Simple Linear Regression (Supervised Machine learning algorithms)

Regression :→ To establish a relationship b/w the two variables / more than two variables.

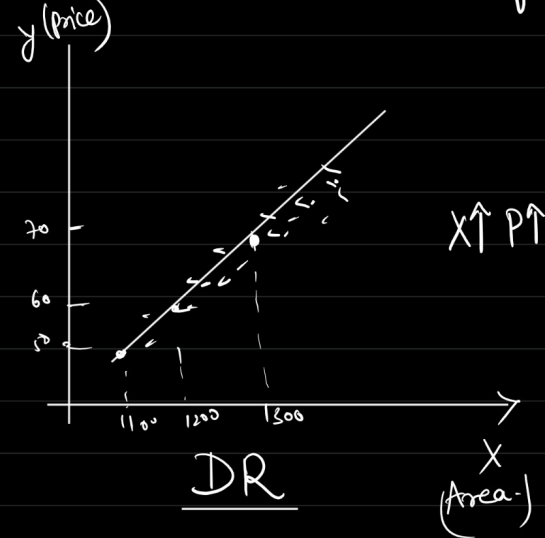
Linear → It establish a linear relationship.

Supervised ML
→ Regression (y as continuous)
Classification (discrete)

* SLR attempts to determine the strength and characteristics of the relationship between one independent variable (X) and another dependent variable (Y)

* Predict the price of a house based on Area of house

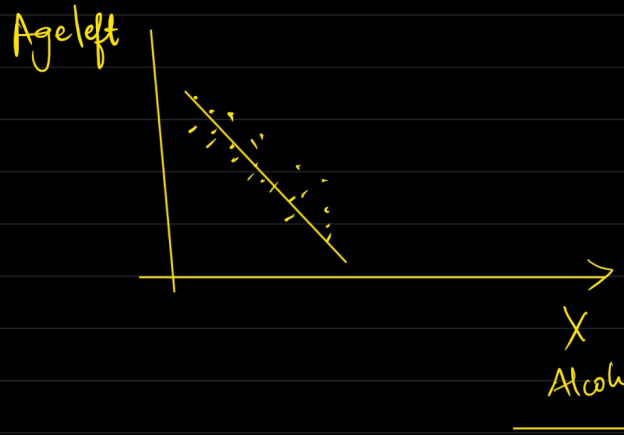
Area of house (X_1)	Price of house (y)
1100	50
1200	60
1300	70



* Selling Price of car wrt age of car

X (age of car)	y (SP)
10	3.1
8	4.5
=	=

Age left Vs Alcohol Consumption



Simple → Only one X and y

Linear → Captures the linear relationship

Regression → Understand the relationship

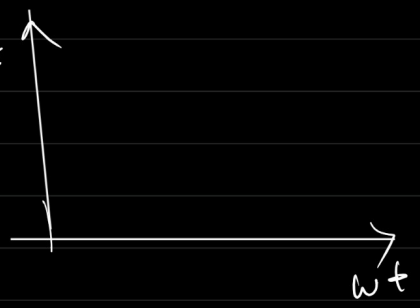
* if multiple IV (multiple X's) → Multiple linear regression.

Example

No of hours studied	Marks obtained
8	72
9	85.5
—	—
—	—
—	—

Example

Weight	Height
65	168cm
70	172cm
—	—
—	—



* To predict price of house based on number of rooms.

y = price of house

x = No of rooms.

# of rooms (x)	Price of house(y)
1	50
1.5	60
2	65
3	75
4	?? 85L

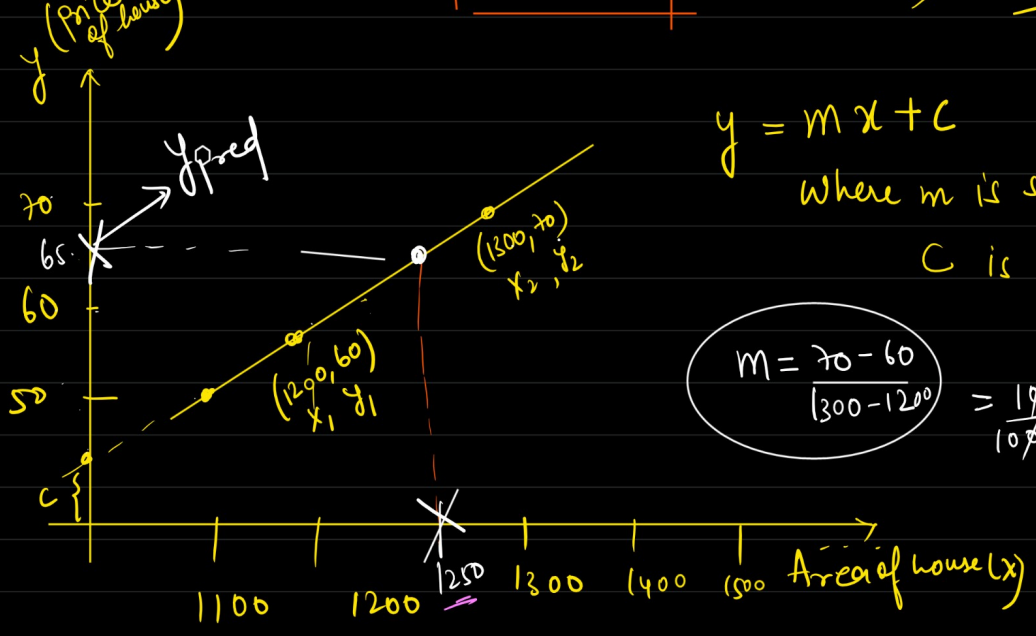
⇒ ML – To understand the mathematical relationship

→ This was 4 dP's that's why you understood the trend!

If many dP's you need to train the model to understand the mathematical relationship.

Train

New no of rooms → [model] → Price of house.



$y = mx + c$

where m is slope $\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$

c is intercept (where it cuts y axis, $x = 0$)

$$m = \frac{70 - 60}{1300 - 1200} = \frac{10}{100} = 0.1$$

$$y = 0.1x + 10$$

$$y = mx + c$$

$$y_{pred} = 0.1x + 10$$

$x = 1250$

y estimator :-

You have estimate the price of a house ← $\hat{y} = mx + c$

based on m, c, and Area of house

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

$$\hat{y} = \beta_0 + \beta_1 x \quad \left.\begin{matrix} \beta_1 = m \\ \beta_0 = c \end{matrix}\right\} \rightarrow \text{Coefficients}$$

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x}$$

↓ many books you will find this format.

hypothesis representation

↳ because we consider a SLR line to be a hypothesis and it passes from these points.

$X \rightarrow IV$

$y \rightarrow DV$

$\left.\begin{matrix} m, c \\ \beta_0, \beta_1 \\ \theta_0, \theta_1 \end{matrix}\right\} \rightarrow$ coefficients