



Data Science and Artificial Intelligence

Machine Learning



Regression

Lecture No. 01

By- SIDDHARTH SABHARWAL SIR



Topics to be Covered



Topic

Basics of ML

Topic

What is ML \Rightarrow Basic terms in ML

Topic

Topic

Topic

About the Faculty

- AIR 1 GATE 2021, 2023 (ECE). ✓
- AIR 3 ESE 2015 ECE. ✓
- M.Tech from IIT Delhi in VLSI.
- Published 2 papers in field of AI-ML.
- Paper 1 ✓ : Feature Selection through Minimization of the VC dimension.
- Paper 2 ✓ : Learning a hyperplane regressor through a tight bound on the VC dimension.



By- SIDDHARTH SABHARWAL SIR



Machine Learning:

- (i) **Supervised Learning:** regression and classification problems, simple linear regression, multiple linear regression, ridge regression, logistic regression, k-nearest neighbour, naive Bayes classifier, linear discriminant analysis, support vector machine, decision trees, bias-variance trade-off, cross-validation methods such as leave-one-out (LOO) cross-validation, k-folds cross-validation, multi-layer perceptron, feed-forward neural network;
- (ii) **Unsupervised Learning:** clustering algorithms, k-means/k-medoid, hierarchical clustering, top-down, bottom-up: single-linkage, multiple-linkage, dimensionality reduction, principal component analysis.

Rules

* Note making
~~PDF~~ ^{Sath Sath.}

Fully Follow Course
and class Material
Sufficient for ML.
Gate exam

* Weekly test

* DPP provide Chapterwise

* Class H.W

* Work Book Chapterwise

Pre requisite

- ① Matrix multiplication
- ② Probability Basic
- ③ Basic d/dx , min/max Concept.

Gate 2025 - Theory | Numerical

35 ← Math } 70.1 marks
35 marks ← { AI
ML



Today's Topic

- ❖ **What is Machine Learning** ✓
- ❖ **What is optimisation** ✓
- ❖ **What is a model** ✓

Gate 2026 Fodenge

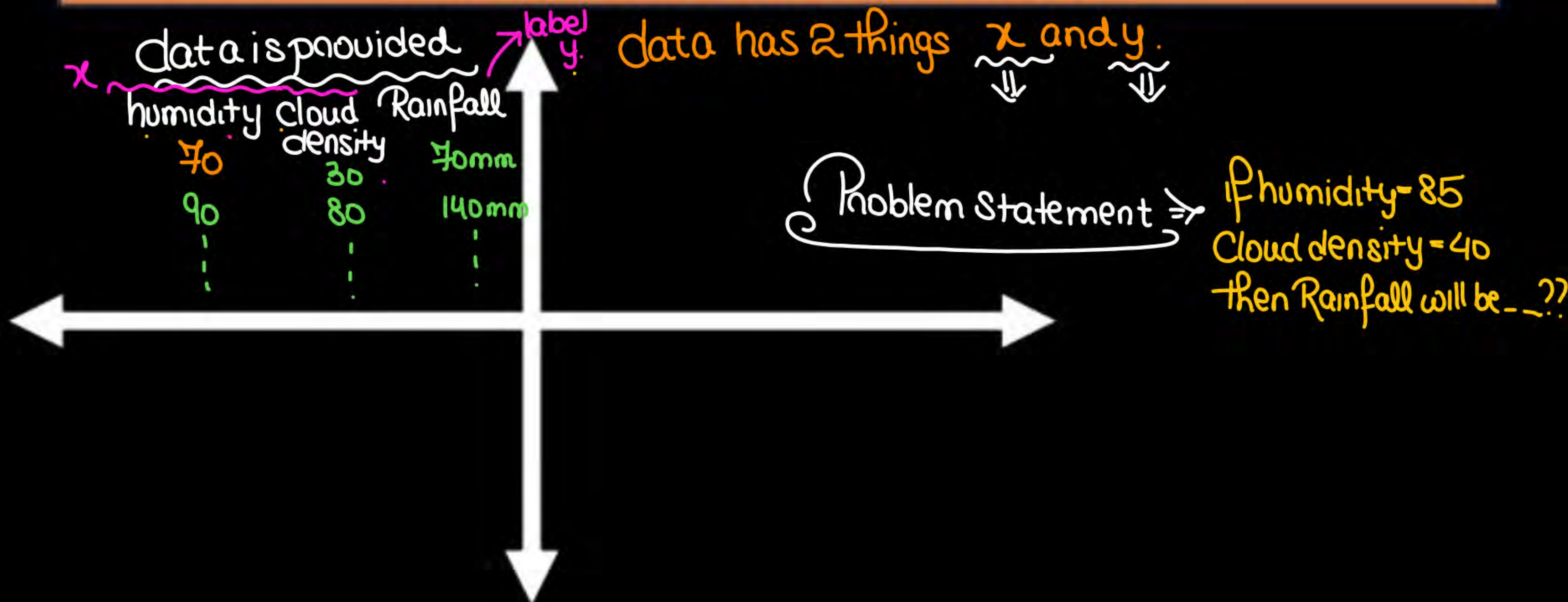
**“Your positive action
combined with
positive thinking
results in success.”**



Basics of Machine Learning

What is the General Problem in Machine Learning

Lets start with a basics



So in the provided data we have large no of columns.

x^1 amt of Sugar	x^2 Heart Rate	x^3 Blood Pressure	y % of Blockage in heart
⋮	⋮	⋮	⋮

Problem statement

Predict/Find % of
Blockage if
amt of sugar is ...
heart rate is ...
Blood Pressure ...

The value that is to be predicted is y , rest others are x

Example

data has only
2 column

x	label y
humidity	amt of Rain
70	30
80	90
50	60
40	80
⋮	⋮
75	??

Predict amt of rain fall
if humidity = 75.

- So we have to do analysis of
given data and find a
relationship b/w y and x .
- $y = f(x)$ So y is a function of x

Once we get the function then simply put $x=75$
and we predict
"y" value

:

