CompSci, Nov 2, 2022

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

Model $\tilde{y} = X\beta$

Design matrix

model parameters

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

$$= \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \ldots + \beta_{p-1} x_i^{p-1}$$

Assumption $y(x_i) = y_i$

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

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

deterministic

noise

$\tilde{y}_i = f_i$

Assess quality of model

Cost/loss/error/risk/... function

$$C(\beta) = MSE = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j\right)^2$$

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

$$= \frac{1}{n} \| (y - X\beta) \|_2^2$$

$$X = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0p-1} \\ x_{10} & & & \\ \vdots & & & \\ x_{n-1\,0} & - & - & x_{n-1\,p-1} \end{bmatrix}$$

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

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^{p-1} \\ 1 & x_1 & & & \vdots \\ 1 & x_2 & & & \vdots \\ \vdots & \vdots & & & \vdots \\ 1 & x_{n-1} & & & x_{n-1}^{p-1} \end{bmatrix}$$

$$\frac{\partial C(\beta)}{\partial \beta} = 0$$

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} C(\beta)$$

optimal β

$$\Rightarrow \hat{\beta} = \underbrace{(X^T X)}_{\mathbb{R}^{p \times p}}{}^{-1} X^T y$$

$$p < n \quad (p \ll n)$$

$$\frac{\partial^2 C(\beta)}{\partial \beta \partial \beta^T} = \frac{2}{n} X^T X$$

$\Rightarrow$ convex optimization (there is a minimum)

- scaling of design matrix by subtracting mean value of each column. Polynomial fitting case:

$$X = \begin{bmatrix} 1 & x_0 & \cdots & x_0^{p-1} \\ 1 & x_1 & & \\ \vdots & \vdots & & \\ 1 & x_{n-1} & \cdots & x_{n-1}^{p-1} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x^{(1)} & x^{(2)} & \cdots & x^{(p-1)} \end{bmatrix}$$

$$X^T = [x_0 \; x_1 \; \cdots \; x_{n-1}]$$

Define mean $\mu^{(i)} = \frac{1}{n} \sum_{j=0}^{n-1} x_j^{(i)}$

$$= \frac{1}{n} \sum_{j=0}^{n-1} x_{ji}$$

$$v = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{n-1} \end{bmatrix} \qquad w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{n-1} \end{bmatrix}$$

$$v \otimes w = \begin{bmatrix} v_0 \cdot w_0 \\ v_1 \, w_1 \\ \vdots \end{bmatrix}$$

$$\beta = \underbrace{(X^T X)}_{\det(X^T X) = 0}^{-1} X^T y$$

this leads to problems.

cheap trick to avoid $\det(X^T X) = 0$

$$X^T X = H = \begin{bmatrix} h_{00} & h_{01} & \dots & h_{0p-1} \\ x & h_{11} & x\ x & x \\ x & x & \dots & x\ x\ x \\ x & x & \dots & h_{p-1\,p-1} \end{bmatrix}$$

$$\varepsilon = 10^{-10}$$

$$\begin{bmatrix} h_{00}+\varepsilon & x & x & x & x \\ x & h_{11}+\varepsilon & & x & x \\ x & x & & & \\ & x & x & \dots & h_{p-1\,p-1}+\varepsilon \end{bmatrix}$$

$$H \Rightarrow H + \lambda \mathbb{I}$$

$$\hat{\beta} = (H + \lambda \mathbb{I})^{-1} X^T y$$

$$= (X^T X + \lambda \mathbb{I})^{-1} X^T y$$

$\hat{\beta}_{Ridge}$ or just Ridge Regression

$$C(\beta) \Rightarrow C(\beta)_{Ridge} =$$

$$\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2$$

$$+ \lambda \sum_{j=0}^{p-1} \beta_j^2$$

$\lambda > 0$ $\qquad$ $\sum_{j=0}^{p-1} \beta_j^2 \leq t$

hyperparameter.

$$C(\beta) = \frac{1}{n} \| (y - X\beta) \|_2^2 + \lambda \|\beta\|_2^2$$

$$\frac{\partial C(\beta)}{\partial \beta} = -\frac{2}{n} X^T (y - X\beta) + 2\lambda\beta = 0$$

$$\Rightarrow X^T X \beta + \underbrace{\lambda n}_{\tilde{\lambda}} \beta = X^T y$$

$$\hat{\beta} = \underbrace{(X^T X + \lambda I)^{-1}}_{\text{invertible.}} X^T y$$

Lasso Regression

$$C(\beta) = \frac{1}{n} \|(y - X\beta)\|_2^2 + \lambda \|\beta\|_1$$

$$\|\beta\|_1 = \sum_{j=0}^{p-1} |\beta_j|$$

$$\frac{d|x|}{dx} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$$

---

*

SVD

any matrix X can be decomposed as

$$X = U \Sigma V^T$$

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

$$UU^T = U^T U = \mathbb{1}$$
$$U \in \mathbb{R}^{n \times n}$$

$$VV^T = V^T V = \mathbb{1}$$
$$V \in \mathbb{R}^{p \times p}$$

$$U = [u_0 \; u_1 \; \ldots \; u_{n-1}]$$
$$u_i^T u_j = \delta_{ij}$$

$$V = [v_0 \; v_1 \; \ldots \; v_{p-1}]$$
$$v_i^T v_j = \delta_{ij}$$

$$\Sigma = \begin{bmatrix} \sigma_0 & & & \\ & \sigma_1 & & 0 \\ 0 & & \ddots & \\ & & & \sigma_{p-1} \\ & & & 0 \end{bmatrix}$$

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