LINEAR REGRESSION

→ build relation b/w multiple feature to estimate target o/p

→ $y = mx + b$

↳ ~~there is room for only one feature~~ in this eqn

* ordinary least square :-
  ↳ helps us to slove for m(slope) and b(intercept) fo

SINGLE FEATURE	X			Y
	x_1	x_2	x_3	y
	Feature 1	feature 2	Feature 3	Target

$$\hat{y} = \beta_0 x_0 + \beta_0 x_1 + \dots + \beta_n x_n$$

denotes prediction (no perfect fit)

Beta coefficient to minimize error

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

best fit

MATHS behind Linear Regression for single feature

$$y = mx + b$$
$$\hat{y} = b_0 + b_1 x$$

$$(y - y) = b_1 (x - \bar{x})$$

$$y = b_1 x + \underbrace{\bar{y} - b_1 \bar{x}}_{b_0}$$

Derive :- (b_0) and (b_1) {Regression coefficient}

$$b_1 = \rho_{x,y} \frac{\sigma_y}{\sigma_x}$$

(r)
ρ : Karl Pearson coefficient
σ_x, σ_y : standard Deviation

$$= \frac{cov(x,y)}{\sigma^2(x)}$$

$$r = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

$$\Rightarrow \boxed{b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}}$$

$$cov(x,y) = \frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})$$

$$\Rightarrow \boxed{b_0 = \bar{y} - b_1 \bar{x}}$$

* once we get b_1, b_0 we will get line of regression for given data points

$$\sigma_x = \sqrt{\frac{\Sigma(x - x_i)^2}{n}}$$

* Limitation of L.R are Anscombe's Quartet
↳ It is group of 4 diff graph which result is same line of regression

classmate
Date ____
Page ____

MULTPLE FEATURE → we use gradient descent to solve a cost function to calculate Beta values

$$\hat{y} = \sum_{i=0}^{n} \cancel{\beta_0} \cdot \beta_i x_i + \cancel{\text{intercept}} \text{ (intercept)}$$

↳ error.

Find set of Beta coeff that minimize the error (cost function)

$$\left[\frac{1}{m} \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 \right]$$

Average square error (Residual sum of square)

* **cost function (J) :-**

$$J(\beta) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

* 1/2 is for convenience for derivative.

$$= \frac{1}{2m} \sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{n} \beta_i x_i^j \right)^2$$

[~~derive and set~~ it equal to zero for minimize

$$\frac{\partial J}{\partial \beta_k}(\beta) = \frac{1}{m} \sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{n} \beta_i x_i^j \right) \left(- x_k^j \right) = 0$$

[complex way]

* **GRADIENT DESCENT :-** [Important]

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \cdots & x_{1n} \\ 1 & x_{21} & x_{22} \cdots & x_{2n} \\ \vdots & & & \\ 1 & x_{m1} & x_{m2} \cdots & x_{mn} \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

GRADIENT OF CF

Dividing cost function in two

$$\nabla_\beta J = \frac{-1}{m}\begin{bmatrix} \sum_{j=1}^{m} y_j x_{0j} \\ \vdots \\ \sum_{j=1}^{m} y_j x_{nj} \end{bmatrix} + \frac{1}{m}\begin{bmatrix} \sum_{j=1}^{m}\sum_{i=0}^{n} \beta_i x_{ij} x_{0j} \\ \vdots \\ \cdots \ \beta_i x_{ij} x_{nj} \end{bmatrix}$$

⇒ only β is unknown in above matrix

↳ now we substitute values of β to minimize $\nabla_\beta J$

RESIDUALS

⇒ Anscombe's Quote

↳ plot residual error against true y values

(actual-y) - (predicted-y)

↳ should be random and close to normal distributed.

↳ residual should be close to zero because it indicates perfit fit

(freq plot) (scatter with y line at 0)

LINEAR REGRESSION FROM SCRATCH

MATHS

$\hookrightarrow \hat{y} = \beta_0 + \beta_1 x_1 + \ldots$

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \hat{y}_5 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 x_{11} & \beta_2 x_{12} & \ldots & \beta_m x_{1m} \\ \beta_0 & \beta_1 x_{21} & \beta_2 x_{22} & \ldots & \\ \vdots & \vdots & \vdots & & \\ & & & & \beta_m x_{nm} \end{bmatrix}$$

simplified matrix

$$= \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1m} \\ 1 & x_{21} & x_{22} & \ldots & x_{2m} \\ 1 & & & & \\ 1 & & & & \\ 1 & & & & x_{nm} \end{bmatrix}$$

$$\hat{Y} = \beta \quad X$$

(Error function) $\rightarrow E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

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

$$E = e^T e$$

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

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

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

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

finding minima to mini Error

$$\frac{\partial E}{\partial \beta} = 0 - 2y^T X + \frac{d}{\partial \beta}\left[X^T X \beta^T \beta\right] = 0$$

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

$$\cancel{2} X^T X \beta^T = \cancel{2} y^T X$$

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

Beta simplified equation $\longrightarrow$ $\boxed{\beta = yX^T (XX^T)^{-1}}$