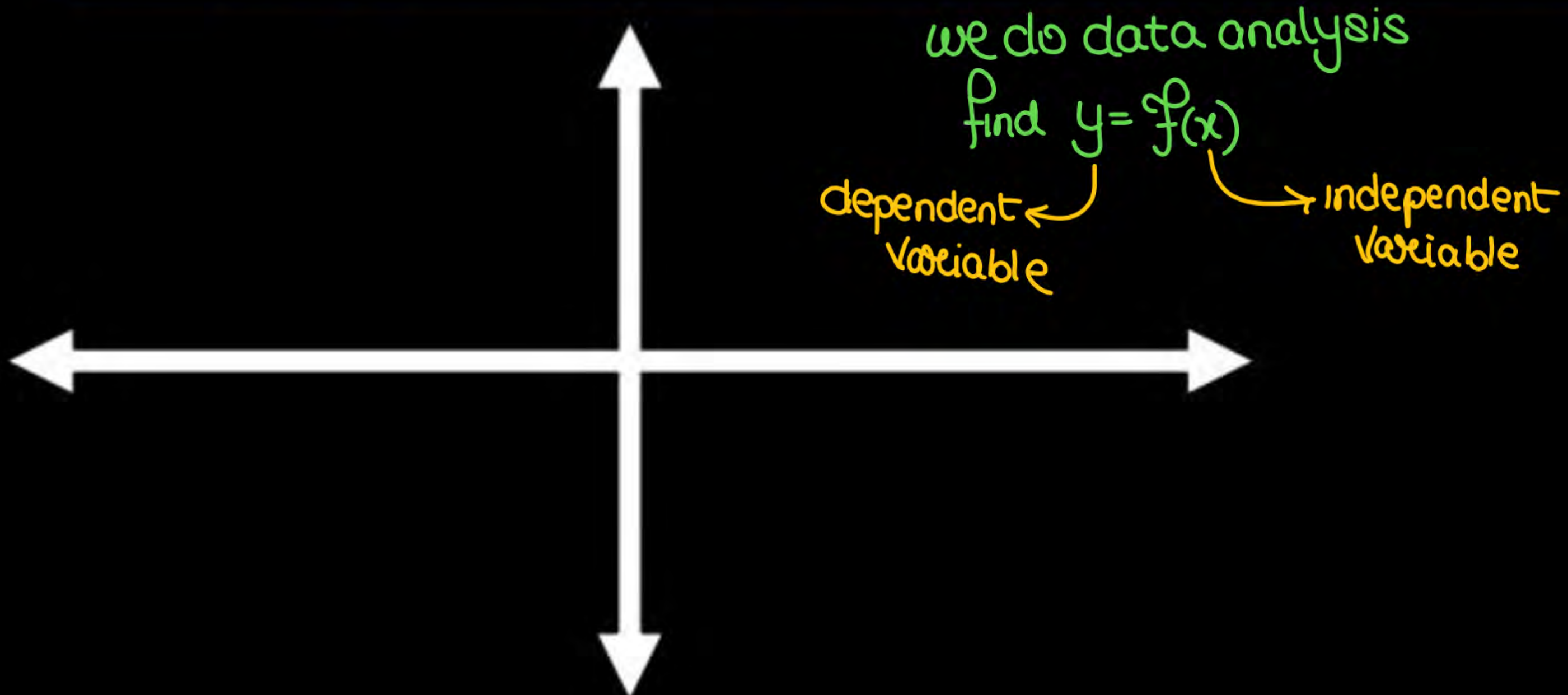


Basics of Machine Learning



What is the General Problem in Machine Learning

How you will predict/ find the value of y for new x





Basics of Machine Learning



What is the General Problem in Machine Learning

This is called a function of Independent variable

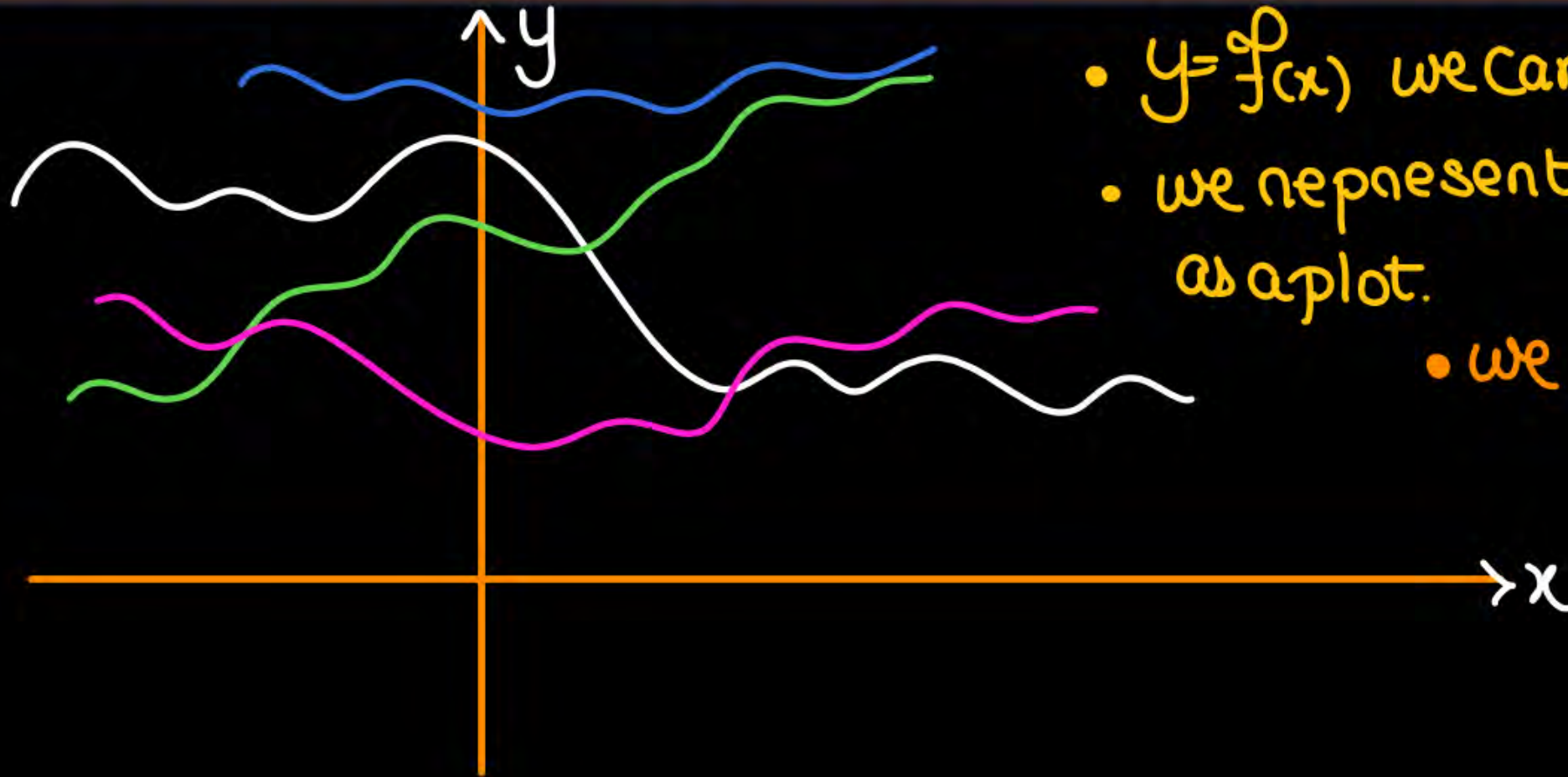
$y = f(x)$,
This matlab



Basics of Machine Learning

What is the General Problem in Machine Learning

These already known points relating Y and X is called _____

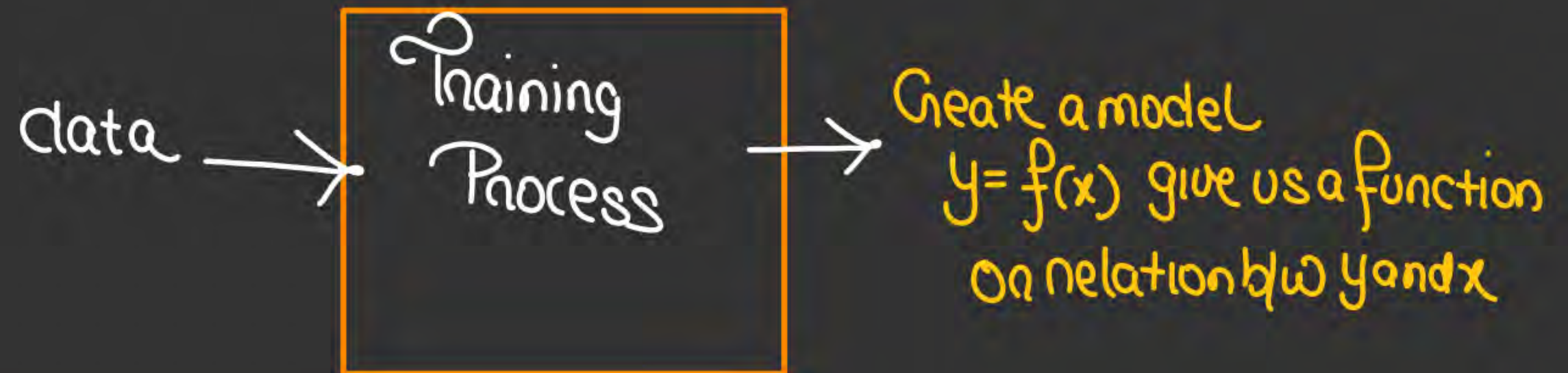


- $y = f(x)$ we can have ∞ Function
- we represent $y = f(x)$ on 2D plane as a plot.
- we can get ∞ plots.

So we have ∞ functions, but which one to Choose.

* So to select the best function (out of ∞ options), the process we follow is called optimization.

So the complete process of data analysis and optimization to find best function representing relation b/w y and x is called training process.





Basics of Machine Learning

What is the General Problem in Machine Learning

So using the data we learn a relation between Y and X and this process is called Training Process.

on Simply.
Training

Optimization is a part of Training



What is the General Problem in Machine Learning

What is Data

data is a set of observation/measurement which is collected to train the ML models.

which is collected to find out the best relation b/w y and x .



