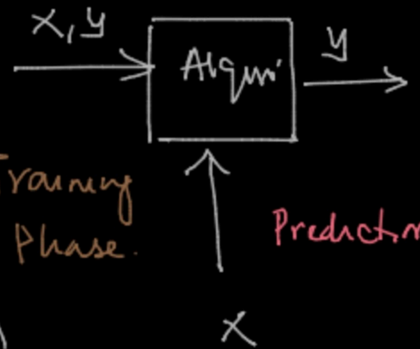
Indep. ← | → predicted
Inp. x | y — dependent.
predictor — o/p.
Features

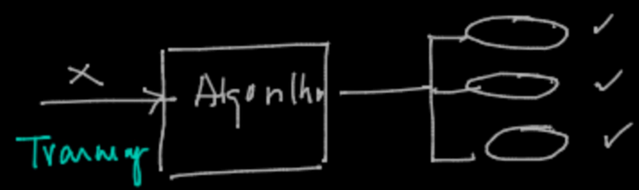
Machine Learning.

Supervised

— Model is trained on X & y.
→ Predictions can be made



unsupervised



— Model is trained on X-only
— No prediction

	← X →			← y →
	Area	loc	Bnilden	Bed.	House price
	✓	✓	✓	✓	
1	—	—	—	—	75 lakh ✓
2	—	—	—	—	43 lakh
3	—	—	—	—	—

1000 Houses

1500 — 3BHK — Prestige — 2BHK

→ 75 lakhs
← y →

Understand the pattern.

$$y = f(x)$$

Model

Regression Models
— 'y' continuous

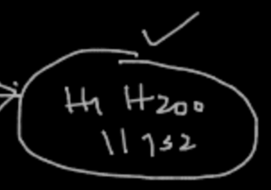
Classification Model
— 'y' is categorical

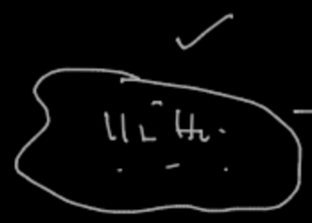
2000 sqft, 3BHK Bungalow

Clustering.

1800 · 3BHK 3BHK Bungalow

H1 —
H2 —
H3 —
H1000





→ < 2000 sft
= 2BHK.
⇒ posh L

3 —

→ Village; Gated

Simple Linear Regression

↳ 'y' is continuous

← one x.

$y = f(x)$

↳ Linear Fn

$y = mx + c$

↓ slope ↓ Intercept.

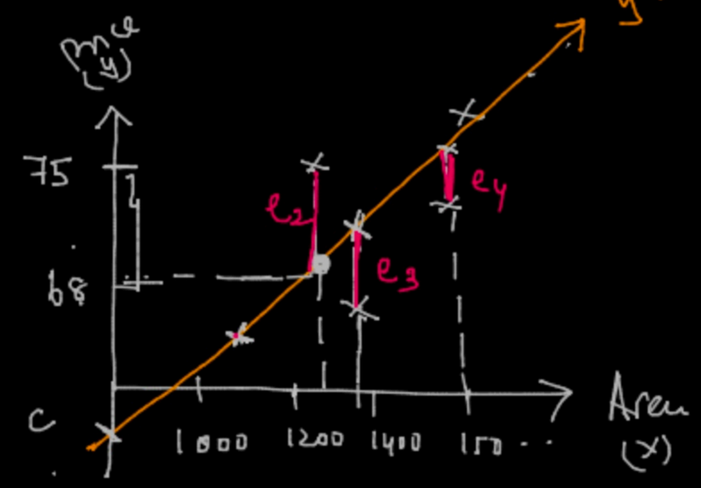
Error/ Residual

x	y	$\hat{y}$	$(y-\hat{y})$	$(y-y)^2$
1200	50	48.5	1.5	2.25 ✓
1800	78	78.9	-0.9	+0.81 ✓
.	.	.		MSE
.	.	.		

Mean = 0

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

← Predicted by the model



Metric:

- MSE = Mean of the square of the error = 3.06/2 = 1.53

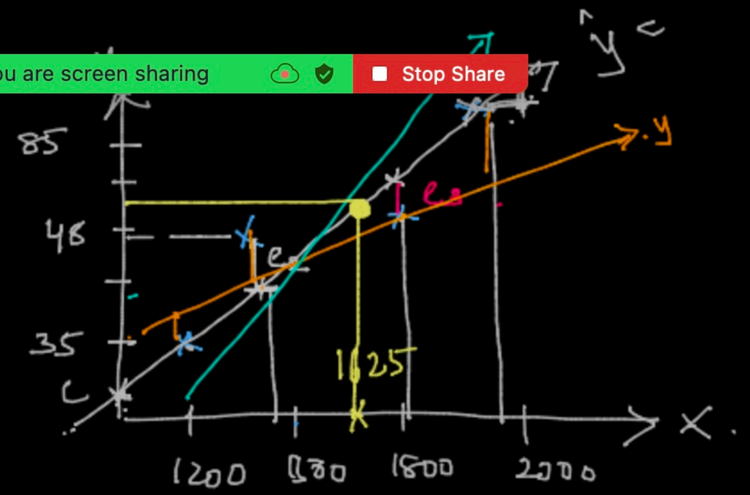
$$RMSE = \sqrt{MSE} = 1.24$$

MAE = Mean of the absolute value of Errors

$$MSE = \frac{1}{5}(e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2)$$

Training:

