FYS-STK 4155, AUGUST 25

- data

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

output

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \in \mathbb{R}^n$$

input

- model

$$\text{Data } y(x) = f(x) + \varepsilon$$

$$\varepsilon \sim N(0,1)$$

$$\tilde{y}(x) = P_d(x) \simeq f(x)$$

$$P_d(x_j) = \sum_{i=0}^{d-1} \beta_i x_j^i$$

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

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

$$\vdots$$

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

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

$$X = \begin{array}{r} n=0 \\ n=1 \\ \\ \\ n-1 \end{array} \begin{bmatrix} 1 & x_0^1 & x_0^2 & \cdots & x_0^{d-1} \\ 1 & x_1^1 & x_1^2 & \cdots & \\ 1 & \vdots & & & \\ \vdots & & & & \\ 1 & x_{n-1}^1 & \cdots & \cdots & x_{n-1}^{d-1} \end{bmatrix}$$

Design / feature matrix

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

$n \times 1$ $\quad$ $n \times d$ $\quad$ $d \times 1$

normally $d << n$

unknown parameters to be determined.

– Cost/loss... function

$$C(y \mid \beta, X)$$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^d}{\arg\min}\; C(y \mid \beta, X)$$

optimal value

$$\frac{\partial C(y \mid \beta, X)}{\partial \beta} = 0$$

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

$$C = \frac{|y - \tilde{y}|}{|y|}$$

$$= \frac{|y - X\beta|}{|y|}$$

$$f(x) = |x|$$

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

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

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

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

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

$$\frac{\partial C}{\partial \beta_j} = 0$$

$$\frac{\partial C}{\partial \beta} = 0 = -\frac{1}{n} 2 X^T (y - X\beta)$$

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

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

known