Lecture Eleven

High-Dimensional ~~R~~ Linear Regression Model

$\{y_t\}$

$$y_t = \beta_0 + \beta_1 (t-1) + \beta_2 \text{ReLU}(t-2) + \beta_3 \text{ReLU}(t-3) + \cdots + \beta_{n-1} \text{ReLU}(t-(n-1)) + \varepsilon_t$$

$\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$

# of beta parameters $= n$ $\parallel$ Data size

① $y_t = \mu_t + \varepsilon_t$

$\mu_t \to$ trend

$$\mu_t = \underbrace{\beta_0 + \beta_1 (t-1)} + \underbrace{\boxed{\beta_2} \underbrace{\text{ReLU}(t-2)} + \cdots + \boxed{\beta_{n-1}} \text{ReLU}(t-(n-1))}$$

② $y = X\beta + \varepsilon$

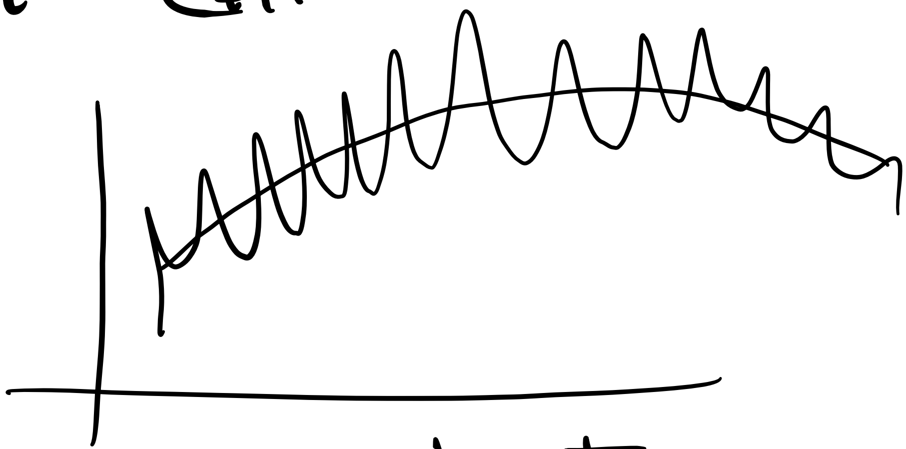
$$\underset{n \times n}{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & & 0 \\ \vdots & 1 & 0 & 0 & & 0 \\ \vdots & 2 & 1 & 0 & & \vdots \\ \vdots & \vdots & 2 & 1 & \cdots & \vdots \\ & \vdots & \vdots & 2 & & 0 \\ & & & \vdots & & 1 \\ 1 & n-1 & n-2 & n-3 & & \end{bmatrix}$$

Least Squares Estimates

$$\sum_{t=1}^{n} \left[y_t - \beta_0 - \beta_1 (t-1) - \beta_2 \text{ReLU}(t-2) - \cdots - \beta_{n-1} \text{ReLU}(t-(n-1)) \right]^2$$

→ Minimize over all parameters $\beta_0, \beta_1, \ldots, \beta_{n-1}$

$$\begin{cases} \beta_0 = y_1, \quad \beta_1 = y_2 - y_1, \quad \beta_2 = (y_3 - y_2) - (y_2 - y_1) \\ \beta_t = (y_{t+1} - y_t) - (y_t - y_{t-1}), \quad t = 2, \dots, n-1 \end{cases}$$



