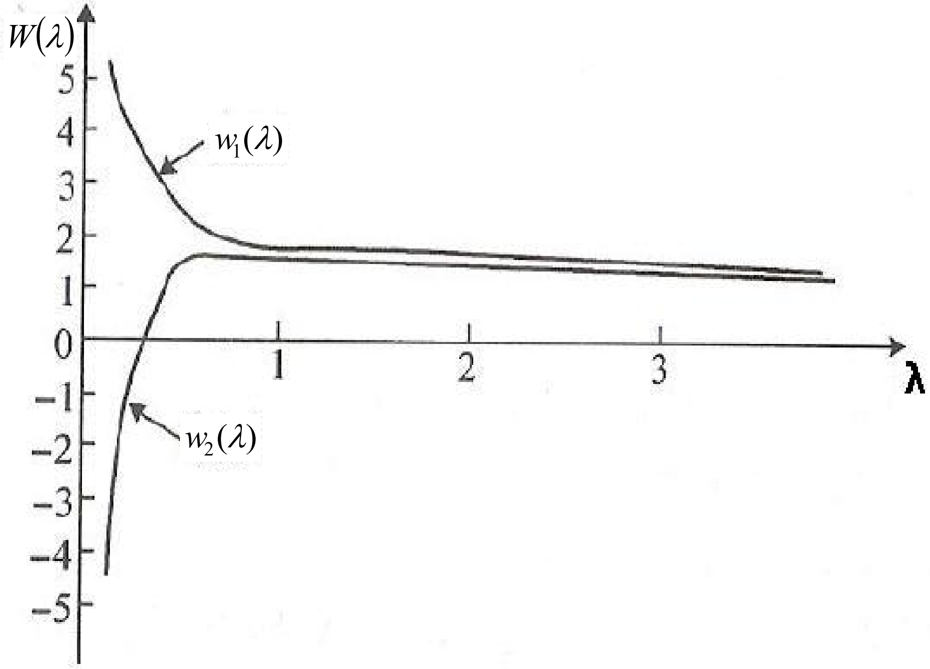

最小二乘解
抛物面等高线投影
(同心椭圆)
w₂
岭回归解
w₁

岭迹图

岭迹图作用：
1）观察λ较佳取值；
一般通过观察，选取喇叭口附近的值，此时各w值已趋于稳定，但总的RSS又不是很大。

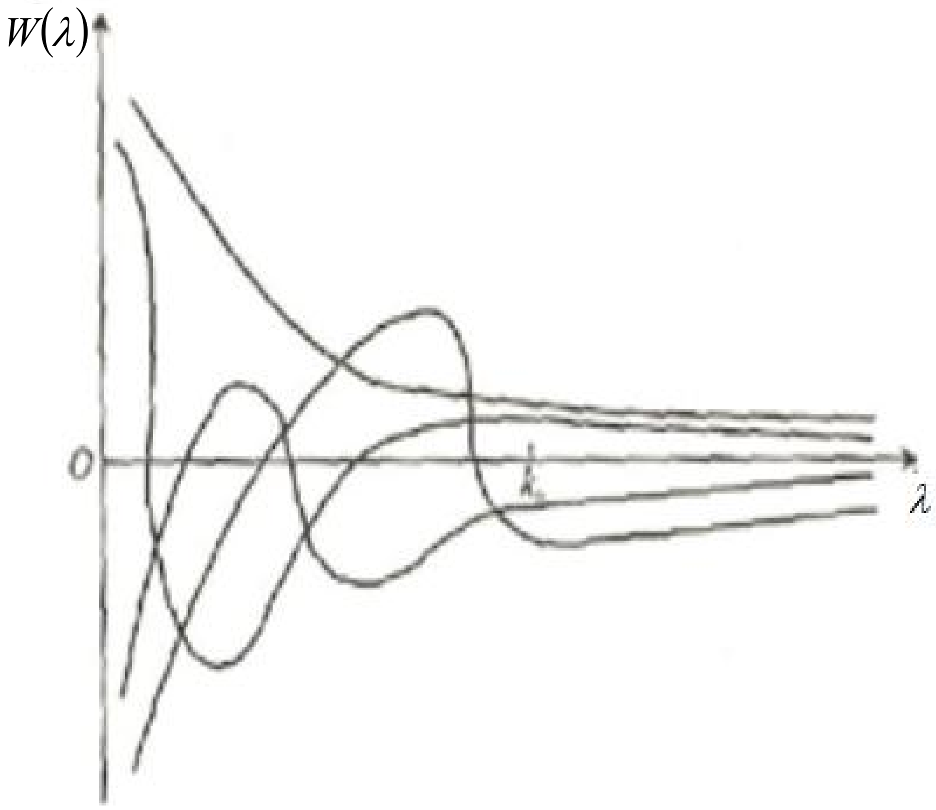
2）观察变量是否有多重共线性
岭迹波动大，说明该变量参数有共线性

3）选择变量
删除那些w取值一直趋于0的变量。



RR参数选择原则

1. 各回归系数的岭估计基本稳定
2. 用最小二乘估计时符号不合理的回归系数，其岭估计的符号变得合理
3. 回归系数没有不合乎实际意义的绝对值
4. 残差平方和增大不太多



