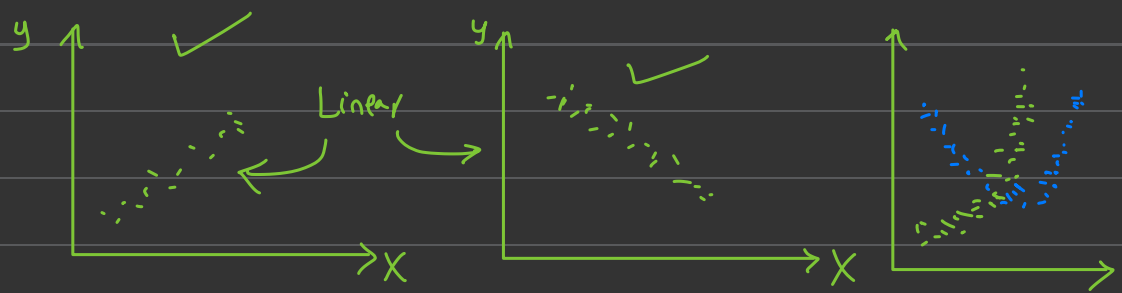
Linear Regression

X ∝ y (Linear)



Simple LR

$y \propto X$

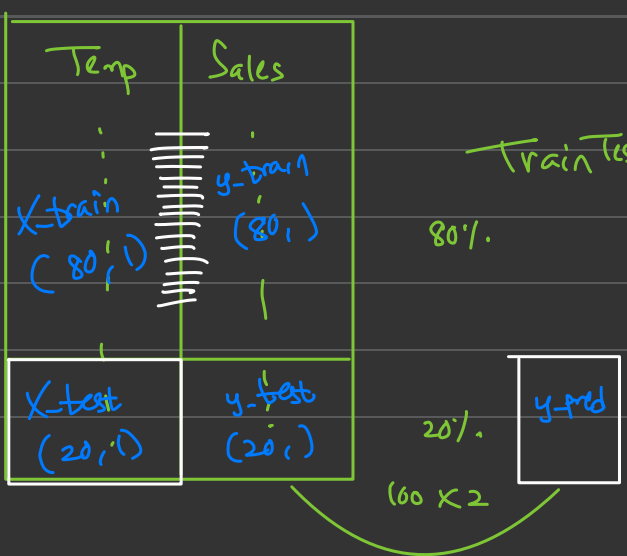
eg: Sales of ∝ Temp
icecream

Multiple LR

$y \propto X_1 + X_2 + \cdots + X_n$

eg: Price of ∝ Length + width + Body type + HP +
car
Mileage

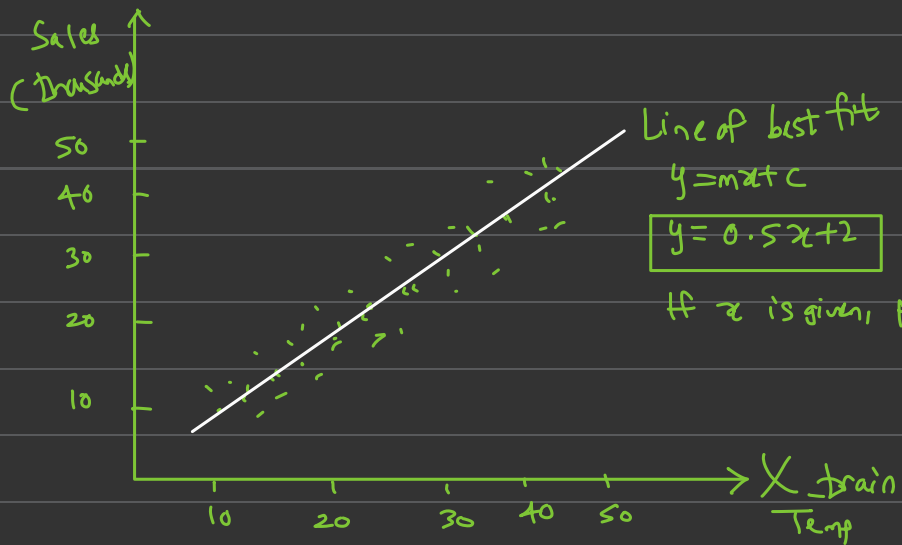
Simple LR:



Train Test Split

80%.

Comparing it & evaluating the goodness of the model.



Line of best fit

$y = mx + c$

$\boxed{y = 0.5x + 2}$

If x is given, predict y.

Accuracy YES

Sales ~ Temp (SLR)

$y \sim X \rightarrow$ 2D $\quad y = mx + c$

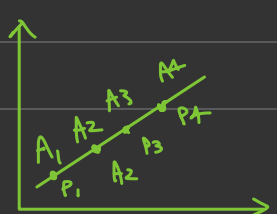
Multiple LR

$y \sim X_1 + X_2 \rightarrow$ 3D $\quad y = m_1x_1 + m_2x_2 + c$

$y \sim X_1 + X_2 + X_3 \rightarrow$ 4D $\quad y = m_1x_1 + m_2x_2 + m_3x_3 + c$

$y \sim X_1 + \cdots + X_{20} \rightarrow$ 21D

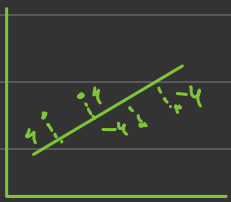
$y \sim X_1 + \cdots + X_n \rightarrow (n+1)$D $\quad \boxed{y = m_1x_1 + m_2x_2 + \cdots + m_nx_n + c}$



$$\text{Mean error} = \sum_{i=1}^{N} (A_i - P_i)$$

$$= \frac{(A_1 - P_1) + (A_2 - P_2) + (A_3 - P_3) + (A_4 - P_4)}{4}$$

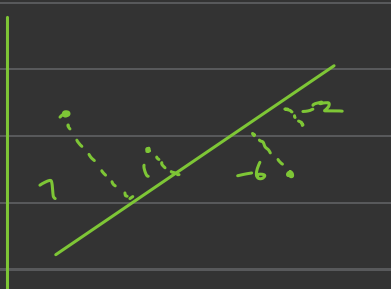
$$= \frac{0+0+0+0}{4} = 0$$



$$ME = \frac{4+4+(-4)+(-4)}{4} = \frac{0}{4} = 0$$

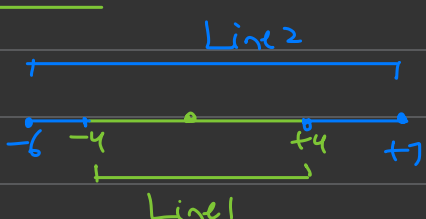
$$MAE = \sum_{i=1}^{N} \frac{|A_i - P_i|}{N}$$

$$= \frac{|4|+|4|+|-4|+|-4|}{4} = \frac{16}{4} = 4$$



$$MAE = \frac{|7|+|1|+|-6|+|-2|}{4}$$

$$= \frac{16}{4} = 4$$



$$MSE = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{N}$$

(1) $= \frac{4^2 + 4^2 + (-4)^2 + (-4)^2}{4} = 16$

(2) $= \frac{7^2 + 1^2 + (-6)^2 + (-2)^2}{4} = 22.5$

RMSE $= \sqrt{MSE}$

(1) $\sqrt{16} = 4$

(2) $\sqrt{22.5} = 4.74$

RMSE is best metric. $\rightarrow$ Regression Pb.

$R^2 \rightarrow [0 - 1]$ $\rightarrow$ higher the better.

0.8 ↑

Adjusted R^2 - $[0 - 1]$