Data Science and Artificial Intelligence

Machine Learning

Regression

Lecture No. 01

By- SIDDHARTH SABHARWAL SIR

Topics to be Covered

- Topic
- Topic
- Topic
- Topic
- Topic

Basic ML

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

Syllabus

GATE 2026 FODENGE

PW

Machine Learning:

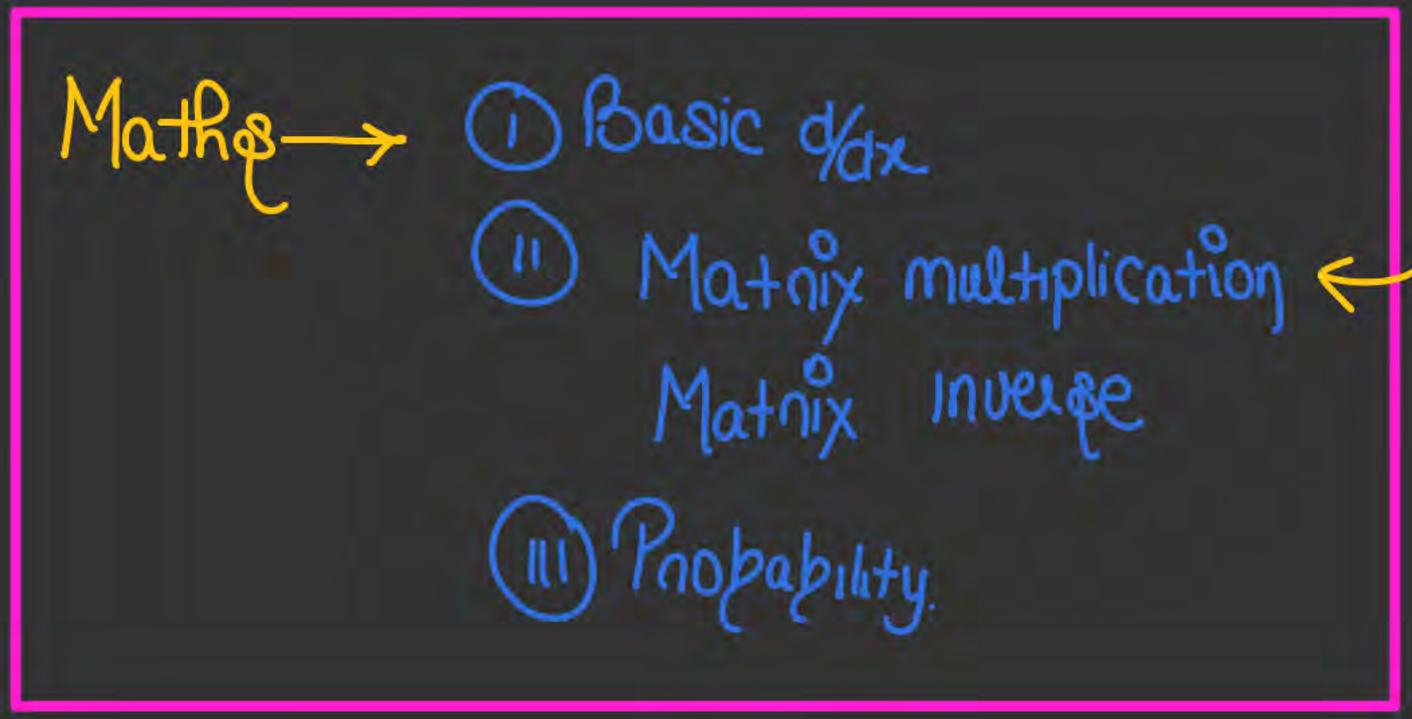
(i) Supervised Learning: regression and classification problems, simple linear regression, multiple linear regression, ridge regression, logistic regression, k-nearest neighbour, naive Bayes classifier, linear discriminant analysis, support vector machine, decision trees, bias-variance trade-off, cross-validaθon 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.

* Note making
* DPP provided
* weekly test

Chapter wise.

To crack the exam
Course material will
be more than sufficient


Maths →
(I) Basic d/dx
(II) Matrix multiplication
Matrix inverse
(III) Probability.
Before Topic Revise.

- GATE 2026 – Theory | Numericals.

