OH : Wed 12-1

email : pyuning@bu.edu

For linear model, $\{X_i, Y_i\}_{i=1}^{n}$, n observations

predictor / feature (X_i), response (Y_i)

p - predictors

$X_i = (X_{i1}, X_{i2}, \cdots, X_{ip})$

goal: what should be the value of response when we observe X_0.

$Y_i \mid X_i \overset{ind}{\sim} N(X_i^T \beta, \sigma^2)$

X_i: i-th observation; β: parameter to be estimated

Maximum likelihood / Ordinary least square

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \underbrace{(Y - X\beta)^T (Y - X\beta)}_{RSS}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p \times 1}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

design matrix

$$X = \begin{pmatrix} X_{11} & \cdots & X_{1p} \\ X_{21} & \cdots & X_{2p} \\ \vdots & & \\ X_{n1} & \cdots & X_{np} \end{pmatrix}_{n \times p}$$

row: i-th observation; column: j-th predictor / feature

square matrix

$$\Longrightarrow \underbrace{X^T X}_{\text{square matrix}} \hat{\beta} = X^T Y$$

$(X^TX)^{-1}$ exists, $\hat{\beta} = (X^TX)^{-1} X^T Y$

$(X^T X)^{-1}$ does not exist, determinant is extremely small

it' hard to solve for $\hat{\beta}$.

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 2 & 2 & 2 & \cdots & 2 & 2 \\ & & & & & \end{pmatrix}_{n \times n} \quad \text{determinant} = 0$$

Or if we want to control the complexity of the model, i.e. # of the predictors,

$$\underset{\beta}{\operatorname{argmin}} (Y - X\beta)^T (Y - X\beta) + \underbrace{\lambda \cdot \text{penalty term}}$$

ridge regression : $\lambda \cdot \beta^T \beta$ Lasso : $\lambda \cdot \sum_{j=1}^{p} |\beta_j|$

λ — penalty parameter

should not choose λ to be too big.

$$\underbrace{(Y - X\beta)^T (Y - X\beta) + \boxed{\lambda \sum_{j=1}^{p} |\beta_j|}}$$

Δ $\uparrow$ Δ

λ, dominate

to minimize the sum, is equivalent to minimize

GLM

$\log \lambda_i = X_i^T \beta$ $\longleftarrow$ Poisson GLM

$\log \frac{p_i}{1 - p_i} = X_i^T \beta$ $\longleftarrow$ Binomial GLM

Logistic regression