RR的问题

- 岭参数计算方法太多，差异太大
- 若不经过变量筛选，岭回归返回的模型包含所有变量
- 根据岭迹图进行变量筛选，随意性太大

LASSO回归

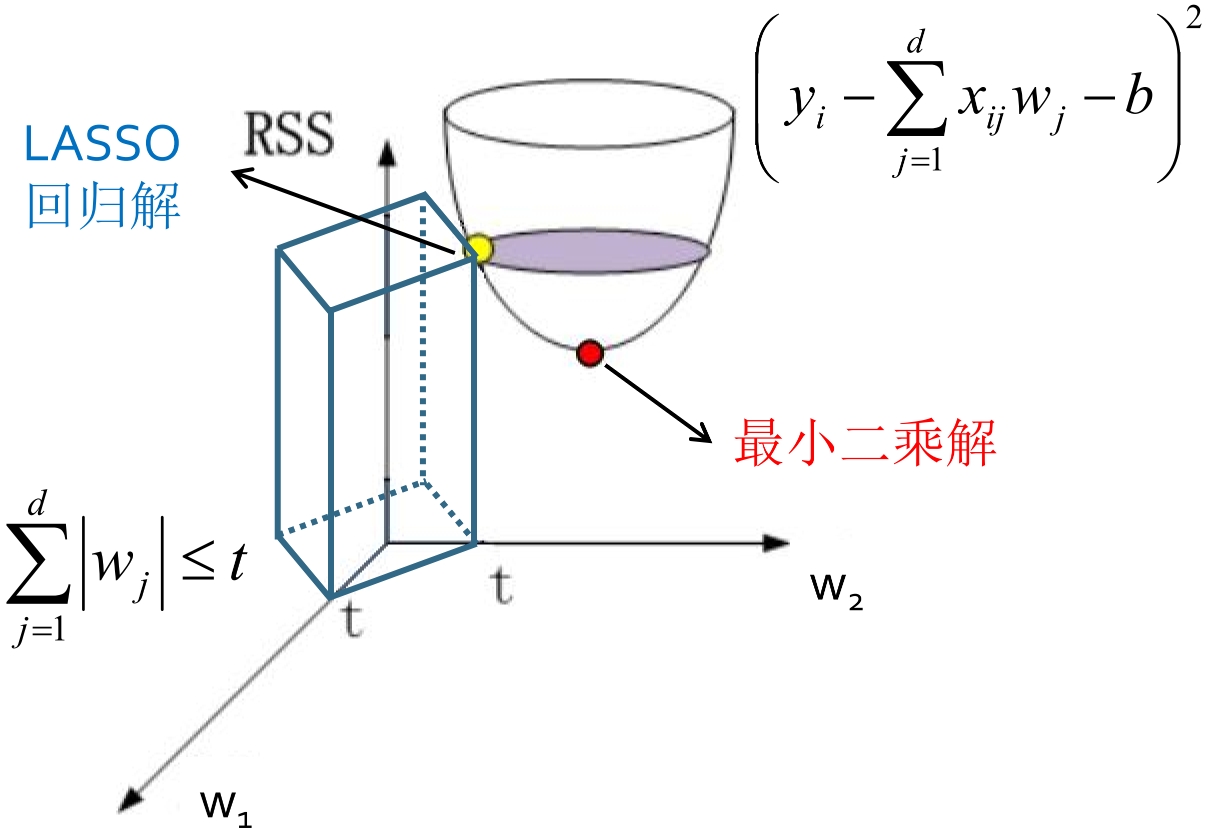
$$(\hat{w}, b) = \underset{\hat{w}, b}{\operatorname{argmin}} \left\{ \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{d} x_{ij} w_j - b \right)^2 + \boxed{\lambda \sum_{j=1}^{d} \left| \mathrm{w}_j \right|} \right\}$$

