Simple Linear Regression

WC AT.

	x	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
1	.				
2	.				
3	.	.			
.					
n	.				

$\underline{\underline{SSE}}$.

↳ 'y' is continuous

$y = f(x)$

$= \beta_0 + \beta_1 x$

$MSE = \frac{1}{n} SSE$

$RMSE = \sqrt{MSE}$

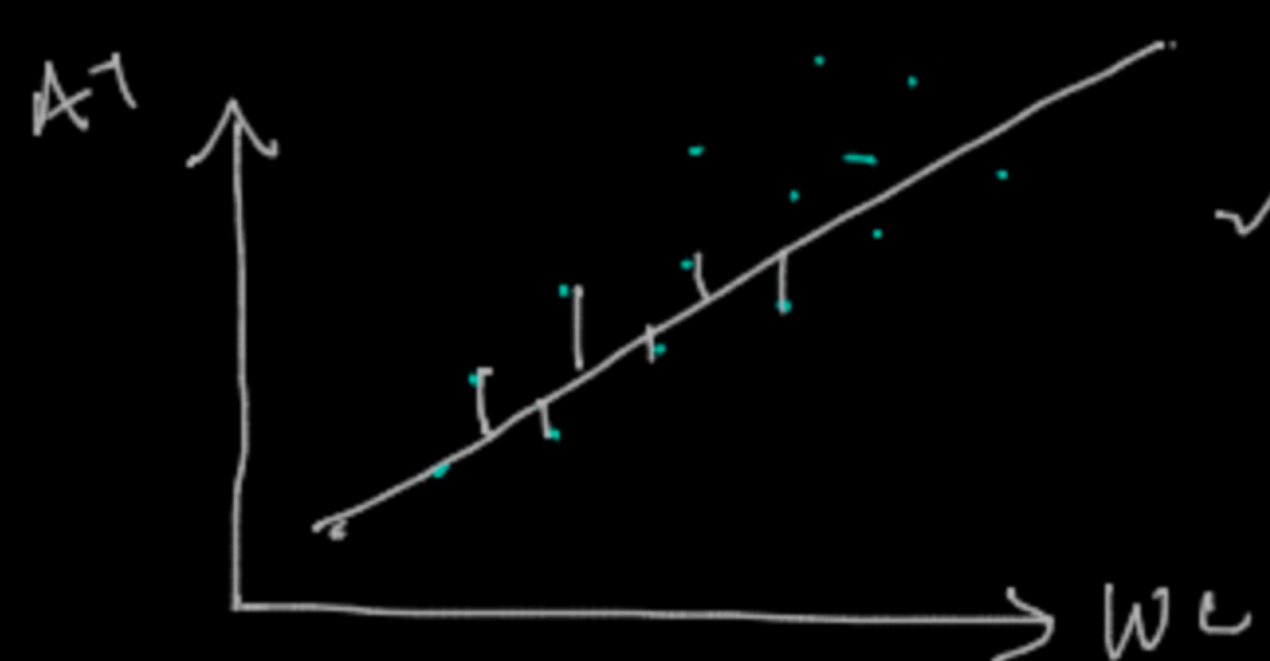
$R^2 = 65\%$

WC.

x ✓	y ✓ AT
-	-
-	.
-	

✓ 1. Correlation betn x & y.

a. Scatter plot



✓

b. Correlation Co-efficient

$r \geq \pm 0.85$ ✓

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

$R^2 \rightarrow$ Definition ✓

$r \rightarrow$

Model params.

$\rightarrow$ What?

Hyperparams $\rightarrow$?

