

Machine Learning

- Understand the rules/patterns from the data

Predictors: Iris Dataset				(y) ✓	
SL	SW	PL	PW		
Features/Independent Variables				Species dependent - Predicted Value	
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Versicolor
3	4.7	3.2	1.3	0.2	Setosa
4	4.6	3.1	1.5	0.2	Setosa
5	5.0	3.6	1.4	0.2	Setosa
6	5.4	4.5	1.5	0.4	Setosa
7	4.5	3.9	1.7	0.4	Setosa
8	4.9	3.7	1.5	0.4	Setosa
9	4.8	3.4	1.6	0.4	Setosa
10	4.8	3.0	1.6	0.4	Setosa
11	5.1	3.8	1.9	0.4	Setosa
12	5.7	4.4	1.5	0.4	Setosa
13	5.1	4.0	1.8	0.4	Setosa
14	5.4	4.5	1.6	0.4	Setosa
15	5.1	4.0	1.7	0.4	Setosa
16	4.6	3.6	1.4	0.4	Setosa
17	5.0	3.4	1.6	0.4	Setosa
18	5.0	3.4	1.6	0.4	Setosa
19	5.2	3.7	1.5	0.4	Setosa
20	5.5	4.2	1.5	0.4	Setosa
21	4.9	3.5	1.4	0.4	Setosa
22	4.8	3.2	1.6	0.4	Setosa
23	5.1	3.8	1.6	0.4	Setosa
24	5.0	3.3	1.4	0.4	Setosa
25	5.4	3.9	1.5	0.4	Setosa
26	5.1	3.5	1.4	0.4	Setosa
27	4.8	3.2	1.3	0.4	Setosa
28	5.0	3.6	1.5	0.4	Setosa
29	5.1	3.4	1.5	0.4	Setosa
30	5.0	3.4	1.5	0.4	Setosa
31	5.2	3.7	1.5	0.4	Setosa
32	5.4	4.0	1.5	0.4	Setosa
33	5.1	3.5	1.4	0.4	Setosa
34	5.0	3.3	1.3	0.4	Setosa
35	5.4	3.9	1.5	0.4	Setosa
36	5.1	3.5	1.4	0.4	Setosa
37	5.0	3.3	1.3	0.4	Setosa
38	5.2	3.6	1.5	0.4	Setosa
39	5.1	3.4	1.5	0.4	Setosa
40	5.0	3.3	1.5	0.4	Setosa

 $n=100$

(x, y) $\xrightarrow{\text{Training the model}}$ $\boxed{\text{Model}}$

$\xrightarrow{\text{Arrives at}}$

$\begin{cases} \text{If } PL < 4.0 \text{ and } SL \geq 3.1 \\ \quad \text{Species} = \text{setosa} \\ \text{if } PL < 3.5 \text{ and } SL \geq 4.3 \\ \quad \text{Species} = \text{versicolor} \end{cases}$