$$\Updownarrow$$

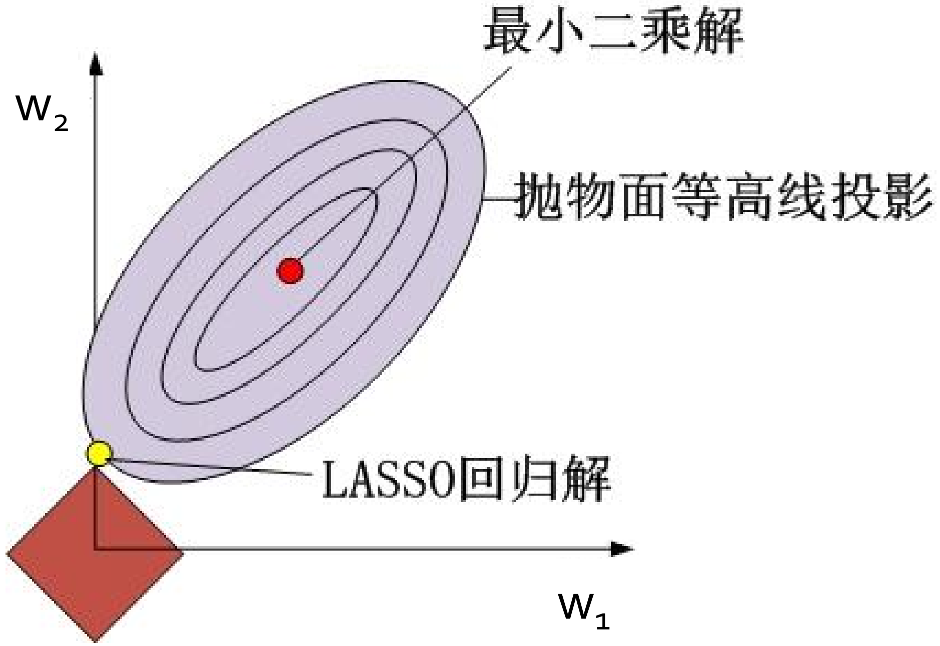
$$(\hat{w}, b) = \underset{\hat{w}, b}{\arg\min} \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{d} x_{ij} w_j - b \right)^2$$

