

03 Sept 2022

Term 4 - ML April 2022

- Data is the new oil.
- ML is a powerful concept & framework for making the best out of it.
- Age of information or
Age of data.

e.g.: Google knows everything about my location.

MMT
(Mumbai & N. Mumbai)



Information
[Darshan + 49 more ppl]

A \longleftrightarrow B
8am 4pm

| <u>Bus Service</u> | <u>Car</u> |
|-----------------------------|---------------------------|
| X | ✓ convenient & easy |
| AC | AC |
| ✓ | Guaranteed Seat |
| P2P | Home - Destn |
| 15min | Time Cons. |
| ₹ 2950/- pm | 10000 - 12000 rs / month |
| No Driving | Driving - Tired & fatigue |
| Sleep, Meetings, Netflix | Home - fresh |

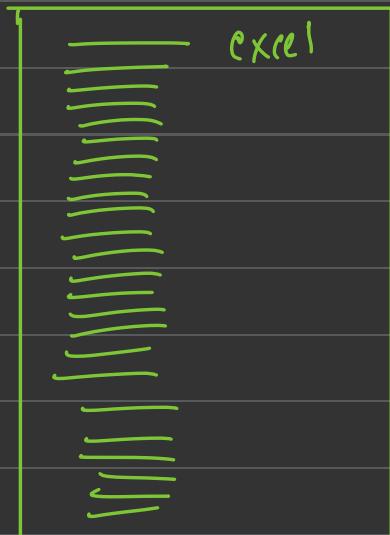
Who is most intelligent lifeform on the planet at the moment?
 Human.

-m/c : To mk data driven decisions at a scale.

your friend (10 yrs)

Will you lend a 1000 rs?
10 yrs (100 interaction)

past
experience



M/C
80 Yes ✓ ; Money Lent
20 No

Human

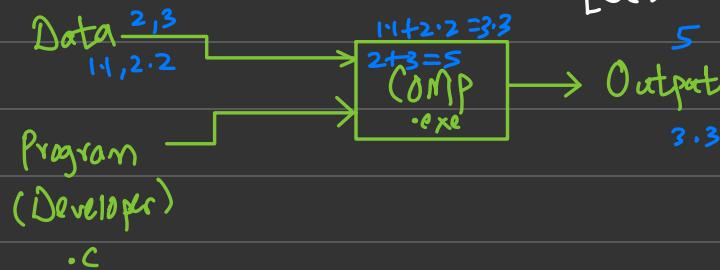
7
Yes ✓
3
No

; Money Lent.

Traditional Programming Paradigm

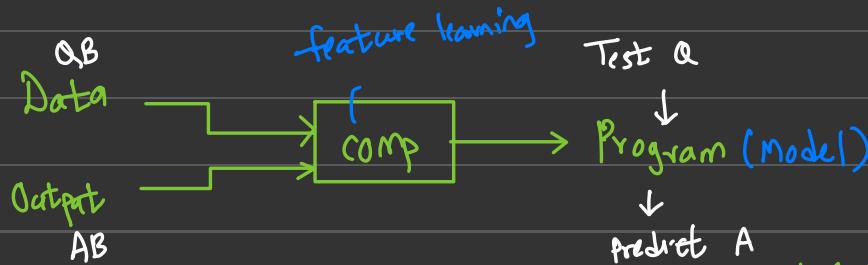
HDFC $\xrightarrow{\text{website}}$ Infosys

Task: Add two nos



[Deterministic]