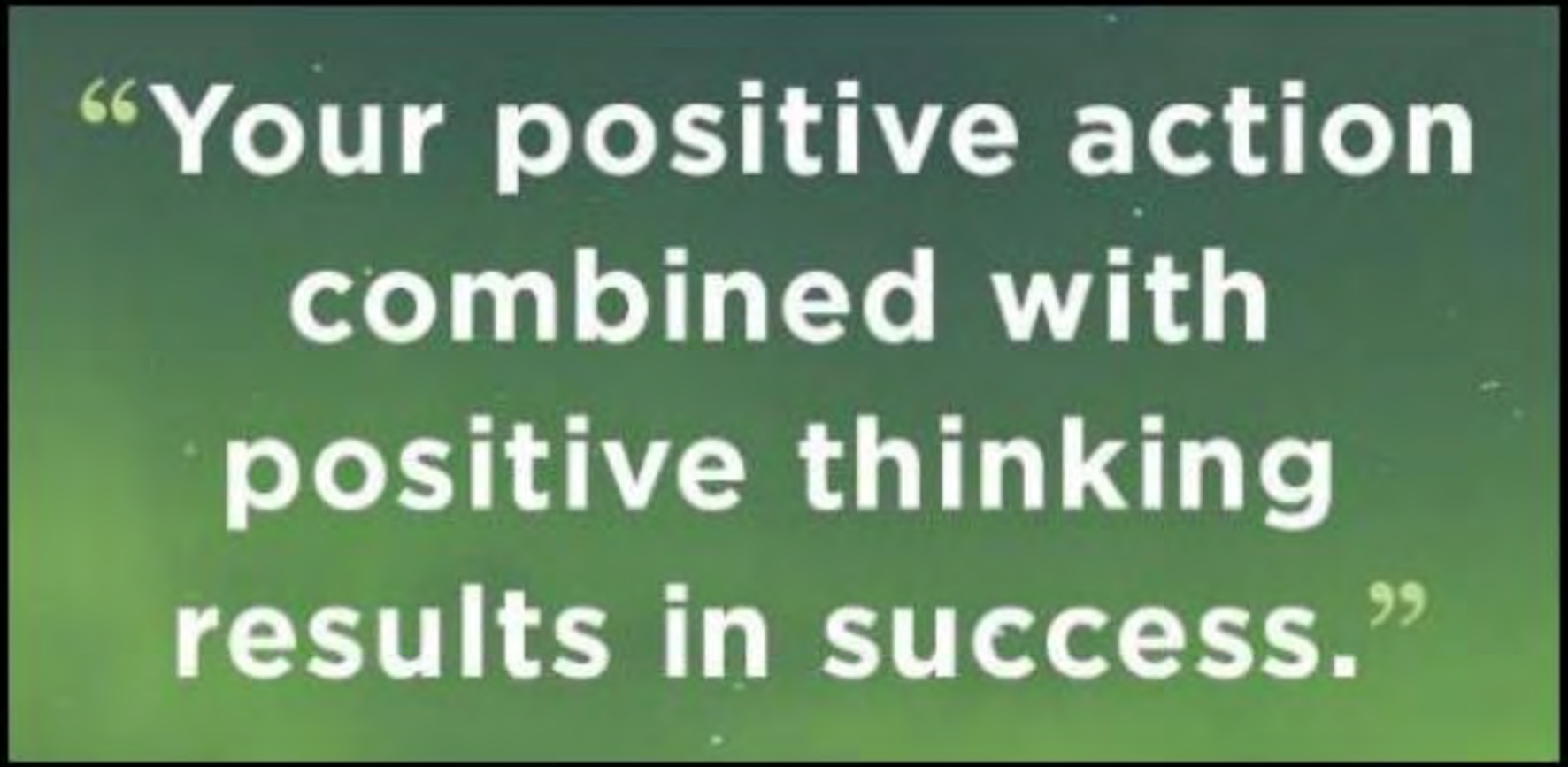
weightage :–

- Maths :– 35 Marks ← done
- AI/ML ⇒ 20-25 marks

- DS-AI ⇒ CS | EC | EE | CE | all are equal

Basics of Machine Learning

- What is Machine Learning
- What is optimisation
- What is a model


"Your positive action
combined with
positive thinking
results in success."

Basics of Machine Learning

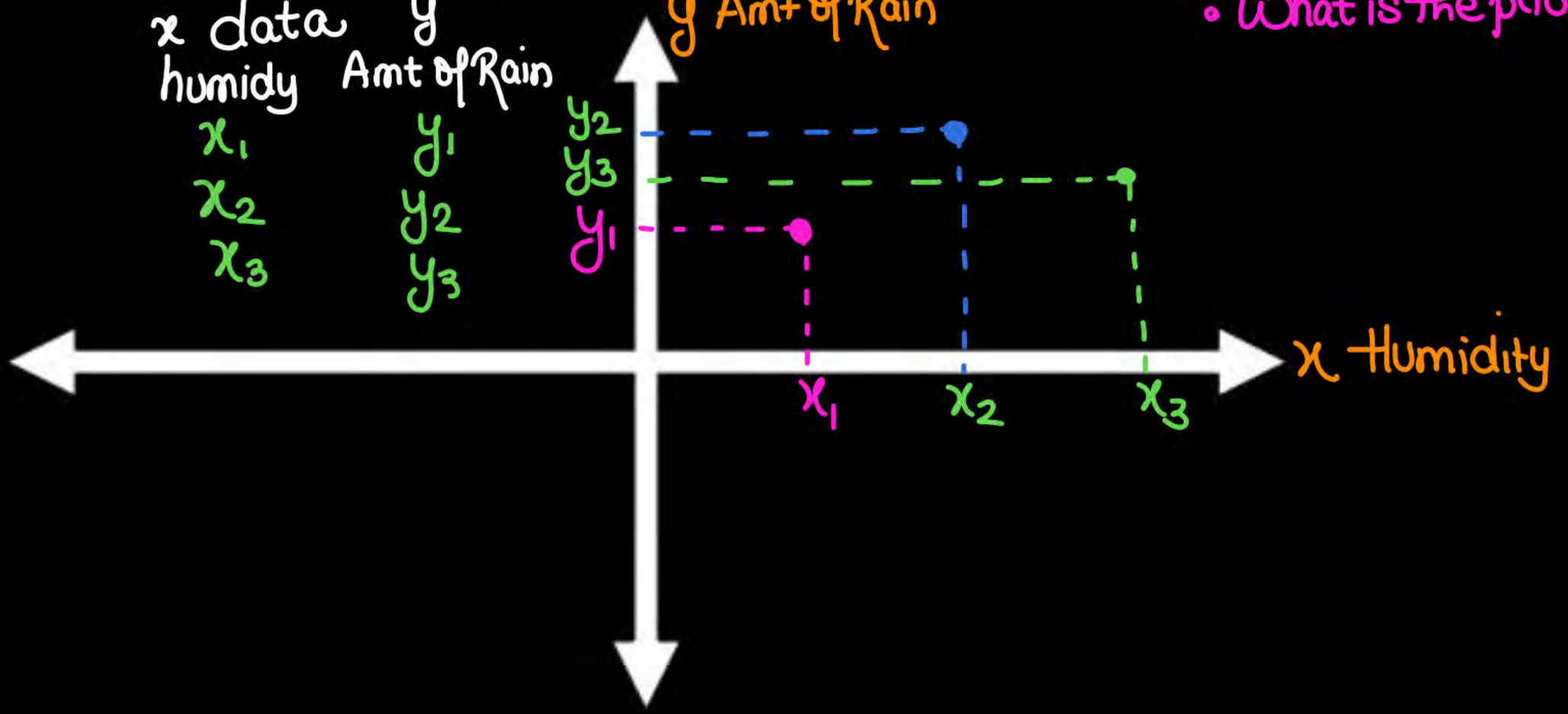
??

