Lecture November 1

Simple example

$X = \in \mathbb{R}^{n \times n}$ $\qquad p = n$

$$X = \left[\begin{array}{cc|cc} 1 & & 0 & \\ & 1 & & \\ 0 & & \ddots & \\ & & & 1 \end{array}\right]$$

$\Big(\tilde{y} = X\beta \Rightarrow \tilde{y}_i = \sum_{j=0}^{p-1} x_{ij} \beta_j$

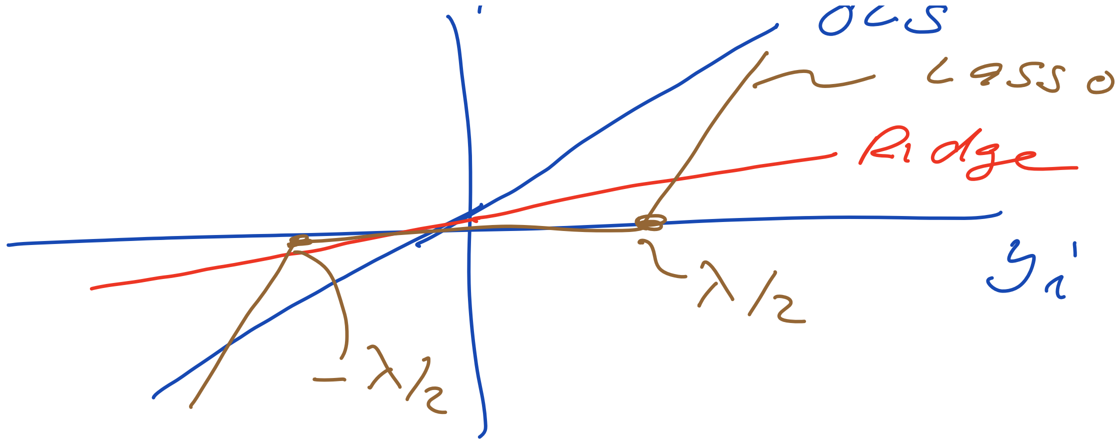
$$\tilde{y}_i = \sum_{j=0}^{p-1} \beta_j x_i^j$$

$$\hat{\beta}_i^{OLS} = y_i$$

$$\hat{\beta}_i^{Ridge} = \frac{y_i}{1+\lambda}$$

$$\hat{\beta}_i^{Lasso} = \begin{cases} y_i - \lambda/2 & y_i > \lambda/2 \\ y_i + \lambda/2 & y_i < -\lambda/2 \\ 0 & |y_i| \leq \lambda/2 \end{cases}$$

β_i



OLS from statistics - assumptions in OLS

$$y = f(x) + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

$f(x)$ is a continuous function.

$$f(x_i) \simeq \tilde{y}(x_i) = \sum_{j=0}^{p-1} x_{ij} \beta_j$$

$$\tilde{y}(x) = X\beta$$

$$\mathbb{E}[y] = \frac{1}{n} \sum_{i=0}^{n-1} y_i = \mu_y$$

$$\left(\mu_y = \int p(y)\, y\, dy\right)$$

$$\bar{\mu}_y \neq \mu_y$$

$$y(x) \simeq \underbrace{X\beta}_{\text{not stochastic}} + \varepsilon$$

$$\mathbb{E}[X\beta] = X\beta$$

$$\mathbb{E}[y_i] = \mathbb{E}[\underbrace{x_{i*}\beta}_{\sum_{j=0}^{p-1} x_{ij}\beta_j}] + \underbrace{\mathbb{E}[\varepsilon]}_{=0}$$

$$\Rightarrow \quad \mathbb{E}[y_i] = x_{i*}\beta$$

$$\Rightarrow \quad y_i \sim N(x_{i*}\beta, \sigma^2)$$

$$\mathrm{var}[y_i] = \sigma^2$$

$$\underline{p(y_i X|\beta)} = \frac{1}{(2\pi\sigma^2)^{1/2}} \times \exp\left[-\frac{(y_i - x_{i*}\beta)^2}{2\sigma^2}\right]$$