Regularized Estimation

Ridge regularization   LASSO regularization

$$\underbrace{\|y - X\beta\|^2}$$

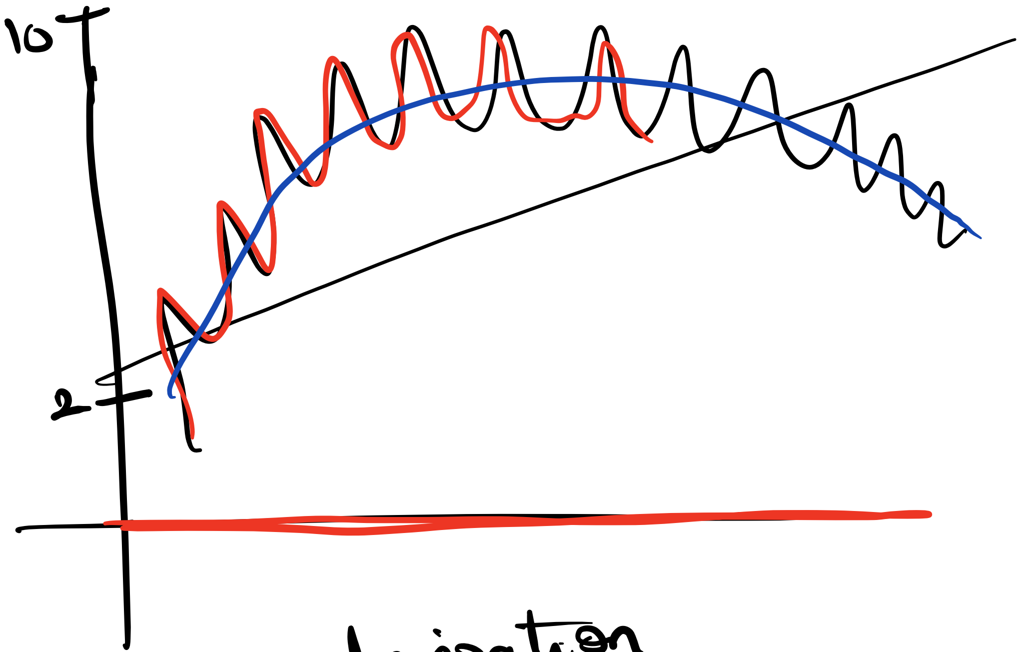
$$= \sum_{t=1}^{n} \left[\boxed{y_t} - \beta_0 - \beta_1 (t-1) - \beta_2 \,\mathrm{ReLU}(t-2) - \dots - \beta_{n-1} \,\mathrm{ReLU}(t-(n-1)) \right]^2$$

$$\min_{\beta_0, \beta_1, \dots, \beta_{n-1}} \left[\|y - X\beta\|^2 + \lambda \left(\beta_2^2 + \dots + \beta_{n-1}^2 \right) \right] = \hat{\beta}^{ridge}(\lambda)$$

Tuning Parameter

① $\hat{\beta}^{ridge}(0)$ = unregularized least squares $\left(y_1, y_2 - y_1, (y_3 - y_2) - (y_2 - y_1), \dots \right)$

② $\hat{\beta}^{ridge}(+\infty)$ = linear regression $\left(\bar{y} - \hat{\beta}_1 \bar{x}, \dfrac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}, 0 \dots 0 \right)$



LASSO regularization

$$\boxed{\|y - X\beta\|^2} + \lambda \left(|\beta_2| + |\beta_3| + \cdots + |\beta_{n-1}| \right)$$

Minimize $\beta_0, \beta_1, \beta_2 \ldots \beta_{n-1}$

$$= \boxed{\hat{\beta}^{LASSO}(\lambda)}$$

$\lambda = 0 \rightarrow$ Unregularized

$\lambda = \infty \rightarrow$ linear regressi

Fitted Values

$$\hat{\mu}^{ridge}(\lambda) = X \hat{\beta}^{ridge}(\lambda)$$

$$\hat{\mu}^{LASSO}(\lambda) = X \hat{\beta}^{LASSO}(\lambda)$$

① **Ridge**

$$\|y - X\beta\|^2 + \lambda \sum_{j=2}^{n-1} \beta_j^2$$

$$= \sum_{t=1}^{n} \left(y_t - \underbrace{\cdots}_{\beta_0 + \beta_1 (t-1) + \beta_2 \mathrm{ReLU}(t-2) + \cdots + \beta_{n-1} \mathrm{ReLU}(t-(n-1))} \right)^2 + \lambda \sum_{j=2}^{n-1} \beta_j^2$$

μ_t

$$\mu_t = \beta_0 + \beta_1 (t-1) + \beta_2 \mathrm{ReLU}(t-2) + \dots + \beta_{n-1} \mathrm{ReLU}(t-(n-1))$$

$t = 1, \dots, n$

$\beta_0 = \mu_1, \quad \beta_1 = \mu_2 - \mu_1$

$\beta_2 = (\mu_3 - \mu_2) - (\mu_2 - \mu_1)$

$\beta_t = (\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})$

$$\hat{\beta}^{\text{ridge}}(\lambda) = \underset{\beta}{\operatorname{argmin}} \left[\|y - X\beta\|^2 + \lambda \sum_{j=2}^{n-1} \beta_j^2 \right]$$

$$\hat{\mu}^{\text{ridge}}(\lambda) = \operatorname{argmin} \left[\sum (y_t - \mu_t)^2 + \lambda \sum_{j=2}^{n-1} \left[(\mu_{j+1} - \mu_j) - (\mu_j - \mu_{j-1}) \right]^2 \right]$$

