Last class

- Linear Regression : Univariate case

$$f_\theta(x) = \theta_0 + \theta_1 x = y$$

y-intercept → $\theta_0 = \bar{y} - \theta_1 \bar{x} = \frac{1}{N}\sum y - \theta_1 \frac{1}{N}\sum x$

slope → $\theta_1 = \dfrac{\sum xy - \sum x \sum y}{\sum x^2 - (\sum x)^2} = \dfrac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2}$

derive .

- Optimal solution that minimizes mean squared error

$$L(f_\theta, D) = \frac{1}{N}\sum_{i=1}^{n} \left(y_n - f_\theta(x_n)\right)^2$$

(1) Illustrative example) $f_\theta(x) = \theta_0 + \theta_1 \cdot x = y$

| $x$ | $y$ |
|---|---|
| 1 | 4.8 |
| 3 | 11.3 |
| 5 | 17.2 |
| 6 | ? |

$$f_\theta(x) = 1.8 + 3.1x$$

$$f_\theta(6) = 1.8 + 3.1 \times 6 = 1.8 + 18.6$$

$$[= 20.4]$$

Today's class — Caesarean Delivery Probability ($y$) → Blood pressure ($x_1$), Number of delivery ($x_2$), Age ($x_3$)

- Generalize to multivariate case

[MSE]

$$f_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n = y$$

Least square solution : Vector / Matrix Notation

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

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \end{bmatrix}$$

[ Code: scikit learn : Linear Regression ]

$\theta = (X^T X)^{-1} X^T y \rightarrow f_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots \theta_n x_n.$

Univariate case $f_\theta(x) = \theta_0 + \theta_1 x = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}^T \begin{bmatrix} 1 \\ x \end{bmatrix}$

$\theta_0 = \bar{y} - \theta_1 \bar{x}$

$$\theta_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

## Vector/Matrix Notation

$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{3\times 1}$ column vector

$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ $3\times 2$

$x^T$

$x = [x_1 \quad x_2 \quad x_3]$ row vector

$x^T = [\;]$

$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$ $2\times 3$

Vector dot product.

$x^T \cdot y = [x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3]_{1\times 1}$

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}_{2\times 2}$

$x \quad y$

$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{3\times 1} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_{3\times 1} = x^T y$

$= [x_1 \; x_2 \; x_3] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ row column.

$$A^{-1} = \frac{1}{\underbrace{ad - cb}_{\text{determinant}}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$D = \{(x_1, y_1), \ldots (x_n, y_n)\}$

$f_\theta(x) = \theta_0 + \theta_1 x_1 + \ldots \theta_n x_n$

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

$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$

Attributes

target / label

$y$

$D = \{(x_i, y_i) : i = 1, \ldots N\}$

$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$f_\theta(x) = \theta_0 + \theta_1 x$

$= \theta_0 \cdot 1 + \theta_1 \cdot x$

$= \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}^T \begin{bmatrix} 1 \\ x \end{bmatrix}$

| $x$ | $y$ |
|---|---|
| 1 | 4.7 |
| 3 | 11.8 |
| 5 | 17.7 |

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

$X = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 5 \end{bmatrix} \rightarrow X^T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 5 \end{bmatrix}$

N = number of d...

$\underline{X^T} \qquad X$

$\begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{bmatrix}_{2\times 3} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}_{3\times 2}$

$X^T X = \begin{bmatrix} \boxed{N} & \boxed{\Sigma x} \\ \boxed{\Sigma x} & \boxed{\Sigma x^2} \end{bmatrix}_{2\times 2}$

$(X^T X)^{-1} = \frac{1}{N \Sigma x^2 - (\Sigma x)^2} \begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & N \end{bmatrix}$

$X^T y = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{bmatrix}_{2\times 3} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_{3\times 1} = \begin{bmatrix} \boxed{\Sigma y} \\ \boxed{\Sigma xy} \end{bmatrix}_{2\times 1}$

$$\theta = (X^T X)^{-1} X^T y = \frac{1}{N\sum x^2 - (\sum x)^2} \begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & N \end{bmatrix} \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

$$(X^T X)^{-1} \qquad (X^T y)$$

loop → hard to parallelize

Vector.

Matrix → Effecient Alogrithm

$$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}_{2 \times 1}$$

$$\underline{\underline{\theta = (X^T X)^{-1} X^T y}} \iff \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \rightarrow \begin{aligned} \theta_0 &= \bar{y} - \theta_1 \bar{x} \\ \theta_1 &= \frac{1}{N} \cdot \frac{\sum xy - \sum x \sum y}{\sum x^2 - (\sum x)^2} \end{aligned}$$

Numpy Library → Overview.

↳ Vectorized form.

Prerequisites

- Numpy library ⟷ scikit learn

⇒ # TODO: send a notebook with introduction to numpy.