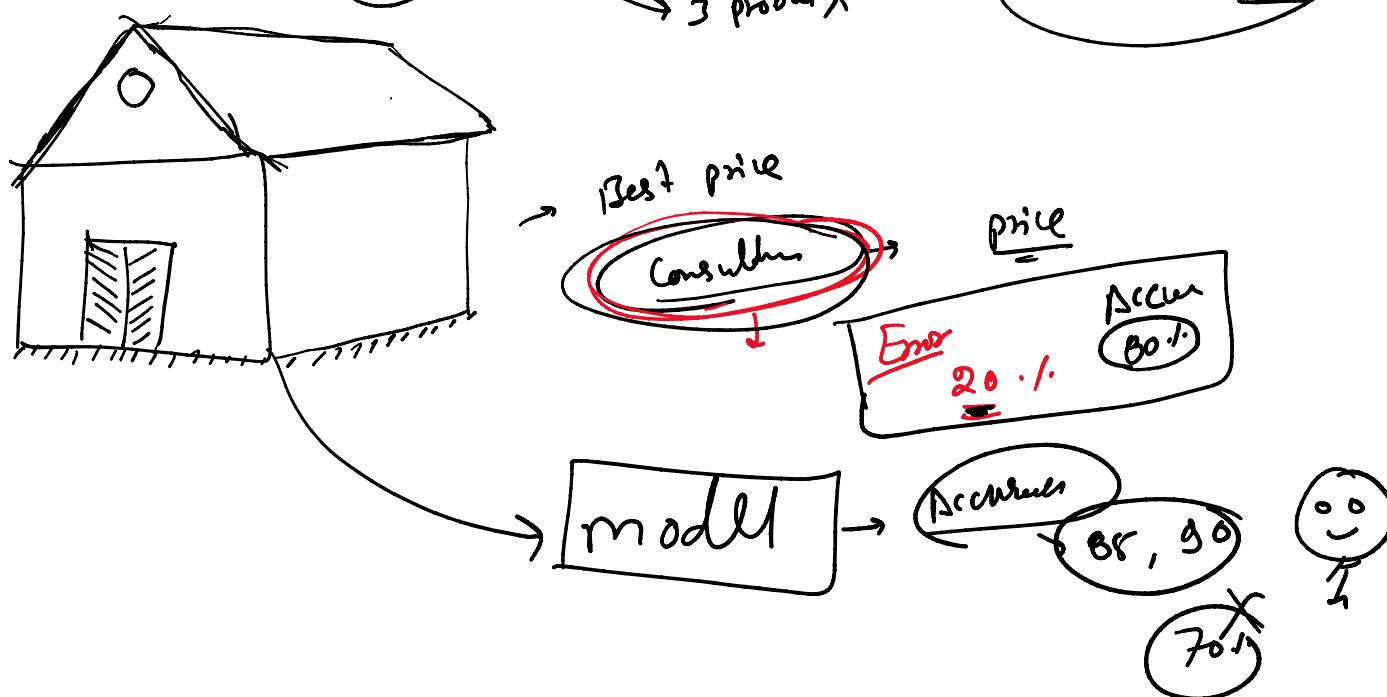
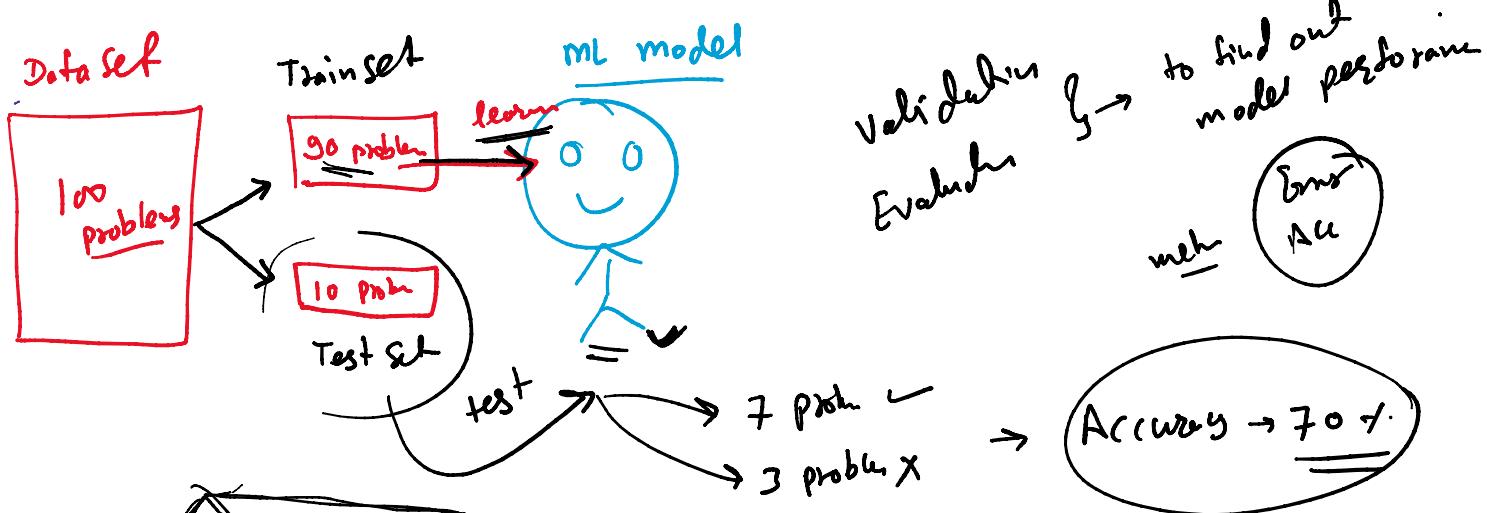