y_i ~ independent & identically distributed

$$D = \{ (x_0\ y_0), \ldots (x_{n-1}\ y_{n-1}) \}$$

$$P(D|\beta) = \prod_{i=0}^{n-1} P(y_i x | \beta)$$

$$\hat{\beta} = \arg\max_{\beta \in \mathbb{R}^p} P(D|\beta)$$

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} (-\log P(D|\beta))$$

$$-\log P(D|\beta) = \frac{n}{2} \log (2\pi\sigma^2) + \frac{\|(y - X\beta)\|_2^2}{2\sigma^2}$$

$$= c(\beta)$$

$$\frac{\partial C}{\partial \beta} = 0 = X^T(y - X\beta)$$

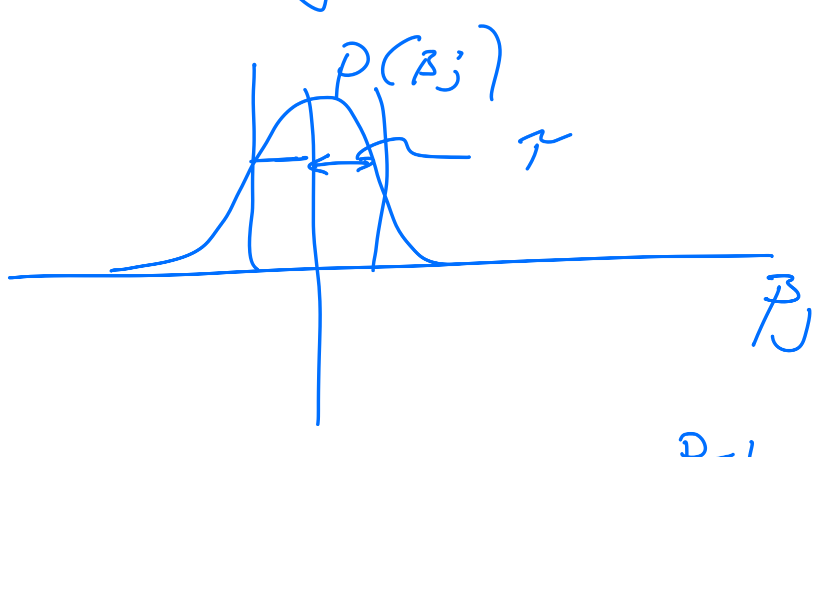
$$\Rightarrow \boxed{\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y}$$

$$p(D|\beta)$$

$$p(\beta|D) \propto p(D|\beta)\,\underline{p(\beta)}$$

$$p(y|x) \propto p(x|y)\,p(y)$$

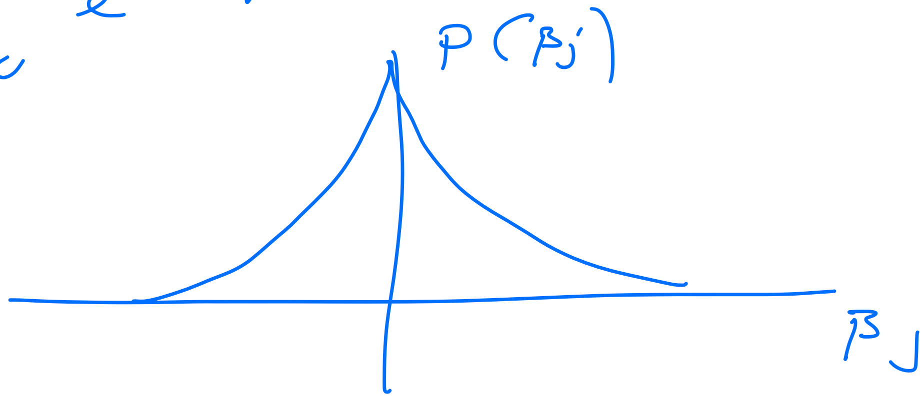
$$p(\beta) = \prod_{j=0}^{P-1} e^{-\frac{\beta_j^2}{2\tau^2}}$$



