OLS (Ordinary Least square)

Its a method in LR used to estimate unknown parameters in a model.

$$y = x.\theta + \epsilon$$

$$y = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_i \\ e_n \end{bmatrix}$$

x = features of House

θ = parameter / Coeff.

ϵ = error term.

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \cdots & x_{1n} \\ x_{21} & x_{22} & & \cdots & & x_{2n} \\ \vdots & & & & & \vdots \\ x_{n1} & \cdots & & & & x_{nn} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ e_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$y = \theta x + \epsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \begin{bmatrix} \quad \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ e_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$y = X\theta + \epsilon$$

$$\epsilon = y - X\theta \longrightarrow \star$$

to avoid (-ve) value of Residual we can take squares.

$$\epsilon_0^2 + \epsilon_1^2 + \quad + \epsilon_n^2 = [\epsilon_0\ \epsilon_1\ \epsilon_2 \cdots \epsilon_n]\begin{bmatrix}\epsilon_0\\ \epsilon_1\\ \vdots\\ \epsilon_n\end{bmatrix} = \epsilon\epsilon^T$$

$\epsilon\epsilon^T$ is called cost function which we want to minimize

$$\epsilon\epsilon^T = (y - x\theta)(y - x\theta)^T$$

$$= (y - x\theta)(y^T - (x\theta)^T)$$

$$\epsilon\epsilon^T = (y - X\theta)(y^T - (X\theta)^T)$$

$$= (y - X\theta)(y^T - X^T.\theta^T)$$

$$= yy^T - yX^T.\theta^T - y^TX\theta + X\theta.X^T\theta^T$$

$$\epsilon\epsilon^T = yy^T - 2y\theta^TX^T + \theta^TX^TX\theta \quad \text{—} (1)$$

~~to~~ ~~find~~ find θ that minimize sum of squared Residual

$$\frac{\partial \epsilon\epsilon^T}{\partial\theta} = \frac{\partial yy^T}{\partial\theta} - 2\frac{\partial y\theta^TX^T}{\partial\theta} + \frac{\partial \theta^TX^TX\theta}{\partial\theta}$$

$$\frac{\partial \epsilon \epsilon^T}{\partial \theta} = -2X^T Y + 2X^T . X\theta \quad \because \frac{\partial \epsilon \epsilon^T}{\partial \theta} = 0$$

$$\cancel{2} X^T Y = \cancel{2} X^T X \theta$$

$$X^T y = X^T . X \theta$$

$$\theta = \frac{\cancel{X}^T y}{X^T . X}$$

$$\theta_* = (X^T Y)(X^T . X)^{-1}$$

$$y = \theta_0 + \theta_1 x$$

$$y = X\theta + \epsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & \vdots \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

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

$$\theta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\theta_1 = \frac{SS_{xy}}{SS_{xx}} = (x_1 - \bar{x})(y_1 - \bar{y})$$

$$= \underline{x_1 y_1} - x_1\bar{y} - \bar{x}y_1 + \bar{x}\bar{y}$$

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\boxed{SS_{xy} = x_i y_i - n\bar{x}\bar{y}}$$

$$SS_{xx} = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$$

x_i = feature

y_i = Target

$\bar{X} = 594$

$\bar{y} = 191083$

$\sum XY = 107819200$

$\sum X_i = 3122829$

Garage Area (X)	Sale price (Y)
548	208500
460	181500
608	223500
836	140000
480	250000
636	143000
Total =	

$\underline{SS_{xy}} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = 5904088$

$y = mx + c$

$y = \underline{\theta_0} + \underline{\theta_1} x$

$SS_{xx} = 249861$

slope $\theta_1 = \dfrac{SS_{xy}}{SS_{xx}} = \dfrac{5904088}{249861} = 23$

$\theta_0 = \bar{y} - \theta_1 \bar{x} = 191083 - 23 \times (594)$

$\theta_0 = 177421 \leftarrow$ (intercept)

$$\boxed{\text{Saleprice} = 177421 + 23\,(\text{Garage area})}$$

$(\theta_0, \theta_1, \theta_2\, \theta_3 \dots \theta_9)$