$\Rightarrow (\text{Versicolor})$

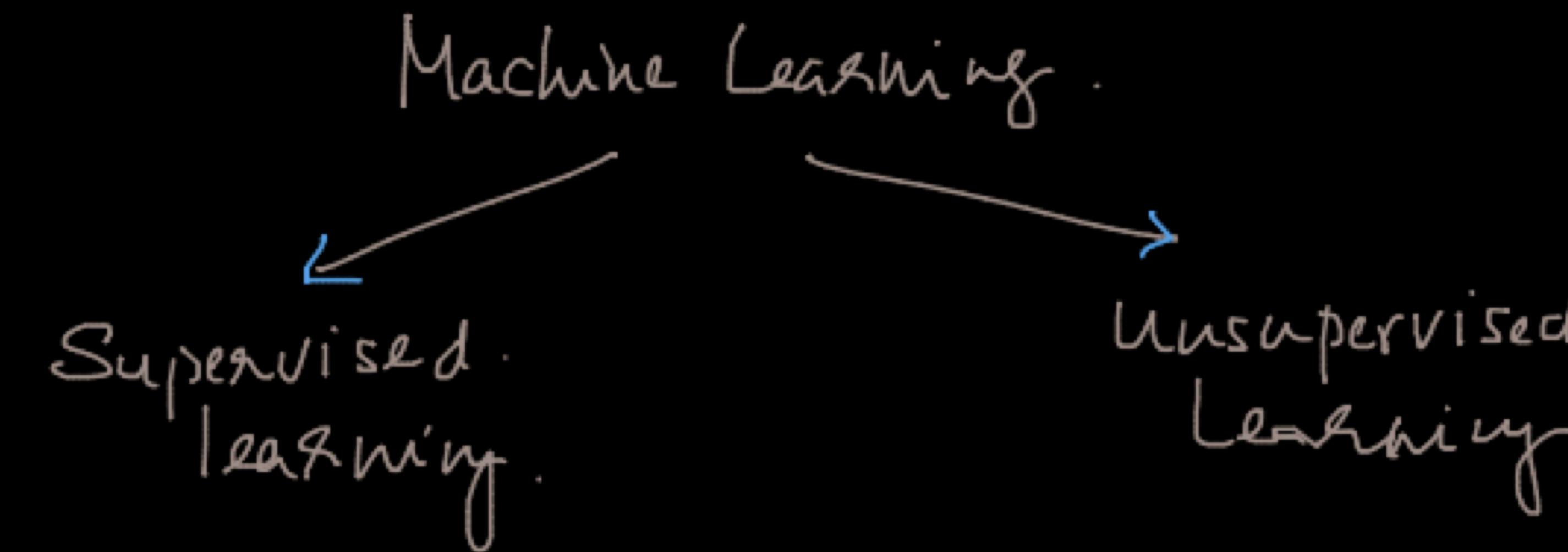


$\rightarrow 3 \text{ Species}$

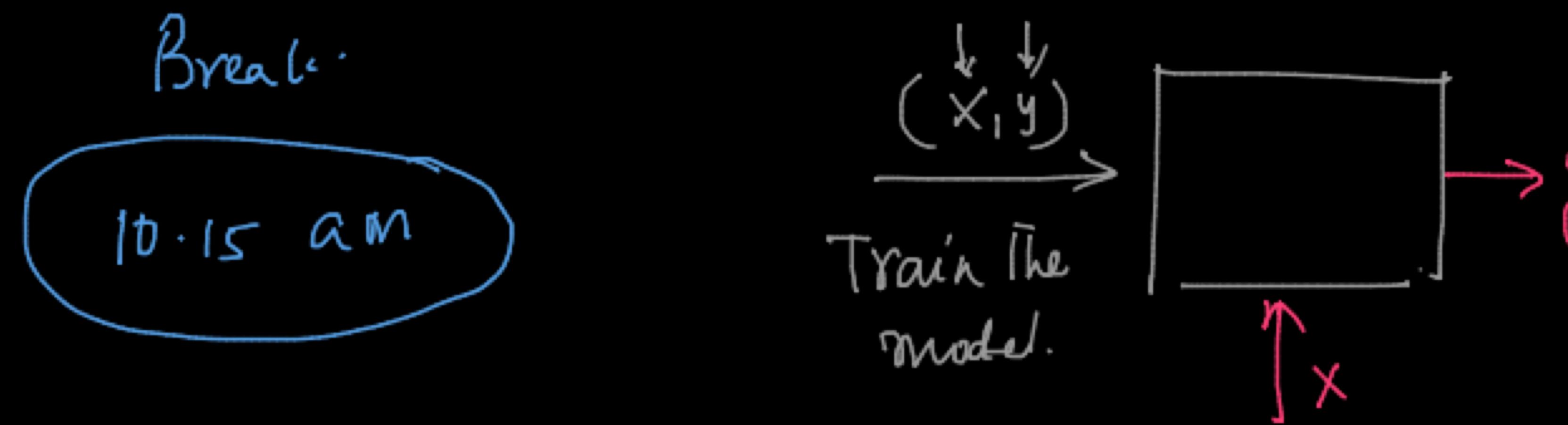
$\rightarrow \text{Setosa, Virginica, Versicolor.}$

$\rightarrow PL = 4.0 \quad SL = 3.0$

$PW = 2.5 \quad SW = 1.2$



— train the model by -
showing both x & y .



- Model is trained by showing only x . Values
- No predictions can be made.
- Grouping / Clustering
- Customer Segmentation

1. Regression Models.
 - y is continuous.
2. Classification Models.
 - y is categorical.

c_1	-	-	-	-	-
c_2	-	-	-	-	-
c_3	-	-	-	-	-
c_4	-	-	-	-	-

— clustering models.





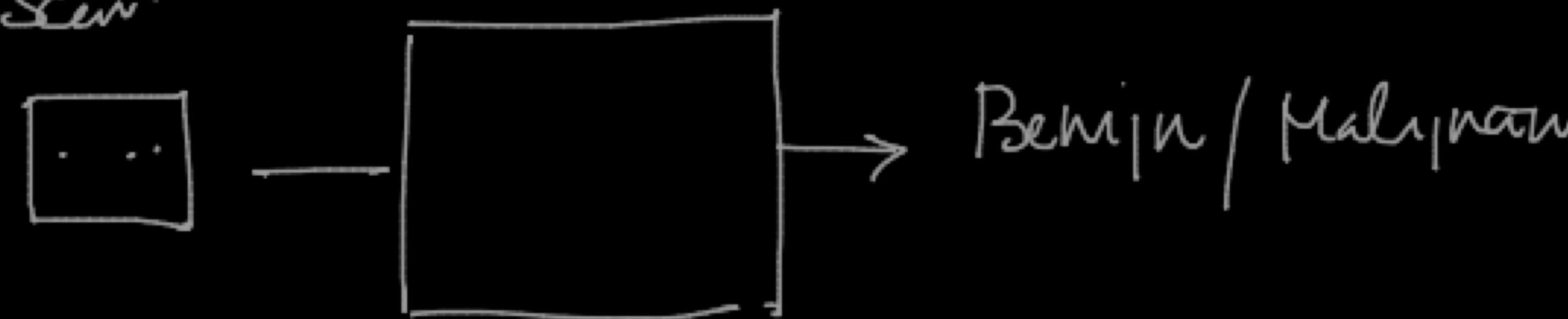
•



3 / 5

- More data means better model
 - Model is only as good as your

Sec.



A graph showing the relationship between draw force (F_A) and shear stress (τ). The x-axis is labeled F_A and the y-axis is labeled τ . Four points are plotted at (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , and (x_4, y_4) . A straight line is drawn through the first three points, and a dashed line extends it to the fourth point. A vertical line is drawn from the fourth point to the x-axis, labeled μ . A circle labeled B is drawn around the third point (x_3, y_3) .

