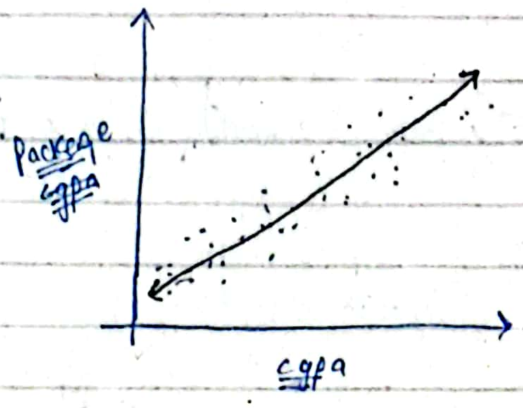
① - Linear Regression (i) Simple LR



How to find m & b value

- Closed form solution
  - (OLS → ordinary least squares)
  - (When we have a certain formulae)
  - → It is used when 1D/2D data is present (i.e cGPA)
- Non-closed form solution
  - (Approximation used)
  - (We used gradient discent)
  - * When higher dimensional data is present

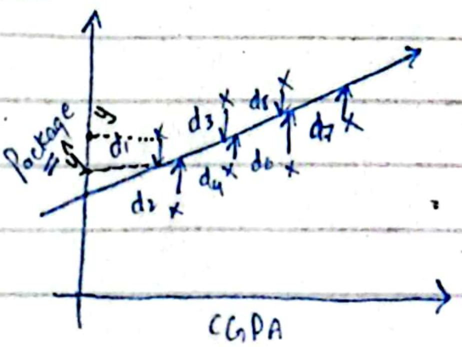
Closed form solution (OLS)

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

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

In Linear regression of Sklearn OLS is used internally.

Derivation



Error $= d_1 + d_2 + d_3 \cdots d_n$

$E = d_1^2 + d_2^2 + d_3^2 \cdots d_n^2$

$$\boxed{E = \sum_{i=1}^{n} d_i^2}$$

↓ We need a line with minimum this value our line corresponding package $\leftrightarrow$ our actual data package

J in ML

$$d_i = y_i - \hat{y}_i$$ → hat is used in ML for predictive parameters.

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

$$E(m,b) = \sum_{i=1}^{n} \left(y_i - (mx_i + b)\right)^2 \qquad \boxed{\hat{y}_i = mx_i + b}$$

(m & b are the only variable parameter's which decide the error value, other's are constant)

Initutive

m → rotation of line (weightage)

b → upward and downward movement of line

For b value

$$\frac{\partial E}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - mx_i - b)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\partial}{\partial b} (y_i - mx_i - b)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} -2(y_i - mx_i - b) = 0$$

$$\Rightarrow \sum (y_i - mx_i - b) = 0$$

$$\frac{\sum y_i}{n} - \frac{\sum mx_i}{n} - \frac{\sum b}{n} = \frac{0}{n}$$

$$\bar{y}_i - m\bar{x}_i - \frac{nb}{n} = 0$$

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

For m value

$$\frac{\partial E}{\partial m} = \frac{\partial}{\partial m} \sum_{i=1}^{n} \left(y_i - m x_i - \bar{y}_i + m\bar{x}_i\right)^2 = 0$$

$$\Rightarrow \sum 2\left(y_i - m x_i - \bar{y} + m\bar{x}\right)\left(-x_i + \bar{x}\right) = 0$$

$$\Rightarrow \sum -2\left(y_i - m x_i - \bar{y} + m\bar{x}\right)\left(x_i - \bar{x}\right) = 0$$

$$\Rightarrow \sum \left(y_i - m x_i - \bar{y} + m\bar{x}\right)\left(x_i - \bar{x}\right) = 0$$

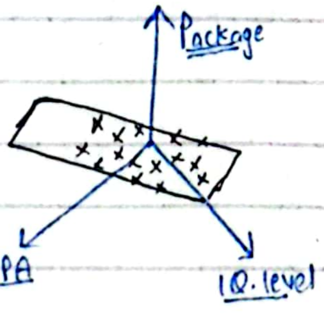
$$\Rightarrow \sum \left[\left(y_i - \bar{y}\right) - m\left(x_i - \bar{x}\right)\right]\left(x_i - \bar{x}\right) = 0$$

$$\Rightarrow \sum \left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right) = m \sum \left(x_i - \bar{x}\right)^2$$

$$\boxed{m = \frac{\sum \left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum \left(x_i - \bar{x}\right)^2}}$$

(ii) Multiple LR

i.e cGPA, IQ level, Package.