— likely values ✓

✓ 2 Build the model

OLS $\Rightarrow$ Try different orientations (β_0, β_1) for the line

$\Rightarrow$ Find the "Line of Best Fit"

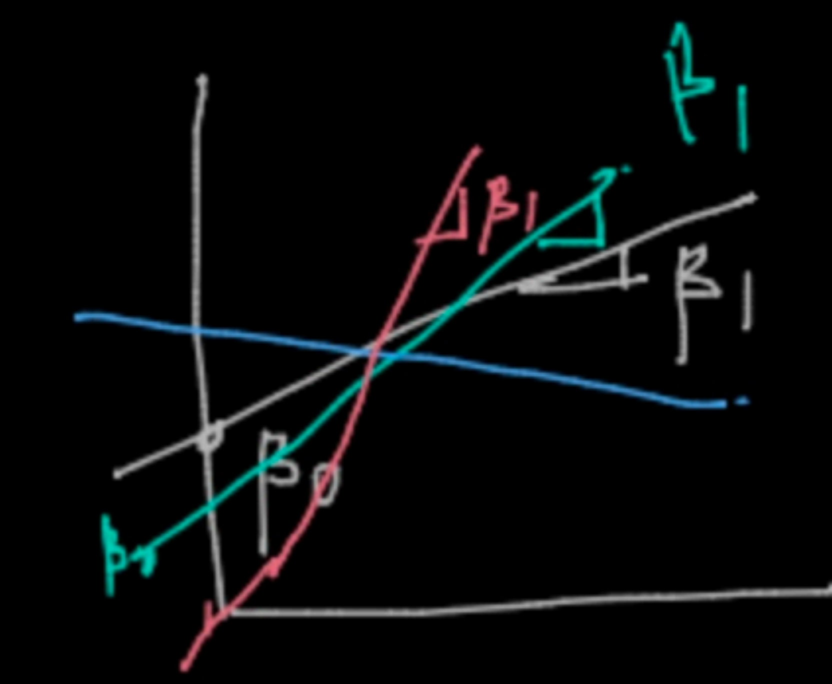
$\Rightarrow$ The line (β_0, β_1) for which the SSE is minimum.

Task of Algorithm:

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

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

$$SSE = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2$$



Find β_0 & β_1 which minimizes SSE

'Ordinary Least Squares'

'Gradient Descent'

Evaluate the Model:

