Ridge Regressn & Lasso

Recall,

RIDGE

$$\text{OLS}: \min_{\beta_0, \beta_1} \|\vec{y} - A\vec{\beta}\|_2^2$$

Accordg to Gauss-Markov Thm: OLS is B.L.U.E

Best Linear Unbiased Estimator
" least variance " $E(\hat{\beta}) = \beta$

OLS may not be appropriate b/c perhaps even the best estimator of the lin. unbiased estimators have high variance. In that case, we are willg to sacrifice unbiased for ~~lower~~ smaller variance.

eg. S = money spent on food, m = incm, t = taxs paid

m & t are highly correlatd $\Rightarrow$

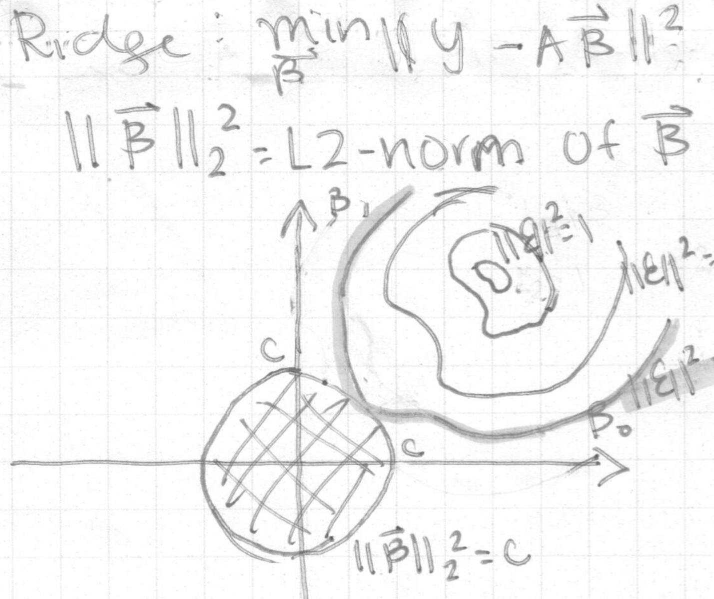
$\hat{\beta}_0$ and $\hat{\beta}_1$ have high variance.

If we took a sample & calculated $\hat{\beta}_0$ & $\hat{\beta}_1$ and then repeatd, the distn of $\hat{\beta}_0$ & $\hat{\beta}_1$ ~~for~~ for each sample would have a large spread

Ridge Regressn (1) introduces bias but (2) minimizs variance

$$\text{Ridge}: \min_{\vec{\beta}} \|y - A\vec{\beta}\|_2^2 \text{ st } \|\vec{\beta}\|_2^2 \le c^2$$

$$\|\vec{\beta}\|_2^2 = \text{L2-norm of } \vec{\beta} = \sqrt{\beta_0^2 + \beta_1^2}$$



$$\min_{\vec{\beta}} \|\vec{y} - A\vec{\beta}\|_2^2 \text{ st } \|\vec{\beta}\|_2^2 \le c^2$$

NB: Purpose of Ridge Regressn for when feats are highly correlated if you only have β_0 and β_1 (two betas) $\Leftrightarrow$ one feat, you DO NOT use Ridge Regressn b/c $\exists$ only one feat.

$$\min_{\vec\beta} \sum (y_i - \beta_0 - \beta_1 x_i)^2 \text{ s.t. } \left[\|\vec\beta\|_2^2 \le c^2 \iff \beta_0^2 + \beta_1^2 \le c^2 \right]$$

Recall, one way to minimize or maximize (ie optimize) a fn subj to a constraint ie ($\min f(x,y,z)$ st. $g(x,y,z)$) is the Method of Lagrange Multipliers

Let $g = \beta_0^2 + \beta_1^2 = c^2$
Let $f = \sum(y_i - \beta_0 - \beta_1 x_i)^2$

(i) $f_{\beta_0} = \lambda g_{\beta_0}$ (iii) $g = c^2$

(ii) $f_{\beta_1} = \lambda g_{\beta_0}$

Notice that the soln to the abv sys of eqns (i) and (ii) is also the soln the

(iii) $\min_{\beta_0,\beta_1} \sum (y_i - \beta_0 - \beta_1 x_i)^2 + \lambda(\beta_0^2 + \beta_1^2)$

Soln to (iii) is soln to

(i) $0 = f_{\beta_0} + \lambda g_{\beta_0}$

(ii) $0 = f_{\beta_1} + \lambda g_{\beta_1}$ (iii) $g = c^2$

Method of Lagrange Multipliers

To optimize (find min & max) of $f(x,y,z)$ subj. to constraint $g(x,y,z) = k$

step 1) solve sys. of eqns:

1. $\nabla f = \lambda \nabla g$
2. $g = k$; λ is any #

step 2) plug in solns (x,y,z) to find min or max.

not atn

$\nabla f = \langle f_x, f_y, f_z \rangle$, $f_x = \frac{\partial f}{\partial x}$; "deriv w/ respect to x"

$$\Rightarrow \boxed{\text{Ridge: } \min_{\beta} \|\vec y - A\vec\beta\|_2^2 + \lambda \|\beta\|_2^2}$$

$$\iff \min_{\beta} \|\vec y - A\vec\beta\|_2^2 \text{ s.t. } \|\vec\beta\|_2^2 \le c^2$$

Back to solvg (i), (ii), and (iii) for $\vec\beta$

⋮

- Use calculus -

⋮

nb as $\uparrow \lambda$, $\Downarrow \vec\beta$

$$\boxed{\vec\beta^R = (A^T A + \lambda I)^{-1} A^T \vec y}$$

Lasso on next page

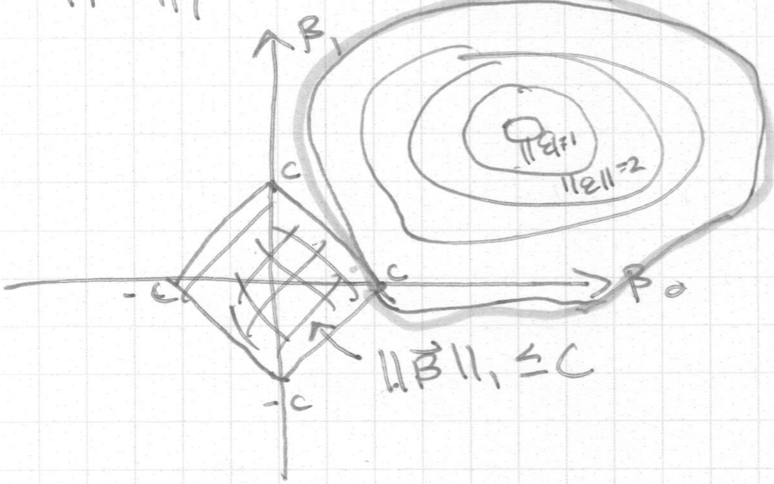
Lasso: $\min_{\vec{\beta}} ||\vec{y} - A\vec{\beta}||_2^2$ s.t. $\lambda||\vec{\beta}||_1$ ← L1-norm

LASSO

$\Updownarrow$

$$\boxed{\text{Lasso}: \min_{\vec{\beta}} ||\vec{y} - A\vec{\beta}||_2^2 + \lambda||\vec{\beta}||_1}$$

Recall $||\vec{\beta}||_1$ = L_1-norm = $\Sigma \beta_i$



NB: Lasso is good for eliminatg features, ie feature selxn, b/c $||\varepsilon||^2$ graphd obj is likely to hit a corner, ie where a β is is $\emptyset$.