

Central Limit Theorem

10K

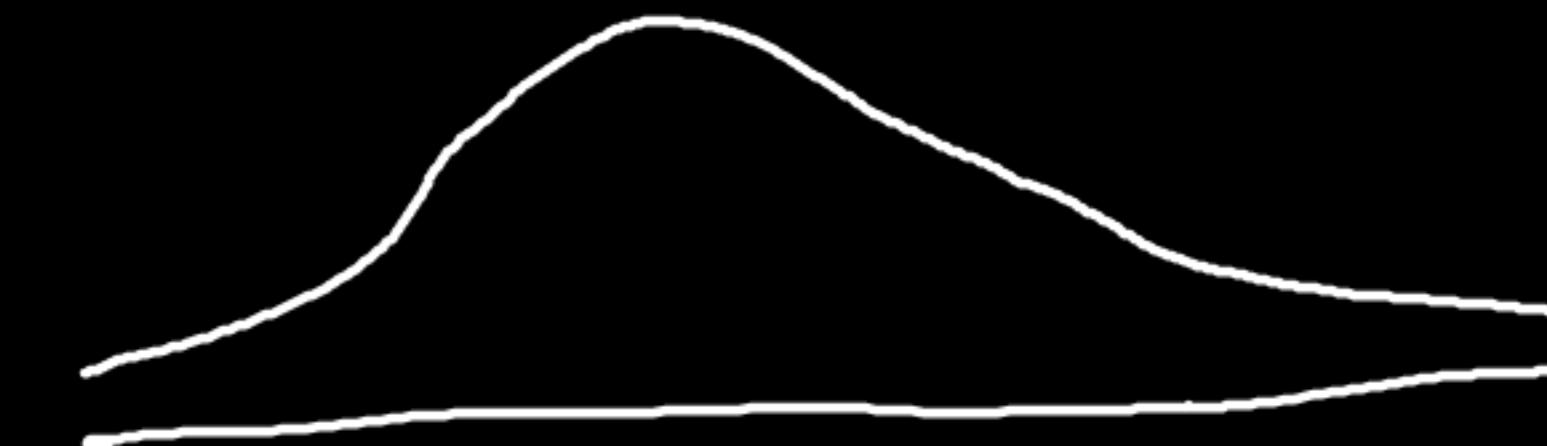
$$\underline{x} \neq G \cdot D / N \cdot D$$

sample size (30 or more than 30)



$$\bar{x}_1 \bar{x}_2 \bar{x}_3$$

$$\bar{x}_{100}$$



100 times

Mean

$$x_1, x_2, x_3, \dots, x_{100} = \bar{x}_1$$

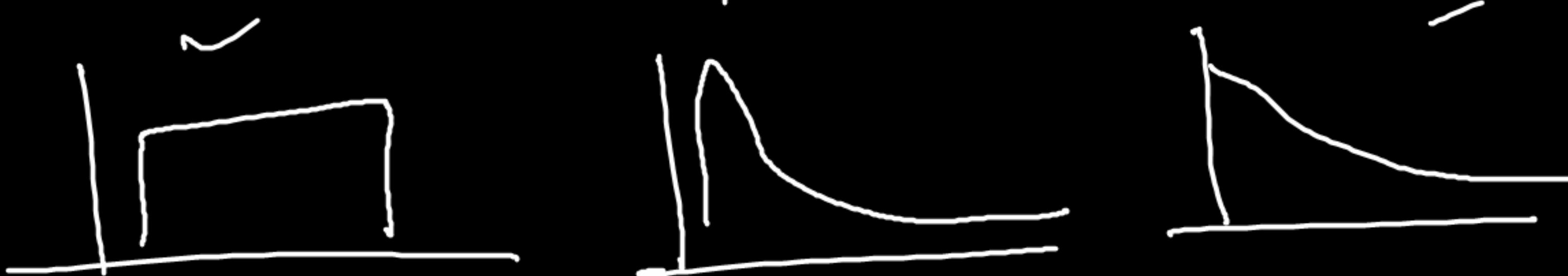
$$x_1, x_3, x_7, x_{88}, x_{100} = \bar{x}_2$$

