LINEAR REGRESSION

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

$\rightarrow y = mx + b$

$\hookrightarrow$ there is room for only one feature in this eqn

* ordinary least square :-
  $\hookrightarrow$ helps us to slove for m(slope) and b (intercept) to

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

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

$\hat{y}$ denotes prediction (no perfect fit)

Beta coefficient to minimize error

MATHS behind Linear Regression for single feature

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

$\hat{\ }$ best fit

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

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

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

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

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

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

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

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

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

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

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

* 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

$$\hat{y} = \sum_{i=0}^{n} \cancel{B_0} \times \beta_i x_i + \cancel{\text{intercept}} \text{ (intercept)}$$

↳ 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

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

[complex way]

* GRADIENT DESCENT :- [Important]

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

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \qquad B = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_n \end{bmatrix}$$

GRADIENT OF CF

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

Dividing cost function in two

⇒ 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

↳ should be random and close to normal distributed.

(actual y) - (predicted y)

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

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