$$subject \quad to \quad \boxed{\sum_{j=1}^{d} \left| w_j \right| \le t}$$

LASSO回归



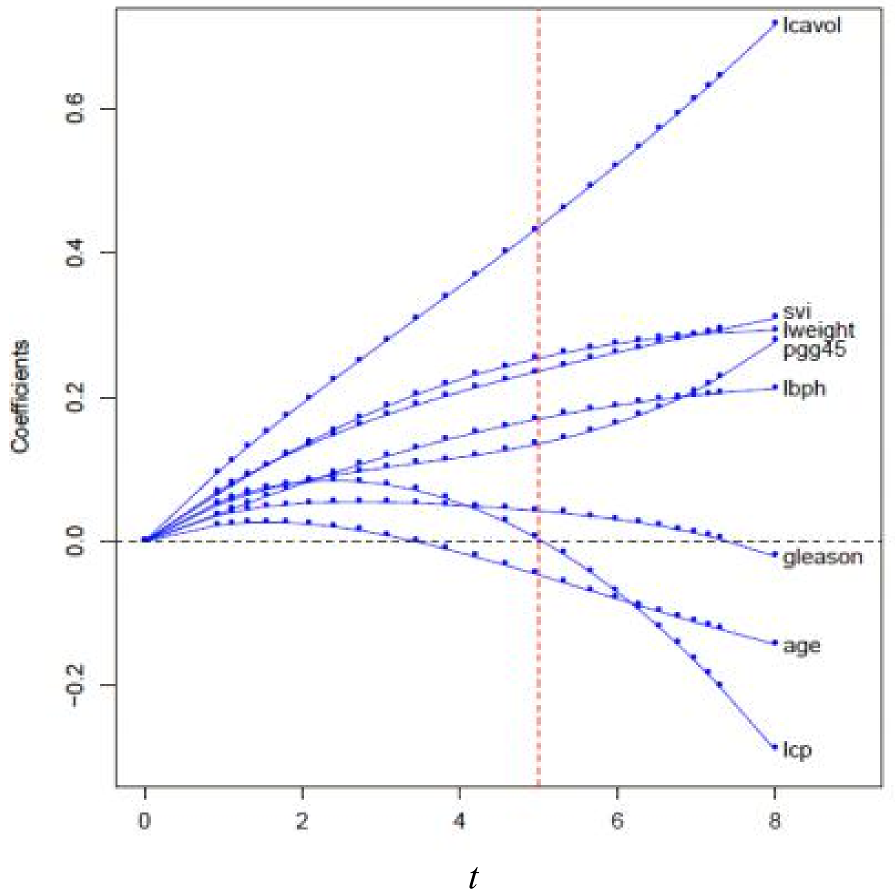
LASSO
回归解
RSS
$\left(y_i - \sum_{j=1}^{d} x_{ij} w_j - b\right)^2$
最小二乘解
$\sum_{j=1}^{d} |w_j| \le t$
t
t
w₂
w₁