What is the General Problem in Machine Learning

How we collect data

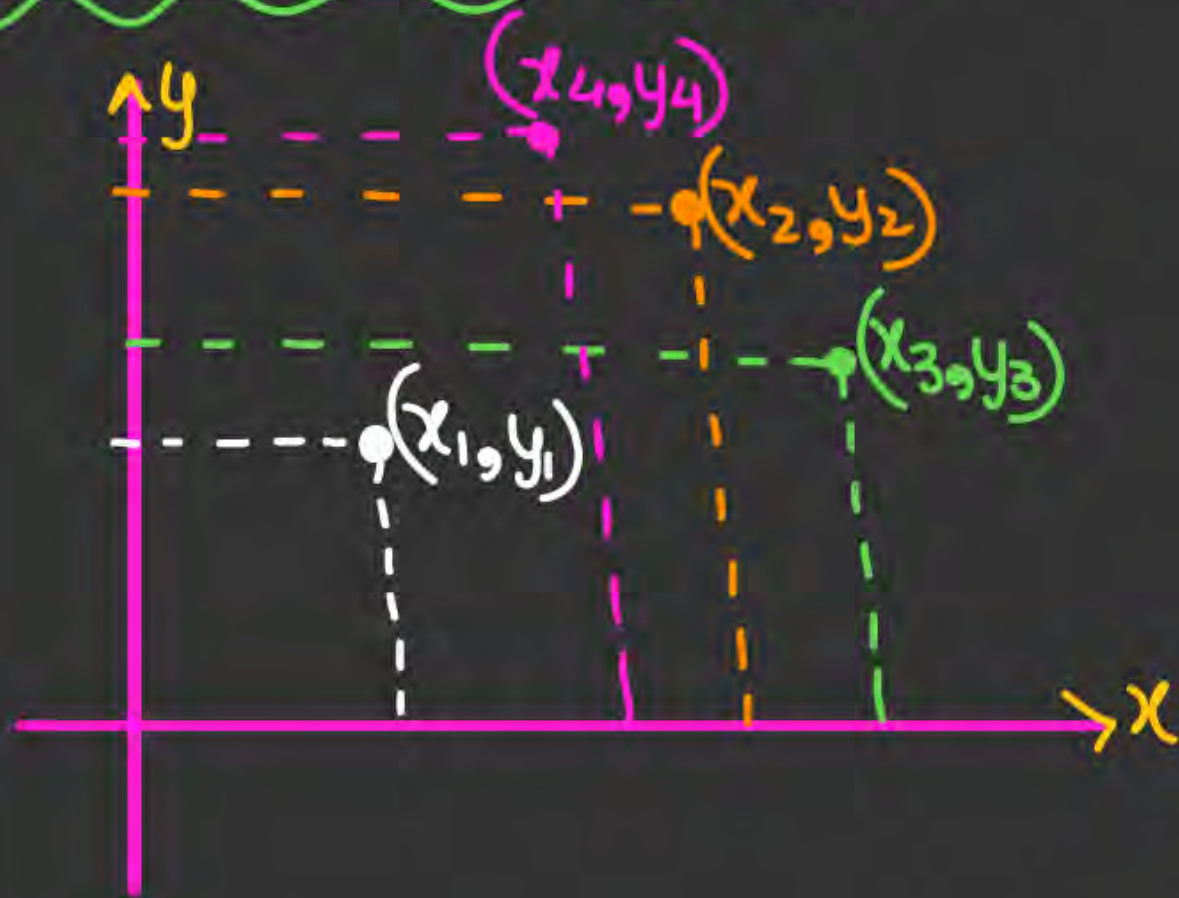
- i) experiment / Surveys.
- ii) Online Sources
- iii) Take data from expert.

What is optimization \Rightarrow

definition:

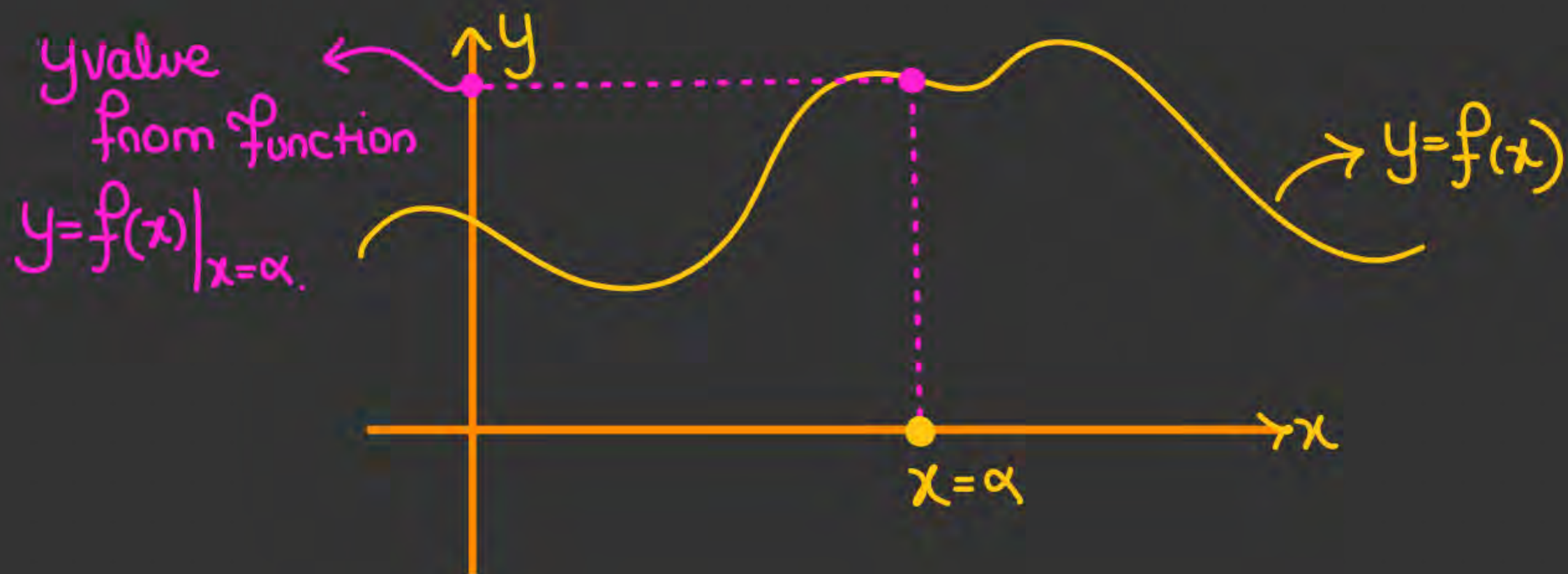
In ML we have to find the relation b/w y and x , optimization is the step in training process that give best fxn b/w. y and x