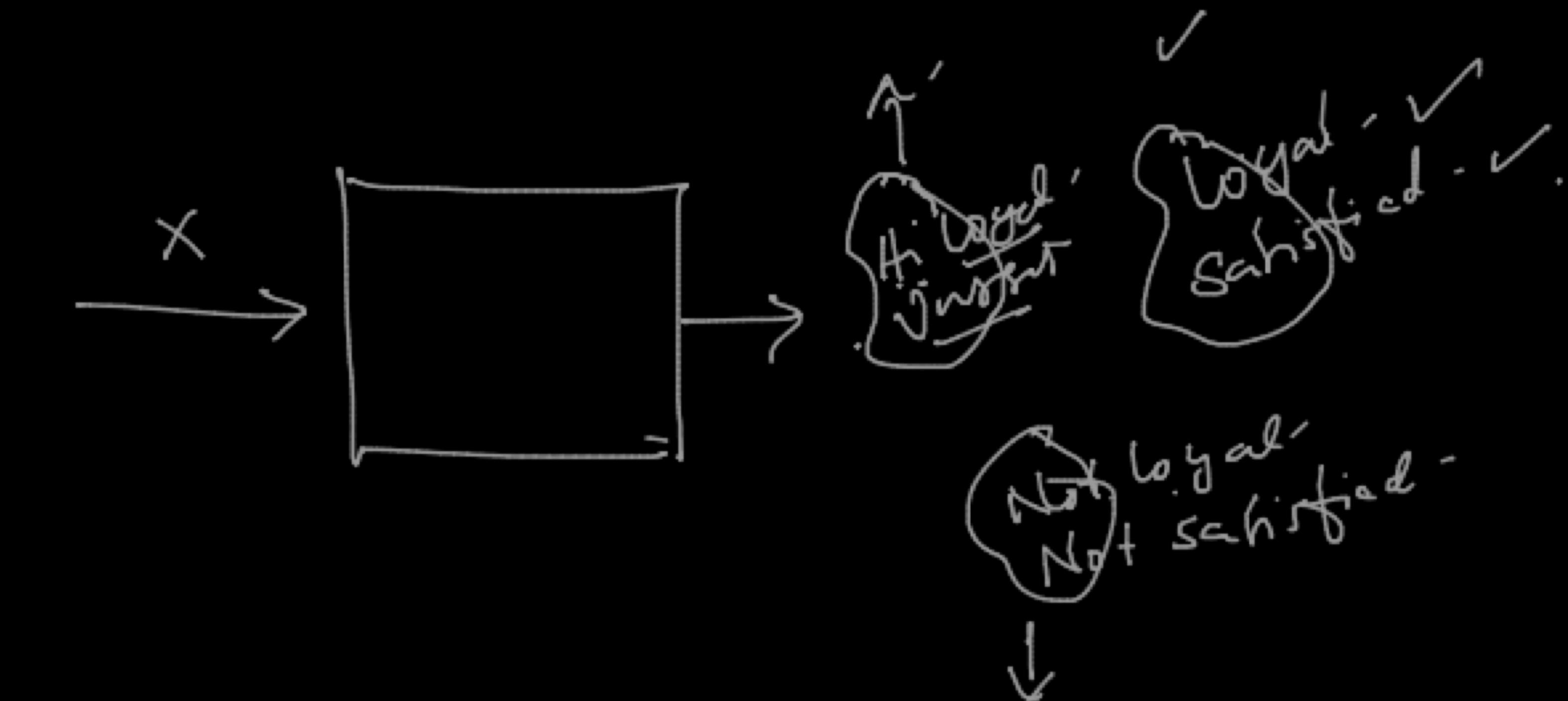
Lashed

Antra.

Unsupervised Learning

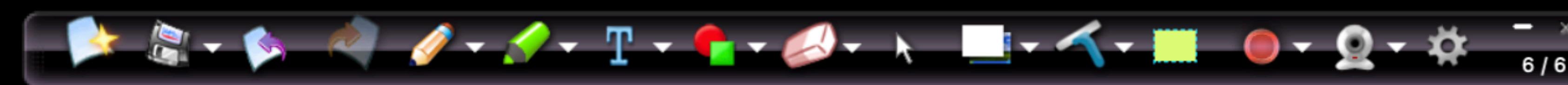
A 2x2 matrix diagram for the F1-F2 model. The vertical axis is labeled Loyalty (F1) and the horizontal axis is labeled Satisfaction (F2).

Loyalty (F1)	Satisfaction (F2)
Loyalty (F1)	Satisfaction (F2)
Satisfaction (F1)	Satisfaction (F2)
Satisfaction (F1)	Loyalty (F2)





•



6 / 6

Supervised Learning -

Regression

Classification -

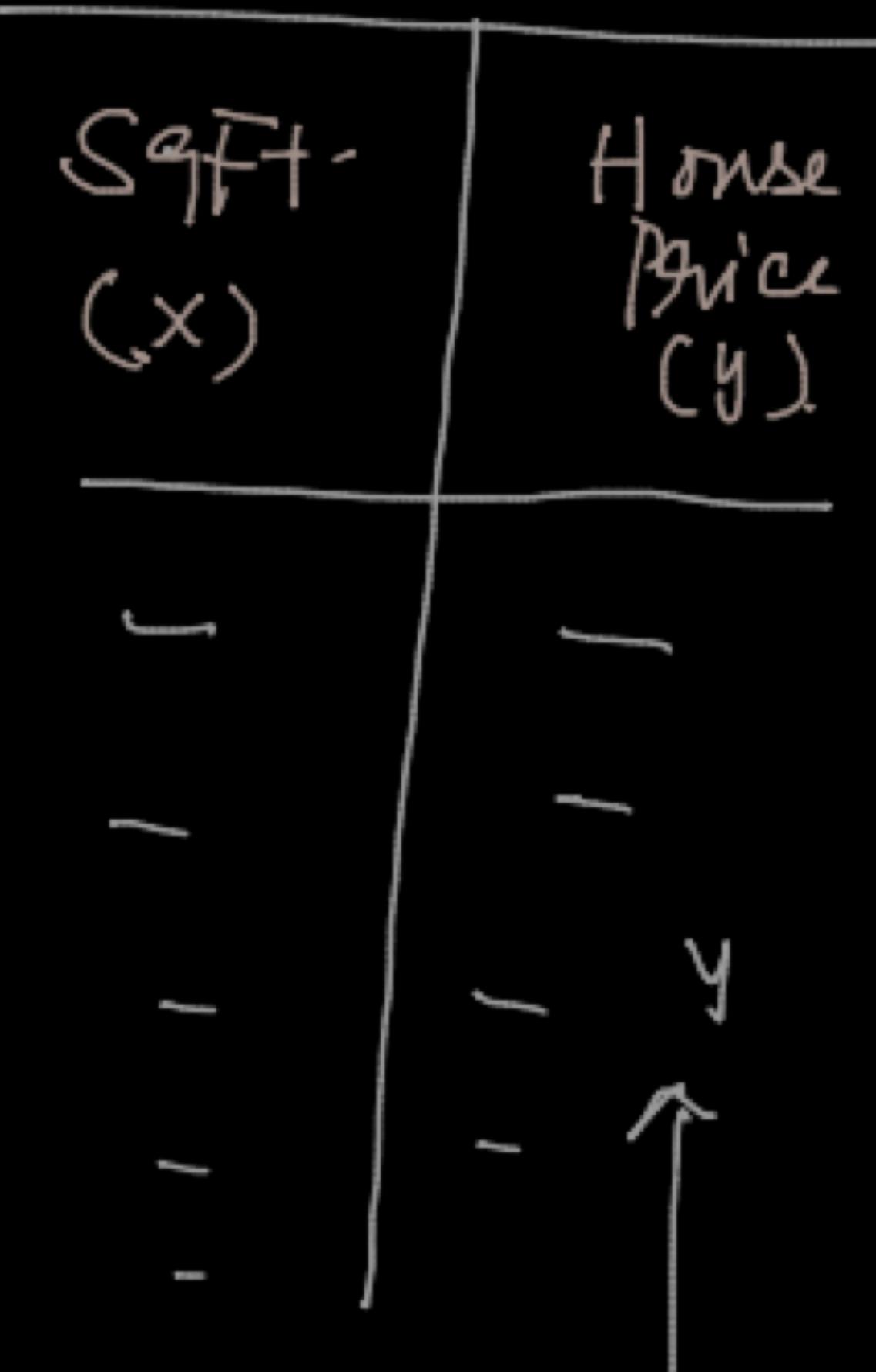
→ Predicts .
Continuous data .