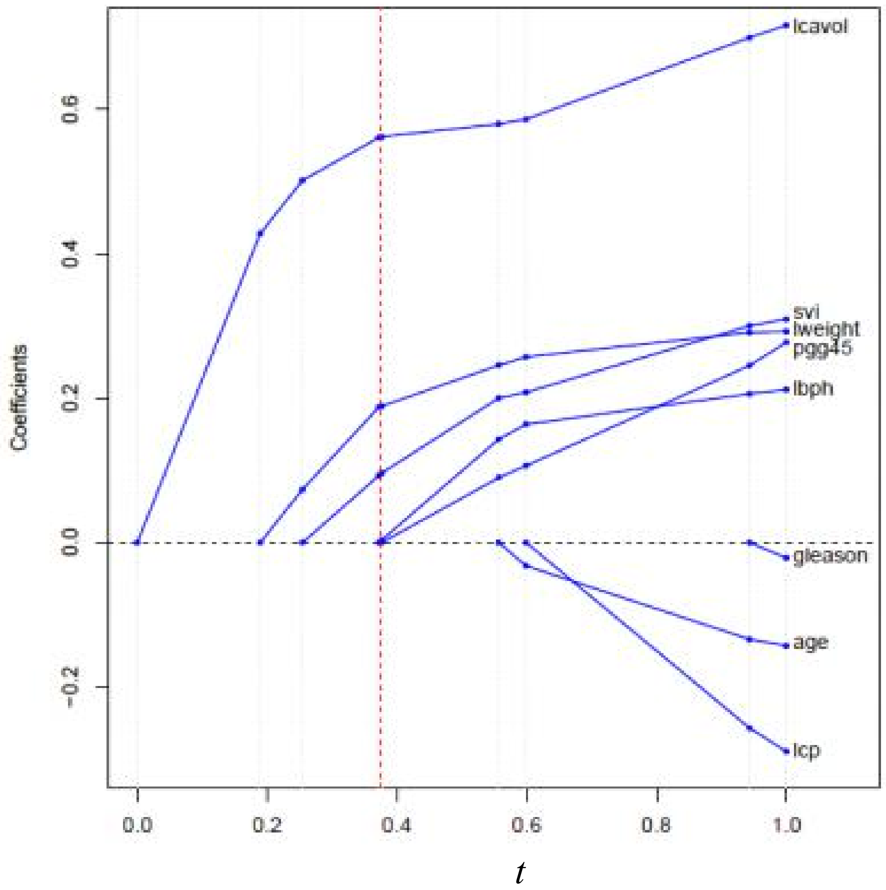
最小二乘解
w₂
抛物面等高线投影
LASSO回归解
w₁

LASSO VS RR

岭回归



LASSO



LASSO VS RR

LASSO回归	岭回归
L1范式 $\|w\|$	L2范式 w^2
使部分系数缩小为0	使所有系数收缩趋向于0
筛选变量	消除共线性

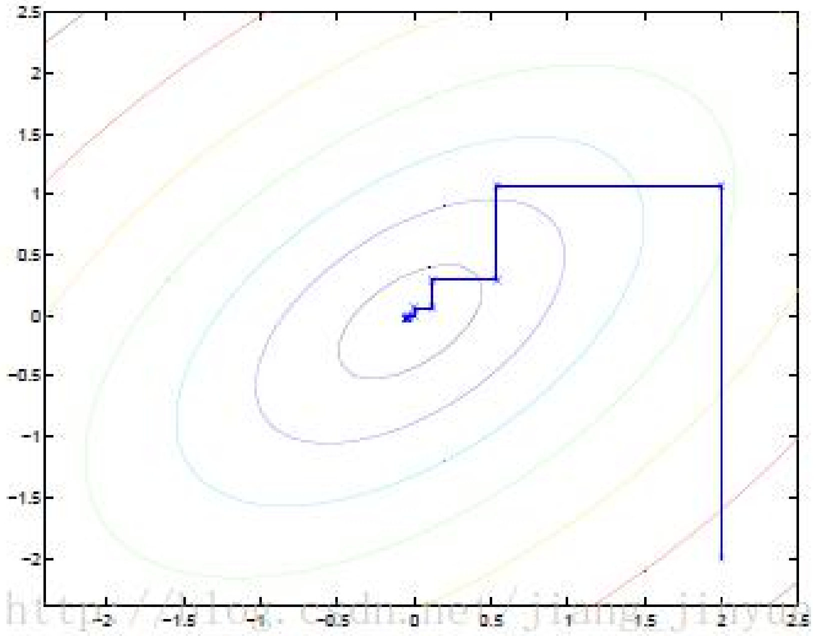
LASSO计算方法

- 坐标轴下降法（Coordinate descent）
- 最小角回归法（Least Angle Regression）