Mathematically



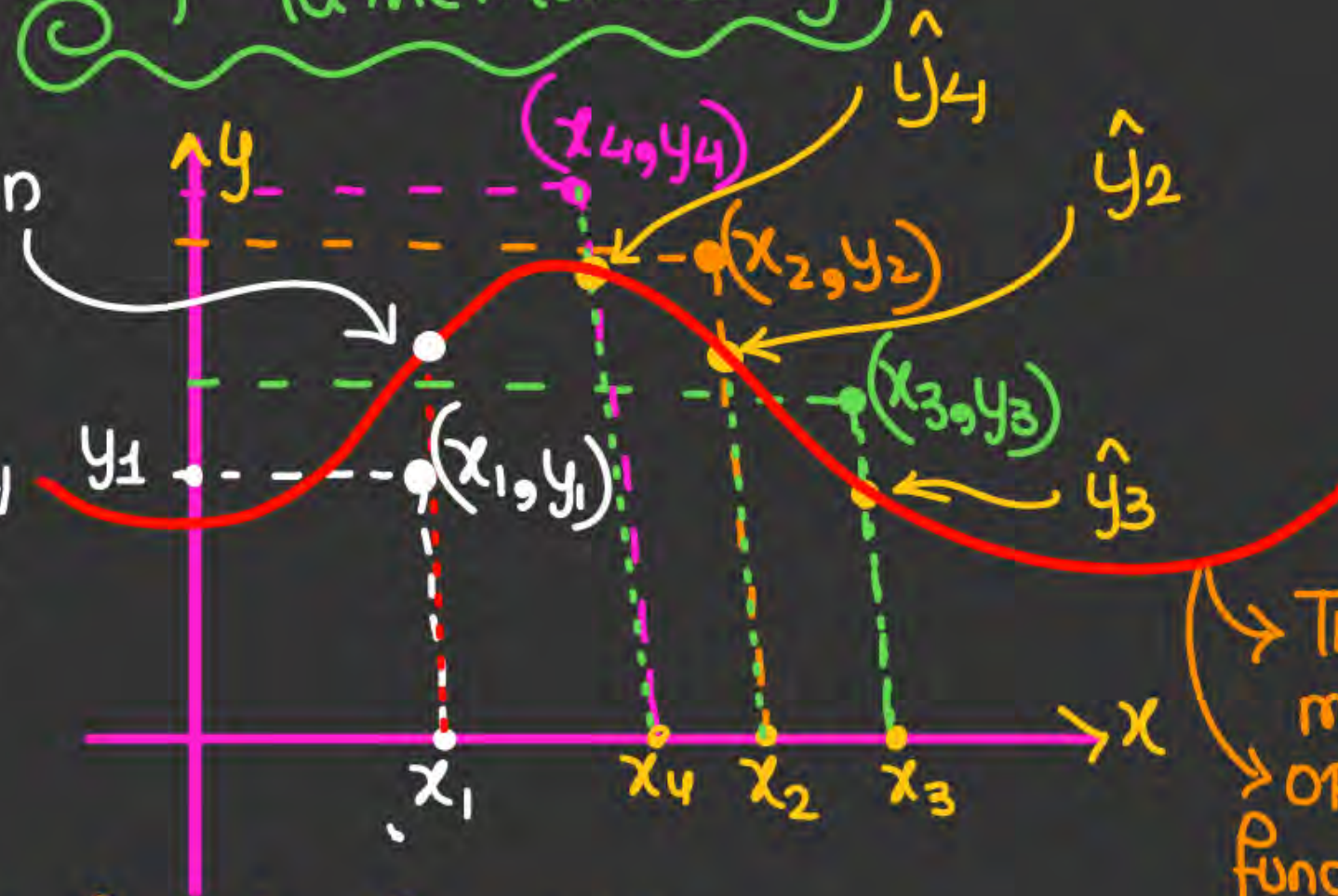
x	y
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4

- These all are actual values of y in data



Mathematically

- Value given by model
- y value Predicted by model \hat{y}_i



How to find that this model is good or not??

→ This is my model
→ OR this is function blw y and x
 $y = f(x)$

data		Pred. by model
x	y	
x_1	y_1	\hat{y}_1
x_2	y_2	\hat{y}_2
x_3	y_3	\hat{y}_3
x_4	y_4	\hat{y}_4

• These all are actual values of y in data

How to find model is good/bad \Rightarrow

The diff b/w actual and predicted value shd be small.

Term Residue \Rightarrow diff b/w y and \hat{y}
actual \leftarrow y \hat{y} \leftarrow Predicted.

We want a model that has

- $\min [(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3) + (y_4 - \hat{y}_4)]$

$$\text{So min} \left[|y_1 - \hat{y}_1| + |y_2 - \hat{y}_2| + \dots \right]$$

$$\text{OR min} \left[(y_1 - \hat{y}_1)^2 + \dots \right]$$

we want
model that
minimize
this.

So out of ∞ function $y=f(x)$ the best function/model is

Which minimize $\sum_{i=1}^N (y_i - \hat{y}_i)^2$ OR $\sum_{i=1}^N |y_i - \hat{y}_i|$