- Predicting House price .
- Stock price .
- Marks

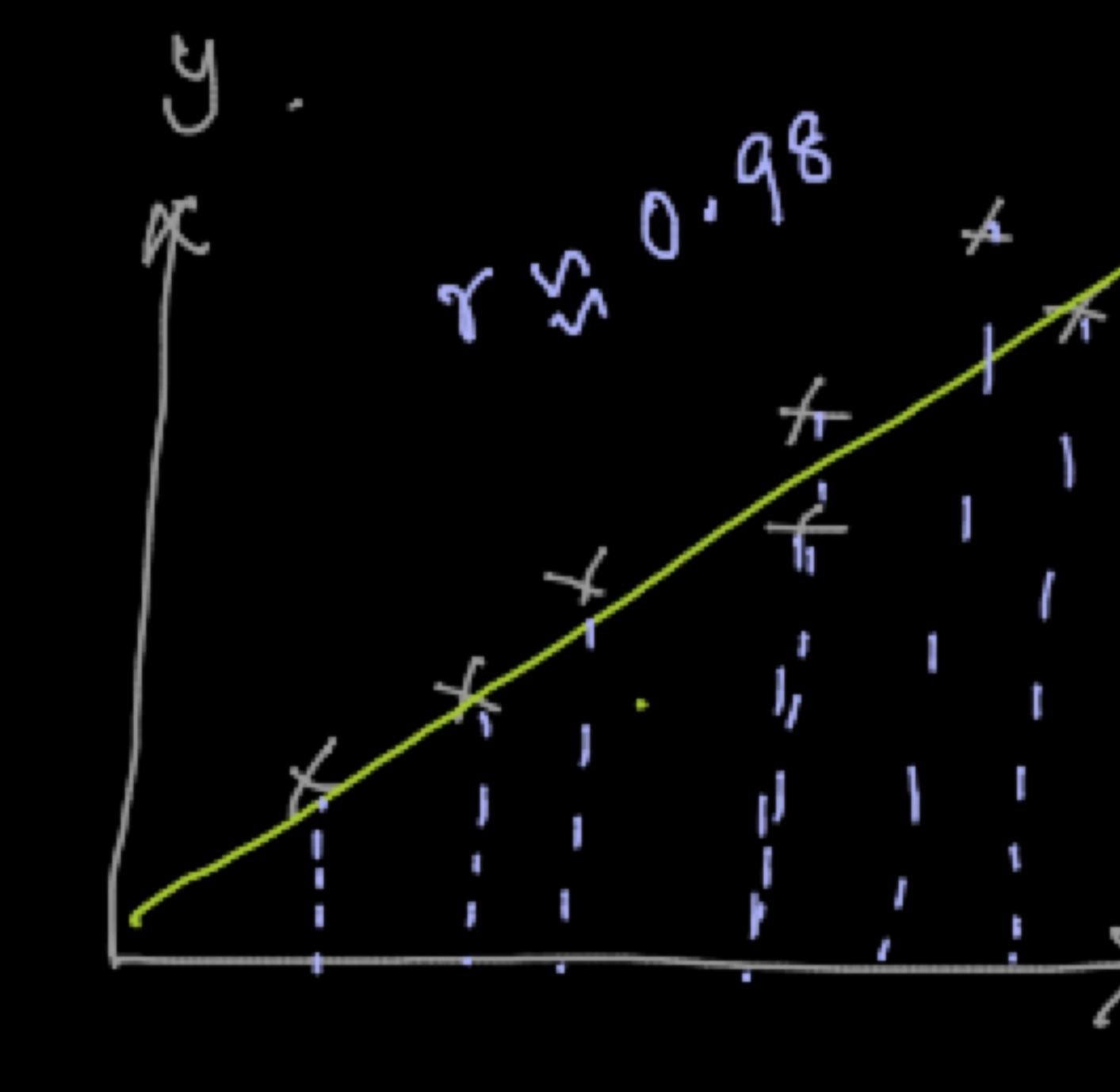
- Predicts .
Categorical Values .
- Tumor Malignant / Benign
- Default / No default
- Has / No Heart
- Cat / Dog / Horse .

Simple Linear Regression

one feature \leftarrow \downarrow predict y
to predict linear relationship between x & y which is continuous.