⋮

$$x_{100}, x_{101}, x_{202}, \dots, x_{901} = \bar{x}_{100}$$

Editing

Raw data (x) = 50K Population \neq Normal Dist
Uniform



Empirical Rule
 μ, σ^2

1000 times Sample 30 dataset mean

$$\begin{aligned} x_1 &= \cancel{x_1} x_2 \cancel{x_3} \\ x_2 &= \dots \\ \vdots & \\ \underbrace{x_n}_{1000} &= \dots \end{aligned}$$

$$x_{30} = \overline{x}_1$$

$$\dots = \overline{x}_2$$

$$\dots = \overline{x}_{1000}$$

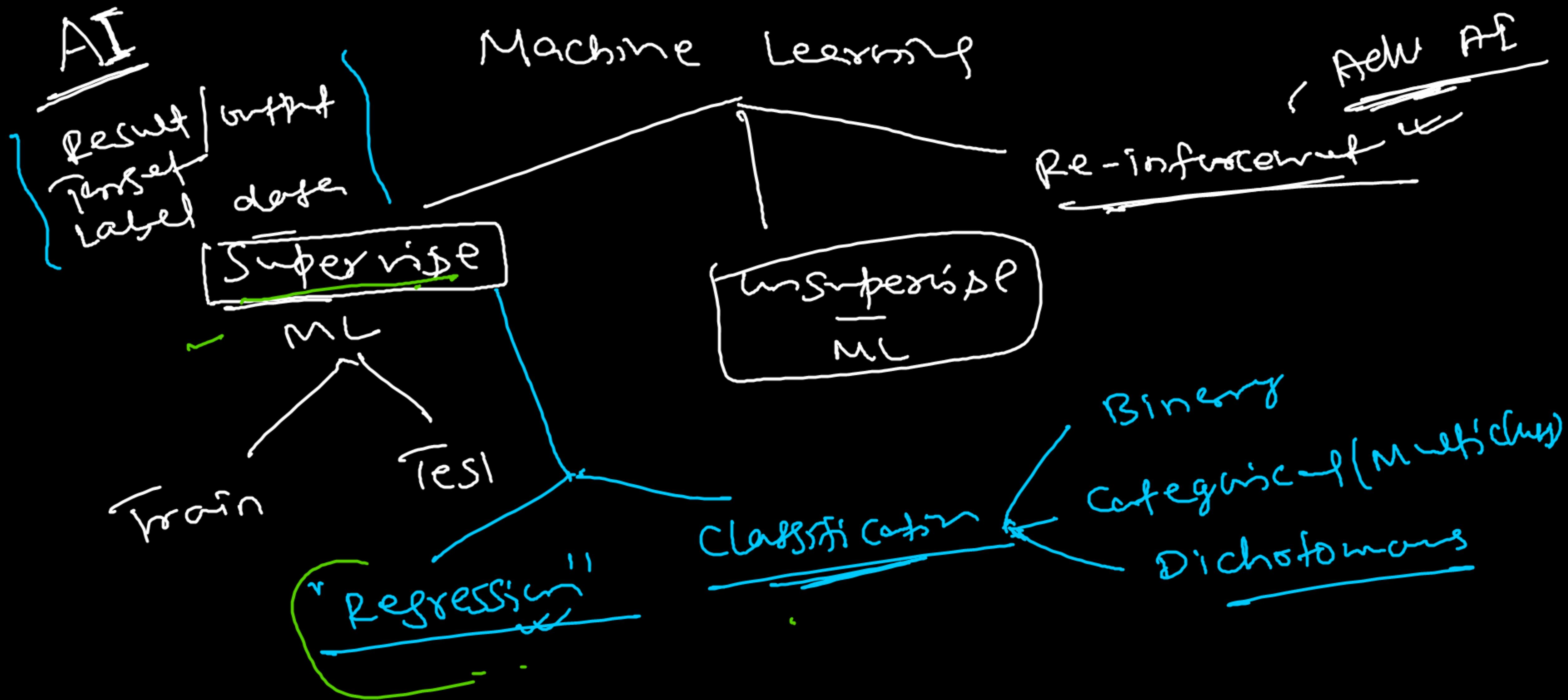


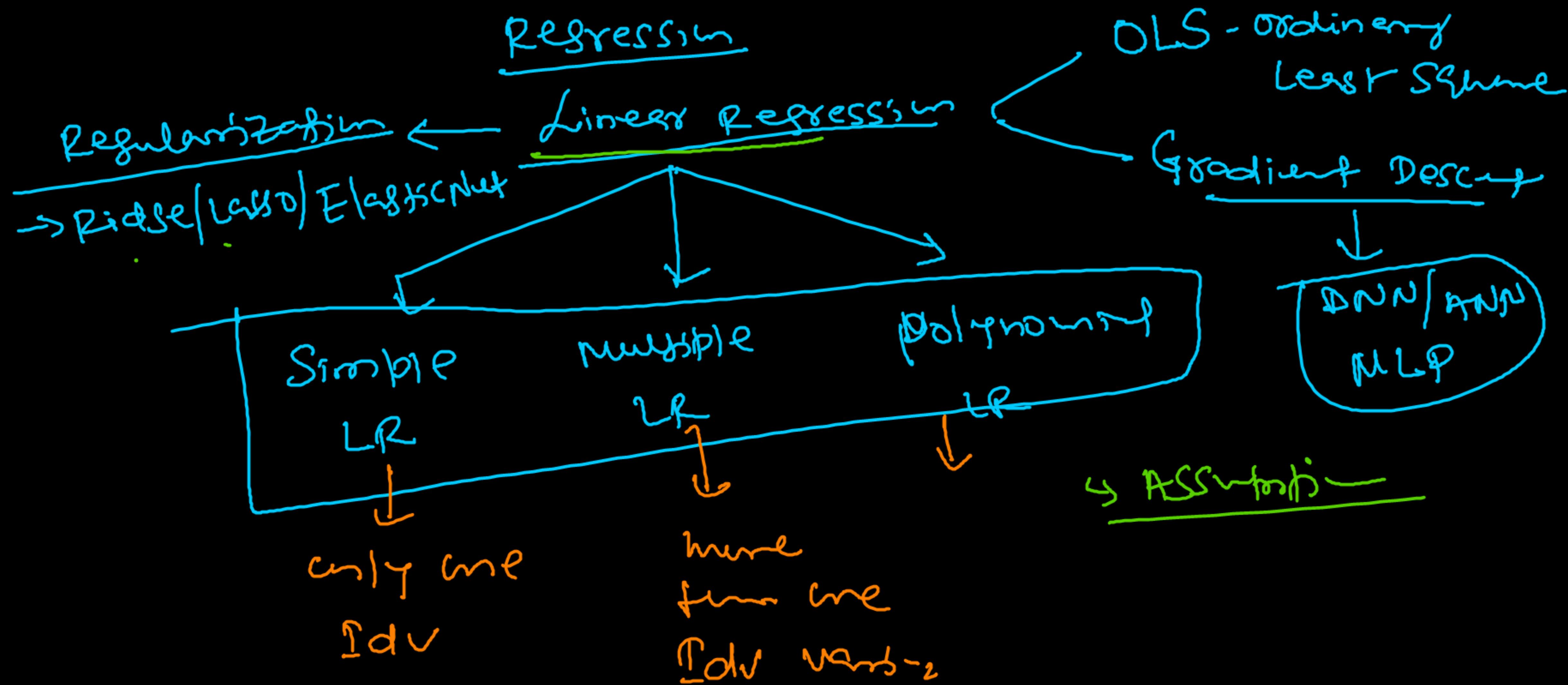
$$I_2 = 3\sigma \pm \mu$$

Sanjoy +1 other raised hands View x

Condition - CLT