In 2D $\rightarrow$ line $\rightarrow$ $y = mx + b$

In 3D $\rightarrow$ Plane $\rightarrow$ $y = mx_1 + mx_2 + b$

$$\boxed{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

For n-dimensions $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \beta_n x_n$

$$\boxed{y = \beta_0 + \sum_{i=1}^{n} (\beta_i x_i)}$$

I.E in 3D case of our example

Package = $\beta_0 + \beta_1 \times cGPA + \beta_2 \times IQ$

* β_1, β_2 → these coefficients represents weightage of cGPA & IQ in calculating the y (package)
  i.e $\beta_1 > \beta_2 \implies$ cGPA matters more than IQ level.

* β_0 → what will the y value if X_1 and X_2 become zero (cGPA & IQ)
  ↳ called as "Offset" (Intercepted value)

How to find $\beta_0, \beta_1, \beta_2 \dots \beta_n$?

* We are interested in finding a plane (i.e 2D and hyperplane) which passes through data points as near as possible.

I.E

cGPA (x_1) / IQ level (x_2) / Gender (x_3) / Package (y) (4D data)

$$\boxed{\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}$$

4D can't be drawn graphically

let's say we have 100 students

In matrix,

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_{100} \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 X_{11} & \beta_2 X_{12} & \beta_3 X_{13} \\ \beta_0 & \beta_1 X_{21} & \beta_2 X_{22} & \beta_3 X_{23} \\ \beta_0 & \beta_1 X_{100(1)} & \beta_2 X_{100(2)} & \beta_3 X_{100(3)} \end{bmatrix} = \hat{Y}$$

Predicted package of all student

If rows → "n" columns = "m"

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 X_{11} & \beta_2 X_{12} & \beta_3 X_{13} & \cdots & \beta_m X_{1m} \\ \beta_0 & \beta_1 X_{21} & \beta_2 X_{22} & \beta_3 X_{23} & \cdots & \beta_m X_{2m} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \beta_0 & \beta_1 X_{n1} & \beta_2 X_{n2} & \beta_3 X_{n3} & \cdots & \beta_m X_{nm} \end{bmatrix}$$

$$\downarrow$$

$$\hat{Y} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1m} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots & & \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}$$

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

Coefficient matrix

Output matrix

Input matrix

$$e = y - \hat{Y}$$

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

<u>In Simple LR</u> $\longrightarrow$ $\boxed{E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$

<u>Here</u>,

$$E = e^T e \Rightarrow \begin{bmatrix} y_1 - \hat{y}_1 & y_2 - \hat{y}_2 & y_3 - \hat{y}_3 & y_n - \hat{y}_n \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ y_3 - \hat{y}_3 \\ y_n - \hat{y}_n \end{bmatrix}$$

$$E \Rightarrow (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_n - \hat{y}_n)^2$$

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

<u>Thus</u>,

$$E = e^T e \Rightarrow (Y - \hat{Y})^T (Y - \hat{Y}) \qquad \left[(A-B)^T = A^T - B^T \right]$$

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

$$= \left[Y^T - (X\beta)^T \right] \left[Y - (X\beta) \right] \qquad (\hat{Y} = X\beta)$$

$$= Y^T Y - Y^T X\beta - (X\beta)^T Y + (X\beta)^T (X\beta)$$

$$\boxed{E = Y^T Y - 2Y^T X\beta + X^T \beta^T X\beta} \qquad \left(\because Y^T X\beta = (X\beta)^T Y \right)$$

To calculate β value

$$\frac{dE}{d\beta} = \frac{d}{d\beta}\left(Y^TY - 2Y^TX\beta + \beta^TX^TX\beta\right) = 0$$

$$= 0 - 2Y^TX + 2X^TX\beta^T = 0$$

$$\cancel{2}X^TX\beta^T = \cancel{2}Y^TX$$

$$\beta^T = \frac{Y^TX}{X^TX}$$

$$\beta^T = Y^TX\,(X^TX)^{-1}$$

$$(\beta^T)^T = \left[Y^TX\,[X^TX]^{-1}\right]^T$$

$$\beta = (Y^TX)^T \bullet \left[(X^TX)^{-1}\right]^T$$

$$\boxed{\beta = X^TY \bullet (X^TX)^{-1}} \qquad \left[(X^TX)^{-1}\right]^T = (X^TX)^{-1}$$

$$(m+1)\times 1 = [(m+1 \times m+1)][(m+1)\times n][n\times 1]$$

$$= [(m+1)\times n][n\times 1]$$

$$(m+1)\times 1 = (m+1)\times 1$$

LHS RHS

Polynomial Regression

In 2D case

$X, Y \quad (1\times2)$

We will select degree for x parameter

degree $\Rightarrow$ hyperparameter (manually tuned)

i.e $2^\circ \rightarrow x^0, x^1, x^2$

$3^\circ \rightarrow x^0, x^1, x^2, x^3$

(for each row)

Equations

3D case (2^{nd} degree)

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2^1 + \beta_4 X_2^2$$

(Low degree $\longrightarrow$ underfitting

High degree $\longrightarrow$ overfitting)