1. MSE, RMSE, R^2, MAE

R-sqr ⇒ Percentage of Variance explained by the model

≥ 85%

Core python

- if - then
- loop
- exception

Third party library

import pandas

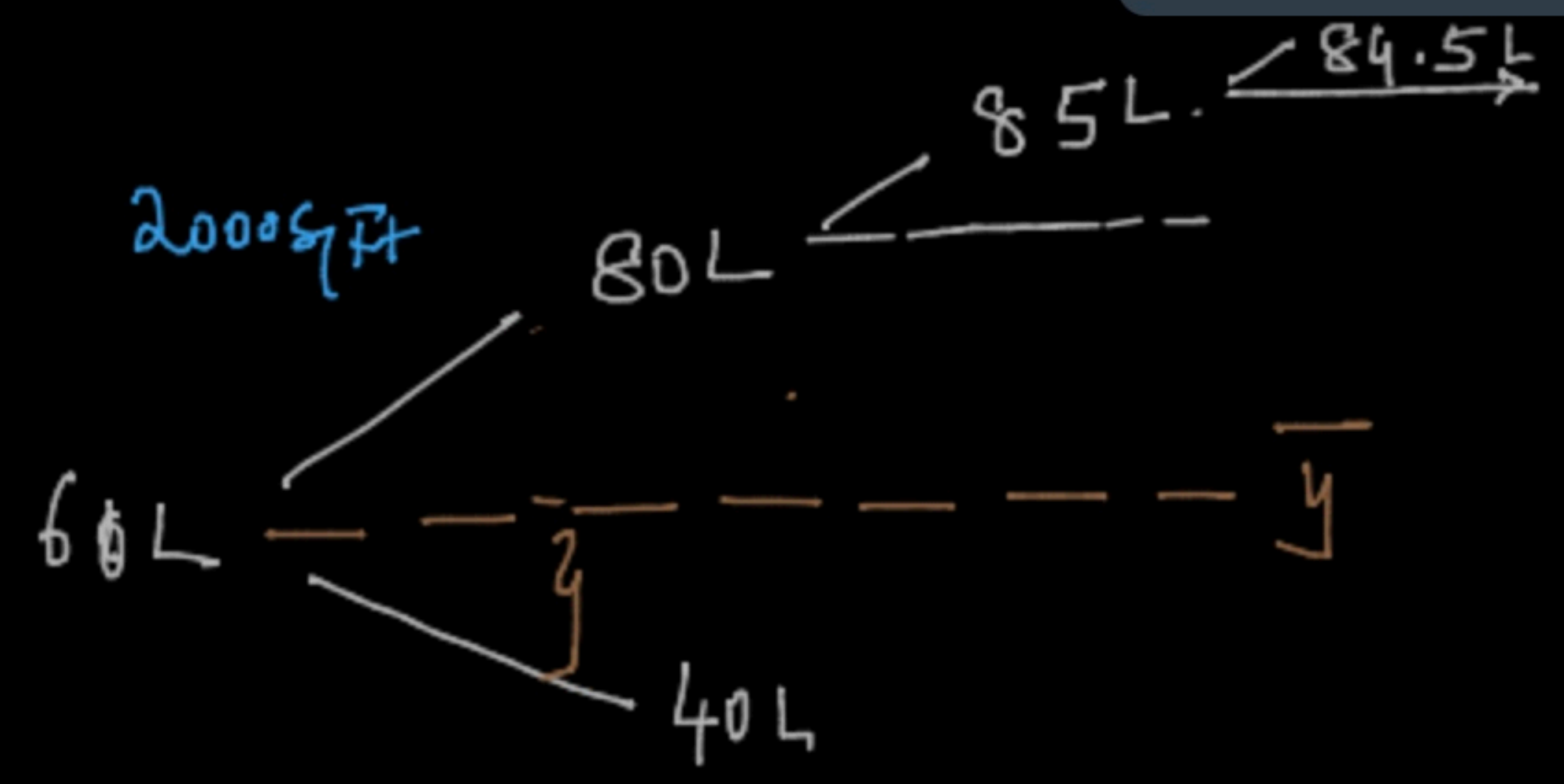
✓ Statsmodels . → ols – function

. sklearn → LinearRegression – class

.
.
.
.
.

x_4	x_3	x_2	x_1	
..	Color	Location	SqFt	'y'
—	—	—	—	53
				35
—	—	—	—	64
—	—	—	—	97
				120
				47

$\bar{y} = 60$



$R^2 \Rightarrow$ 85%   100% →

$\Rightarrow \dfrac{\text{Explained Variance}}{\text{Total Variance}}$

$R^2 \Rightarrow \dfrac{\text{Explained Variance}}{\text{Explained Var} + \text{unexplained Var}}$