ML programming Paradigm



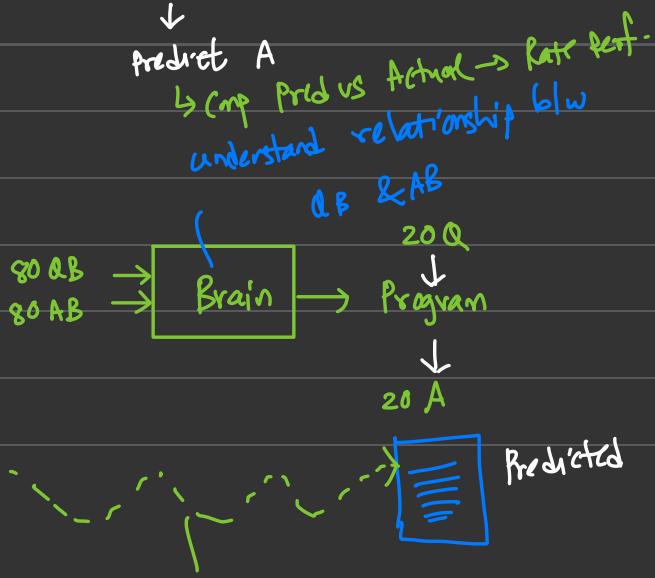
predict A
 \hookrightarrow Comp pred vs Actual \rightarrow Rate perf.
 understand relationship b/w

| QB | AB |
|-----|-----|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| : | : |
| 100 | 100 |

Study 80

20 Test

Actual (True)



|

Compare Actual & Predicted
answers & rate the performance
of a student.



04 Sept 2022

y
Dependent
Variable (1)

X

Independent Variables (1 or more)

① Price ~ Brand + EnginePower + Mileage + Transmission Type
of car + Color + Variant + FuelType + BodyType + ABS +
EBD + ADAS + HP + Torque

② Price of ~ Brand + RAM + Storage + Battery + ScreenSize +
Mobile Camera + OS + Safety + OLED | AMOLED + Processor

③ Apple vs ~ Color + Taste + Shape + Texture + NutritionValue
Orange + CalorieValue + Price