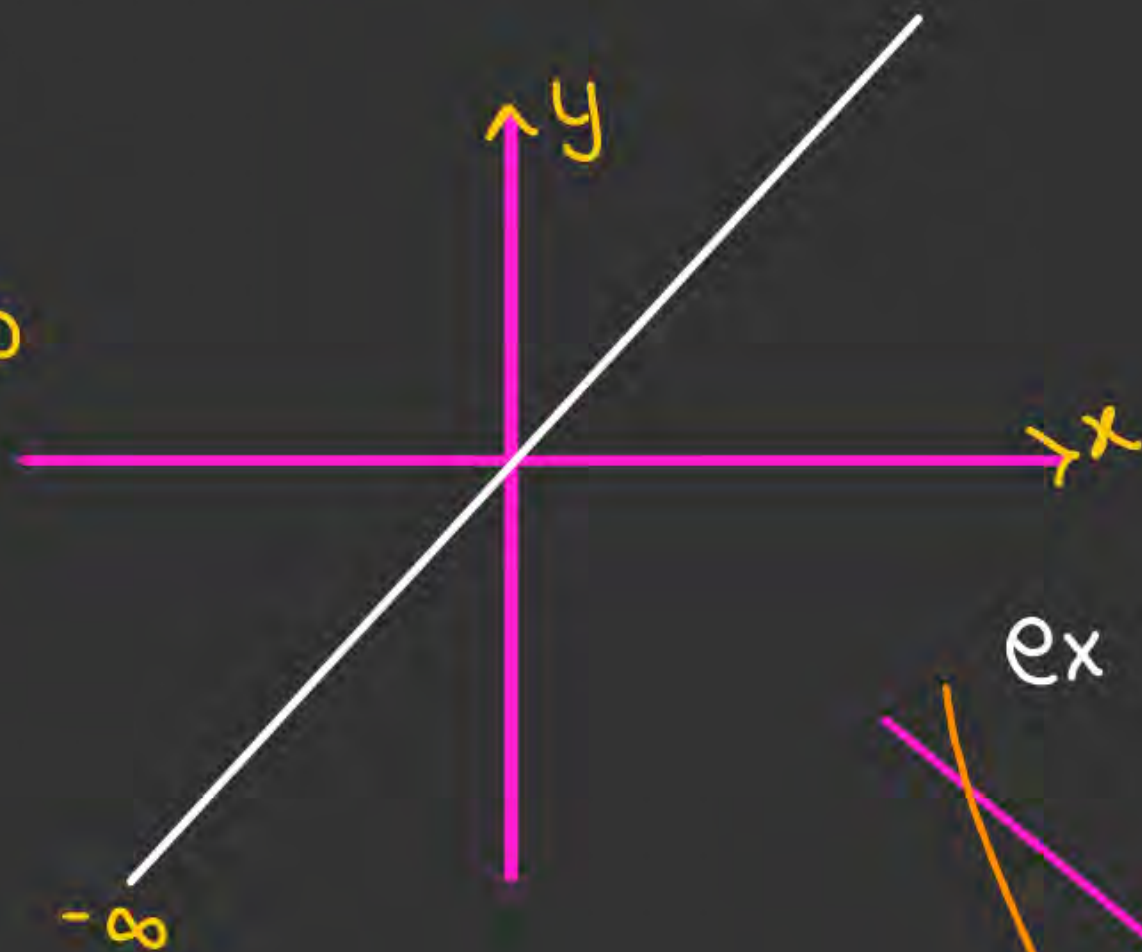
N data points

Kaunsa
data hai

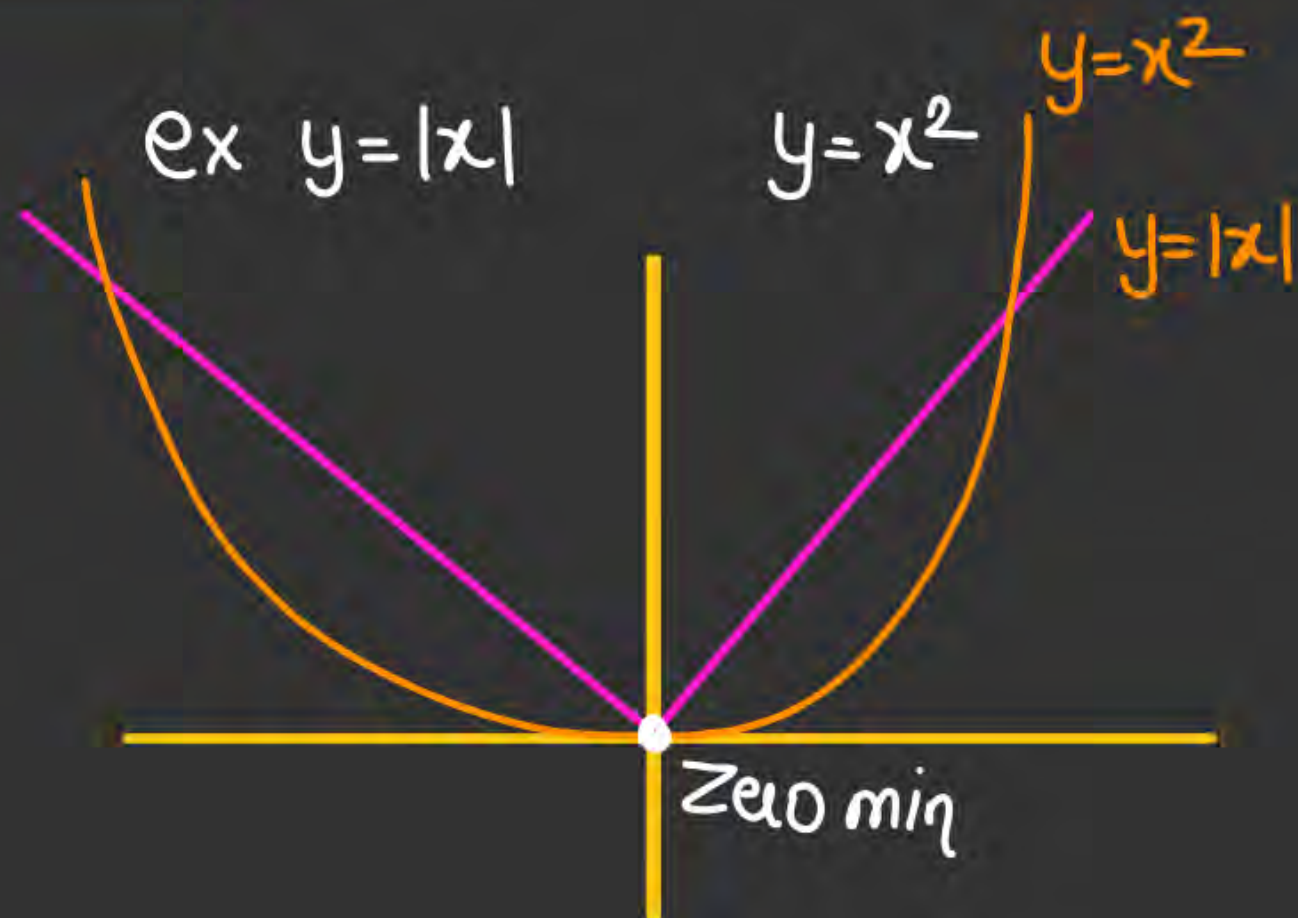
⇒ Training data

ex $y = x$

$\min(y) = -\infty$



ex $y = |x|$





Basics of Machine Learning

What is the General Problem in Machine Learning

Lets take an example of predicting rain



Basics of Machine Learning

What is the General Problem in Machine Learning

What is the problem Statement

- In ML we want to find model or relation b/w y and x , using given data
So that we can predict y for any new value of x



What is the General Problem in Machine Learning

What is a Model

- The function / relation b/w y and x
we get after training process



What is the General Problem in Machine Learning

What is the Optimization

- Part of training process
def: done



Basics of Machine Learning

data has label
↳ Supervised Learning



Basic Understanding – Predict Rain in Your city

What should be the format of the data ?

Humidity	Cloud density	^y Rainfall
40	80	40
50	90	100
60	20	50
⋮	⋮	⋮
⋮	⋮	⋮

No. of Study hrs	Concentration level	^y Pass
5	80%	Yes
10	20%	No
8	50%	Yes
⋮	⋮	⋮
⋮	⋮	⋮

y label discrete nature, y can take 2 value only.

We have two type of Data :-

- Continuous nature
numerical data \Rightarrow y label
Can take any real value
- Categorical nature \Rightarrow
y label can take
only few discrete values.

→ y label is of continuous nature
→ It can take any real value



Basic Understanding – Predict Rain in Your city

What should be the format of the data ?

**We have two type
of Data : -**



Basic Understanding – Predict Rain in Your city

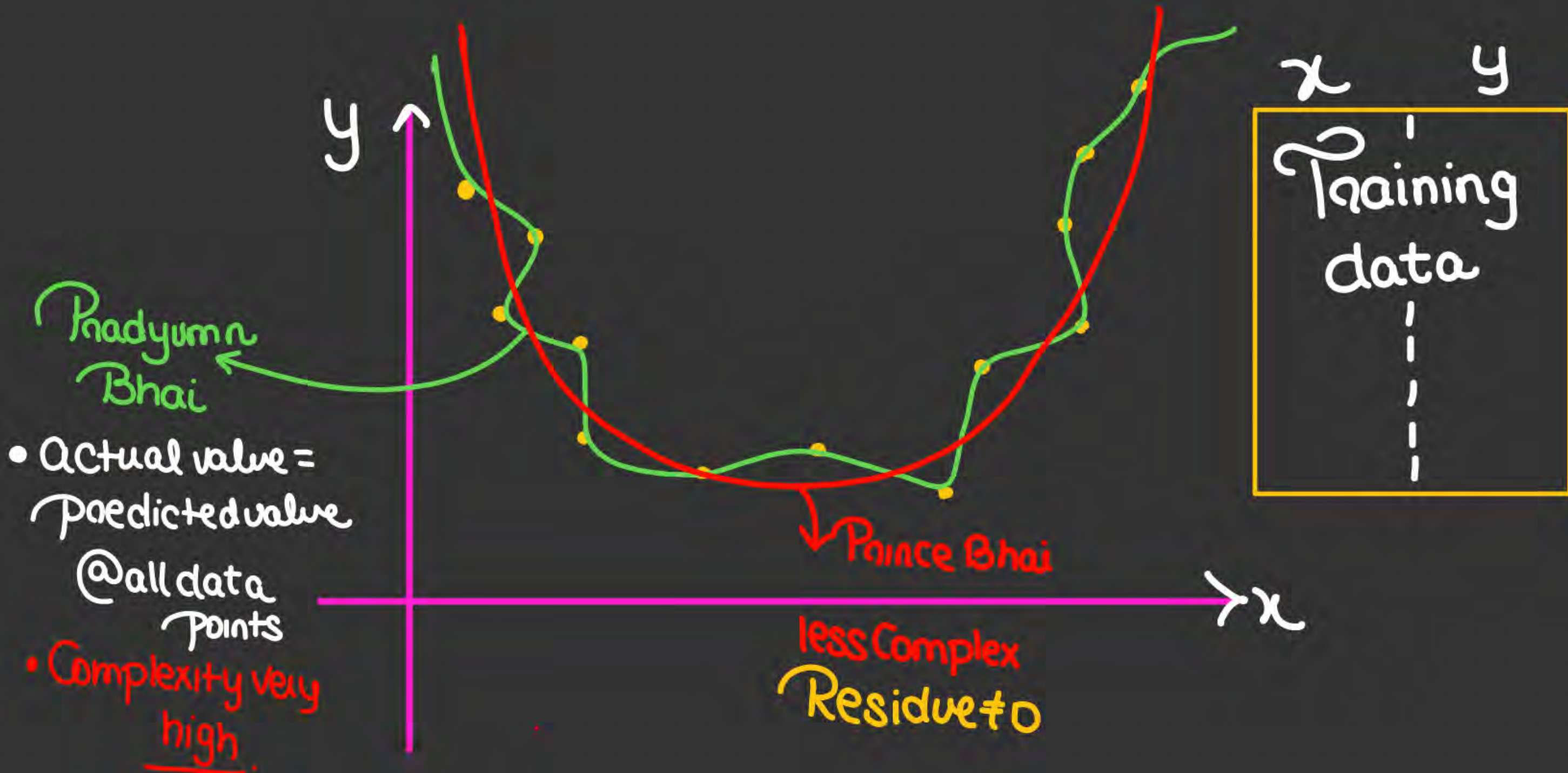
- Now you must create a mathematical model to predict rain in your city?

* So basically model will try to learn the pattern of the data ??