↳ lesser

overfitting

$\overline{\text{RMSE}}$ = ±1 lakh ✓
⇒ small Error

R^2, MSE

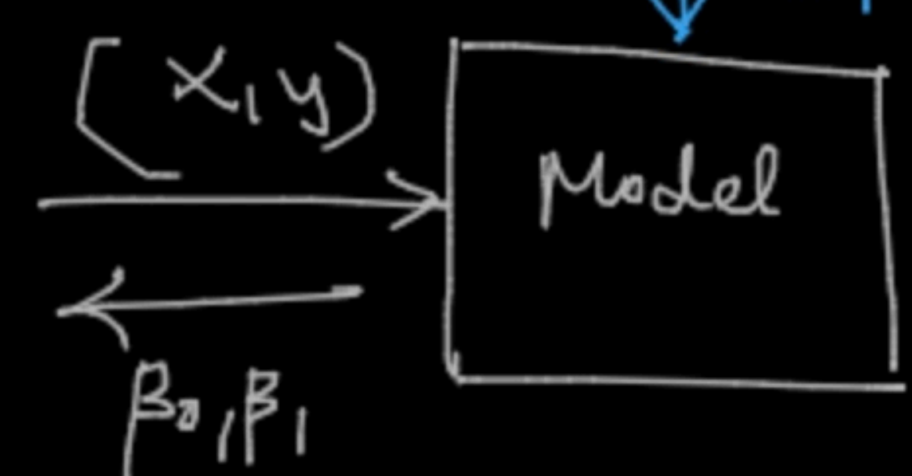
Standardized Metric

Marks = ±70   0 → 100   ±5
→ large Error

84%

1. Find if there is correlation betn x & y
2. Build the model
3. Train the model
