Lecture September 10

OLS

$$\hat{\beta}_{OLS} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|(y - X\beta)\|_2^2$$

Ridge

$$\hat{\beta}_{Ridge} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|(y - X\beta)\|_2^2 + \underline{\lambda \|\beta\|}_2^2$$

$$\|\beta\|_2^2 = \sum_{j=0}^{p-1} \beta_j^2 \leq t$$

$$\lambda \geq 0$$

$$\hat{\beta}_{Lasso} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|(y - X\beta)\|_2^2 + \lambda \sum_{j=0}^{p-1} |\beta_j|$$

<u>SVD</u> $\quad X \in \mathbb{R}^{n \times p}$

$$X = \underset{\mathbb{R}^{n\times n}}{U}\ \underset{\mathbb{R}^{n\times p}}{\Sigma}\ \underset{\mathbb{R}^{p\times p}}{V^T}$$

$$(X^T X)\, V = V\, \Sigma^2$$

$$\tilde{y}_{OLS} = X \hat{\beta}_{OLS} = \sum_{j=0}^{p-1} u_j u_j^T y$$

$$\tilde{y}_{Ridge} = X \hat{\beta}_{Ridge} = \sum_{j=0}^{p-1} \left(u_j u_j^T \frac{\sigma_j^2}{\sigma_j^2 + \underline{\lambda}} \right) y$$

$$\sigma_0 > \sigma_1 > \sigma_2 \ldots > \sigma_{p-1} > 0$$

---

statistical analysis

Basic assumption:

$$y_i = f(x_i) + \varepsilon_i$$

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

$$f(x_i) \simeq x_{i*}\beta$$

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

$$\mathrm{var}[y_i] = ?$$

$$\mathrm{var}[x] = \int dx\,(x-\mu)^2 p(x)$$

$$\mu = \mathbb{E}[x] = \int dx\, x\, p(x)$$

sample mean

$$\bar{\mu} = \frac{1}{n}\sum_{i=0}^{n-1} x_i$$

sample variance

$$\mathrm{var}(x) = \frac{1}{n}\sum_{i=0}^{n-1}(x_i - \bar{\mu})^2$$

$$\mathrm{cov}(x,y) = \frac{1}{n}\sum_{k=0}^{n-1}(x_k - \bar{\mu}_x)(y_k - \bar{\mu}_y)$$

iid = independent and identically distributed

$$\operatorname{cov}(x, y) = 0$$

$$\operatorname{var}[y_i] = \mathbb{E}[y_i^2] - (\mathbb{E}[y_i])^2$$

$$(\; y_i = x_{i*}\beta + \varepsilon_i \;)$$

$$= \mathbb{E}[(x_{i*}\beta + \varepsilon_i)^2] - \cancel{(x_{i*}\beta)^2}$$

$$\mathbb{E}[\cancel{(x_{i*}\beta)^2} + \underset{\downarrow\, 0}{2\varepsilon_i}\, x_{i*}\beta + \varepsilon_i^2]$$

$$= \mathbb{E}[\varepsilon_i^2] = \sigma^2$$

y_i have mean value $x_{i*}\beta$ and variance σ^2

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

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i * \beta - y_i)^2}{2\sigma^2}}$$

$$= p(y_i, x \mid \beta)$$

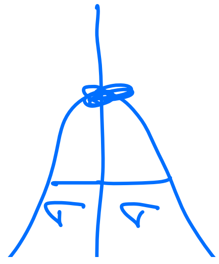
$$D = [(x_0, y_0), (x_1, y_1) \ldots, (x_{n-1}, y_{n-1})]$$

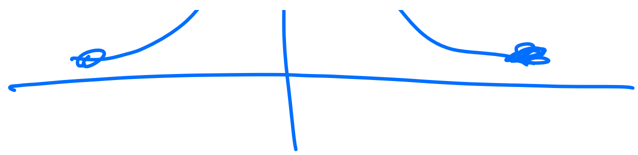
$y_i \sim$ iid

$$\underline{p(D \mid \beta)} = \prod_{i=0}^{n-1} p(y_i x \mid \beta)$$

$$= \prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i - x_i * \beta)^2 / 2\sigma^2}$$

How do we find $\hat{\beta}$?





$$\log P(D|\beta) = C(\beta)$$

max problem

$-\log P(D|\beta)$ min problem

$$\Rightarrow C(\beta) = -\log P(D|\beta)$$

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

$$= \frac{1}{2} n \log(2\pi\sigma^2)$$

$$+ \boxed{\sum_{i=0}^{n-1} \frac{(y_i - x_i * \beta)^2}{2\sigma^2}}$$

$$\sum_{i=0}^{n-1} \frac{1}{2} \log(2\pi\sigma^2)$$

$$C(\beta) = \|(y - X\beta)\|^2 / 2\sigma^2$$

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

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

$$X\beta = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0p-1} \\ x_{10} & & & \\ \vdots & & & \\ x_{i0} & x_{i1} & \cdots & x_{ip-1} \\ \vdots & & & \\ x_{n-1\,0} & \cdots & \cdots & x_{n-1\,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

