# Linear Regression.

OLS (ordinary least square) implemented in Sklearn

SGD Regressor (Gradien descent approach)

## Ordinary Least Square.

**Simple Linear Regression** (one input feature / column)

$\bar{y} = m\bar{x} + b$.

$\bar{y}$ → mean of target col.

$\bar{x}$ → mean of input col.

$\Rightarrow \underline{b} = (\bar{y} - m\bar{x})$

and $\underline{m}$ can be calculated using the loss $f^n$, $Eq^n$, $E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$$E(m,b) = \sum (y_i - mx_i - b)^2$$

→ argmin of $\underline{(m, b)}$

$$\boxed{m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \quad \text{Eq}^n \; (1)$$

$$\boxed{b = \bar{y} - m\bar{x}} \quad \text{Eq}^n \; (2)$$

# Multiple Linear Regression

$\hat{y}_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_m x_{1m}$

$\hat{y}_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_m x_{2m}$

$\vdots$

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

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

$$\boxed{\hat{y} = X\beta}$$

loss, $E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$= e^T e. \qquad e = (y_i - \hat{y})$

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

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

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

$y^T \hat{y} \to$ scaler, $\hat{y}^T y \to$ scaler } always symmetric

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

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

$E(\beta) = (y^T y - 2y^T X\beta + \beta^T X^T X\beta) \to \text{argmin } \beta.$

$\frac{\partial}{\partial \beta}(\beta^T X^T X \beta) = 2\beta^T X^T X$ if $(X^T X) \to$ symmetric

$\frac{\partial E}{\partial \beta} = -2y^T X + 2\beta^T X^T X = 0$

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

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

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

$[y^T X (X^T X)^{-1}]^T$

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

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

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

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

$X \to n \times (m+1)$

$y \to n \times 1$

$\beta \to (m+1) \times 1$

$$\boxed{\hat{y} = X\beta = X(\beta_1, \beta_2, \beta_m) + \beta_0}$$

$\beta' = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_m \end{bmatrix}$

$$\boxed{\hat{y} = \beta_0 + X\beta'}$$