Training data =
Given data

Model is Simply mathematical function that try to minimize the **Residue** on the

Training data



y

x y

Training data

Follow all data points

* Role learning of data

Residue = 0

Complexity of model = high

overfitting

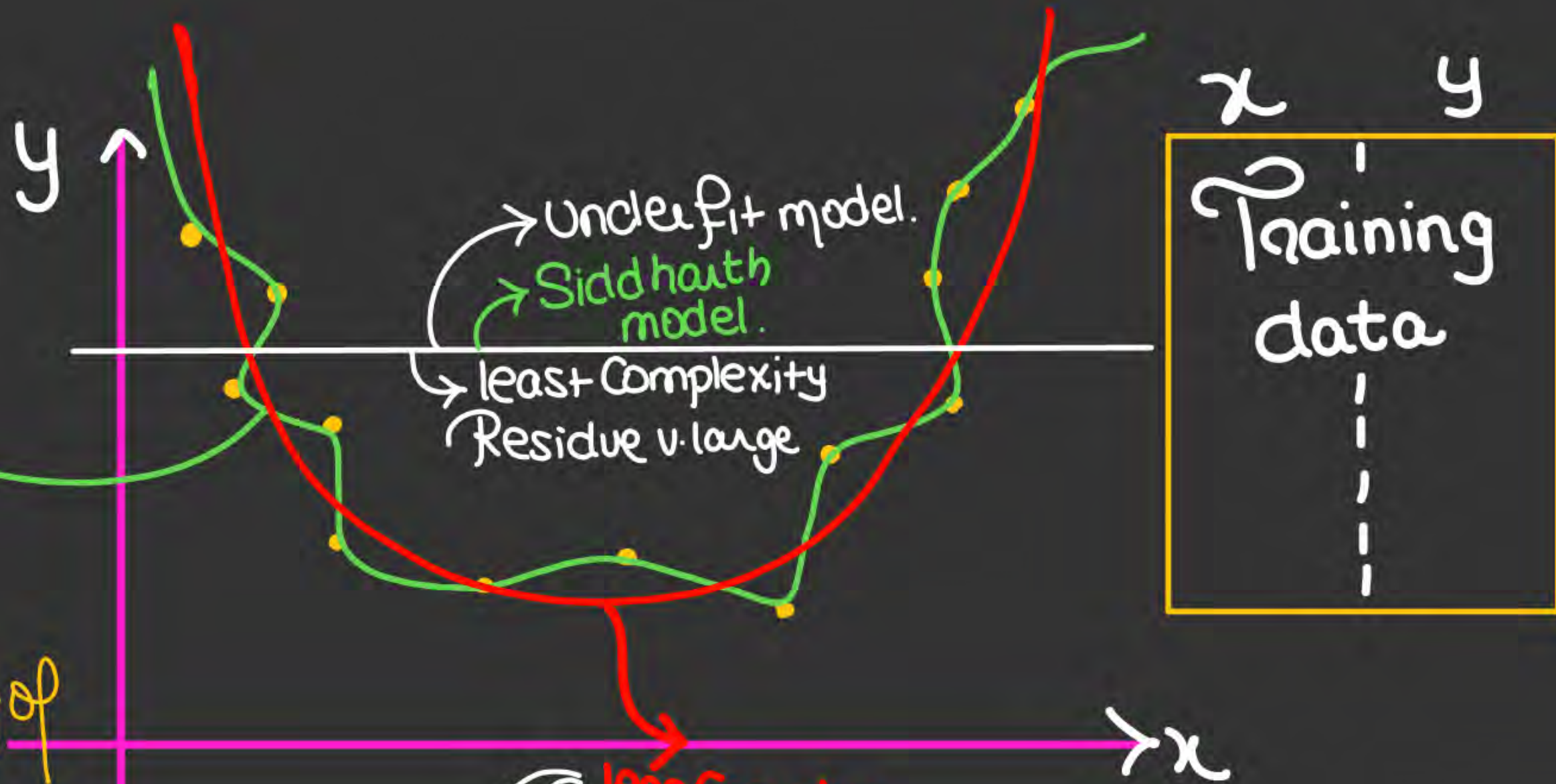
⇒ Capture noise of data

less Complex

Residue $\neq 0$, less Residue

→ Understand pattern of data

Best fit



Follow all data points

* Role learning of data

Residue = 0

Complexity of model = high

overfitting

⇒ Capture noise of data

less Complex

Residue $\neq 0$, less Residue

→ Understand pattern of data

Best fit

Simple \mathcal{P}_1 $y = c$ ✓


\mathcal{P}_2 $y = mx + c$

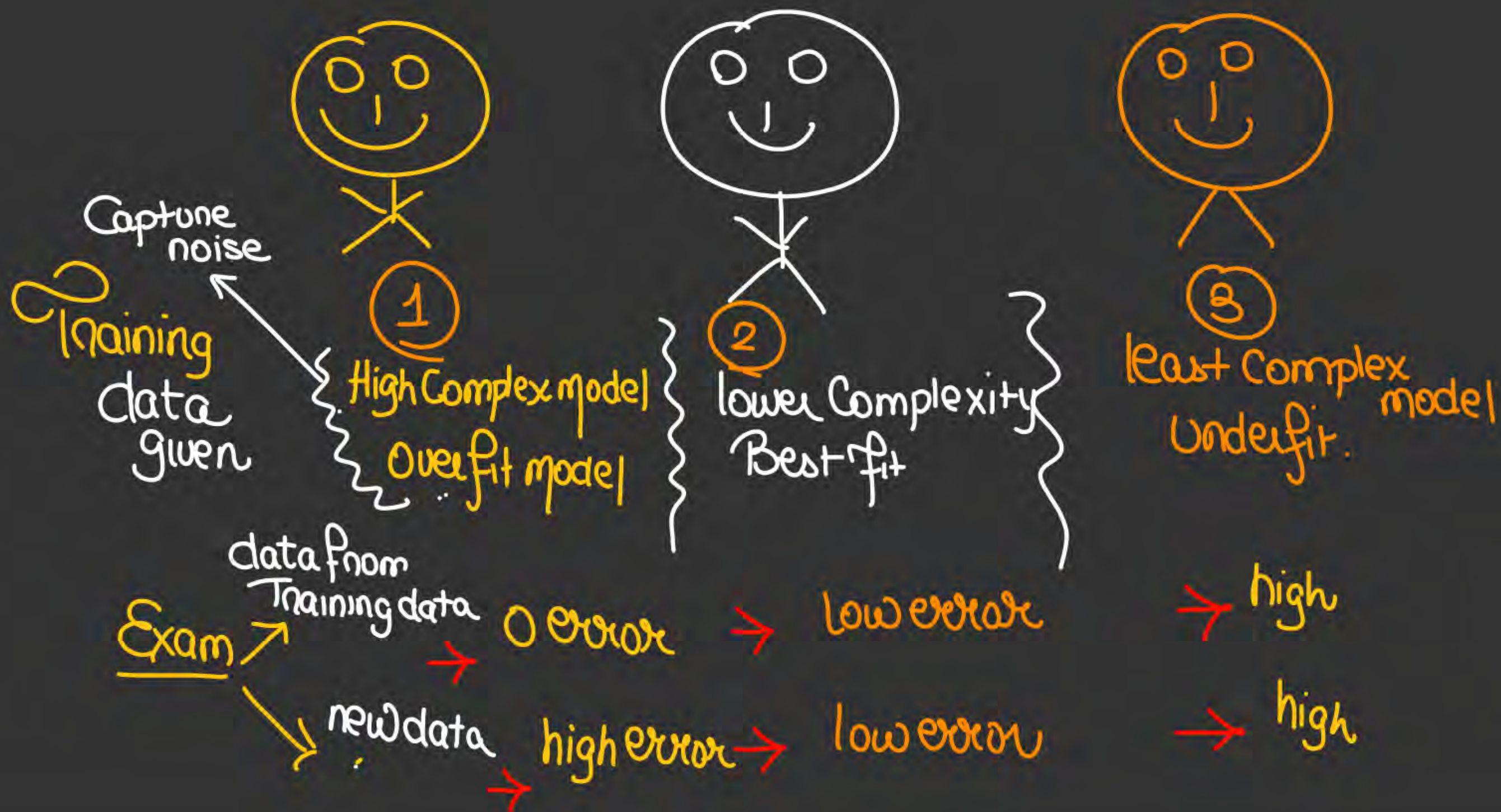
\mathcal{P}_3 $y = ax^2 + bx + c$

\mathcal{P}_4 $y = ax^3 + bx^2 + cx + d$

\mathcal{P}_5 $y = ax^4 + bx^3 + cx^2 + dx + e$

Complex \mathcal{P}_6 $y = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$







Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

- **Example of a model**



Basics of Machine Learning



Basic Understanding – Predict Rain in Your city

- **We can have a simple model**

Problem = ?? H.W.