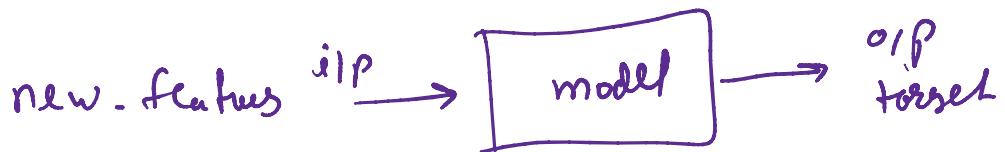
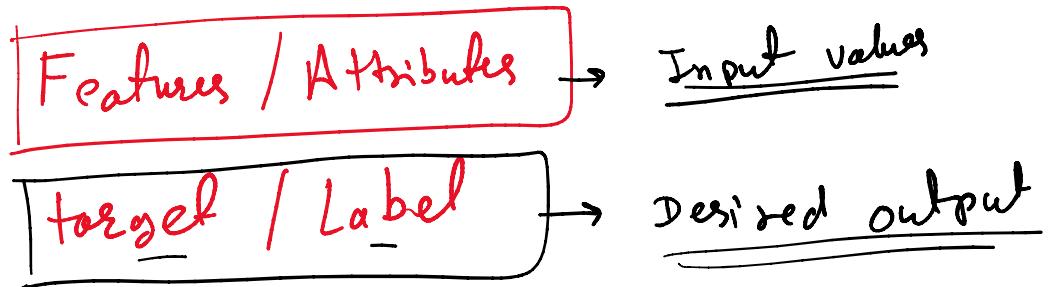
Model → ML program  
DataSet → complete data (Population)

Train DataSet → part of dataset which we will use to train our Algoithms

Test DataSet → part of DataSet which is never exposed to ML model during training and it is used to Evaluate model's performance.