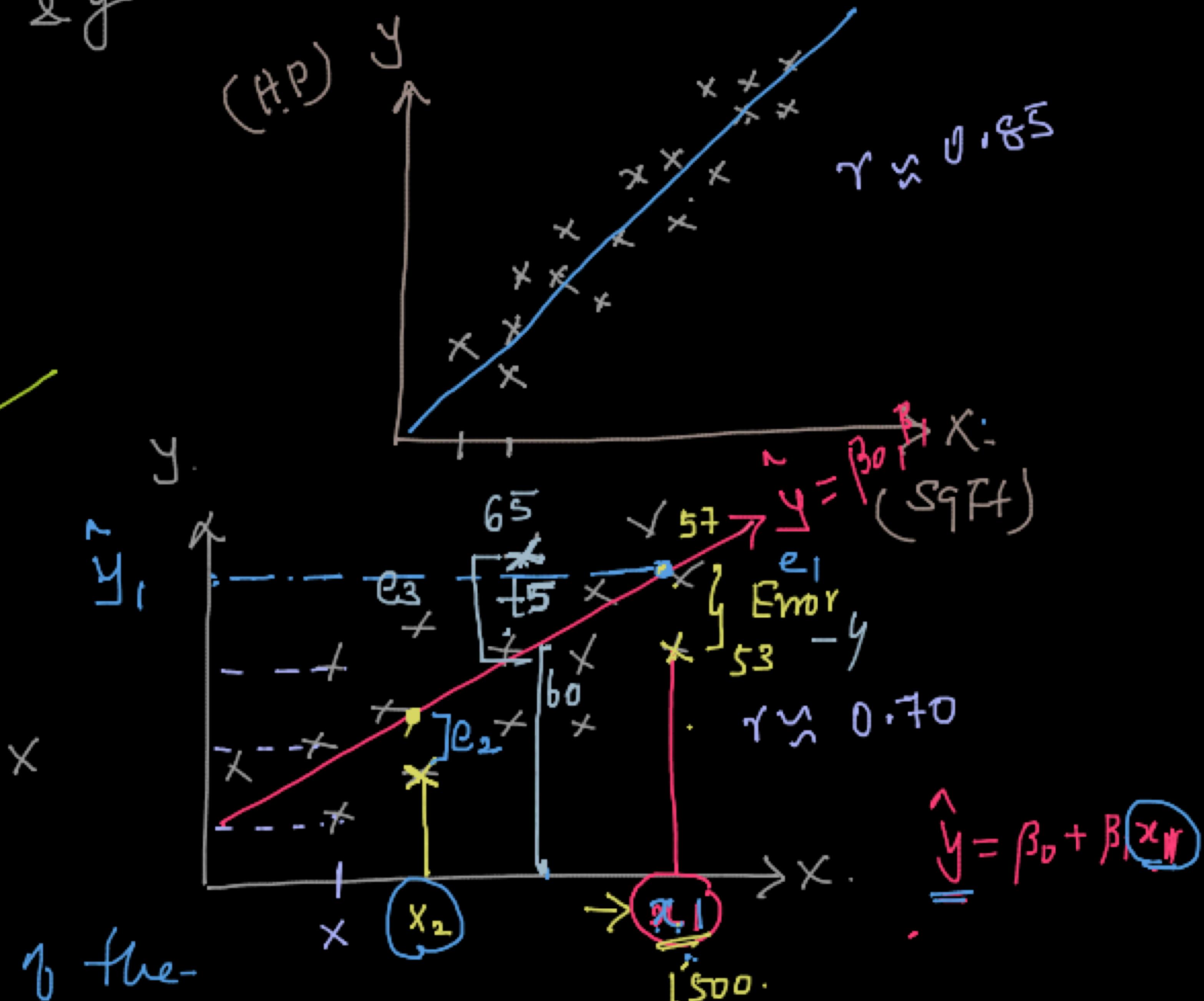


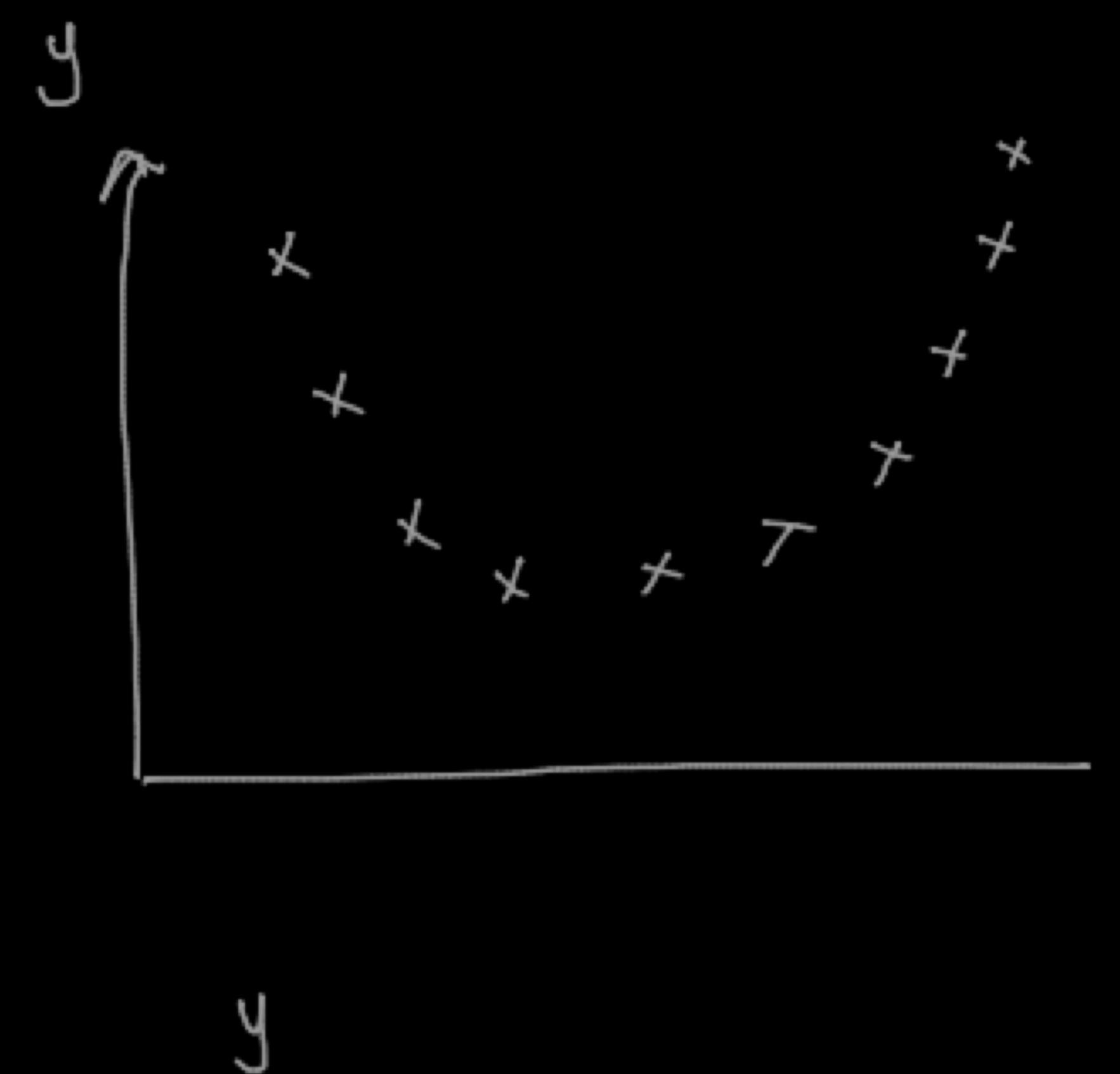
1. Scatter plot



2. Quantify the strength of the relationship between x & y .

Pearson's Correlation Coefficient \rightarrow Strength of the LINEAR relationship between x & y .