What is the General Problem in Machine Learning

Lets start with a basics

x data humidy: x_1, x_2, x_3

y Amt of Rain: y_1, y_2, y_3



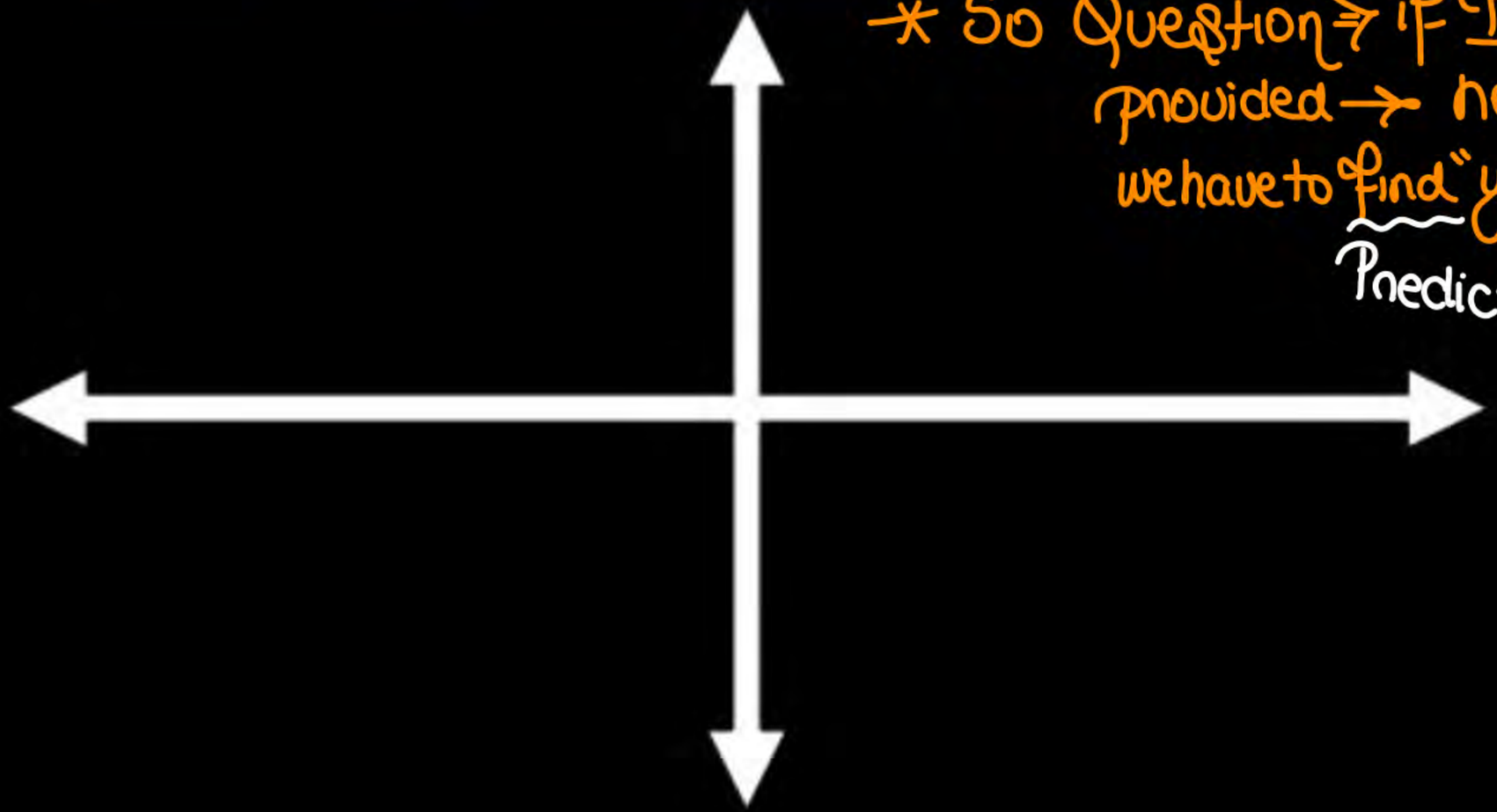
- What is the problem Statement?

Basics of Machine Learning

What is the General Problem in Machine Learning

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

* So Question ⇒ if I Provide new x value
Provided → new humidity value
we have to find "y" → amt of Rain
Predict




data given
y, x
data ki analysis
and task is to predict
y for new x

Basics of Machine Learning

What is the General Problem in Machine Learning

This is called a function of "independent" variable

$y = f(x)$

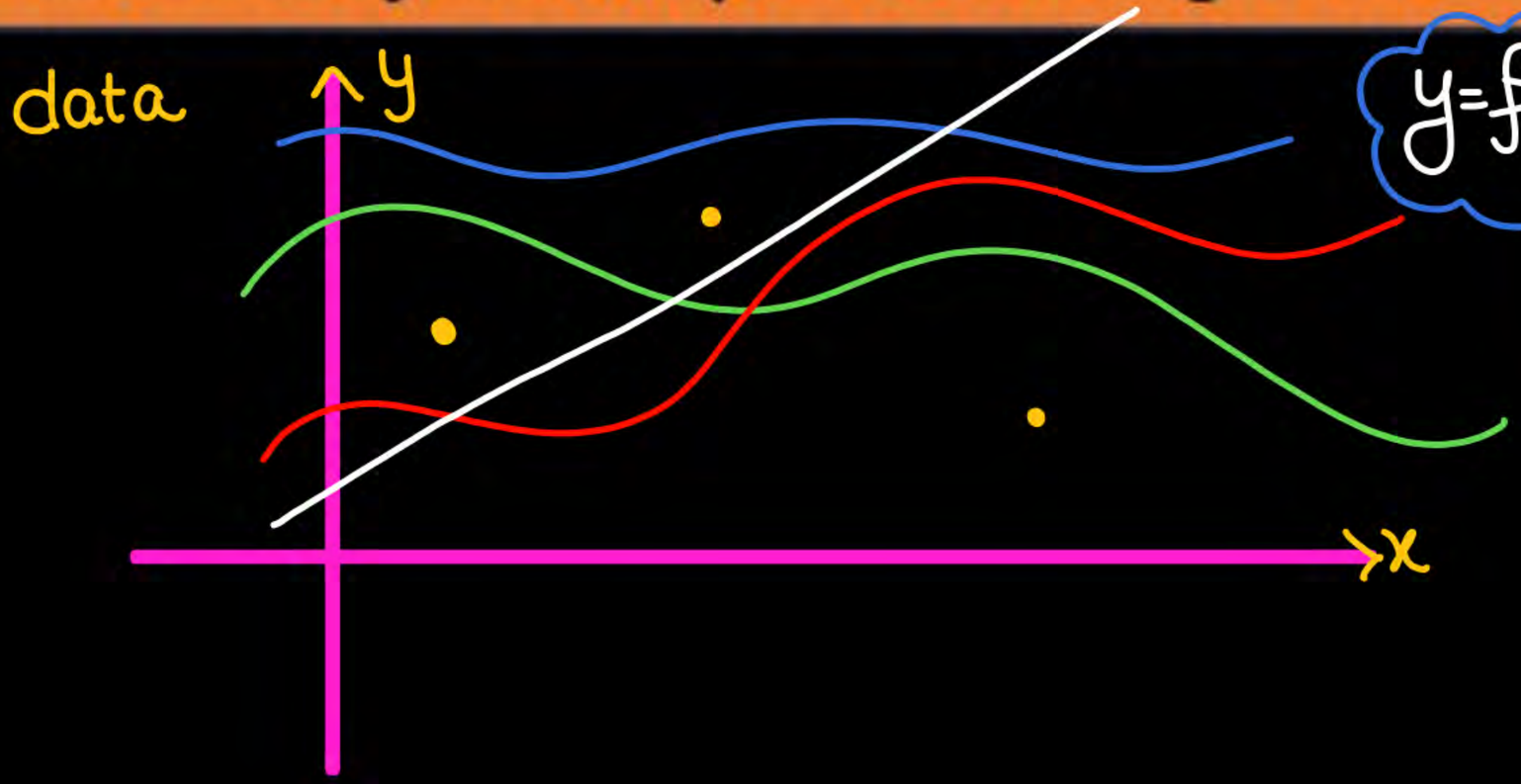
- y → * dependent variable
- x → * Independent variable