Under fit,
model is unable to Capture
Pattern of data.

- **We can have a very complicated ~~problem~~**

Problem H.W

model



Basics of Machine Learning



Basic Understanding – Predict Rain in Your city

We can have a simple model

done



Basics of Machine Learning



Basic Understanding – Predict Rain in Your city

Problem in Simple Models ?

done



Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

We can have a very complicated problem

done



Basics of Machine Learning



Basic Understanding – Predict Rain in Your city

Problem in Complex Models ??

done✓

**Bacche
Wala
Example**



Basics of Machine Learning

Problem 1 – Predict Population of bacteria in a lab

We must create a model with following data

Time	Population
0	50
10	200

- Find out the linear fxn $y = mx + c$, best for the data
we have to find m and c , for the model.

Now predict the population at $t = 20$

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

$$\min \left[(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 \right]$$

$$\min \left[(50 - c)^2 + (200 - 10m - c)^2 \right]$$

⇒ we have to find m, c
Such that this
is minimize

x	y	$\hat{y} = mx + c$
$x_1 = 0$	$y_1 = 50$	$\hat{y}_1 = c$
$x_2 = 10$	$y_2 = 200$	$\hat{y}_2 = 10m + c$

we have a fxn, with variable
 α, β

$f(\alpha, \beta)$

To minimize f

$$\frac{\partial f}{\partial \alpha} = 0, \quad \frac{\partial f}{\partial \beta} = 0$$

we take β as const, we take α as const.

$$f(m, c) = \left[(50 - c)^2 + (200 - 10m - c)^2 \right]$$

$$\frac{\partial f}{\partial m} = \frac{\partial}{\partial m} \left[\cancel{(50 - c)^2} + (200 - 10m - c)^2 \right]$$

$$2(200 - 10m - c)(-10) = 0$$

$$200 - 10m - c = 0$$

$$\begin{aligned} \frac{\partial f}{\partial c} &= \frac{\partial}{\partial c} \left[(50 - c)^2 + (200 - 10m - c)^2 \right] \\ &= 2(50 - c)(-1) + 2(200 - 10m - c)(-1) = 0 \end{aligned}$$