$x_{i*}\beta$

$$y = X\beta + \varepsilon$$

$$y_i = x_{i*}\beta + \varepsilon_i$$

Continuous distribution

$p(x\,y)$

$$p(x) = \underbrace{\sum_y p(x, y=y)}$$

$$= \sum_y p(x \mid Y=y)\, p(y)$$

$$\int P(x|y)P(y)\,dy = P(x)$$

<u>Example</u> :

you have (sensitivity) cancer if test is positive

$$\underline{P(x=1 \mid \underline{y=1}) = 0.8}$$

$$\underline{P(y=1 \mid x=1)}$$

$$= \frac{\underline{P(x=1 \mid y=1)} P(y=1)}{P(x=1 \mid y=1)P(y=1) + P(x=1 \mid y=0) P(y=0)}$$

$$P(y=1) = \underline{\underline{0.004}}$$

$$P(x=1 \mid y=0) = 0.1$$

$$= \frac{0.8 \times 0.004}{\ldots}$$

$0.8 \times 0{,}004 + 0{,}1 \times 0{,}496$

$= 0{,}03 \Rightarrow$
3 %.

Ridge and Lasso?

OLS: $y \sim N(X\beta, \sigma^2)$

$$\underbrace{P(D|\beta)}_{\text{Likelihood}} = \prod_{i=0}^{n-1} p(y_i, x_i | \beta)$$

$$P(\beta|D) \propto P(D|\beta)\, P(\beta)$$

prior

Model for $P(\beta)$

$P(\beta) \sim N(0, \tau^2)$

$$\prod_{j=0}^{p-1} e^{-\beta_j^2/2\tau^2}$$

$$= \prod_{j=0} e$$

$$P(\beta|D) =$$

$$\prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - x_i * \beta)^2}{2\sigma^2}} \prod_{j=0}^{p-1} e^{-\beta_j^2/2\tau^2}$$

$$C(\beta) = -\log P(\beta|D)$$

$$= -\sum_{i=0}^{n-1} p(y_i x_i|\beta) + $$

$$-\sum_{j=0}^{p-1} \log p(\beta_j)$$

$$= \frac{n}{2} \log(2\pi\sigma^2) +$$

$$\|(y - X\beta)\|_2^2 / 2\sigma^2$$

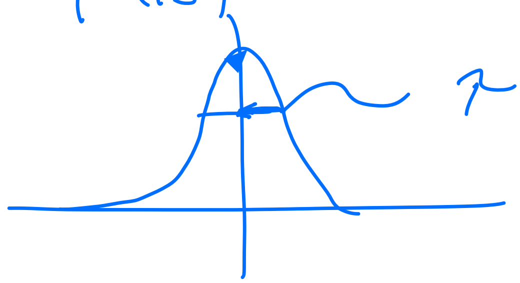
$$+ \|\beta\|_2^2 / 2\tau^2$$

$1/2\tau^2 = \lambda$, skip constant

$$C(\beta) = \frac{\|(y - X\beta)\|_2^2}{2\sigma^2} + \lambda \|\beta\|_2^2$$

Ridge

$p(\beta_j)$



$$\lambda = \frac{1}{2\tau^2}$$

$$\frac{\sigma_j^2}{\sigma_j^2 + \lambda}$$

SVD

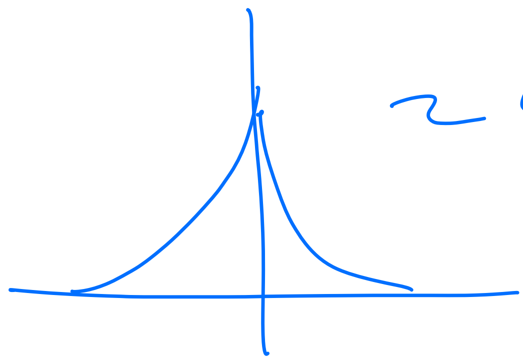
Ridge : $p(\beta) = \prod_{j=0}^{p-1} e^{-\beta_j^2/2\tau^2}$

Lasso :

$$C(\beta) = \|(y - X\beta)\|_2 + \lambda \sum_{j=0}^{p-1} |\beta_j|$$

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

$$\lambda = \frac{1}{\tau}$$



~ Laplace