x	y	$\hat{y}$	e	e^2
1200	35	35	0	0
1500	48	45	3	9
1600	72	74	2	4
2000	84	84	0	0
			MSE	



m $\beta_1 \Rightarrow$

c $\beta_0 \Rightarrow$

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

1575 sq ft

$$\hat{y} = (\beta_0 + \beta_1 \times 1575)$$

$\hat{y} \Rightarrow$ 1625

$$\frac{1}{4}(e_1^2 + e_2^2 + e_3^2 + e_4^2)$$

MSE $\Rightarrow$ $\frac{1}{4}(9+4) = 13/4 \Rightarrow 3.2$

RMSE $= \sqrt{3.2}$

$$MSE = \frac{1}{4}(e_1^2 + e_2^2 + e_3^2 + e_4^2)$$

$$MSE = \frac{1}{4}(e_1^2 + e_2^2 + e_3^2 + e_4^2)$$

'Line of Best Fit'

$m =$ $c =$

$$y = mx + c$$

'Ordinary least squares'

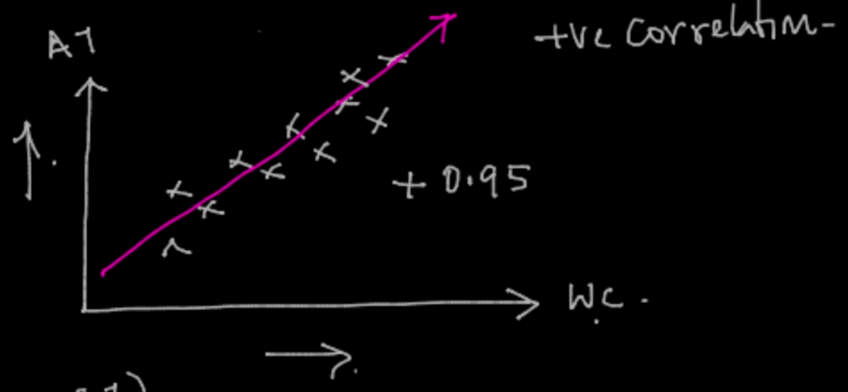
↳ m, c

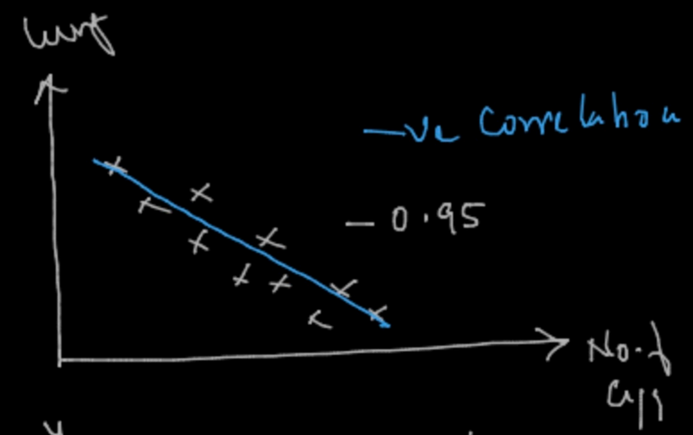
β_1, β_0

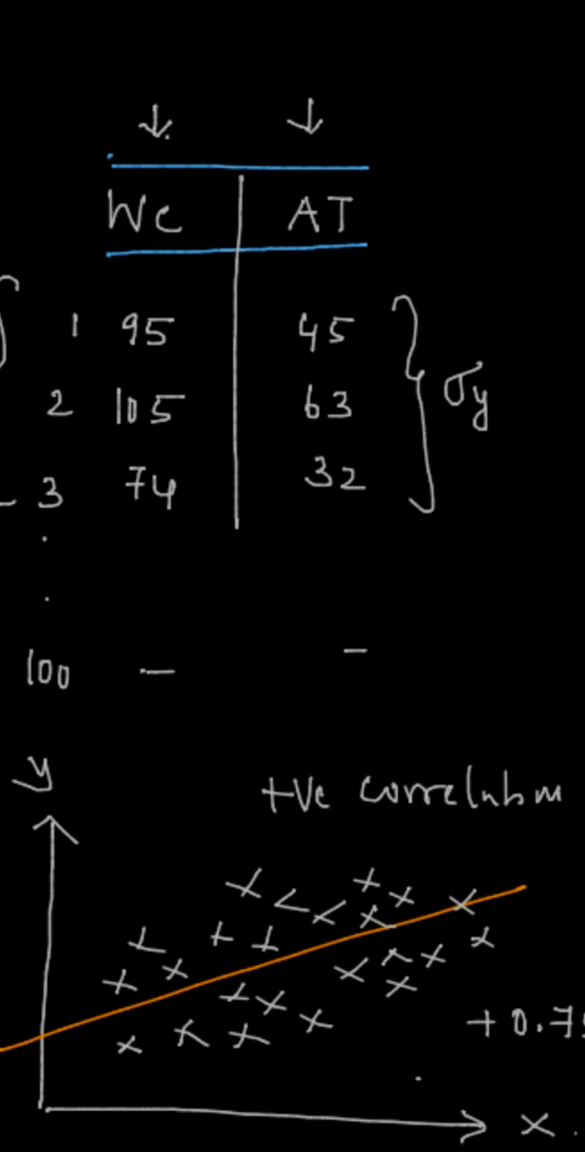
	WC	AT
1	95	45
2	105	63