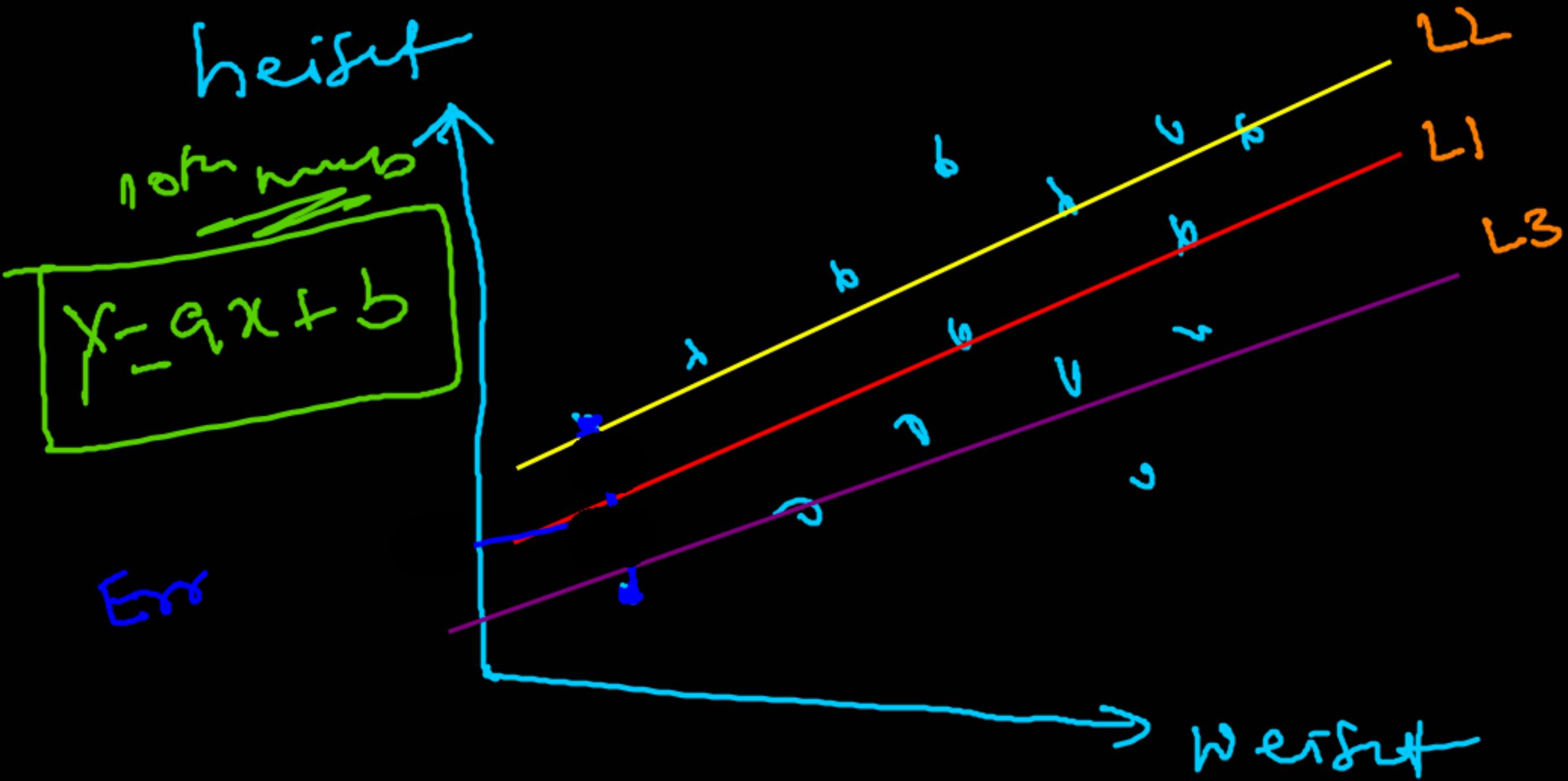
- ① The Sample size is large enough, typically greater than or equal to 30.
- ② The Sample is drawn from a finite population or an infinite poplⁿ. with a finite Variance.
- ③ The random variable in the Sample are independent & identically distributed.





