现代统计方法测试试题

1. 由于它在导数等于0的点有最优解，

求导并令导数等于0就有

-（XTy-XTXw）+λ\*sign（w）=0

由于λ>0,所以

S=w-sign（w）

所以s=sign（w）（w-λ），

其中，w=XTy

2、线性模型中的数据中，x这个向量里面有很多数都是0，而非零的向量远小于x的维度（x向量的长度）。LASSO是一种压缩的方法，他是很多变量的取值为0，所以产生了稀疏性.

3、交叉验证的核心思想, 就是在模型拟合的过程中并不适用全部的历史数据,而是保留了一部分数据,用来模拟未来的数据,对模型进行检验.

首先选用不同的λ对模型进行拟合，然后选另一部分数据，对λ的格点值进行交叉验证，选用误差最小的lambda，最后用得到的λ重新拟合数据。

4、广义线性模型

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 108.36259 7.58761 14.282 < 2e-16 \*\*\*

tobacco 0.08524 0.21843 0.390 0.6966

ldl -0.08336 0.47872 -0.174 0.8618

adiposity 0.31860 0.21118 1.509 0.1321

famhist -1.20533 1.86615 -0.646 0.5187

typea -0.08215 0.09120 -0.901 0.3682

obesity 0.35129 0.31246 1.124 0.2615

alcohol 0.07889 0.03658 2.157 0.0315 \*

age 0.35994 0.08770 4.104 4.81e-05 \*\*\*

chd 2.63619 2.08357 1.265 0.2064

一般线性模型

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 108.36259 7.58761 14.282 < 2e-16 \*\*\*

tobacco 0.08524 0.21843 0.390 0.6966

ldl -0.08336 0.47872 -0.174 0.8618

adiposity 0.31860 0.21118 1.509 0.1321

famhist -1.20533 1.86615 -0.646 0.5187

typea -0.08215 0.09120 -0.901 0.3682

obesity 0.35129 0.31246 1.124 0.2615

alcohol 0.07889 0.03658 2.157 0.0315 \*

age 0.35994 0.08770 4.104 4.81e-05 \*\*\*

chd 2.63619 2.08357 1.265 0.2064

逐步回归

Df Sum of Sq RSS AIC

<none> 158589 2705.4

+ chd 1 383.7 158205 2706.3

+ obesity 1 300.1 158289 2706.5

+ typea 1 133.8 158455 2707.0

+ tobacco 1 116.4 158473 2707.1

+ famhist 1 71.4 158518 2707.2

+ ldl 1 0.2 158589 2707.4

- alcohol 1 1728.9 160318 2708.4

- adiposity 1 3818.2 162407 2714.4

- age 1 8333.9 166923 2727.1

LASSO

Length Class Mode

a0 66 -none- numeric

beta 594 dgCMatrix S4

df 66 -none- numeric

dim 2 -none- numeric

lambda 66 -none- numeric

dev.ratio 66 -none- numeric

nulldev 1 -none- numeric

npasses 1 -none- numeric

jerr 1 -none- numeric

offset 1 -none- logical

call 3 -none- call

nobs 1 -none- numeric