$y = mx + b$

$\Rightarrow \quad \mathcal{E} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$\Rightarrow \quad \boxed{\mathcal{E} = \frac{1}{n} \sum_{i=1}^{n} (y_i - mx_i - b)^2}$

$\Rightarrow$ Adjusted R2 score $= 1 - \left[\frac{(1 - R_2)(n-1)}{n - 1 - k} \right]$

$k \rightarrow$ no of col.

$n \rightarrow$ no. of rows.

$\Rightarrow$ Simple linear regression:-

$y = mx + b$

$\boxed{b = \bar{y} - m\bar{x}}$

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

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

$$\Rightarrow \boxed{y = \beta_0 + \sum_{i=1}^{m} \beta_i X_i} \qquad (y = mx + b)$$

X_1	X_2	X_3	$\cdots$	X_m
$\vdots$				
X_n				

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 X_{11} & \beta_1 X_{12} & \beta_1 X_{13} & \cdots & \beta_m X_{1m} \\ \beta_0 & \beta_1 X_{21} & \beta_2 X_{22} & \beta_2 X_{23} & \cdots & \beta_m X_{2m} \\ \vdots & & & & & \\ \beta_0 & \beta_1 X_{n1} & \beta_2 X_{n2} & \beta_3 X_{n2} & \cdots & \beta_m X_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1m} \\ 1 & X_{21} & X_{22} & \cdots & X_{2m} \\ 1 & \vdots & & & \\ \vdots & \vdots & & & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$$

$\cancel{\hat{y}_{1\times m} = X_{m\times(m+1)}}$

$$\boxed{\hat{Y}_{m\times 1} = X_{m\times(m+1)}\, B_{(m+1)\times 1}}$$

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

$$e^T e = [(y_1-\hat{y}_1)\,(y_2-\hat{y}_2) \cdots (y_m-\hat{y}_m)]$$

$$e^T e = (Y-\hat{Y})^T (Y-\hat{Y})$$

$$= (Y^T - \hat{Y}^T)\,(Y-\hat{Y})$$

$$= (Y^T - (XB)^T]\,[Y - XB]$$

$$= Y^T Y - Y^T XB - (XB)^T Y + (XB)^T XB$$

Now, $Y^T XB = (XB)^T Y$ (Proof):-

let $Y = A$, $XB = B$

$$\boxed{A^T B = B^T Y}$$

(

$$\boxed{\begin{aligned} y &= A^T X A \\ \frac{dy}{dA} &= 2XA^T \end{aligned}}$$ ← Imp.

Dt.:
Pg.:

$\Rightarrow e^T e = Y^T Y - Y^T X B - (XB)^T Y + (XB)^T (XB)$

$= Y^T Y - 2Y^T X B + (XB)^T X B$

$$\boxed{\mathcal{E} = Y^T \cancel{Q} Y - 2Y^T X B + B^T X^T X B}$$

$$\frac{d\mathcal{E}}{dB} = -2Y^T X + \frac{d}{dB}(B^T X^T X B)$$

$\Rightarrow -2Y^T X + 2X^T X B^T = 0$

$\Rightarrow \cancel{2} X^T X B^T = \cancel{2} Y^T X$

$$\Rightarrow \boxed{B^T = \frac{Y^T X}{X^T X}}$$

$B^T = Y^T X (X^T X)^{-1}$

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

$$\boxed{B = [(X^T X)^{-1}]^T X^T Y}$$

$$\boxed{B = [(X^T X)^{-1}] X^T Y}$$

⇒ Gradient Descent :-

$$\boxed{X_{new} = X_{old} - \underbrace{\eta}_{\text{learning rate}} \frac{dE}{dX}}$$

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_m X_m$$

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad \text{let } \beta_0 = 0 \; (\beta_1, \beta_2 = 1)$$

$$\beta_0' = \beta_0 - \eta \; \text{slope} \rightarrow \frac{dL}{d\beta_0}$$

$$\beta_1' = \beta_1 - \eta \; \text{slope}$$

$$\boxed{\begin{aligned} \hat{y}_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} \\ \hat{y}_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} \end{aligned}}$$

$$\Rightarrow \quad E = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

$$= \frac{1}{2} \sum_{i=1}^{2} \left[y_i - (\beta_0 + \beta_1 X_1 + \beta_2 X_2) \right]^2$$

$$\frac{\partial E}{\partial \beta_0} = \frac{1}{2} \sum_{i=1}^{2} 2 (y_i - \beta_0 - \beta_1 x_1 - \beta_2 x_2)(-1)$$

$$\boxed{\frac{\partial E}{\partial \beta_0} = \frac{-2}{2} \sum_{i=1}^{2} \left[y_i - \hat{y}_i \right]} \leftarrow \text{for 2 terms}$$

$$\boxed{\frac{\partial E}{\partial \beta_0} = \frac{-2}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)} \leftarrow \text{for m terms}$$

Minhaz

Dt.:
Pg.:

$$\boxed{\frac{\partial E}{\partial \beta_1} = \frac{-2}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i) X_{i1}}$$

X_1	X_2	Y
x_{11} a	x_{12} b	c
x_{21} c	x_{22} d	f

$$\boxed{\frac{\partial E}{\partial \beta_2} = \frac{-2}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i) X_{i2}}$$

MINHAZ

⇒ Types of Gradient Descent :-

- Batch GD → full dataset is taken as i/p
- Stochastic GD → only one row is " " "
- Mini GD → takes some pts (not all) as i/p.