- So in ML, we analyse the data and find a relation b/w y and x
- $y = f(x)$ → Once we get this fxn → then for any x we can find y.

Basics of Machine Learning

What is the General Problem in Machine Learning

These already known points relating Y and X is called Data



$y = f(x)$ → this can be seen/visualised as a plot b/w y and x.

→ So we have ∞ option/possibilities of fxn

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.

- So using the data our ML-algorithm will give us only one fxn $y=f(x)$ from ∞ possibilities

Step in training Process

- So optimization is the process followed by ML algorithm to give us the best fxn or best Relation b/w y and x. for given data set.

What is the General Problem in Machine Learning

What is Data

It refers to the set of observations or measurements that can be used to train a machine-learning model.

- data is simply values of y and x used to find relation b/w y and x, $y = f(x)$ after training Process.

Basics of Machine Learning

What is the General Problem in Machine Learning

How we collect data

faltu

(i) By Measurement/Survey

(ii) Buy data from Source

(iii) Consult expert.

Basics of Machine Learning

What is the General Problem in Machine Learning

Lets take an example of predicting rain

done

Basics of Machine Learning

What is the General Problem in Machine Learning

Lets take an example of predicting rain

done

Basics of Machine Learning

GATE 2026 FODENGE

PW

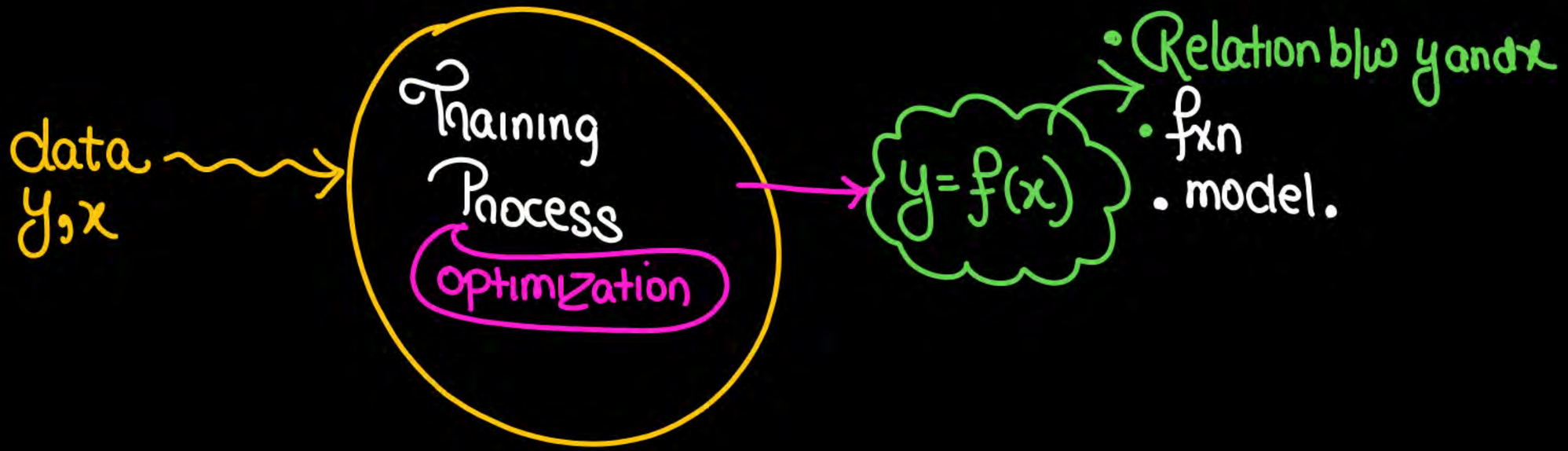
What is the General Problem in Machine Learning