坐标轴下降法（Coordinate descent）

求解步骤：

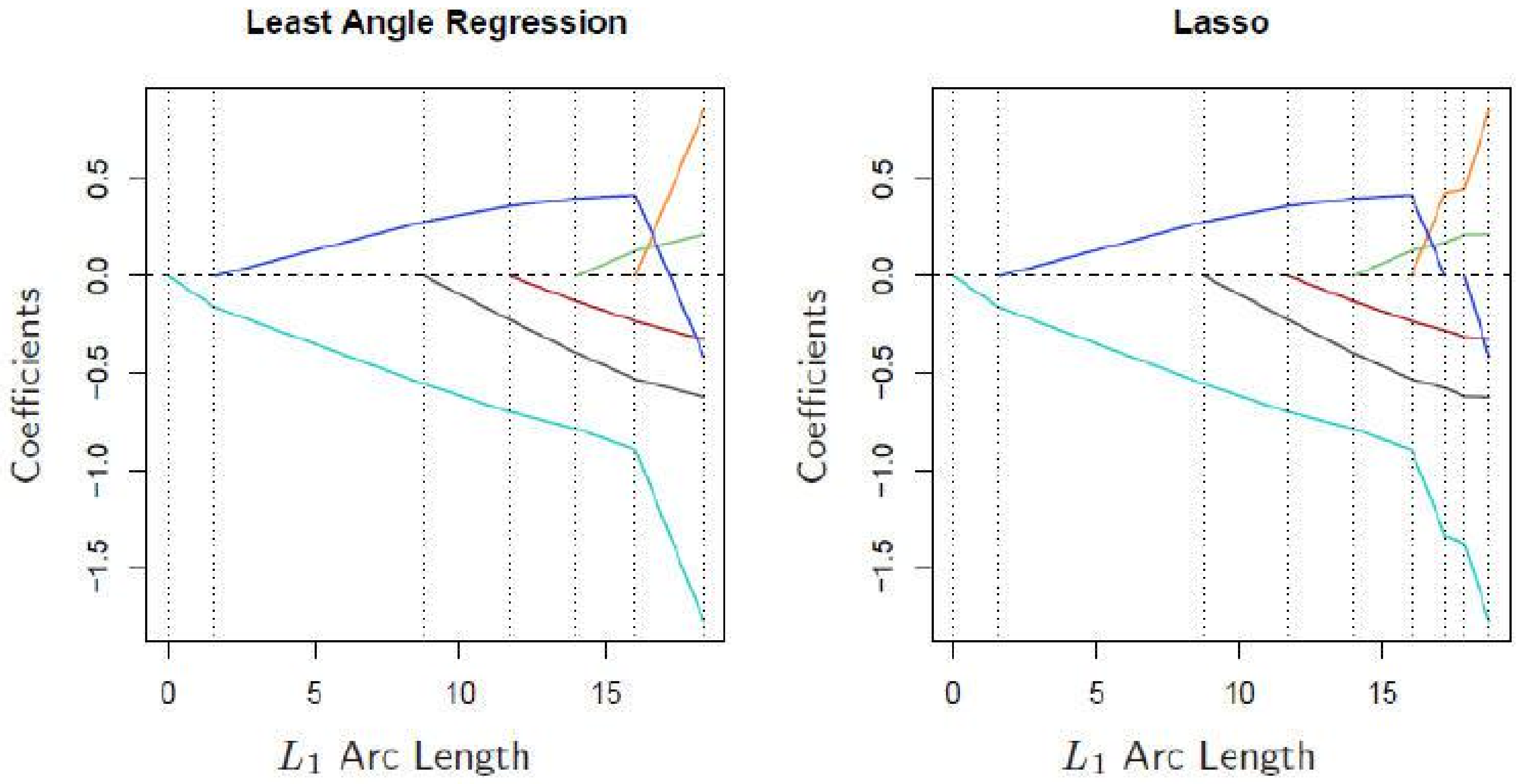
1. 给定初始点(x1, x2, ..., xn)
2. 固定除xi以外其他维度的点，以xi为自变量获取最小值
3. 换个维度，重复2



LASSO计算方法

- 坐标轴下降法（Coordinate descent）
- 最小角回归法（Least Angle Regression）

最小角回归LAR



Least Angle Regression
Lasso
Coefficients
0.5
0.0
−0.5
−1.0
−1.5
0
5
10
15
L1 Arc Length