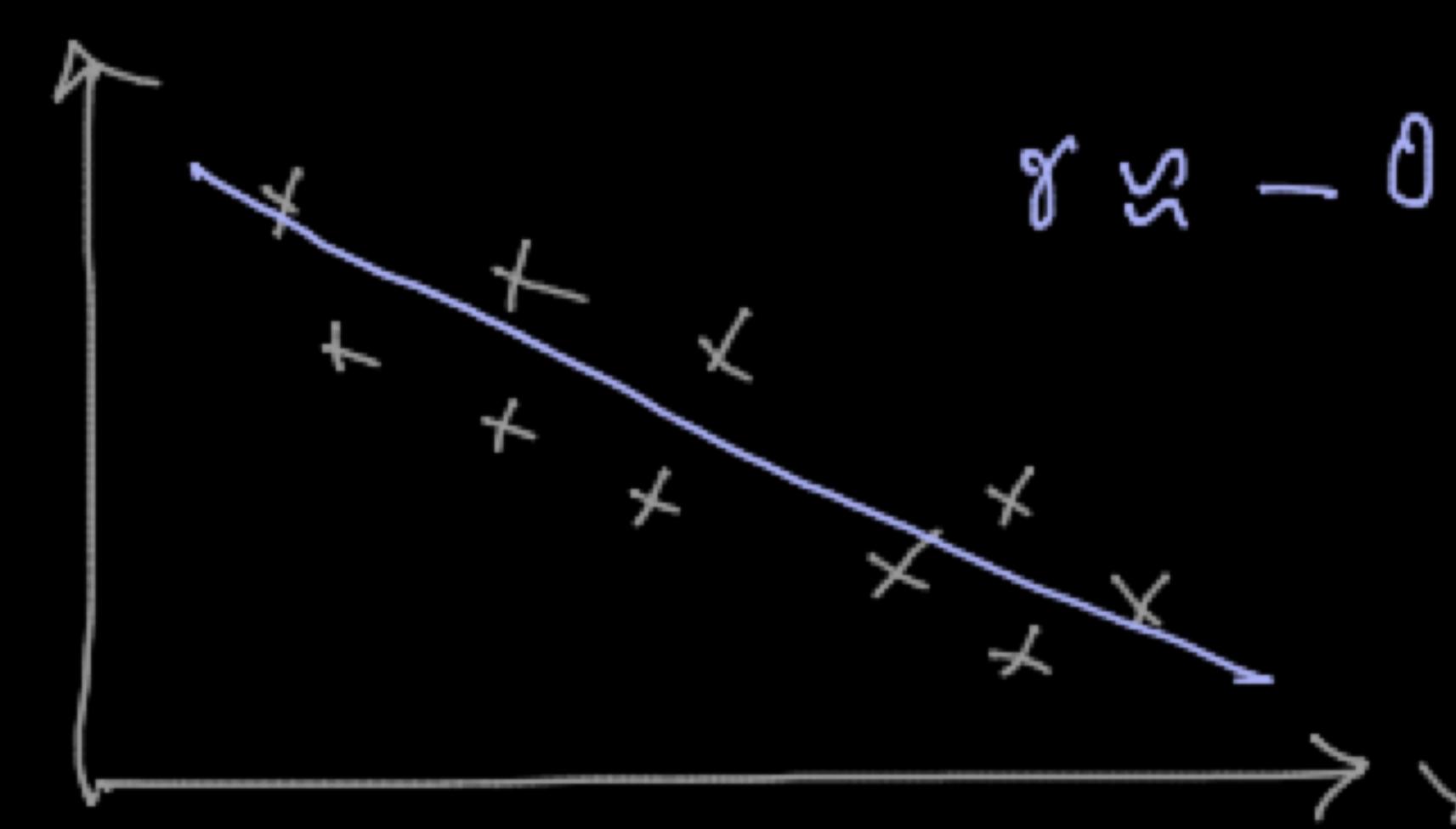


$\gamma \rightarrow$ Cannot predict this relationship.