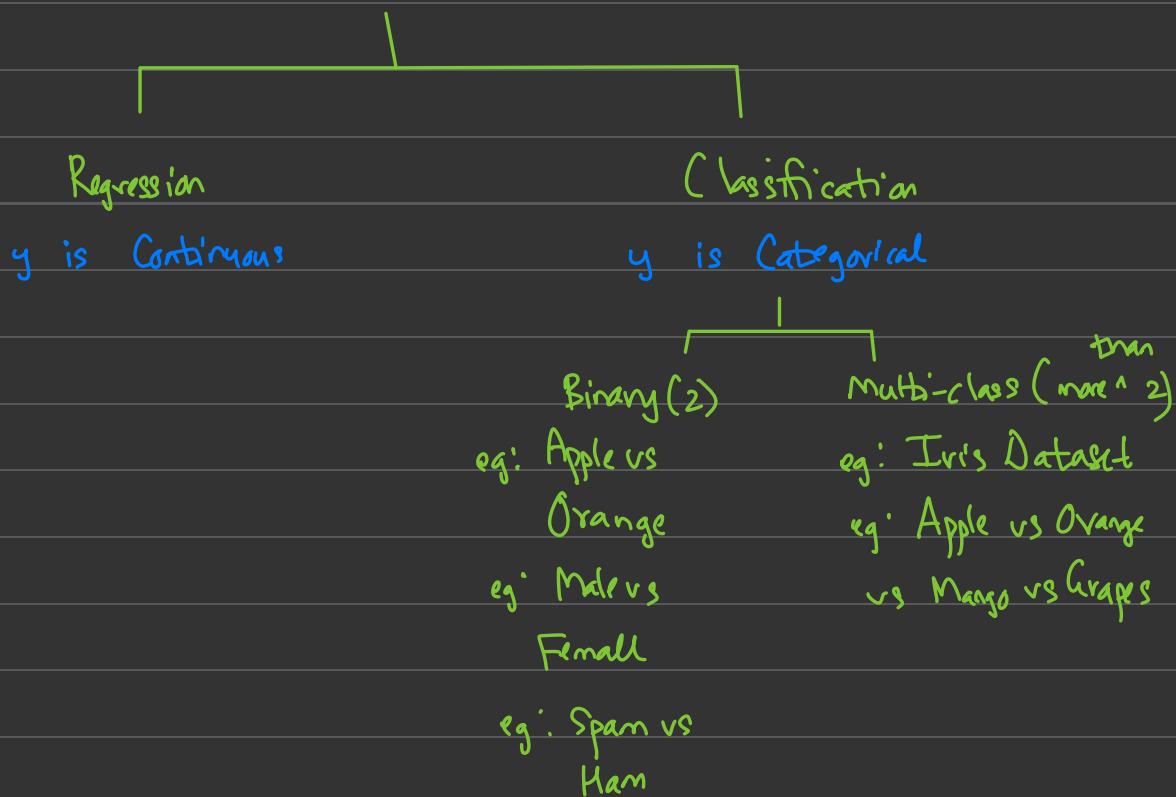
Classification

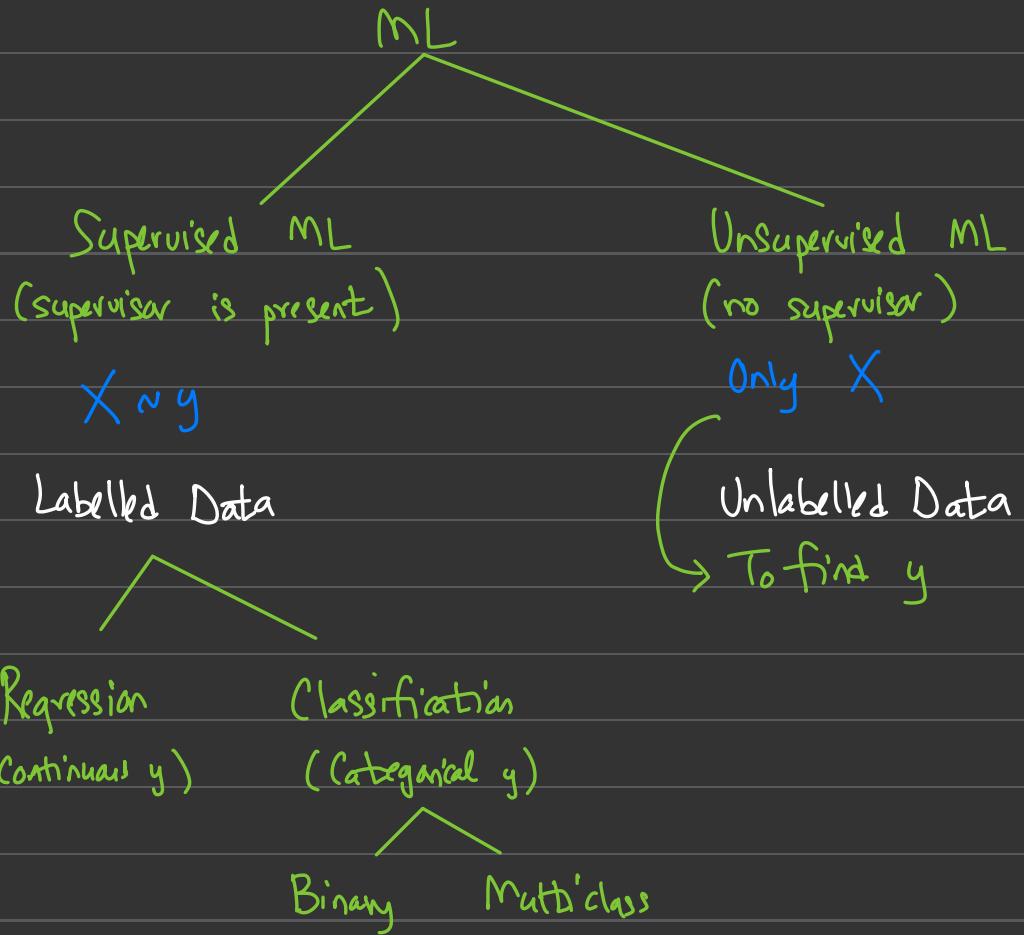
④ Iris (Setosa ~ SepalLength + SepalWidth + PetalLength
vs Versicolor + PetalWidth
vs Virginica)

https://miro.medium.com/max/1000/1*IFC_U5j_Y8IXF4Ga87KNVg.png

https://miro.medium.com/max/638/0*2c7voFri9cIXGrc4

ML Problems





<https://www.aplustopper.com/wp-content/uploads/2017/05/Correlation-1.jpg>

Take Home Task: Difference b/w Pearson's,
Spearman Rank & Kendall's
Correlation Coefficient + Applications

