LM. Eq. 1

Sunday, May 17, 2020 6:05 PM

$$y = \beta_0 + \beta_1 x$$

x	y	$\hat{y}$	e	e^2
.	.	$\hat{\beta}_0 + \hat{\beta}_1 x_1$	$y_1 - \hat{y}_1$	$(y_1 - \hat{y}_1)^2$
.	.	$\hat{\beta}_0 + \hat{\beta}_1 x_2$	$y_2 - \hat{y}_2$	$(y_2 - \hat{y}_2)^2$
⋮	⋮	⋮	⋮	⋮
p	.			
				e^2

$\hat{\beta}_0 , \hat{\beta}_1$:

$$\hat{y} = \hat{\beta}_0 + x\hat{\beta}_1$$

$$e = y - \hat{y}$$

$$e^2 = (y - \hat{y})^2$$

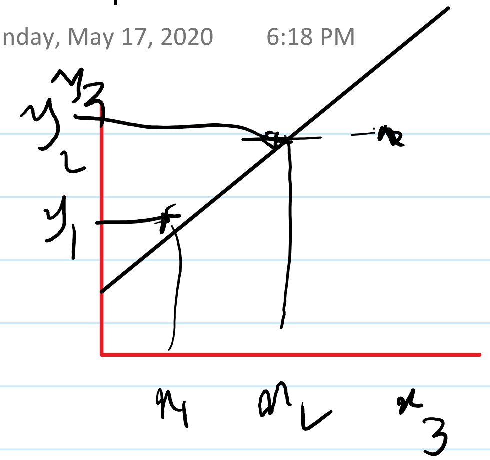
$$RSS \leftarrow S = \sum_{i=1}^{p} e_i^2 = \sum_{i=1}^{p} (y_i - \hat{y}_i)$$

OLS : Ordinary least squares

$$\frac{\partial S}{\partial \beta_0} = 0 , \quad \frac{\partial S}{\partial \beta_1} = 0$$

LM. Eq. 2

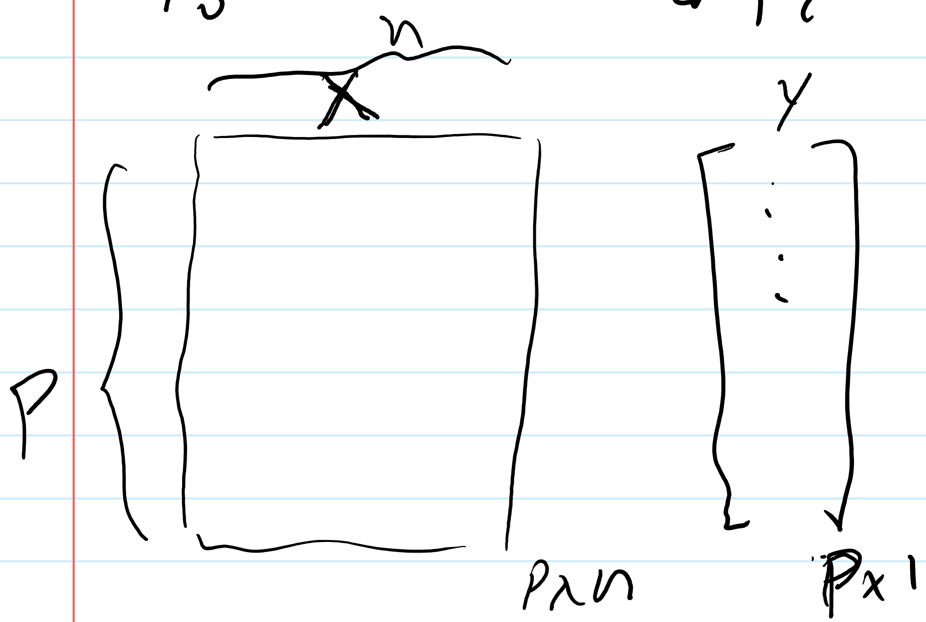
Sunday, May 17, 2020 6:18 PM



n	y	$\hat{y}$	e
1	2	$\beta_0+\beta_1$	$\beta_0+\beta_1-2$
2	4	$\beta_0+2\beta_1$	$\beta_0+2\beta_1-4$
3	4	$\beta_0+3\beta_1$	$\beta_0+3\beta_1-4$

$$S = (\hat{\beta}_0+\hat{\beta}_1-2)^2 + (\hat{\beta}_0+2\hat{\beta}_1-4)^2 + (\hat{\beta}_0+3\hat{\beta}_1-4)^2$$

$$\frac{\partial S}{\partial \hat{\beta}_0} = 0 \qquad \frac{\partial S}{\partial \hat{\beta}_1} = 0$$