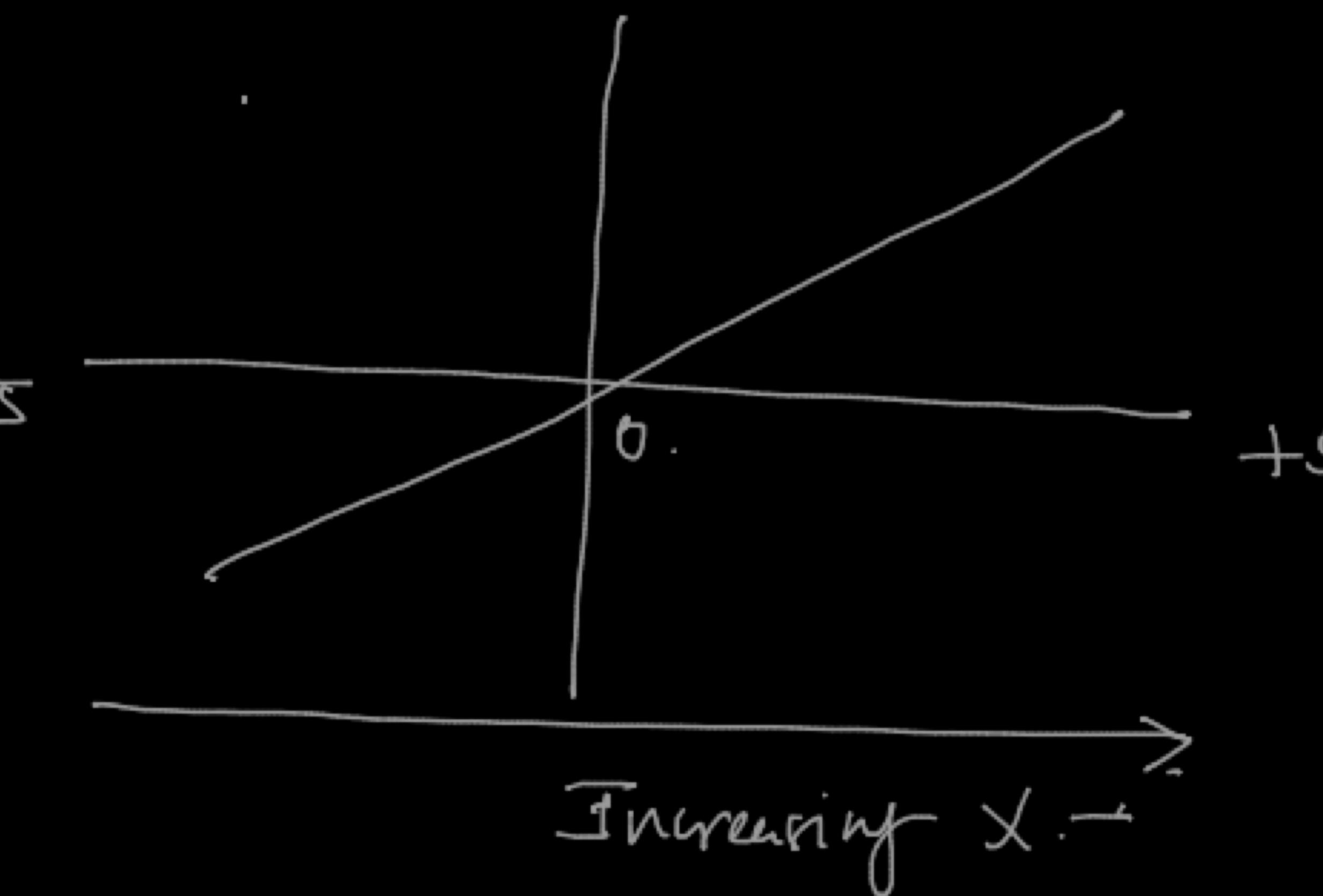
$$\boxed{\gamma \geq \pm 0.85}$$

$$\boxed{\gamma = -1 \text{ to } +1}$$

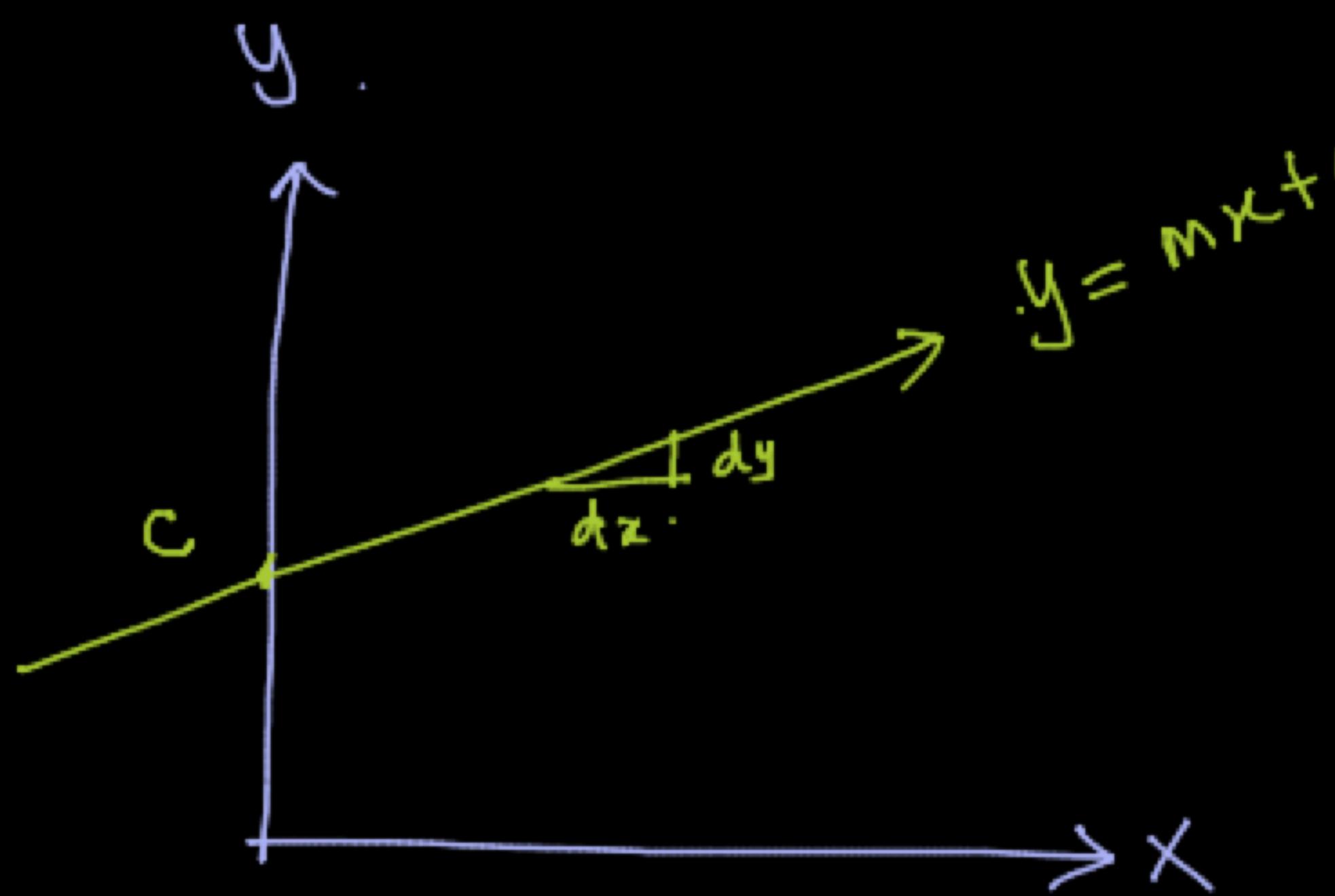
↳ We can build a good linear model



$$\gamma \approx -0.92$$



Increasing $x \rightarrow$



What is a Model.

$$\boxed{y = f(x)} \rightarrow \text{Model}$$

Parameters
Hyper params
— We decide before building the model

$y = mx + c$ → Linear Model
— Estimated by the model during training

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

$\beta_0, \beta_1 \rightarrow \text{Model Parameters.}$

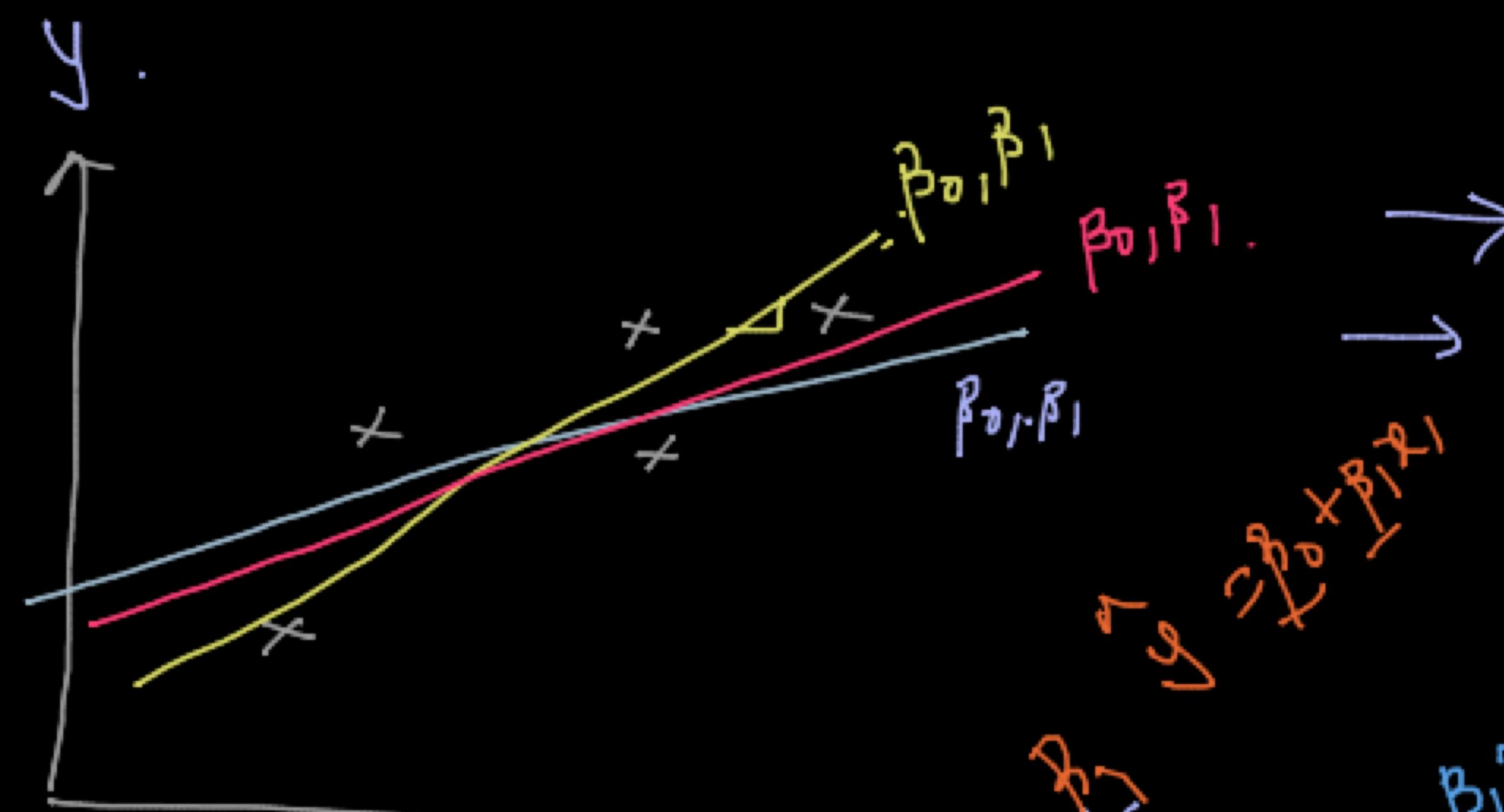
Machine Learning.