<https://www.mathsisfun.com/data/correlation.html>

To calculate the Slope:

Divide the change in height by the change in horizontal distance

Slope
 $m=0.5$

If we move 1 unit on X-axis,
we climb 0.5 units on Y-axis

$$\text{Slope} = \frac{\text{Change in Y}}{\text{Change in X}} = \frac{y_2 - y_1}{x_2 - x_1}$$

y-intercept $c=2$

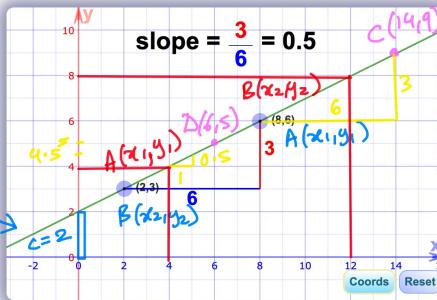
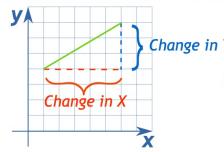
Eqn. of Line

$$y = mx + c$$

$$y = 0.5x + 2$$

Given x , we can find y .
 If $x=4$, $y = 0.5(4) + 2 = 4$, If $x=12$, $y = 0.5(12) + 2 = 8$

Have a play (drag the points):

Equation of a
Straight Line

$$m = \frac{6-3}{8-2} = \frac{3}{6} = 0.5$$

$$m = \frac{3-6}{2-8} = \frac{-3}{-6} = 0.5$$

$$m = \frac{9-5}{14-6} = \frac{4}{8} = 0.5$$

$$m = \frac{5-9}{6-14} = \frac{-4}{-8} = 0.5$$



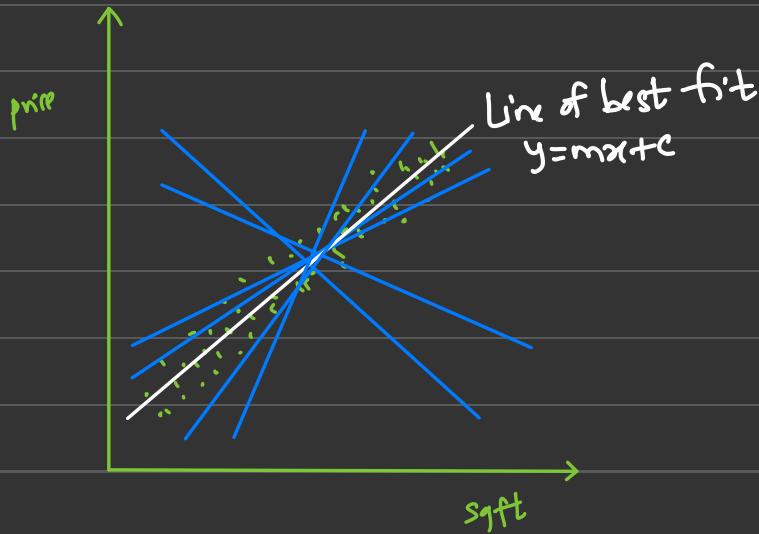
10th Sept 2022

Linear Regression

Simple LR ($X_1 y$)

Multiple LR ($X_1 + X_2 + \dots + X_n y$)

Pb Statement: Predict price of house depending on its sqft area.





$$ME = \frac{\sum_{i=1}^n (A_i - P_i)}{n}$$

$$\frac{A_1 P_1 \quad A_2 P_2 \quad A_3 P_3 \quad A_4 P_4}{0 \quad 0 \quad 0 \quad 0} \quad ME = \frac{0 + 0 + 0 + 0}{4} = 0$$

Mean Absolute error:

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

① MAE = $\frac{|+4| + |+4| + |-4| + |-4|}{4} = \frac{16}{4} = 4$.