What is the problem Statement

done

What is the General Problem in Machine Learning

What is a Model



Basics of Machine Learning

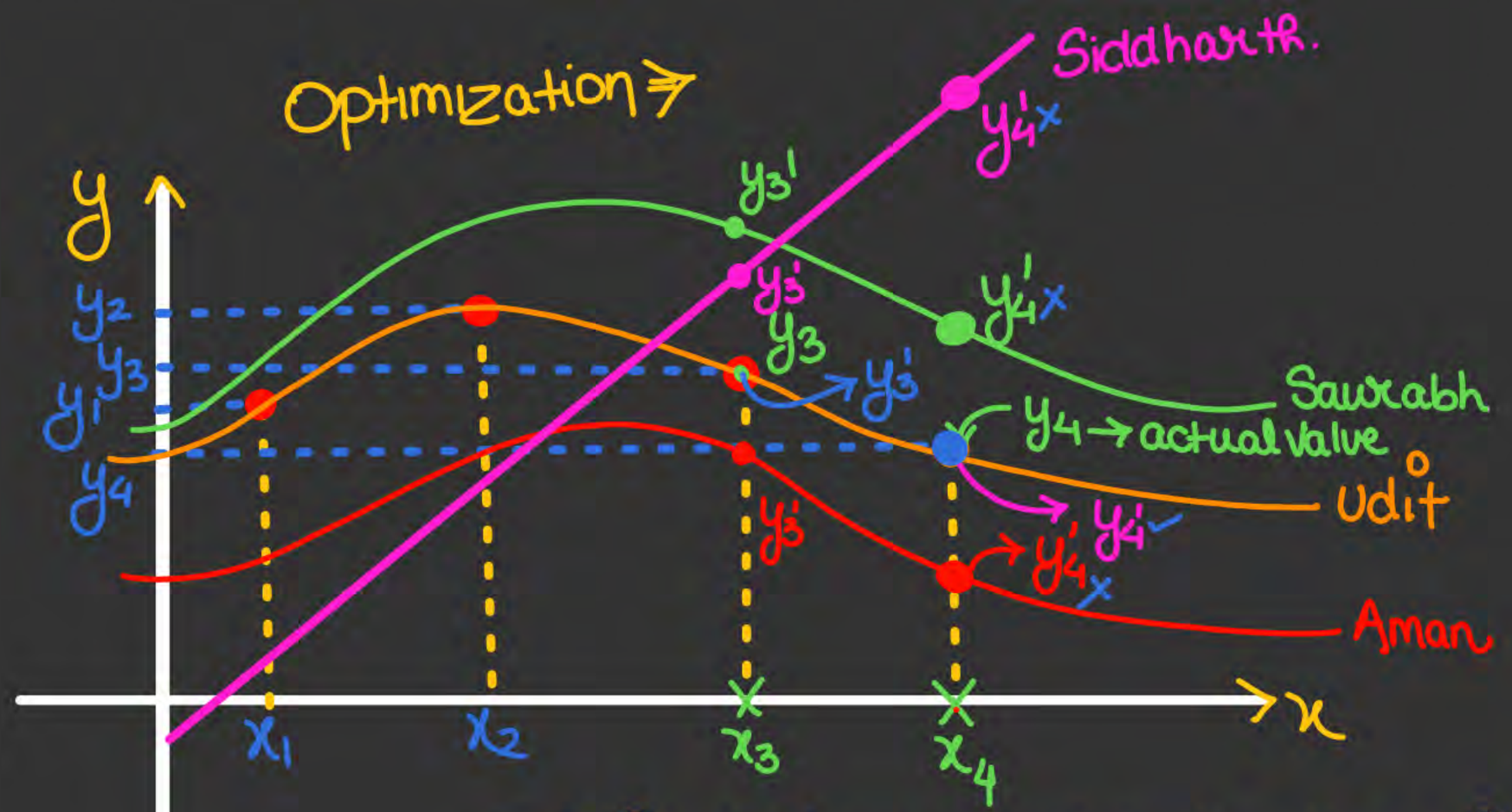
What is the General Problem in Machine Learning

What is the Optimization

(i) Part of training Process

(ii) using this we find best model for given data — Kaise ??

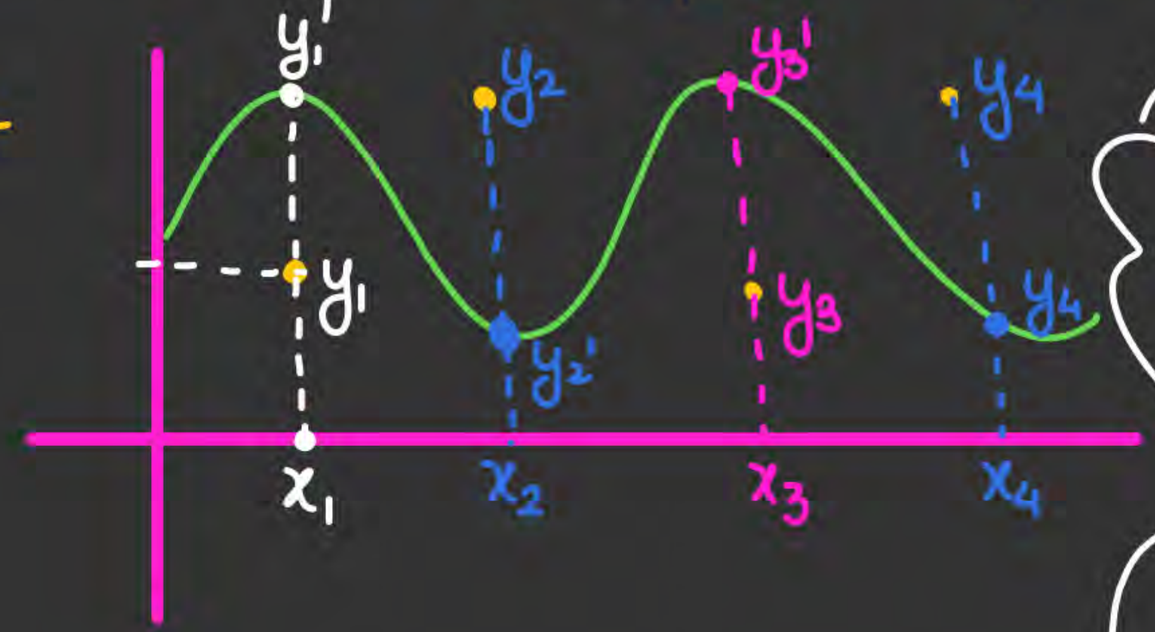
Optimization →



→ So we choose model which has min value of gap b/w actual and predicted values.

Model ⇒ $\min \left|(y_1 - y_1')\right| + \left|(y_2 - y_2')\right| + \left|(y_3 - y_3')\right| + \left|(y_4 - y_4')\right|$

4 data point



So to check which model is best we chose model which has minimum

$$\Rightarrow \sum_{i=1}^{N} \left|y_i - y_i'\right|$$

$$\min \Rightarrow \sum_{i=1}^{N} (y_i - y_i')^2$$

So the best model is which minimizes the sum of |error| blw actual and predicted values of y for Trainingdata/availabledata.

Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

What should be the format of the data ?

Later

We have two type of Data : -

Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

What should be the format of the data ?