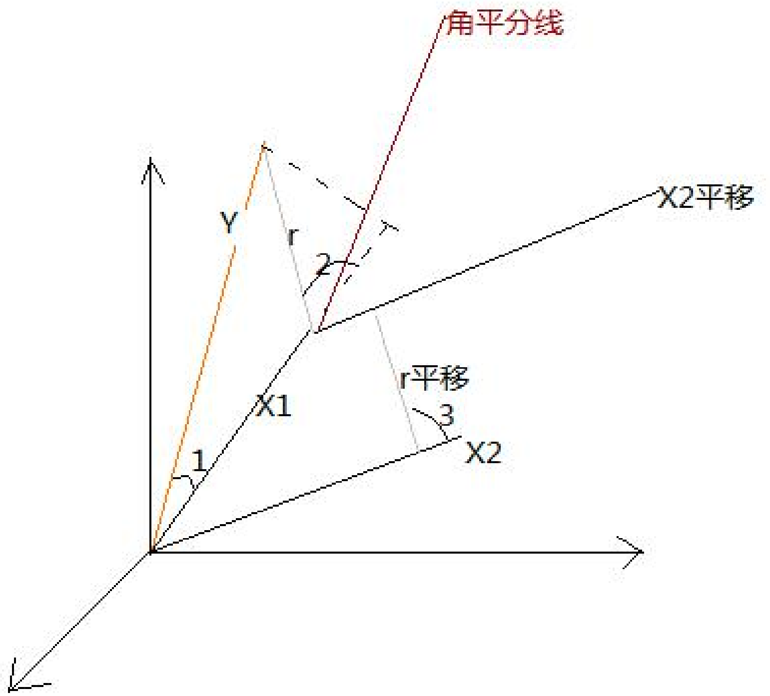
最小角回归LAR

思想：
LAR是每次先找出和因变量相关度最高的那个变量, 再沿着LSE的方向一点点调整其相关系数，在这个过程中，这个变量和残差的相关系数会逐渐减小，等到这个相关性没那么显著的时候，就要选进新的相关性最高的变量，然后重新沿着LSE的方向进行变动。而到最后，所有变量都被选中，就和LSE相同了。

最小角回归LAR

1）找到和Y向量夹角最小的向量X1，记最初夹角为角1。Y与X1的局部最小二乘解为Y到X1的距离。从原点出发，沿着X1，向这个局部最小二乘解移动。

2）在这个变化的过程中，总有某一时刻，另一个变量X2与r之间的相关系数，与X1和r之间的相关系数一样大。这个时候把X2加入，前进的方向修正为r与X2夹角的角平分线方向



最小角回归LAR

优点：

1）特别适合于特征维度n 远高于样本数m的情况

2）算法的最坏计算复杂度和最小二乘法类似，但是其计算速度几乎和前向选择算法一样

3）可以产生分段线性结果的完整路径，这在模型的交叉验证中极为有用

缺点：

由于LAR的方向是根据目标的残差而定，所以该算法对样本的噪声极为敏感。

LASSO回归

- 对数据要求低，应用程度广
- 收缩系数，筛选变量
- 降低模型复杂度

弹性网络（Elastic Net）

一种使用L1和L2先验作为正则化矩阵的线性回归模型

岭回归和lasso回归的结合

用于只有很少的权重非零的稀疏模型

$$(\hat{w}, b) = \underset{\hat{w}, b}{\operatorname{argmin}} \left\{ \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{d} x_{ij} w_j - b \right)^2 + \lambda \sum_{j=1}^{d} \left[\boxed{\alpha} w_j^2 + \boxed{(1-\alpha)} |w_j| \right] \right\}$$

弹性网络（Elastic Net）

岭回归	一定程度上可以拟合模型，但回归结果易失真
LASSO回归	刻画模型代表的现实情况，但模型过于简单，不符实际
弹性网络	达到对重要特征选择的目的，删除对因变量影响较小的特征

课后练习

sklearn包，利用交叉验证法确定参数、带cv校准过程的回归模型

岭回归 ：linear_model.RidgeCV

LASSO ：LassoCV and LassoLarsCV

弹性网络：ElasticNetCV

谢 谢