②

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



Mean Squared Error = $\frac{\sum_{i=1}^n (A_i - P_i)^2}{n}$

$$\textcircled{1} \quad \text{MSE} = \frac{(+4)^2 + (+4)^2 + (-4)^2 + (-4)^2}{4} = 16$$

$$\textcircled{2} \quad \text{MSE} = \frac{(+7)^2 + (+1)^2 + (-6)^2 + (-2)^2}{4} = \frac{90}{4} = 22.5$$

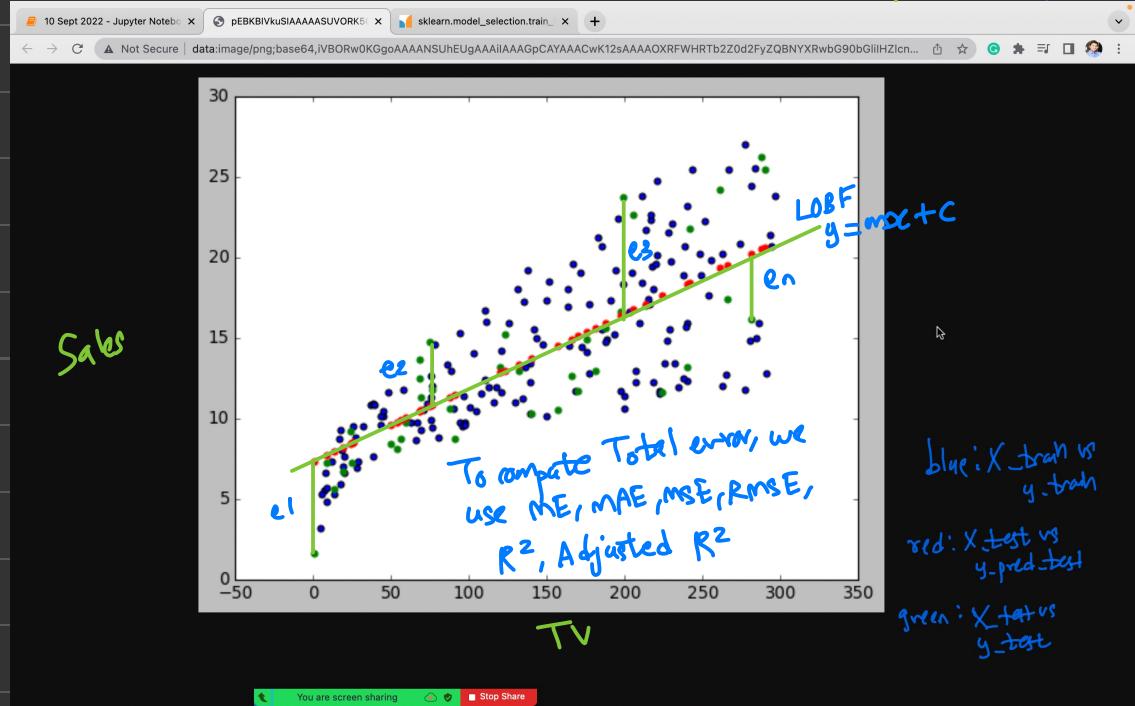
Root MSE = $\sqrt{\text{MSE}}$

$$\textcircled{1} \quad \sqrt{16} = 4$$

$$\textcircled{2} \quad \sqrt{22.5} = 4.74$$

Lv Train Test 160 40

| X | y |
|---------|-------------|
| x-train | y-train |
| fit | |
| x-test | y-test |
| | 40 |
| | y-pred-test |
| | 40 |



— X —

Maths is Fun - Slope (Gradient) of a Straight Line

Graph Index

Slope = $\frac{\text{Change in Y}}{\text{Change in X}}$

Equation of a Straight Line

Algebraically

Given $y = mx + c$

$$y = 0.5x + 2$$

$$\text{If } x = 2, y = 0.5(2) + 2 = \underline{\underline{3}}$$

$$\text{If } x = 10, y = 0.5(10) + 2 = \underline{\underline{7}}$$

Geometrically

Simple LR
compute unknown y using ① Algebraic intuition i.e. $y = mx + c$

or
② Geometric Examples: e.g. geometric intuition

Have a play (drag the points):