3	74	32
.	.	.
.	.	.
100	—	—

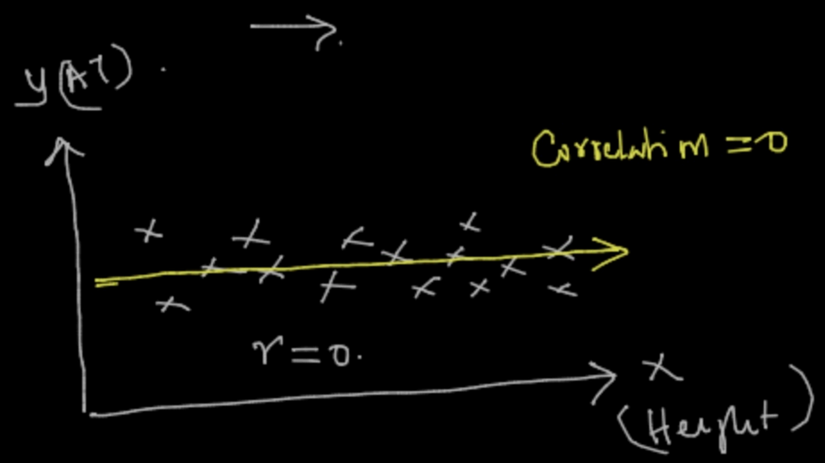
σ_x σ_y

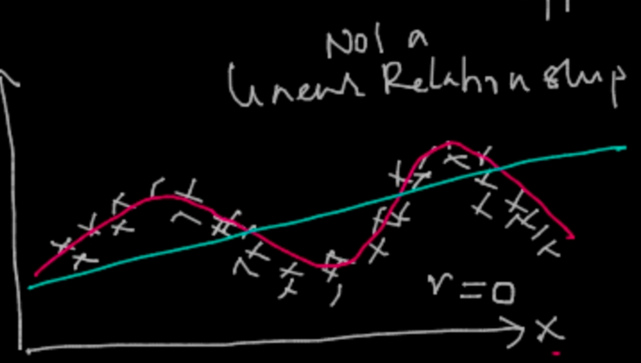
1. Correlation Analysis



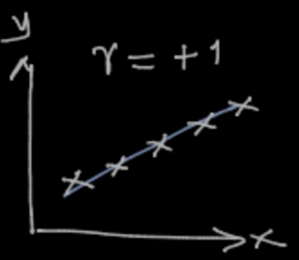








$y = \sin(x)$



Pearson's Correlation Co-efficient (r) $\Rightarrow \dfrac{Cov(x,y)}{\sigma_x \sigma_y}$

$r = -1$ to $+1$

— Strength of the LINEAR Relationship betn x & y

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

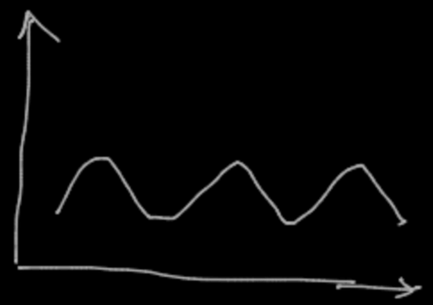
First step:

a Visually look at the data

b. Calculate the correlation co-efficient (r)

$r \geq \pm 0.85$

'x' to predict 'y'



Correlation → Standardized Covariance

Covariance → 474 ✓ · — 63

→ −1 & +1

- Simple Linear Reg → OLS
- Neural Network → gradient

- Decision Tree
- Random Tree
-