$P-1$

$$P(\beta|D) \propto P(D|\beta) \prod_{j=0}^{p-1} e^{-\beta_j^2/2\tau^2}$$

$$\prod_{j=0}^{p-1} e^{-|\beta_j|/\tau}$$



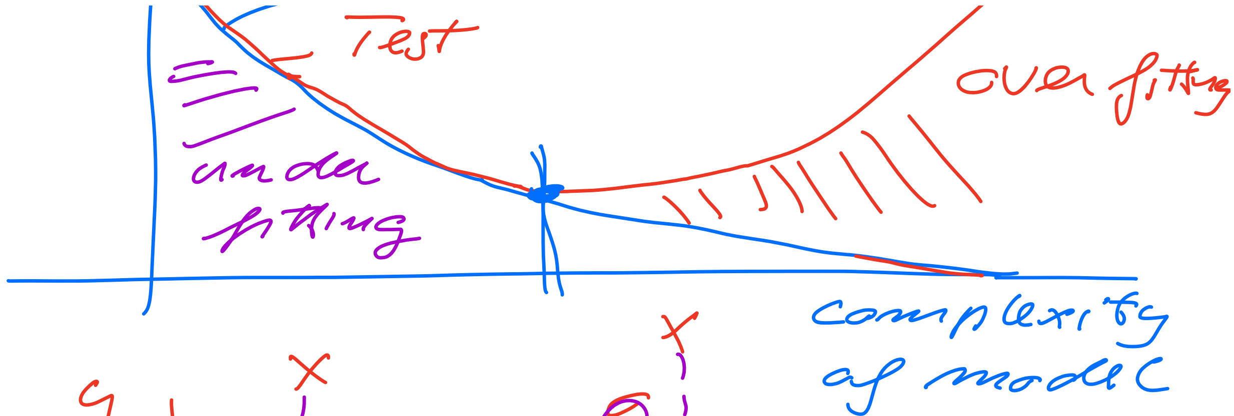
Resampling

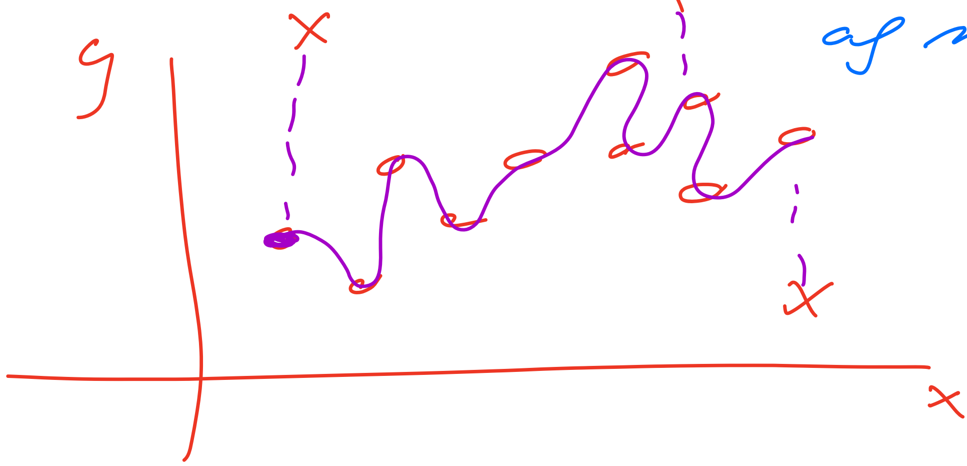
$$\mu_x = \mathbb{E}[x] = \frac{1}{n}\sum_{i=0}^{n-1} x_i \neq \text{exact}$$

$$\sigma_x^2 = \mathbb{E}[(x-\mu_x)^2]$$

$$= \frac{1}{n}\sum_{i=0}^{n-1} (x_i - \mu_x)^2 \neq \text{exact}$$

we want a as reliable as possible estimate of expectation values





$$MSE = \mathbb{E}\left[(y-\tilde{y})^2\right]$$

Resampling techniques promise to give a reliable estimate of various expectation values

- Bootstrap (small n)
- Cross-validation
- Blocking
- Jack knife

- - .

Bootstrap:

sample $D = \{x_0\ x_1\ \ldots\ x_{n-1}\}$

- Draw new sample D^* with n events with replacement.
  - compute selected expectation values Y_i
- repeat M-times,

$$E[Y] = \frac{1}{M} \sum_{i=0}^{M-1} Y_i$$