Later

We have two type of Data : -

Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

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

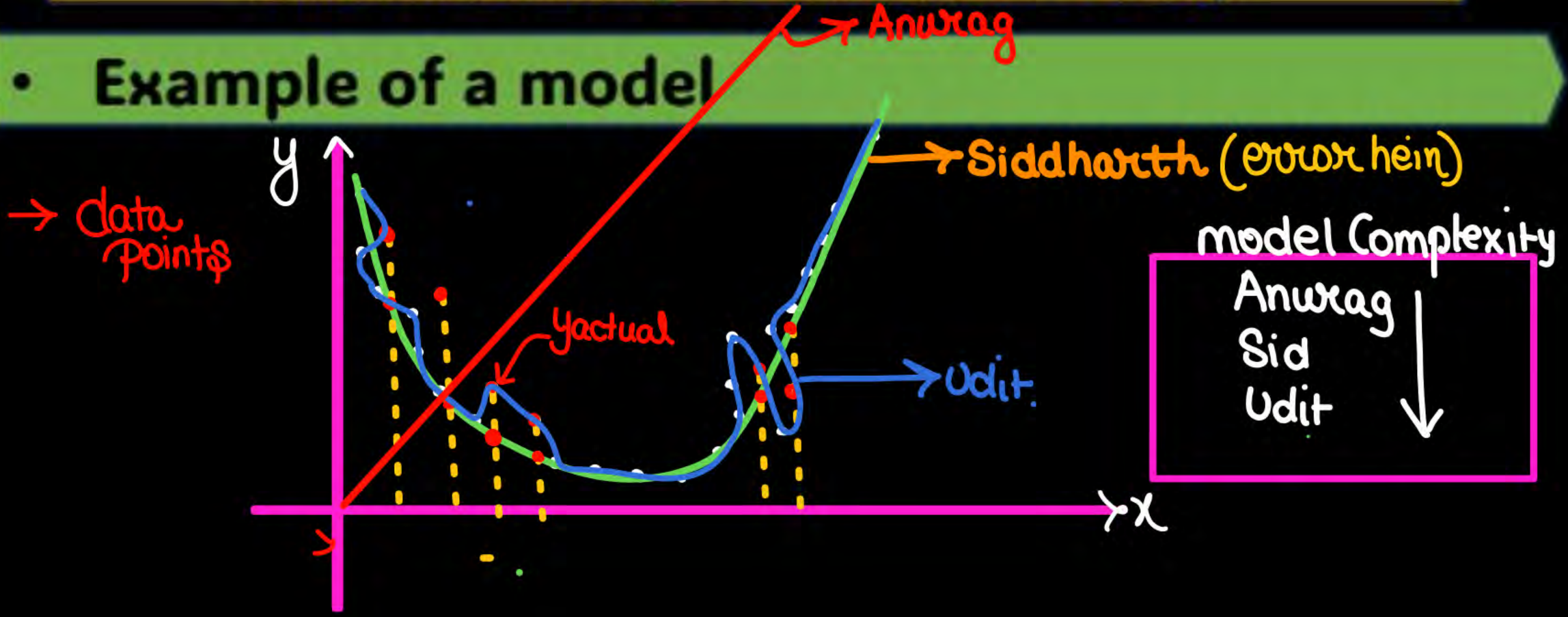
So basically model will try to learn the pattern of the data

Model is Simply mathematical function that try to minimize the error on the Training data/available data.

Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

- **Example of a model**



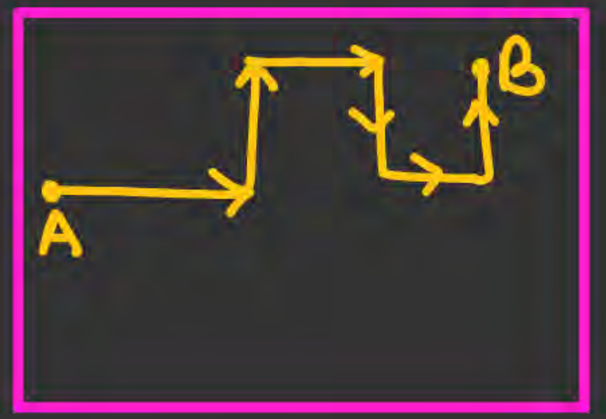
$y = ax + b$ ← Simplest

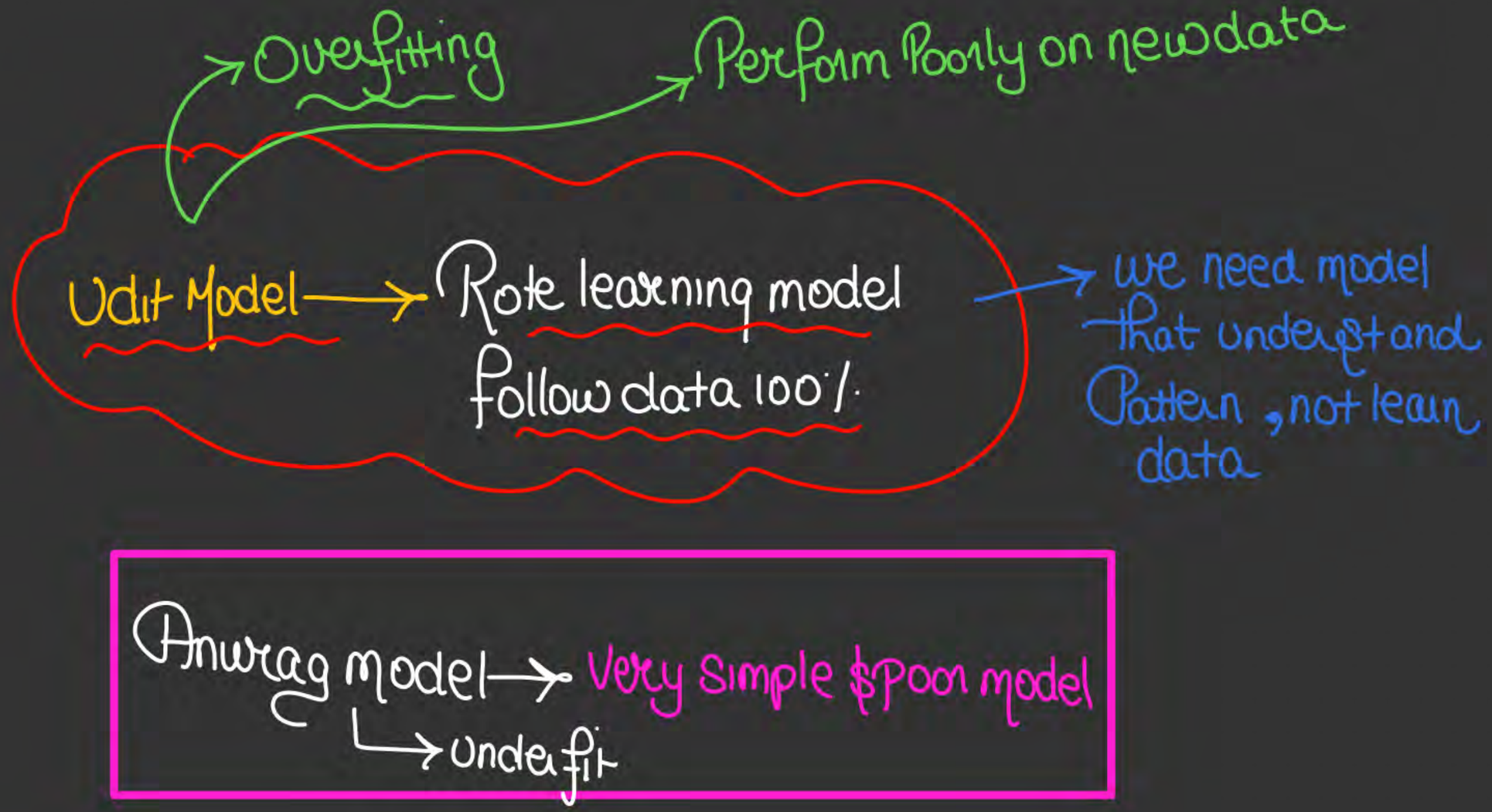
$y = ax^2 + bx + c$

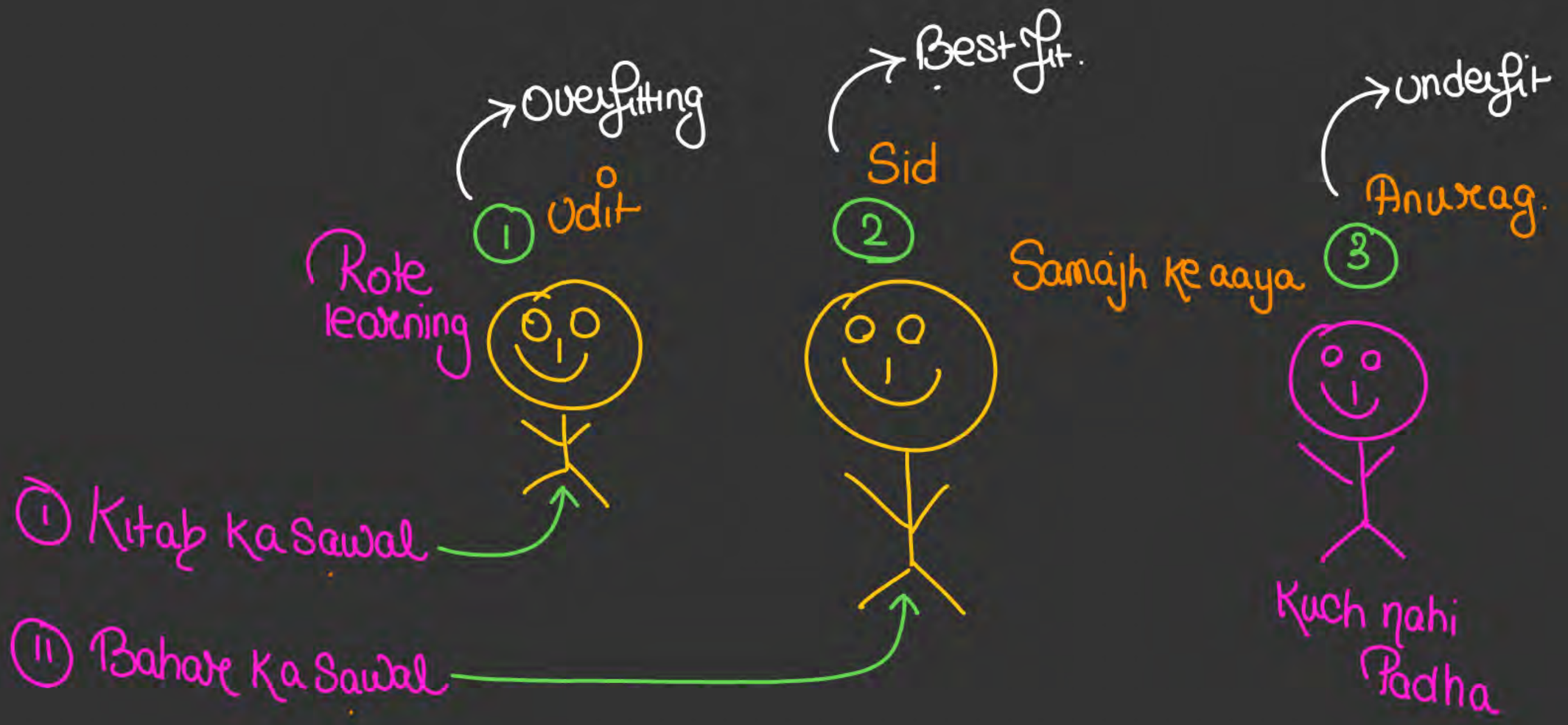
$y = ax^3 + bx^2 + cx + d$

$y = ax^4 + bx^3 + cx^2 + dx + e$ ← most Complex.

Complexity inc




Overfitting
Perform Poorly on new data
Udit Model
Rote learning model
follow data 100%
we need model that understand Pattern, not learn data
Anurag model
Very Simple $Poor model
underfit


Overfitting
Best fit.
Underfit
Udit
Sid
Anurag.
1
2
3
Rote learning
Samajh ke aaya
Kuch nahi Padha
① Kitab ka sawal
② Bahar ka sawal

Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

- We can have a very simple model → Underfit
  Kuch bhi nahi Padha
- We can have a very complicated problem
  → Rote learning / Data Yaad not Data pattern.

Basics of Machine Learning

Basic Understanding – Predict Rain in Your city

We can have a simple model

done

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

GATE 2026 FODENGE

PW

Basic Understanding – Predict Rain in Your city

Problem in Complex Models ??

done

- We want to create a st. line fxn

$y = mx + c$

for given data

$x = 5$ $y_{actual} = 15$

$y_{predic} = 5m + c$

$x = 10$ $y_{actual} = 25$

$y_{predicted} = 10m + c$

Bacche Wala Example

Basics of Machine Learning

Problem 1 – Predict Population of bacteria in a lab

We must create a model with following data

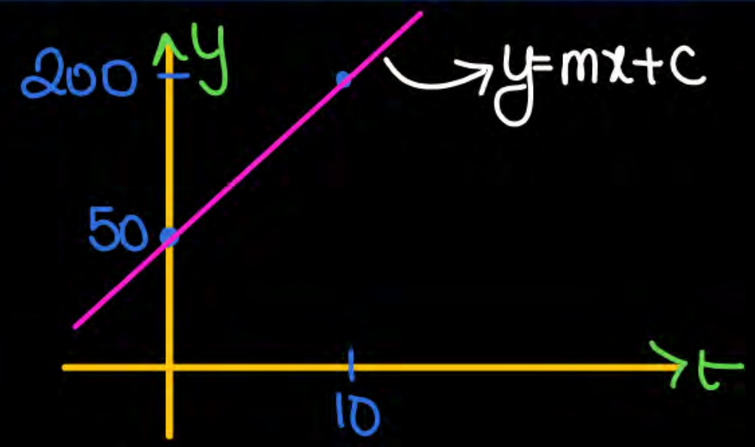
Since we have to Predict POP → y
Time → x

