FYS-STK3155/4155 Aug 24/2023

1) input $x^T = [x_0 \, x_1 \, \dots \, x_{n-1}]$

$x \in \mathbb{R}^n$

output $y^T = [y_0 \, y_1 \, \dots \, y_{n-1}]$

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

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

$$P(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$$

Model

$$f(x_i) \simeq \tilde{y}(x_i) = \tilde{y}_i$$

polynomial

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

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

$p \leq n$

3) assess fit.

y_i and $\tilde{y}_i$

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

$$= C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} L_i$$

cost function

loss function

- Digression

$$\mathbb{E}[x^m] = \int_{x \in D} x^m p(x) dx$$

$m = 1$ $\quad \mathbb{E}[x] = \mu_x$

$$\left(= \sum_i x_i^m p(x_i) \right)$$

Sample average

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

$$\bar{\mu} \neq \mu_x$$

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

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

$$\vdots$$

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

$$\tilde{y}^T = [\tilde{y}_0 \; \tilde{y}_1 \; \dots \; \tilde{y}_{n-1}]$$

$$\beta^T = [\beta_0 \; \beta_1 \; \dots \; \beta_{p-1}]$$

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

X = Feature matrix
Design - - - -

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

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

$$= \frac{1}{n}\sum_{i=0}^{n-1}\Big(y_i - \underbrace{\sum_{j=0}^{p-1} x_{ij}\beta_j}_{\tilde{y}_i}\Big)^2$$

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

optimal $\beta_{opt} = \hat{\beta} = ?$

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

$$\nabla_\beta C(\beta) = 0$$

$$\frac{\partial C}{\partial \beta_j} = -\frac{2}{m}\left[\sum_{i=0}^{m-1} x_{ij}\right.$$

$$\times \left(y_i - \beta_0 x_{i0} - \beta_1 x_{i1} - \right.$$

$$\left.\left. \cdots \quad - \beta_{p-1} x_{ip-1} \right) \right]$$

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

$$= 0$$

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

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

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

$$\frac{\partial^2 C}{\partial \beta \, \partial \beta'} = \frac{2}{m} X^T X$$

$$X = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{bmatrix}$$

$$X^T = \begin{bmatrix} x_{00} & x_{10} \\ x_{01} & x_{11} \end{bmatrix}$$

$$X^T X = \begin{bmatrix} x_{00}^2 + x_{10}^2 & x_{00}x_{01} + x_{10}x_{11} \\ x_{01}x_{00} + x_{01}x_{10} & x_{01}^2 + x_{11}^2 \end{bmatrix}$$

$$X_0 = \begin{bmatrix} x_{00} \\ x_{10} \end{bmatrix} \qquad X_1 = \begin{bmatrix} x_{01} \\ x_{11} \end{bmatrix}$$

$$\mathrm{Cov}(X_i X_j) =$$

$$\frac{1}{n}\sum_{k=0}^{n-1}(x_{ki}-\overset{0}{\mu})(x_{kj}-\overset{0}{\mu})$$

$$= \frac{1}{n} X^T X$$

$$var_0 = cov(x_0 x_0)$$

$n=2$

$$= x_0^2 + x_{10}^2$$

Derivatives of vectors and matrices

$$y = Ax$$

$$y \in \mathbb{R}^m \quad x \in \mathbb{R}^n$$

$$A \in \mathbb{R}^{m \times n}$$

$$\frac{\partial y}{\partial x} = ?$$

$$y_i = \sum_{k=1}^{n} a_{ik} x_k$$

$$\frac{\partial y_i}{\partial x_j} = a_{ij} \text{ for}$$

all $i = 1, 2, \ldots, m$

$j = 1, 2, \ldots, n$

$$\Rightarrow \frac{\partial y}{\partial x} = A$$

Define a scalar

$$\alpha = y^T A x$$

$$y \in \mathbb{R}^m \quad x \in \mathbb{R}^n$$

$$A \in \mathbb{R}^{m \times n}$$

$$\frac{\partial \alpha}{\partial y} = x^T A^T$$

$$\frac{\partial \alpha}{\partial x} = y^T A$$

$$w^T = y^T A$$

$$\alpha = w^T x \implies$$

$$\frac{\partial \alpha}{\partial x} = w^T = y^T A$$

$$w^T = y^T A$$

$$\frac{\partial \alpha}{\partial y} = ?$$

$$\alpha = \alpha^T = \underbrace{x^T A^T}_{w^T} y$$

$$\frac{\partial \alpha}{\partial y} = x^T A^T$$

$$\alpha = x^T A x$$

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

$$\alpha = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

$$\frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i$$

$$\forall k = 1, 2, \ldots n$$

$$\frac{\partial \alpha}{\partial x} = x^T (A^T + A)$$

$$A^T = A$$

$$= 2x^T A$$

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

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

$$-2X^T(y - X\beta)$$