ERASMUS+, October 3, 2022

Basic elements

+ Data, input (x_i), output (y_i)

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

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

+ Model

+ Function to assess quality of model.

Linear regression

Basic assumption

$$y(x) = \underbrace{f(x)}_{\text{Deterministic function.}} + \varepsilon$$

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

$f(x)$ is a continuous

Model : $f(x) \simeq \tilde{y}(x) = X\beta$

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

Data entries

features/predictors

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

Polynomial fit:

$$f(x) \simeq \tilde{y}(x) \rightarrow$$

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

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

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

Design/feature matrix

$$y(x_i) \simeq \sum_{j=0}^{p-1} x_{ij}\beta_j + \varepsilon_i$$

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

- Assess quality of model

Mean squared error

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

$$= C(\beta) \quad \text{(Loss function, cost function)}$$

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

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

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

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

$$
\begin{array}{c}
\\
i \\
\\
\end{array}
\begin{bmatrix}
x_{00} & x_{01} & \cdots & x_{0p-1} \\
\vdots & & & \\
x_{n-10} & x_{n-11} & \cdots & x_{n-1p-1}
\end{bmatrix}
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_{n-1}
\end{bmatrix}
$$

optimization part

optimal parameter $\hat{\beta}$

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

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

$$X^T X \hat{\beta} = X^T y$$

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

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

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

The normal (most common) case $p \ll n$ or up to $p \leq n$

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

$$= \text{Hessian}$$

$C(\beta)$ depends only on one β.

Minimize $C(\beta)$, what are the requirements on $\frac{\partial^2 C}{\partial \beta^2}$?

$$\frac{\partial^2 C}{\partial \beta^2} > 0$$

$\Rightarrow$ $X^T X$ must be

positive definite, $\lambda_i > 0$

Derivatives of $y = f(x)$, $c(\beta)$ etc.

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & & \end{bmatrix}$$

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

Define $y = Ax$

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

$$\frac{\partial y}{\partial x} = A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \\ a_{m1} & \cdots & & a_{mn} \end{bmatrix}$$

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

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

Define a scalar

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

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

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

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

proof: define $w^T = y^T A$

$$\alpha = w^T x$$

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

$$\alpha = \alpha^T \quad (\text{scalar})$$

$$= x^T A^T y$$

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

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

A is a quadratic matrix $A \in \mathbb{R}^{n \times n}$

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

Proof =

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

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

for all x from $i, j = 1, 2, \ldots n$

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

$$= x^T(A^T + A)$$

if A is symmetric

$$= 2x^T A$$

$$\alpha = (y - Ax)^T (y - Ax)$$

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

Define $w = y - Ax$

$$\alpha = w^T w$$

$$\frac{\partial \alpha}{\partial x} = 2 w^T \frac{\partial w}{\partial x}$$

$$\frac{\partial w}{\partial x} = -A \quad \Rightarrow$$

$$\frac{\partial \alpha}{\partial x} = -2(y - Ax)^T A$$

$$\left(\frac{\partial \alpha}{\partial x}\right)^T = -2A^T(y - Ax)$$

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

X does not depend on β $\Rightarrow$

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

Linear regression and ordinary least squares (OLS)

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

Nuclear data:

$$BE(A) = a_1 A + a_2 A^{2/3} + a_3 A^{-1/3} + a_4 A^{-1}$$

$$\begin{bmatrix} A_1 & A_1^{2/3} & A_1^{-1/3} & A_1^{-1} \\ A_2 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ A_n & \cdot & & \end{bmatrix}$$