⇒ linear regression ⇒ $y = mx + b$

⇒ multiple linear regression ⇒ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$

⇒ Polynomial linear regression ⇒ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \dots \beta_m x_1^m$

x_1	x_2	y

$$\Rightarrow y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2$$

Dt.:
Pg.:

$$\Sigma y_i = n\beta_0 + \beta_1(\Sigma x_i) + \beta_2(\Sigma x_i^2) + \cdots$$

$$\Sigma y_i x = \beta_0 \Sigma x_i + \beta_1 \cdot \Sigma x_i^2 + \beta_2 \Sigma x_i^3 + \cdots$$

$$\Sigma y_i x^2 = \beta_0 \Sigma x_i^2 + \beta_1 \Sigma x_i^3 + \cdots$$

$$\Sigma y_i x^3 = \beta_0 \Sigma x_i^3 + \beta_1 \Sigma x_i^4 + \cdots$$

Q:- find the quadratic regression model for

x	3	4	5	6	7
y	2.5	3.2	3.8	6.5	11.5

$$y = \beta_0 + \beta_1 x^2 + \beta_2 x^2$$

Ans $\boxed{y = 12.4285 - 5.512x + 0.7642x^2}$

Dt.:
Pg.:

7) L1 & L2 Regularization:- (methods to avoid overfitting)

L1 = lasso regression
L2 = Ridge regression

- L1:-

Make coefficient of higher degree zero.

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots$

$\beta_3 \downarrow$ make zero.

$$\boxed{L \text{ or } J \text{ (loss)} = \sum_{j=0}^{n} (y_j - \hat{y}_j)^2 + \lambda \sum_{j=1}^{n} |\beta_j|}$$

- L2 :-

$$\boxed{J \text{ or } L = \sum_{j=0}^{n} (y_j - \hat{y}_j)^2 + \lambda \sum_{j=1}^{n} \beta_j^2}$$

Dt:
Pg.:

Delta

⇒ Elastic-Net Regression :- (Consider both L1 & L2) :-

$$\boxed{L \text{ or } J = \sum_{j=0}^{m} (y_j - \hat{y}_j)^2 + a(\beta_j)^2 + b\|\beta_j\|}$$

$\|\beta_j\|^2$

$$\lambda = a + b$$

$$l_1\text{-ratio} = \frac{a}{a+b}$$

$$l_1\text{-ratio} = \frac{a}{\lambda}$$

while training model :-

$\boxed{\lambda = 1 \;\&\; l_1\text{-ratio} = 0.5}$

$$L = \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 + \lambda \sum_{j=1}^{m} \beta_j^2$$

$$= \sum_{j=0}^{m} (y_j - \hat{y}_j)^2 + \lambda m^2$$

$$= \sum_{j=0}^{m} (y_j - (mx_j + b))^2 + \lambda m^2$$

$$L = \sum_{j=0}^{m} [y_j - (mx_j + \bar{y} - m\bar{x})]^2 + \lambda m^2 \qquad (b = \bar{y} - m\bar{x})$$

$$\frac{\partial L}{\partial m} = \sum_{j=0}^{m} 2(y_j - \hat{y}_j)(-1) \times (x_j - \bar{x}) + 2\lambda m$$

$$= 2\sum_{j=0}^{m} (y_j - \hat{y}_j)(\bar{x} - x_j) + 2\lambda m$$

$$\frac{\partial L}{\partial m} = 0$$

$$2 \sum_{j=1}^{m} (y_j - \hat{y}_j)(\bar{x} - x_j) + 2\lambda m = 0$$

$$\cancel{2} \sum_{j=1}^{m} [y_j - \bar{y} - m x_j + m\bar{x}](\bar{x} - x_j) + \cancel{2}\lambda m = 0$$

$$\lambda m - \sum_{j=1}^{m} [(y_j - \bar{y}) - m(x_j - \bar{x})](x_j - \bar{x}) = 0$$

$$\lambda m - \sum_{j=1}^{m} (y_j - \bar{y})(x_j - \bar{x}) + \sum_{j=1}^{m} m (x_j - \bar{x})^2 = 0$$

$$\lambda m + m \sum_{j=1}^{m} (x_j - \bar{x})^2 = \sum_{j=1}^{m} (y_j - \bar{y})(x_j - \bar{x})$$

$$\boxed{m = \frac{\sum_{j=1}^{m} (y_j - \bar{y})(x_j - \bar{x})}{\lambda + \sum_{j=1}^{m} (x_j - \bar{x})^2}}$$