X: Features, variables, predictors, covariates

y: target, Response

$\boxed{p > n}$ ✓

$\boxed{p < n \quad \cancel{X^{-1}}}$

LM. Eq. 3

Sunday, May 17, 2020 6:26 PM

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n$$

$$S = \sum_{i=1}^{p} e^2 = \sum_{i=1}^{p} (y_i - \hat{y}_i)^2$$

$$e = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ \hat{y}_p - \hat{y}_p \end{bmatrix} \qquad S = e^T \cdot e$$

$$S = \begin{bmatrix} e_1 & e_2 & \cdots & e_p \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}$$

$$X = \underset{p \times n}{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & & \\ 1 & x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}} \underset{n \times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}} = \underset{p \times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}}$$

$$X \cdot \underline{\hat{\beta}} = \underline{y}$$

LM. Eq. 4

Sunday, May 17, 2020 6:33 PM

$$S = e^T e = (y - \hat{y})^T (y - \hat{y})$$

$$= (y - X\hat{\beta})^T (y - X\hat{\beta})$$

$A^2 = A^T A$

$(AB)^T = B^T A^T$

$A^{-1} A = I = 1$

$$= (y^T - \hat{\beta}^T X^T)(y - X\hat{\beta})$$

$$= y^T y - y^T X\hat{\beta} - \hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta}$$

$$\frac{\partial S}{\partial \beta} = -2X^T y + 2X^T X\hat{\beta} = 0$$

$$X^T X\hat{\beta} = X^T y$$

$$(X^T X)^{-1} X^T X \hat{\beta} = (X^T X)^{-1} X^T y$$

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$$

LM. Eq. 5

Sunday, May 17, 2020 6:39 PM

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & \vdots & & & \\ \vdots & \vdots & \ddots & & \\ \vdots & & & \ddots & \\ 1 & & & & x_{pn} \end{bmatrix}$$

numpy

$$(y^T x)^T = x^T y$$

$$\hat{\beta} = (x^T x)^{-1} x^T y$$

Closed form Solution

LM. Eq. 5

Sunday, May 17, 2020 6:44 PM

closed for solution:

$$ax^2 + bx + c = 0$$

$$x_1, x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Numerical Solution

$\beta_0^0, \beta_1^0, \beta_2^0 \dots \beta_n^0 : e_0$

$e_0^2 = \checkmark$

$e_1 < e_0$

$\beta_0^1, \beta_1^1 \dots \beta_n^1 \quad e_1$

$e_2 < e_1$

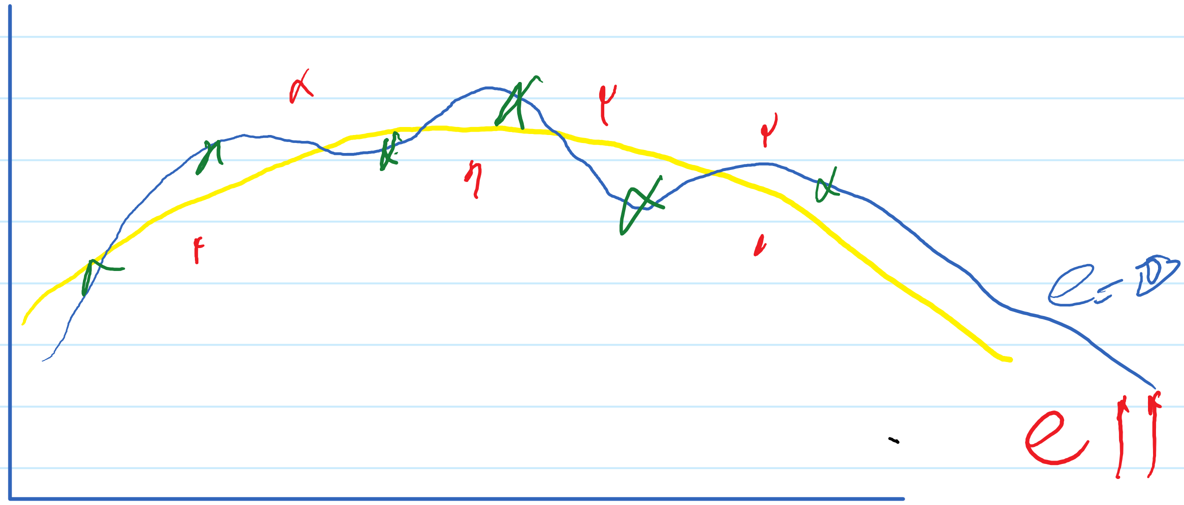
$\beta_0^2, \beta_1^2, \dots \beta_n^2 \quad e_2$