$\{y_t\}$ $\{\mu_t\}$

$\mu_{t+1} - \mu_t \approx \mu_t - \mu_{t-1}$

for all t on average

Smooth Trend Estimation

(Hodrick-Prescott Filter

Cubic spline smoothing)

$$\hat{\mu}^{\text{LASSO}}(\lambda) = \operatorname{argmin} \left[\sum (y_t - \mu_t)^2 + \lambda \sum_{j=2}^{n-1} \left| (\mu_{j+1} - \mu_j) - (\mu_j - \mu_{j-1}) \right| \right]$$

(Trend Filtering)

$\boxed{\text{cvxpy}}$

Selection of λ by CV

$$\sum_{t=1}^{n}\left[y_t - \beta_0 - \beta_1(t-1) - \beta_2 \mathrm{ReLU}(t-2) - \cdots - \beta_{n-1}\mathrm{ReLU}(t-(n-1))\right]^2 + \lambda\left(\beta_2^2 + \cdots + \beta_{n-1}^2\right)$$

minimize to get $\hat{\beta}^{ridge}(\lambda)$

$$t = 1, \ldots, n$$

T_{train} $\quad$ T_{test} $\rightarrow$ SPLIT

e.g.: last 20% $\rightarrow T_{test}$

first 80% $\rightarrow T_{train}$

$$\hat{\beta}^{ridge}_{train}(\lambda) = \underset{\beta_0, \beta_1, \ldots, \beta_{n-1}}{\text{minimize}} \left[\sum_{t \in T_{train}} (\quad)^2 + \lambda\left(\beta_2^2 + \cdots + \beta_{n-1}^2\right)\right]$$

For every $t \in T_{test}$, calculate

$$\hat{\beta}_0^{ridge, train}(\lambda) + \hat{\beta}_1(t-1) + \hat{\beta}_2 \mathrm{ReLU}(t-2) + \cdots + \hat{\beta}_{n-1}\mathrm{ReLU}(t-(n-1))$$

$$\hat{\beta}^{ridge}_{train}(\lambda)$$

$$= \hat{y}_t(\lambda)$$

$$\mathrm{MSE}(\lambda, \text{split}) = \sum_{t \in T_{test}} \left(y_t - \hat{y}_t(\lambda)\right)^2$$

$split_1, split_2, \ldots, split_5$

$$\boxed{MSE(\lambda, \text{all splits})} = \sum_{i=1}^{5} MSE(\lambda, split_i)$$

Candidate λs : $\{10^{-6}, 10^{-5}, \ldots, 10^{5}, 10^{6}, 10^{7}\}$

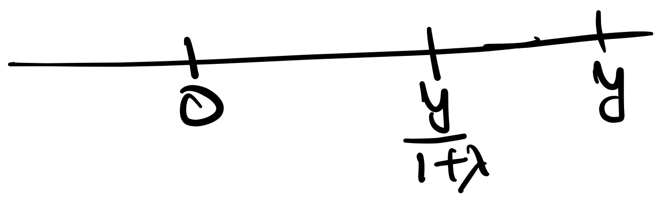
Shrinkage (Ridge)

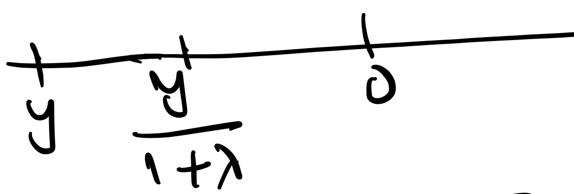
vs

Sparsity (LASSO)

$y \in \mathbb{R}$

$$\min_{\beta} \left[\underbrace{(y-\beta)^2 + \lambda \beta^2} \right] \rightarrow \hat{\beta} = \boxed{\frac{y}{1+\lambda}}$$





$\lambda > 0$

$$\min_{\beta} \left[(y-\beta)^2 + \lambda |\beta| \right]$$

Proof in notes

$$\text{Ans : } \hat{\beta} = \begin{cases} y - \frac{\lambda}{2} & y > \frac{\lambda}{2} \\ y + \frac{\lambda}{2} & y < -\frac{\lambda}{2} \\ 0 & \boxed{-\frac{\lambda}{2} \leq y \leq \frac{\lambda}{2}} \end{cases}$$

SOFT-THRESHOLDING $S_{\frac{\lambda}{2}}(y)$

Sparse