HW3 5241 CM3700

Q_1: $n=2,\ p=2,\ x_{11}=x_{12},\ x_{21}=x_{22},\ y_1+y_2=0,\ x_{11}+x_{21}=0,\ x_{12}+x_{22}=0,\ \hat{\beta}_0=0$

a) Minimize $(y_1-\hat{\beta}_1x_{11}-\hat{\beta}_2x_{12})^2+(y_2-\hat{\beta}_1x_{21}-\hat{\beta}_2x_{22})^2+\lambda(\hat{\beta}_1^2+\hat{\beta}_2^2)$

b) $\frac{df}{d\beta_1}=0=\beta_1(x_1^2+x_2^2+\lambda)+\beta_2(x_1^2+x_2^2)-y_1x_1-y_2x_2$

$\frac{df}{d\beta_2}=0=\beta_1(x_1^2+x_2^2)+\beta_2(x_1^2+x_2^2+\lambda)-y_1x_1-y_2x_2$

$\frac{df}{d\beta_1}=\frac{df}{d\beta_2}=0 \Rightarrow \beta_1(x_1^2+x_2^2+\lambda)+\beta_2(x_1^2+x_2^2)=\beta_1(x_1^2+x_2^2)+\beta_2(x_1^2+x_2^2+\lambda)$

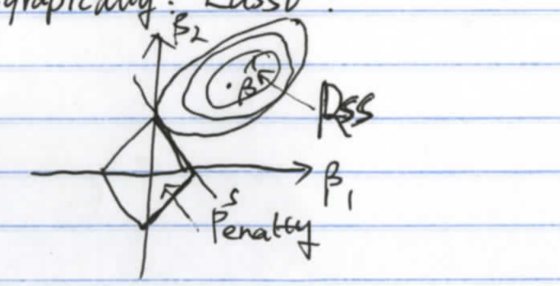
$\beta_1\lambda=\beta_2\lambda$

$\beta_1=\beta_2$

c) Minize $(y_1-\hat{\beta}_1x_{11}-\hat{\beta}_2x_{12})^2+(y_2-\hat{\beta}_1x_{21}-\hat{\beta}_2x_{22})^2+\lambda(|\hat{\beta}_1|+|\hat{\beta}_2|)$

d) For Lasso: $\hat{\beta}_1+\hat{\beta}_2<s$

Geograpically. Lasso:



For this problem: Least squared coefficients $=(y_1-\hat{\beta}_1x_{11}-\hat{\beta}_2x_{12})^2+(y_2-\hat{\beta}_1x_{21}-\hat{\beta}_2x_{22})^2$

minimize this $\Rightarrow \min[2(y_1-(\hat{\beta}_1+\hat{\beta}_2)x_{11}]$

This is achieved at $\hat{\beta}_1+\hat{\beta}_2=y_1x_{11}$

Therefore, just like my graph shows, the solution for this optimization is

$\hat{\beta}_1+\hat{\beta}_2=s$ where $\hat{\beta}_1,\hat{\beta}_2>0$

The solution is not unique since any data set matches

$\hat{\beta}_1+\hat{\beta}_2=s$ and $\hat{\beta}_1,\hat{\beta}_2>0$ is a good answer.

Q_2:

a) $\hat{g}_2$ will have a smaller training RSS

Because $\hat{g}_2$'s penalty term has a higher order derivative than $\hat{g}_1$'s

b) $\hat{g}_1$ will have a smaller test RSS

Because $\hat{g}_2$ is possible to have overfitting problem due to the extra degree of freedom.

c) if $\lambda = 0$

$\hat{g}_1 = \hat{g}_2$

if $\lambda \to 0$ and $\lambda \neq 0$

Since $\hat{g}_2$ has a higher order than $\hat{g}_1$

$\hat{g}_2$ will have a smaller error.