$\vdots$

stopping criteria $\begin{cases} \text{Max iter} \\ S < tol \end{cases}$

LM. Lp Regularization

Sunday, May 17, 2020 6:49 PM



overfit

$e > e$

$e < e$

CV

$$S = \sum_{i=1}^{p} (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{n} |\beta_j|^p$$

$0 \leq \alpha < \infty$

$\alpha = 0 : S \rightarrow lm$

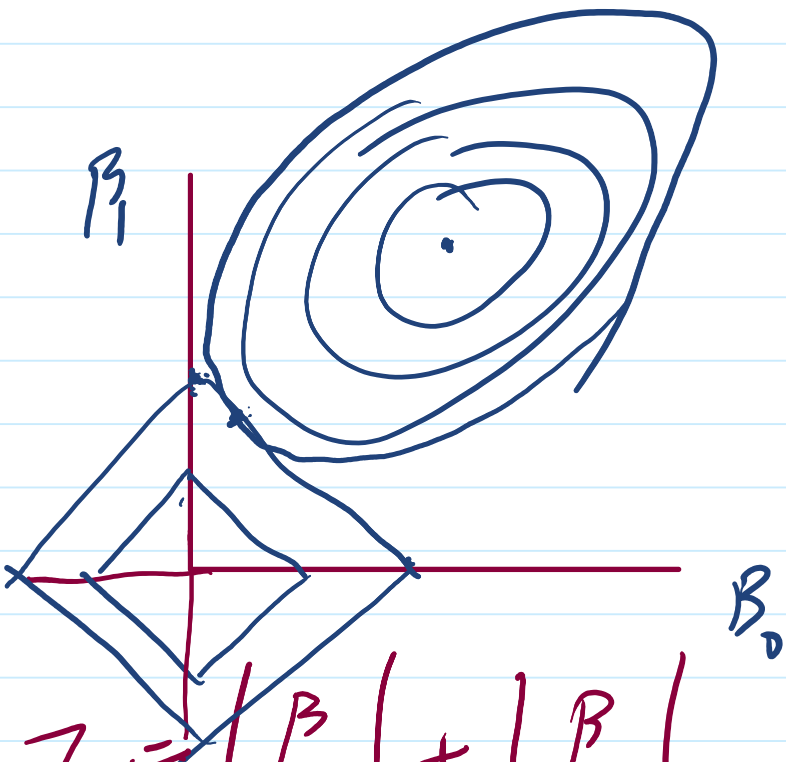
$\alpha > 0 : e^p$

LASSO / L1

Sunday, May 17, 2020 6:55 PM

$l_1 \Rightarrow \cancel{p} = 1$

$$S = \sum_{i=1}^{p} (y_i - \hat{y}_i)^2 + \alpha \sum_{j=0}^{n} |\beta_j|$$



$Z = |\beta_0| + |\beta_1|$

$l_q \rightarrow \beta_0 \rightarrow 0$

Ridge / L2

Sunday, May 17, 2020 6:59 PM

ℓ_2

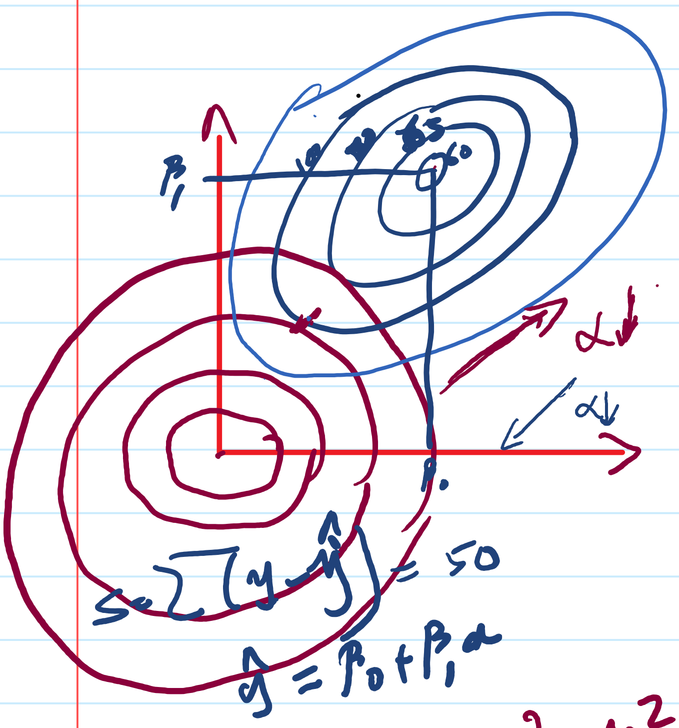
$$S = \sum_{i=1}^{p} (y_i - \hat{y}_i)^2 + \alpha \sum_{i=0}^{n} \beta_i^{2}$$