Regression Analysis

Dependent variable &
one or more than one Poly var



Predict height $\in \mathbb{R}$

Given weight, colour, size, ethnicity, Gender

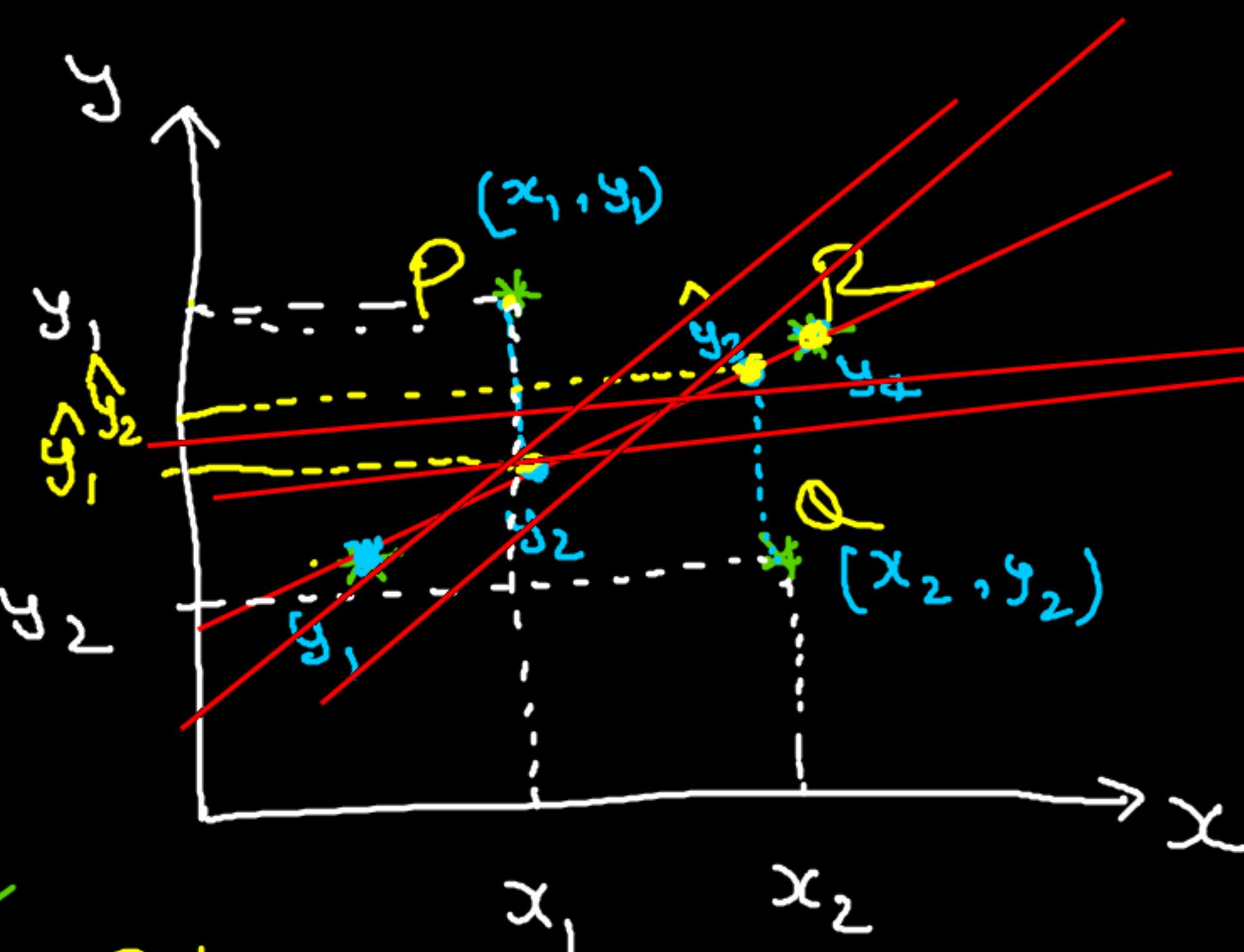
Linear Reg :- Find a line/plane that

fits the given data point

& we can get less error (best fit line)



find a line/plane that best fits for the dataset



best fit

$$\min \sum_{i=1}^n \text{error}_i^2$$

$y_1 = \text{zero error (Residual)}$

$y_4 = \text{zero error (Actual - Predict)}$

$$P = (\text{Actual} - \text{Predicted})^2$$

$$P = ((y_1 - \hat{y}_1))^2 = (+ve)^2 = +ve$$

$$Q = ((y_2 - \hat{y}_2))^2 = (-ve)^2 = +ve$$

$$R = ((y_4 - \hat{y}_4))^2 = (0)^2 = 0$$

Reason :-

You are screen sharing Stop Share

