ML ERASMUS, SEPT 26, 2022

Linear Regression: ordinary least squares (OLS)

(i) Domain of data

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

input output

(ii) Model

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

continuous function.

normally distributed noise

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

mean = 0

Model $f(x) = \tilde{y}(x)$

$$y(x) = \tilde{y}(x) + \varepsilon$$

$$y(x_i) = y_i = \tilde{y}_i + \varepsilon_i$$

polynomial in x

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

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

$$p \leq n$$

(iii) Assess the model.

$$C(\beta) = ?$$

$\tilde{y}_i$ to be compared with y_i

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

cost/loss function

$$\text{relative error} = \frac{|y_i - \tilde{y}_i|}{|y_i|} \qquad \tilde{y}_i = X_i \beta$$

$$y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} \in \mathbb{R}^n$$

$$MSE = \frac{1}{n} \| y - \tilde{y} \|_2^2$$

$$\|x\|_2 = \sqrt{\sum_i x_i^2}$$

$$\|x\|_2^2 = \sum_i x_i^2$$

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

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \in \mathbb{R}^p$$

$$\tilde{y}_0 = \beta_0 + \beta_1 x_0 + \beta_2 x_0^2 + \dots + \beta_{p-1} x_0^{p-1}$$

$$\tilde{y}_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \dots \beta_{p-1} x_1^{p-1}$$

$$\vdots$$

$$\tilde{y}_{n-1} = \beta_0 + \beta_1 x_{n-1} + \dots$$

LHS = vector $\tilde{y}$

RHS = vector β

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

linear in β

X = Design matrix/feature matrix

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

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

n: # data entries

p: # features

$$\tilde{y}_i = \sum_{j=0}^{p-1} x_{ij} \beta_j = X_{i*} \beta$$

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

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

$$\tilde{y}_i = \beta_0 + \beta_1 x_i$$

linear polynomial

$$X \in \mathbb{R}^{n \times 2}$$

normal case $p \leq n$

often $p << n$

$$\frac{\partial C}{\partial \beta_0} = \frac{\partial C}{\partial \beta_1} = \cdots = \frac{\partial C}{\partial \beta_{p-1}} = 0$$

$$\Rightarrow \quad \frac{\partial C}{\partial \beta} = 0$$

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

$$\frac{\partial C}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\frac{1}{n}(y - X\beta)^T * (y - X\beta)\right]$$

$C(\beta)$ is a scalar

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

$$X^T y = X^T X \beta \implies$$

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

optimal

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