$\rightarrow P = 2$

$0 < \alpha < \infty$

$\alpha = 0 \longrightarrow$ Ridge $\rightarrow$ Lm

$\alpha > 0 \longrightarrow \ell_2$ regularization



ℓ_2

$y = \beta_1 x + \beta_0$

$Z = \beta_0^2 + \beta_1^2$

$$S = \sum (y - \hat{y})^2 + (\hat{\beta}_0^2 + \hat{\beta}_1^2)_2$$

Elastic Net / L1,L2

Sunday, May 17, 2020 7:02 PM

$\ell_1 - \ell_2$

$$J = \sum_{i=1}^{p} \left(y_i - \hat{y}_i \right)^2 + \alpha \gamma \sum_{i=1}^{n} |\hat{\beta}_i|$$

$$+ \alpha (1-\gamma) \sum_{i=1}^{n} \hat{\beta}_i^2$$

α

$\gamma = \ell_1 / \ell_2$

Application

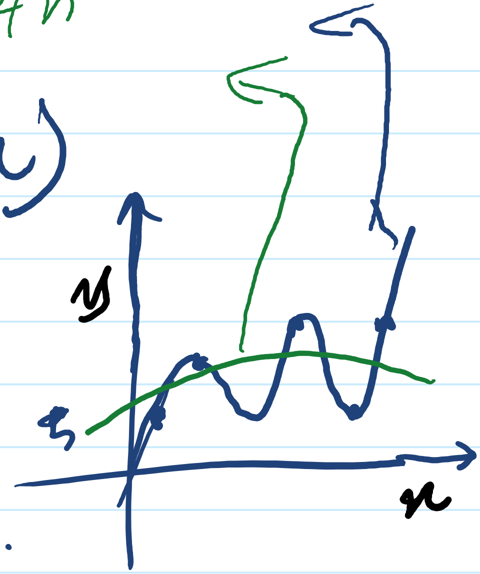
Sunday, May 17, 2020 7:12 PM

w/o Regulirization

with

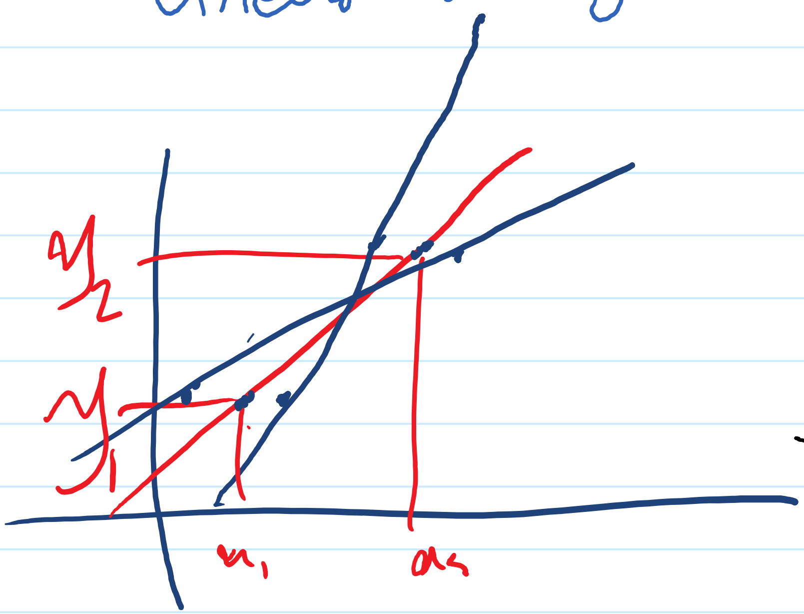
1. Reduce complexity $y = f(x)$

i.e.: reduce over fitting



2. More robust to noise

Linear Regression Ridge



3. l_1 can make $\beta_i = 0$