- Machine learns the model parameters from the data.

— Gradient Descent
→ Ordinary least squares (OLS).

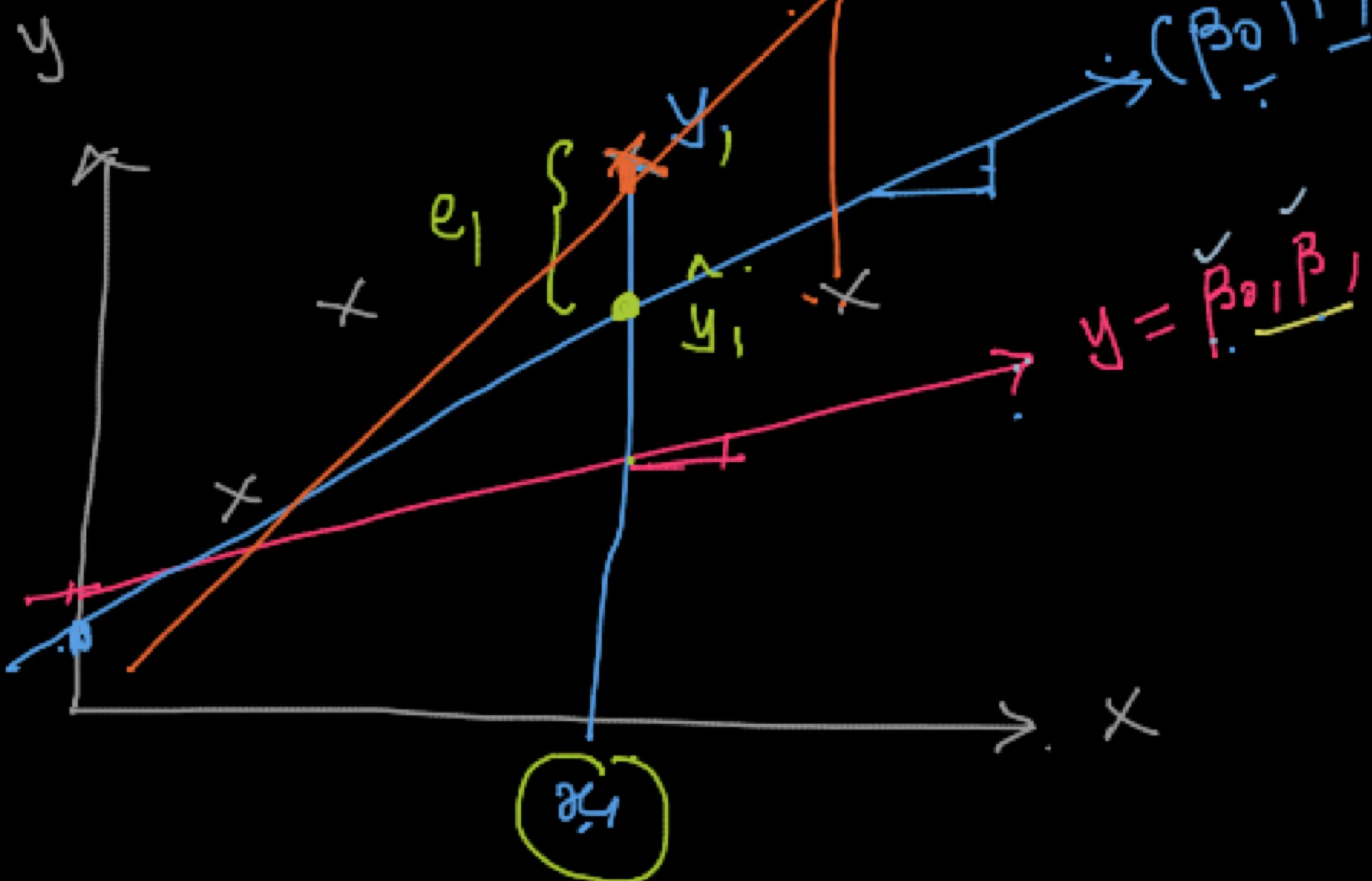
Training the model

- Model is learning the β 's (model parameters) from the data.



→ OLS.

→ Which is the best β_0 & β_1
(best line)?



Min SSE

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

↳ Predicted
Value of

y
Sum $\sum e_i^2 = e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2$
Squared
Errors
(SSE)

$$SSE = e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2 \quad \beta_0, \beta_1$$

$$SSE = e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2 \quad \beta_0, \beta_1$$

$$\begin{aligned} \hat{y} &= (\beta_0 + \beta_1)x, \text{ Form} \\ \hat{y}_1 &= (y_1 - e_1) + \hat{y}_1 \\ \hat{y}_2 &= (y_2 - e_2) + \hat{y}_2 \\ e_3 &= y_3 \end{aligned}$$

x_1 x_2 x_3 x_n

y_1 y_2 y_3 y_n

Task: Find those β_0, β_1 s.t. the SSE is minimum.

$$\begin{array}{llll} x & y & \hat{y} & \varepsilon \\ x_i & y_i & \hat{y}_i & e_i \end{array}$$