Simple linear Regression:-

$$y = \beta_0 + \beta_1 x + \varepsilon$$

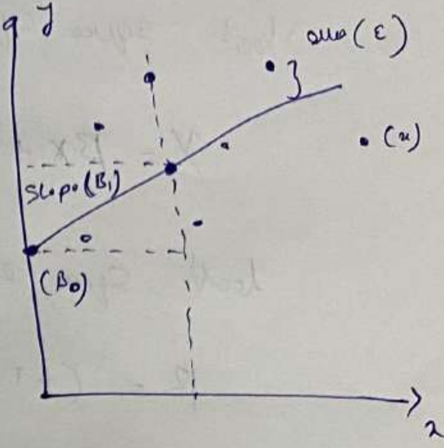
y = dependent variable

β_1 = Regression coefficient

β_0 = intercept, value of y for x=0

x = Independent variable

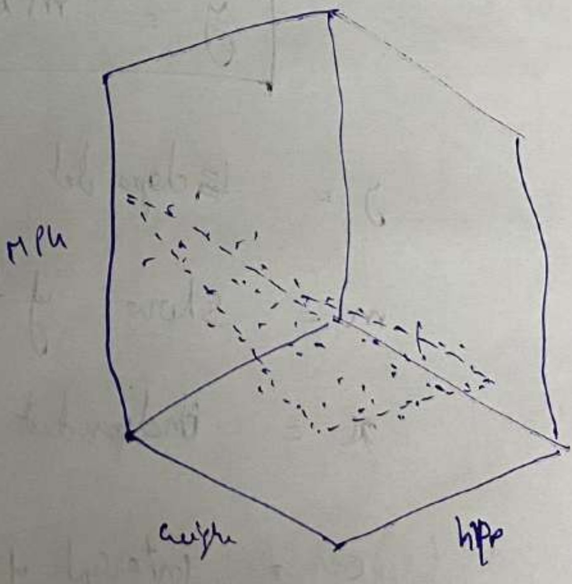
ε = error of the estimation



Multiple linear Regression

Relationship b/w two dependent variable and two or more independent variable

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n + \varepsilon$$



Assumptions:-

* Linearity
* No endogeneity
* Normality and homoscedasticity
* No autocorrelation
* No multicollinearity.

least Square linear Equation :-

$$f(x) = b + mx.$$

least square equation :-

$$y = \beta x + \varepsilon$$

least square estimator -

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

<u>Simple linear Regression</u> (head brain. csv).

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

$$\boxed{y = mx + c}$$

y = dependent variable

m = slope of the line (+ve) and (-ve)

x = Independent variable

c = intercept of y when $x = 0$.

for i in Range(m):   m = len(x)

237

numer += (x[i] - mean_x) * (y[i] - mean_y)

dnom += (x[i] - mean_x) ** 2

b1 = numer / dnom

b0 = mean_y - (b1 * mean_x)

$$\boxed{y = b_1 x + b_0}$$

$$y = 0.263(x) + 325.57$$

y ~~pred~~ =

$$\boxed{R^2 = \frac{\sum (y_p - \bar{y})^2}{1 - \sum (y - \bar{y})^2}}$$

$$\frac{\sum (y_p - \bar{y})^2}{(y - \bar{y})^2}$$

y_p = x(1, 2, 3, 4, 5)   RMSE

$y_p = b_1(1) + b_0$ Root mean square deviation.

$y_p = b_1(2) + b_0$

$$\boxed{\sqrt{\frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n}}}$$

$y_p = b_1(3) + b_0$

$\boxed{\text{Rmse} = 0}$ => decay RMSE => 0

for i in Range(m)   m = len(x)

$$\boxed{y\_pred = b_0 + b_1(x)}$$

y_pred = b0 + b1 * x[i]

Mean Square error:-

Distance between actual & predicted values

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y - y\_pred)^2}{n}}$$

Salary_Data (example)

(30, 2) Shape. (0 missing value).

$$y = mx + c$$

Step:-

① b_1

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

$$\boxed{b_0 = \bar{y} - (b_1 * \bar{x})}$$

To print b_1 and b_0.

Step ② To form a scatter plot with reg line.

$$y = b_0 + (b_1 * x)$$

plt. plot (x, y, color = 'green')

plt. scatter (x, y, color = 'blue')

plt. show()

Step (3) -

To calculate RMSE distance b/w actual - predicted.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} (y - \hat{y}_p)^2}{N}}$$

Step (4) :-

To calculate R^2 value.

$$R^2 = \frac{\sum_{i=1}^{n} (y - \hat{y}_p)^2}{\sum_{i=1}^{n} (y - \bar{y})^2}$$

~~SS_tot~~. SS_Res

~~SS_Res~~ SS_tot