The Slope of this line = $\frac{3}{3} = 1$
So the Slope is equal to 1

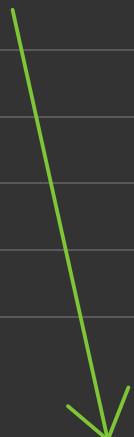
© 2018 MathsIsFun.com v9.0.81

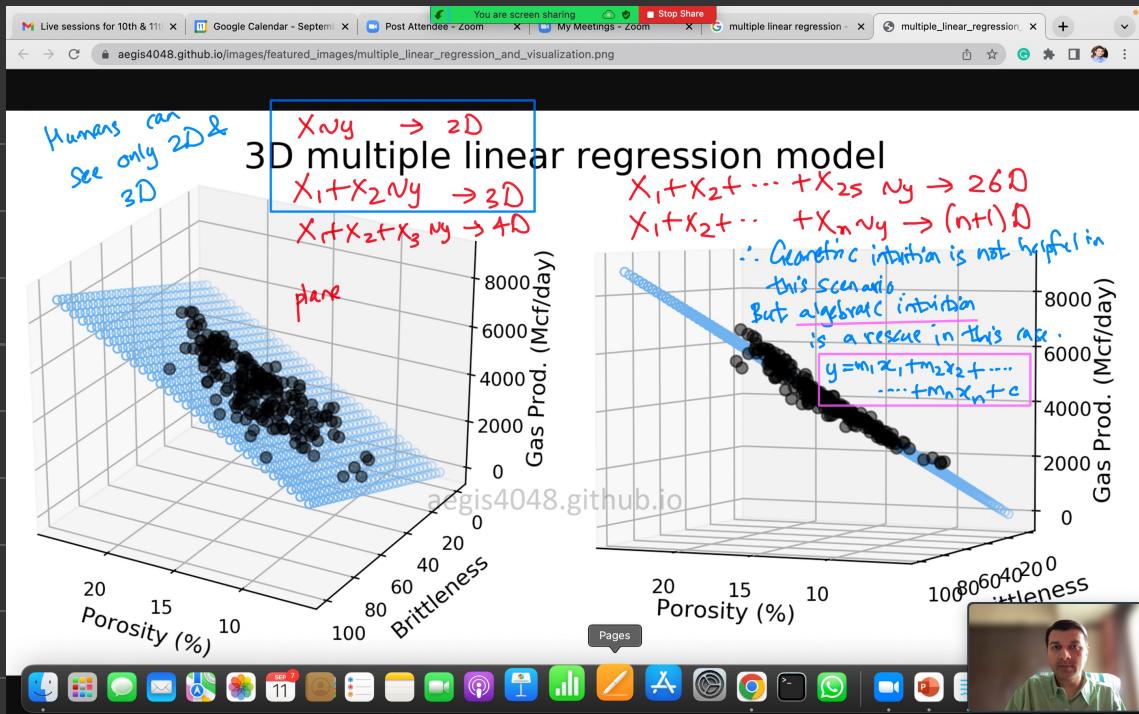
Multiple LR:

$$X_1 + X_2 \sim y$$

Algebraic intuition: $y = m_1 X_1 + m_2 X_2 + c$

Geometric intuition:





Code for Simple LR:

1. read dataset
 2. X vs y Separation
 3. TTS ($70:30$)
X_train, X_test, y_train, y-test
 - + import LR
 5. lr = LR() # LR object
 6. lr.fit(X_train, y_train)
 7. y-pred-test = lr.predict(X-test)
y-pred-train = lr.predict(X-train)
 8. Check error metrics like MAE, MSE, MAPE, RMSE, R², AR².
- Model 1 → MAE = 1.42

How can I create more models?

Repeat all 8 steps as before with a different
random_state like (0, 1, 2, ...)

or change TTS ratio (75:25, 80:20, ...)

Hyperparameter Tuning