   — estimating the "Model parameters"

↓ Hyperparameter
we give



Model parameter
model gives

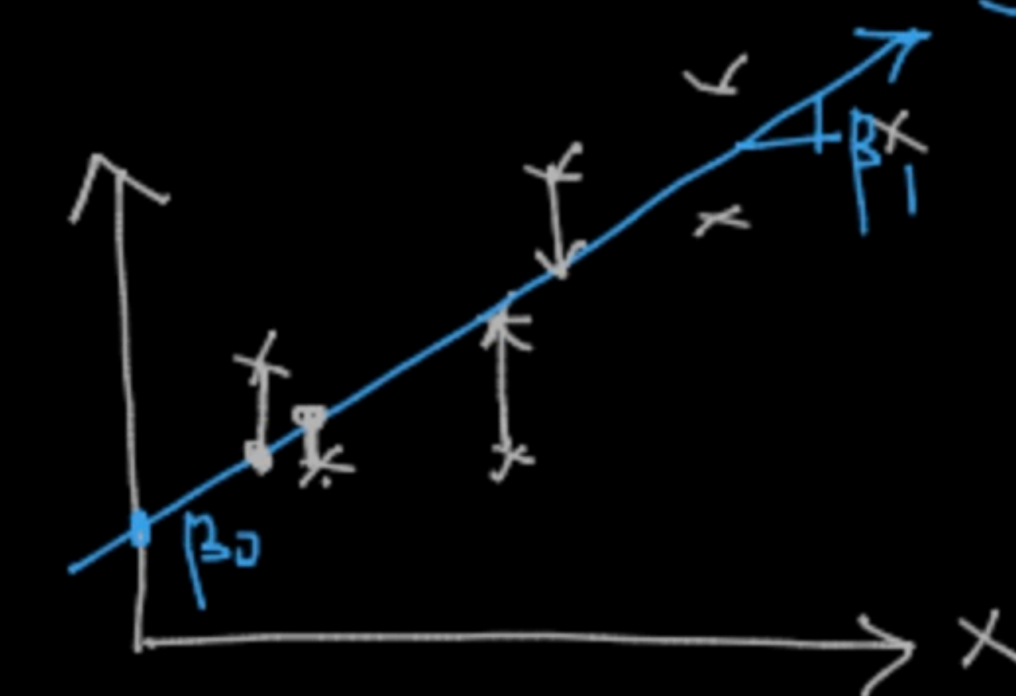
β_0, β_1

4. predict y
5. Evaluate the model

$\boxed{MSE,\ RMSE,\ R^2}$

	X	y	$\hat{y}$
1	–	–	–
2	–	–	–
3	–	–	–
.			
m	–	–	–

