

HW3 S24 CM3700

Q1: $n=2, p=2, x_{11}=x_{12}, x_{21}=x_{22}, y_1+y_2=0, x_{11}+x_{21}=0, x_{12}+x_{22}=0, \hat{\beta}_0=0$

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

b) $\frac{df}{d\hat{\beta}_1} = 0 = \hat{\beta}_1(x_{11}^2 + x_{12}^2 + \lambda) + \hat{\beta}_2(x_{11}^2 + x_{12}^2) - y_1 x_{11} - y_2 x_{12}$

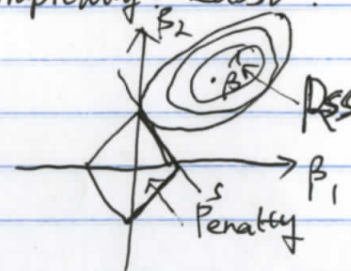
$\frac{df}{d\hat{\beta}_2} = 0 = \hat{\beta}_1(x_{11}^2 + x_{12}^2) + \hat{\beta}_2(x_{11}^2 + x_{12}^2 + \lambda) - y_1 x_{11} - y_2 x_{12}$

$\frac{df}{d\hat{\beta}_1} = \frac{df}{d\hat{\beta}_2} = 0 \Rightarrow \hat{\beta}_1(x_{11}^2 + x_{12}^2 + \lambda) + \hat{\beta}_2(x_{11}^2 + x_{12}^2) = \hat{\beta}_1(x_{11}^2 + x_{12}^2) + \hat{\beta}_2(x_{11}^2 + x_{12}^2 + \lambda)$
 $\hat{\beta}_1 \lambda = \hat{\beta}_2 \lambda$
 $\hat{\beta}_1 = \hat{\beta}_2$

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

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

Geographically: Lasso:



For this problem: Least squared coefficients $= (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2$
 minimize this $\Rightarrow \min [2(y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_{11})^2]$

This is achieved at $\hat{\beta}_1 + \hat{\beta}_2 = y_1/x_{11}$

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

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

The solution is not unique since any data set matches $\hat{\beta}_1 + \hat{\beta}_2 = s$ and $\hat{\beta}_1, \hat{\beta}_2 \geq 0$ is a good answer.

Q2: \hat{g}_1 will have a smaller ^{training} RSS

a) \hat{g}_2 will have a smaller 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

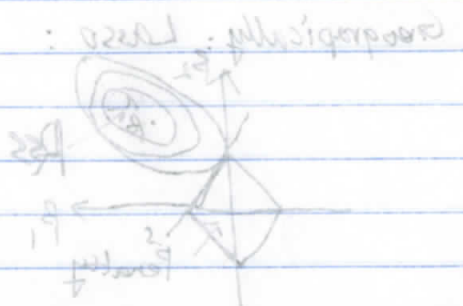
Because \hat{g}_2 is possible to have overfitting problem due to the extra degree of freedom.

c) if $\lambda = 0$
 $\hat{g}_1 = \hat{g}_2$

if $\lambda \rightarrow 0$ and $\lambda \neq 0$

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

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



For this problem, least-squares regression coefficients are $\hat{\beta}_1 = 1$ and $\hat{\beta}_2 = 1$.
Minimize this $\Rightarrow \min_{\beta_1, \beta_2} \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_2 x_i^2)^2$
This is achieved at $\hat{\beta}_1 = 1$ and $\hat{\beta}_2 = 1$.
Therefore, just like any other regression problem, the solution for this optimization is $\hat{\beta}_1 = 1$ and $\hat{\beta}_2 = 1$.
The solution is not unique since some data set matrices are not full rank.