Multiple Linear Regression

Wednesday, 6 March 2024 2:02 pm

→ Multiple input columns, 1 output column

→ $\hat{y}_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_m x_{nm}$

↳ Hyperplane in n-dim coordinate system.

	x_1	$x_2 \cdots x_m$	y_n	$\hat{y}$
1	x_{11}	$x_{12} \cdots x_{1m}$	y_1	$\hat{y}_1$
2	x_{21}	$x_{22} \cdots x_{2m}$	y_2	$\hat{y}_2$
⋮	⋮	⋮ ⋯ ⋮	⋮	⋮
n	x_{n1}	$x_{n2} \cdots x_{nm}$	y_n	$\hat{y}_n$

→ $\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_m x_{1m} \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_m x_{2m} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_m x_{nm} \end{bmatrix}_{n\times 1}$

$= \begin{bmatrix} 1 & x_{11} & x_{12} \cdots & x_{1m} \\ 1 & x_{21} & x_{22} \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \cdots & x_{nm} \end{bmatrix}_{n\times(m+1)} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}_{(m+1)\times 1}$ (β_0 → offset)

$= X \quad \beta$

① $\hat{y} = X\beta$

→ $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}$ $\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{n\times 1}$

$e = y - \hat{y} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$

→ $e^T e = \begin{bmatrix} y_1 - \hat{y}_1 & y_2 - \hat{y}_2 & \cdots & y_n - \hat{y}_n \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$

$= (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_n - \hat{y}_n)$

② $E = e^T e = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ → minimize

↳ Loss function

→ $E = e^T e = (y - \hat{y})^T (y - \hat{y})$

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

$= y^T y - y^T \hat{y} - \hat{y}^T y + \hat{y}^T \hat{y}$

$y^T \hat{y} = \hat{y}^T y$ let $y = A$, $\hat{y} = B$

$A^T B = B^T A$

$(A^T B)^T = B^T (A^T)^T$

$= B^T A$

$A^T B = (A^T B)^T$ let $A^T B = C$

$C = C^T$ ↓ Prove symmetric

$A^T B$ is symmetric

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

$= (1\times n)\ n\times(m+1)\ (m+1)\times 1$

$= 1\times n\ \ n\times 1$

$= 1\times 1$ → scalar ∴ $A^T B$ is symmetric

③ $E = y^T y - 2\hat{y}^T \hat{y} + \hat{y}^T \hat{y}$

→ $E = y^T y - 2y^T \hat{y} + \hat{y}^T \hat{y}$ → sub $\hat{y} = X\beta$

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

④ $E = y^T y - 2y^T X\beta + \beta^T X^T X\beta$

$E(\beta)$

* find such value of β matrix for which E is minimum

$\frac{dE}{d\beta} = 0$

→ $E = y^T y - 2y^T X\beta + \beta^T X^T X\beta$

↳ let $A = X^T X$, given A is symmetric

$\frac{dE}{d\beta} = 0 - 2y^T X + 2\beta^T A = 0$

$(X^T X)^T = X^T (X^T)^T$

$= X^T X$

$= 0 - 2y^T X + 2\beta^T X^T X = 0$

$2\beta^T X^T X = 2y^T X$

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

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

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

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

$\beta = \left[y^T X (X^T X)^{-1} \right]^T$

A B

$\beta = \left[(X^T X)^{-1} \right]^T (y^T X)^T$

$= \left[(X^T X)^{-1} \right]^T X^T y$

$(X^T X)^{-1}$ is symmetric (Prove)

$\left[(X^T X)^{-1} \right]^T = (X^T X)^{-1}$

let $X^T X = A$

$A A^{-1} = I$

$(A^{-1})^T A^T = I$

$(A^{-1})^T A = I$

$(A^{-1})^T A A^{-1} = I A^{-1}$

$(A^{-1})^T = A^{-1}$

$\therefore [(X^T X)^{-1}] = (X^T X)^{-1}$

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

OLS Method

⑤

from sklearn.datasets import ...

What is the problem with OLS?

OLS → $(X^T X)^{-1}$

$O(n^3)$ slow / computationally intensive for high dimensionality data.

one of the reasons to opt for gradient descent