$$200 - 10m - c = 0$$

$$2(50 - c)(-1) + 2(200 - 10m - c)(-1) = 0$$

$$2(50 - c) = 0$$

$$c = 50$$

$$200 - 10m - 50 = 0$$

$$m = 15$$



Basics of Machine Learning

Problem 1 – Predict Population of bacteria in a lab

Because data is too small to predict so we call an expert



Basics of Machine Learning

Problem 1 – Predict Population of bacteria in a lab

So we created a mathematical model

**What are Parameters
of a Model**



Basics of Machine Learning

Problem 1 – Predict Population of bacteria in a lab

So we created a mathematical model

What is optimisation



Problem 2 – Predict Sale of I-phone based on Age of customer

We must create a model with following data

Age	Sale of I-Phone (in a month)
30	300
40	400

Now predict the Sale of I-Phone at Age = 20

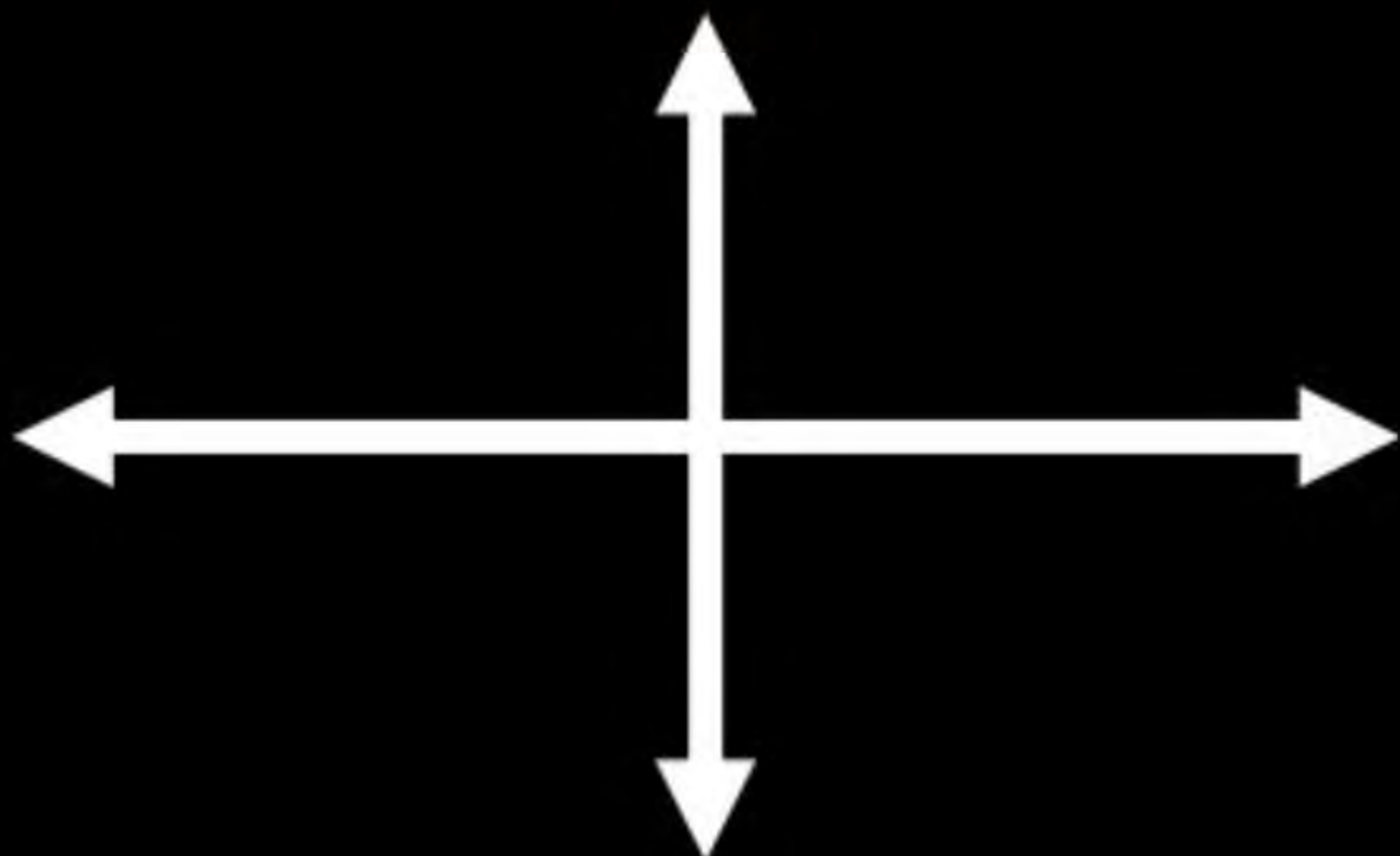


Basics of Machine Learning

Problem 2 – Predict Sale of I-phone based on Age of customer

**We don't have any expert now, and data has only two Points.
So _____**

**What is the
best model
now ?**





Problem 3 – Predict Sale of I-phone based on Age of customer

We must create a model with following data

Age	Sale of I-Phone (in a month)
30	300
40	400
50	300

Now predict the Sale of I-Phone at Age = 20

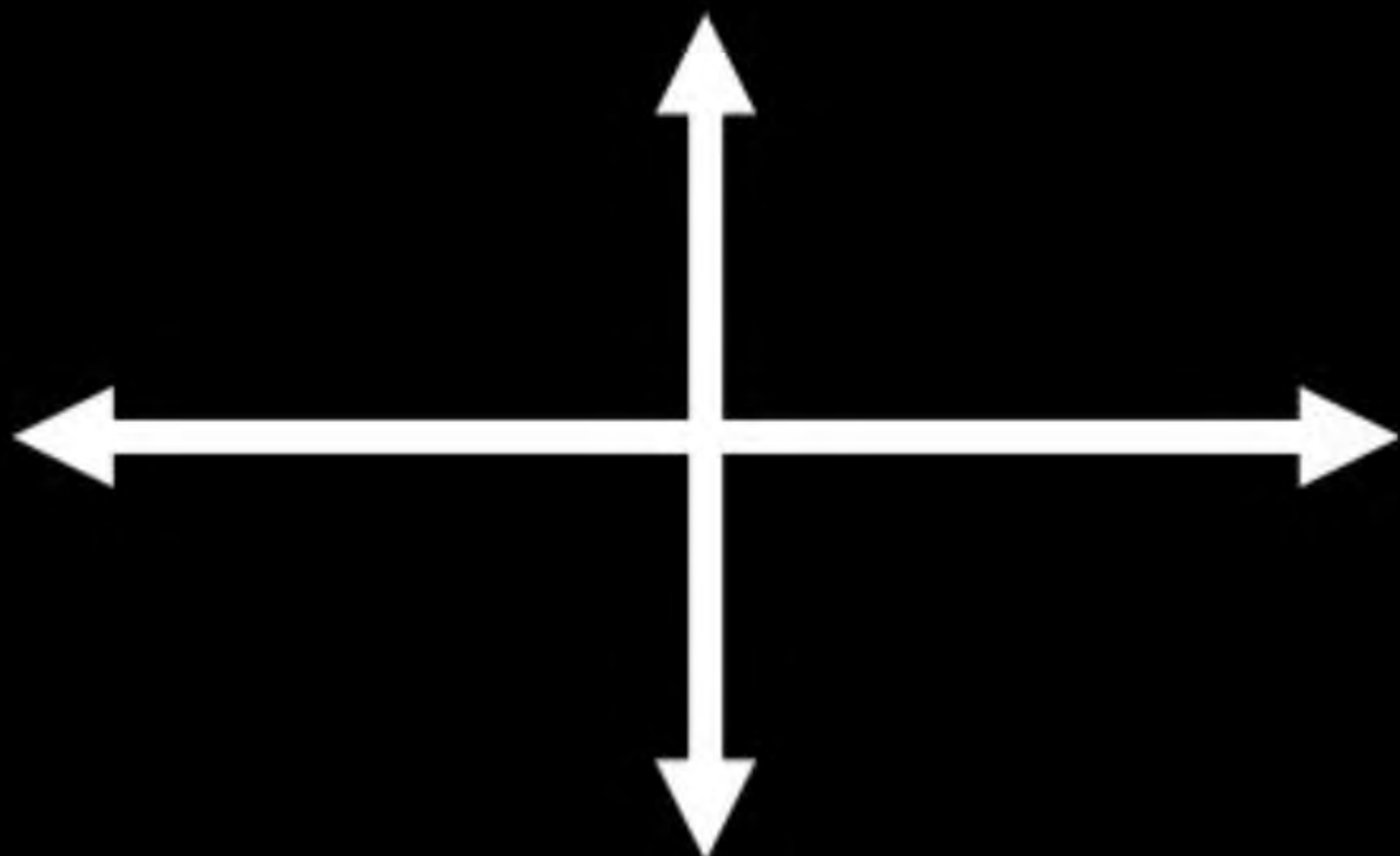


Basics of Machine Learning

Problem 3 – Predict Sale of I-phone based on Age of customer

We don't have any expert now, and data has only two Points. So

**What is the
best model
now ?**





Basics of Machine Learning

Problem 3 – Predict Sale of I-phone based on Age of customer

We don't have any expert now, and data has only two Points. So

**But we will
try to find the
linear model
only.**

**So, we must find the model
that try to**

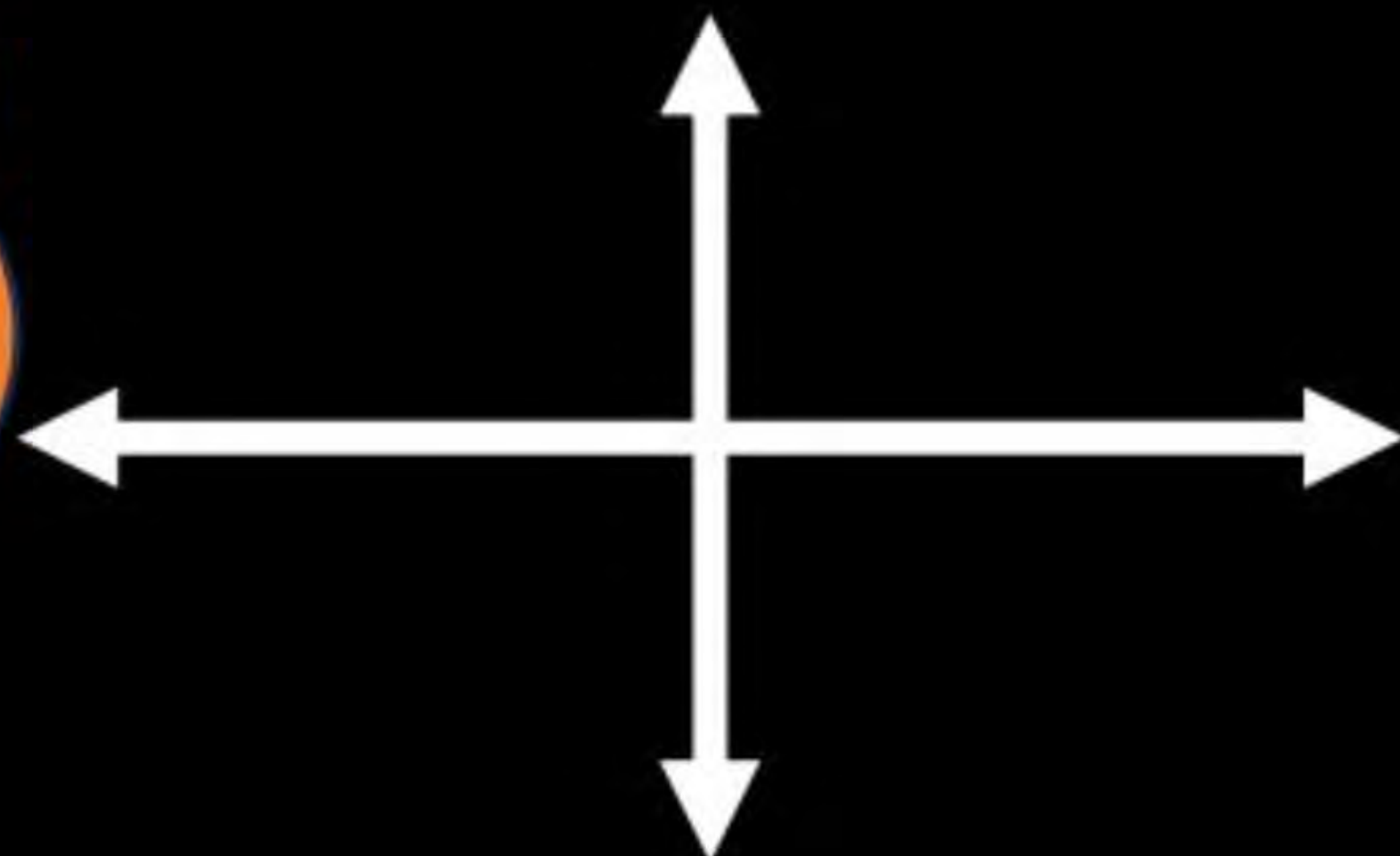


Basics of Machine Learning

Problem 3 – Predict Sale of I-phone based on Age of customer

Creating the best model

Now we have that is
called the predicted value
of input.





Basics of Machine Learning

Problem 3 – Predict Sale of I-phone based on Age of customer

Creating the best model

**Loss Functions ?? (RSS-
Residual Sum of Squares)**

The residual sum of squares (RSS), also known as the sum of squared residuals (SSR) or the sum of squared estimate of errors (SSE), is the sum of the squares of residuals



Basics of Machine Learning

Problem 3 – Predict Sale of I-phone based on Age of customer

Now how to find the best parameters ??

**Formulae to find
direct value of m
and c**



Basics of Machine Learning

Problem 3 – Predict Sale of I-phone based on Age of customer

Now how to find the best parameters ??

Example

Obtain a linear regression for the data in below table assuming that y is the independent variable.

x	1.0	2.0	3.0	4.0	5.0
y	1.00	2.00	1.30	3.75	2.25



Basics of Machine Learning



We can expect
one
Question from
here in
GATE exam

THANK - YOU