If Straight line eq
$y = mx + c$

Time	Population
0	50
10	200

x Yactual

Time	Actual	Predicted
0	50	c
10	200	10m + c



Now predict the population at t = 20

We need that model $\Rightarrow$ $\min \sum_{i=1}^{N} (y_i^o - y_i')^2$

$$\min \left[(y_1 - y_1')^2 + (y_2 - y_2')^2 \right]$$

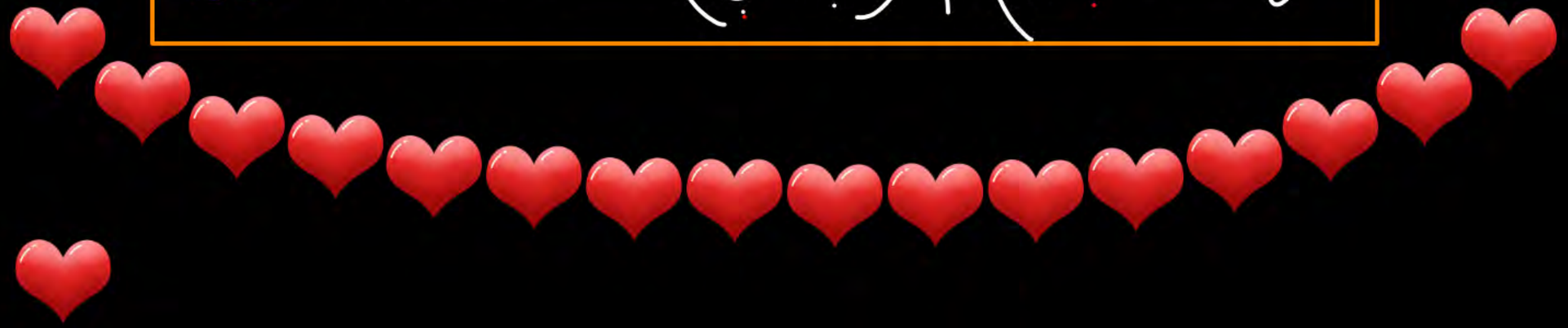
$$\Rightarrow \min \left[(50 - c)^2 + (200 - (10m + c))^2 \right]$$

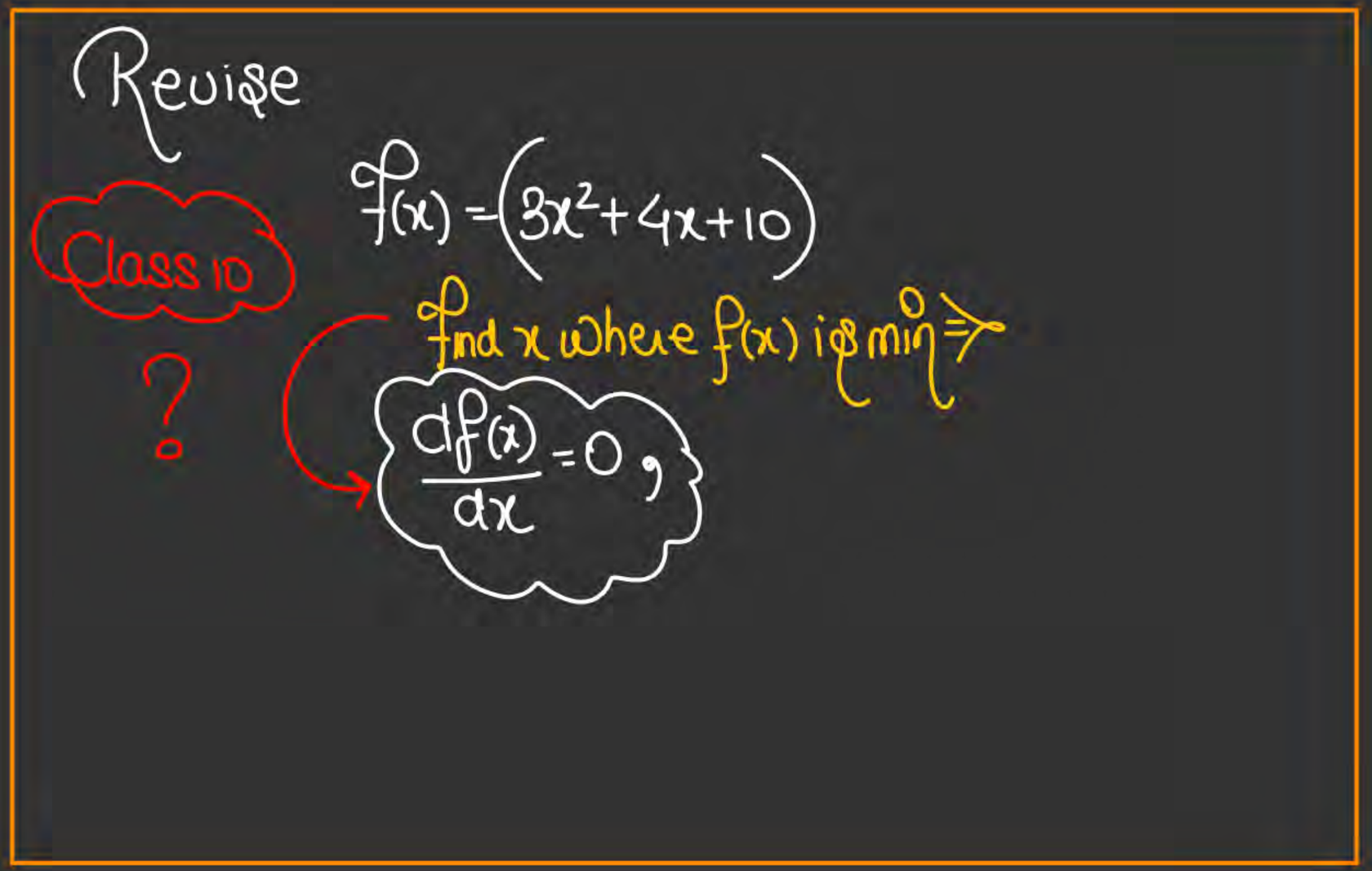
Basics of Machine Learning

Problem 1 – Predict Population of bacteria in a lab

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

- So Minimize $\Rightarrow$ $(c-50)^2 + (10m+c-200)^2$




Revise
Class 10
?
$f(x) = (3x^2 + 4x + 10)$
Find x where f(x) is min ⇒
$\frac{df(x)}{dx} = 0$,

THANK - YOU