$\hat{y} = \beta_0 + \beta_1 x$



Hyperparameters
→ We have to choose

$\hat{y} = \beta_0 + \beta_1 x$

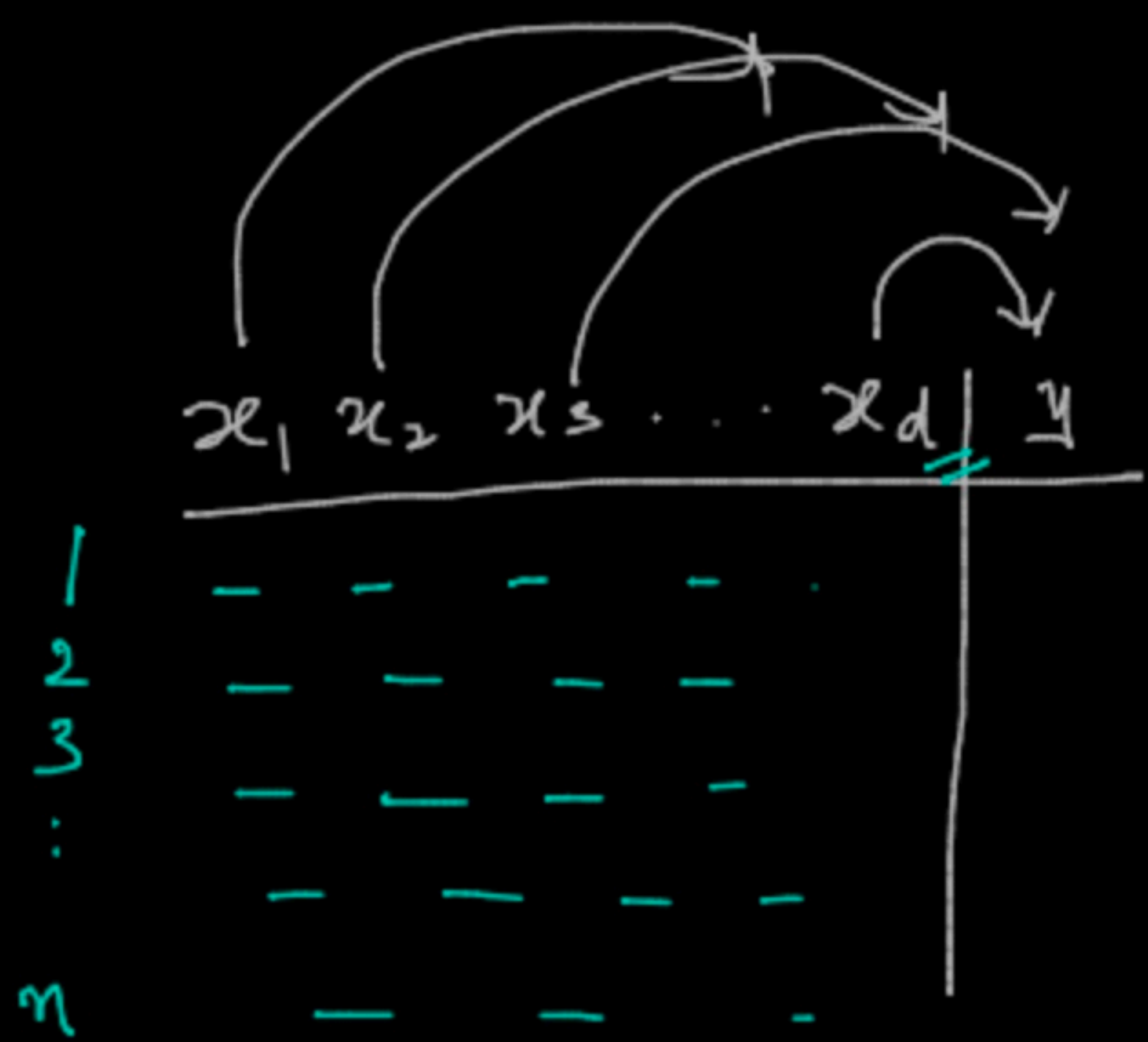
✓ $AT = \beta_0 + \beta_1 \times WC$

Intercept

Multi linear Regression

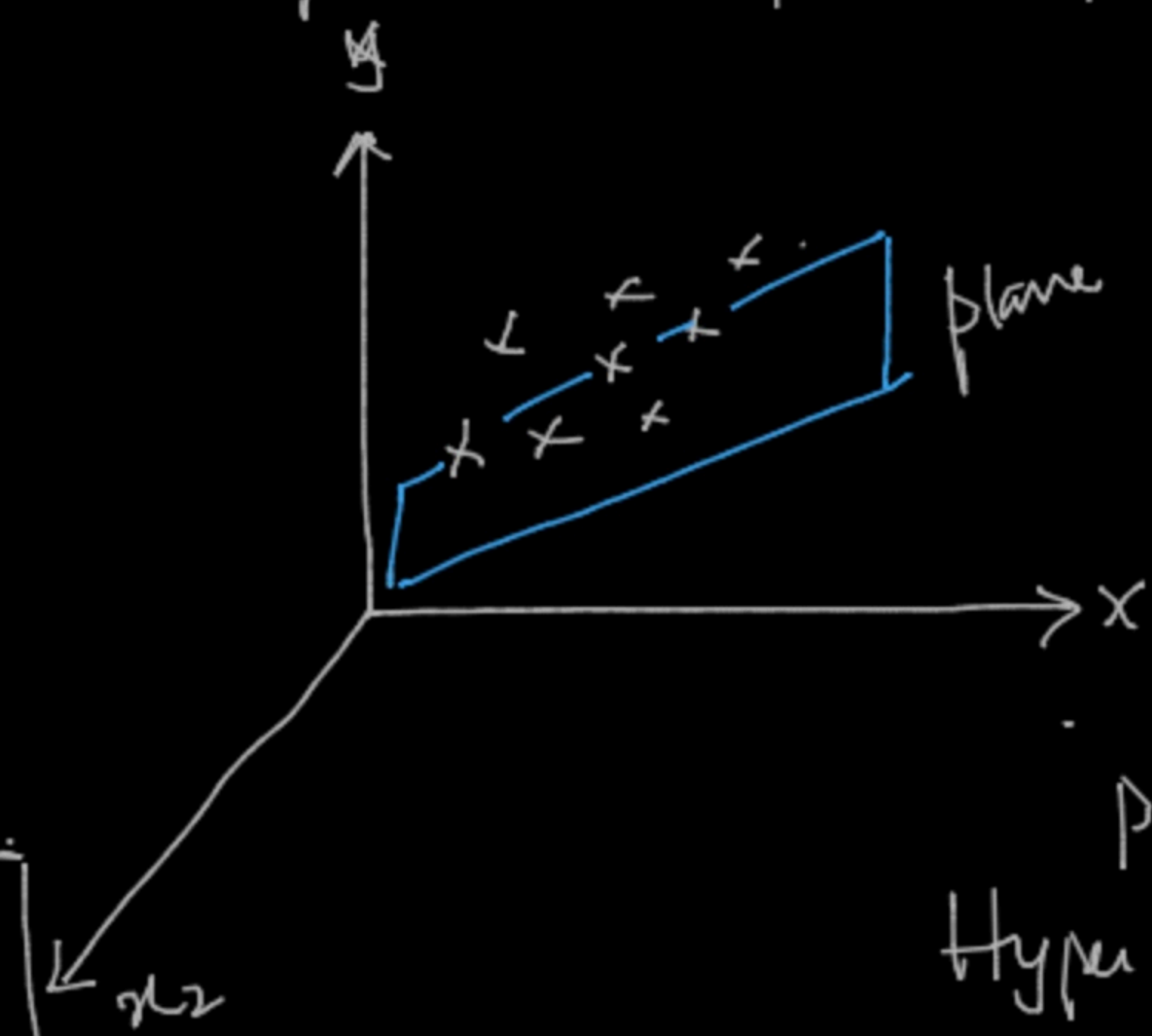
↳ 'y' is continuous.

Multiple features



(1)

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_d x_d .$$

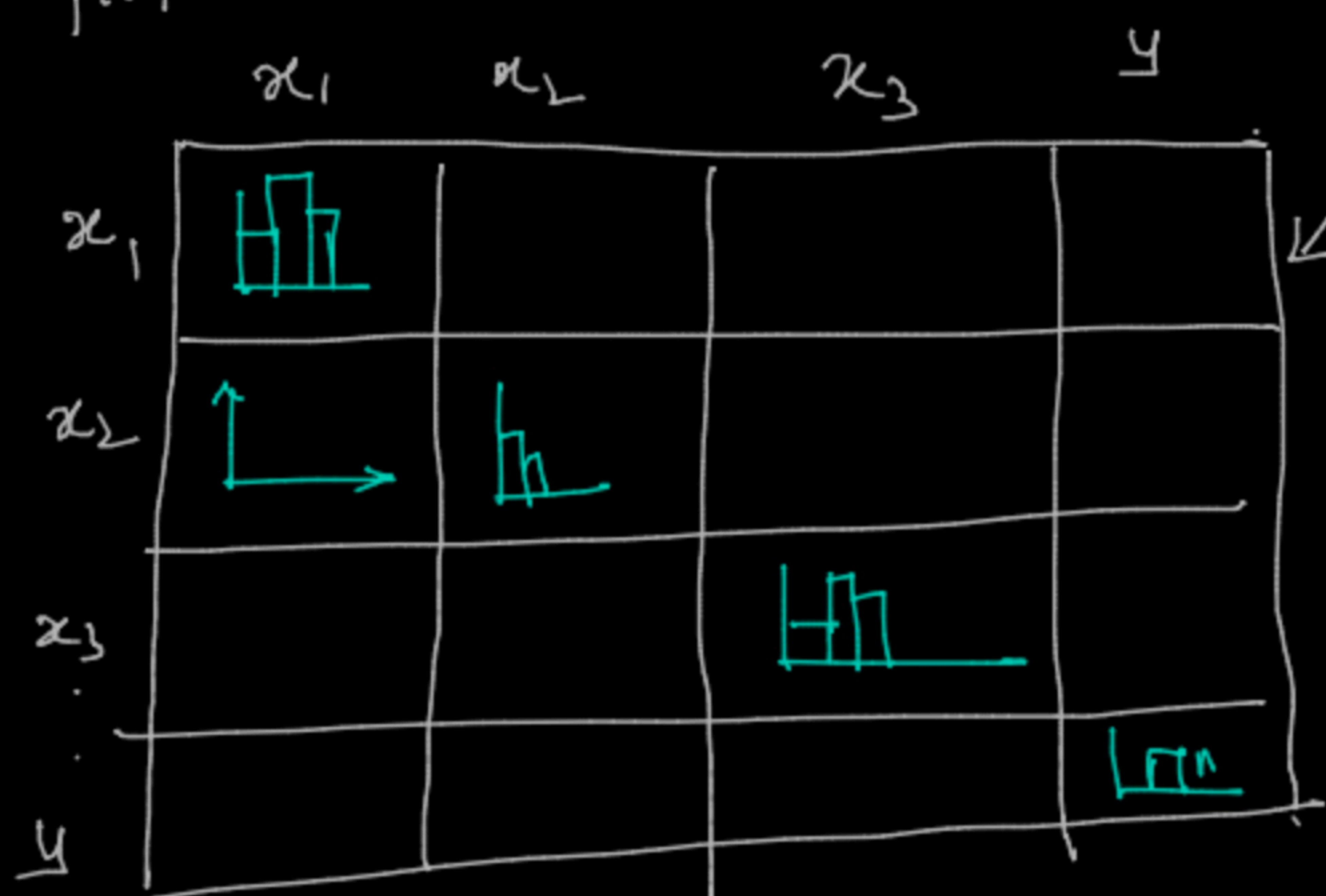


line $\beta_0 + \beta_1 x$

plane $\beta_0 + \beta_1 x_1 + \beta_2 x_2 .$

Hyperplane → $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots$

Pairplot



"Assumptions of Linear Regression"