Data Leakage





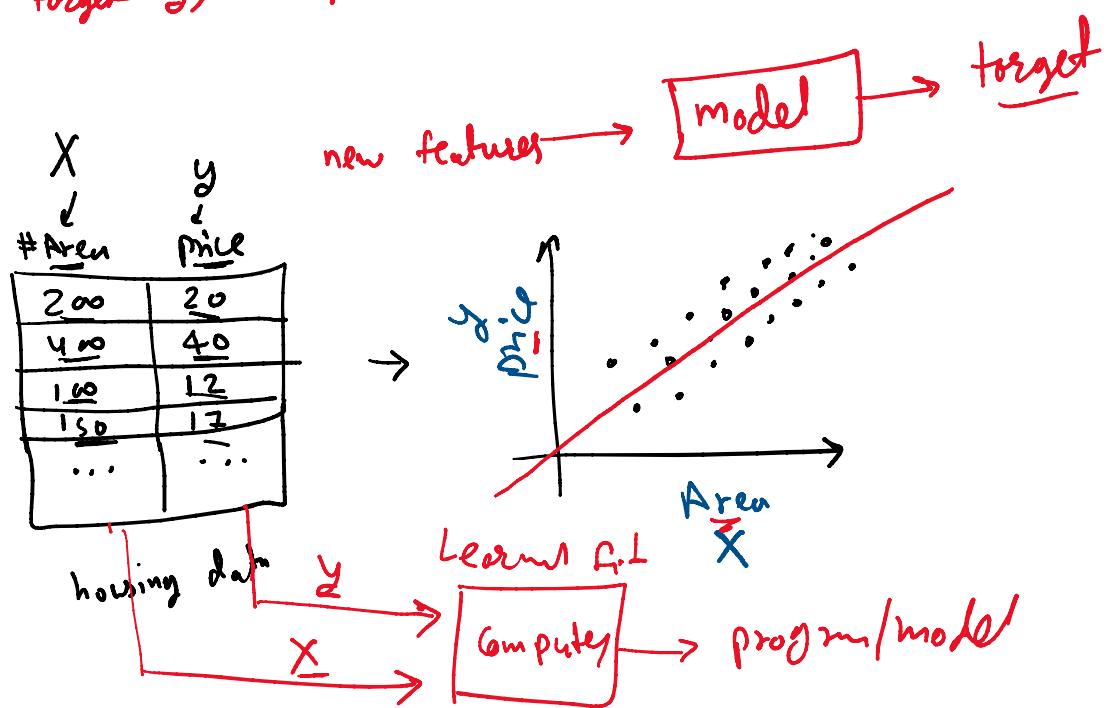
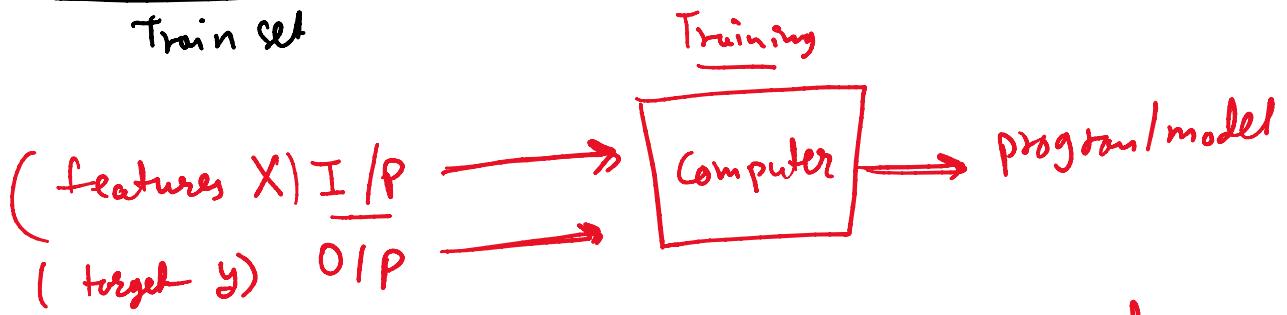
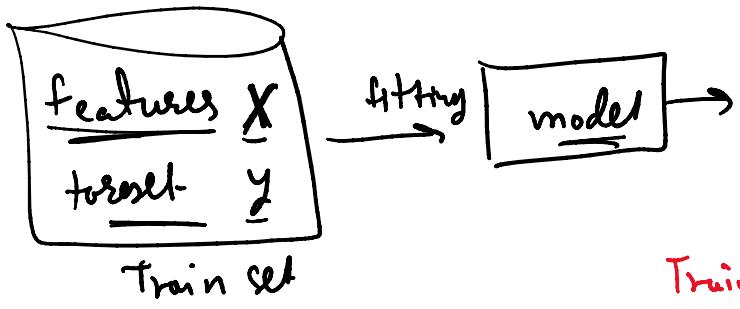
## # House Price Prediction

features  
age, crime, area, location, population, education  
furnished or not furnished, no of rooms, no bedrooms  
floor, pollution, ...

target / Label  
price

## ① Supervised ML

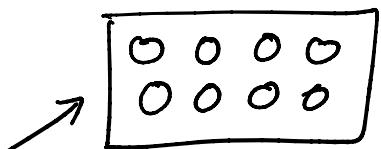
Training Data Set Contains features/attributes  
as well as label or target.

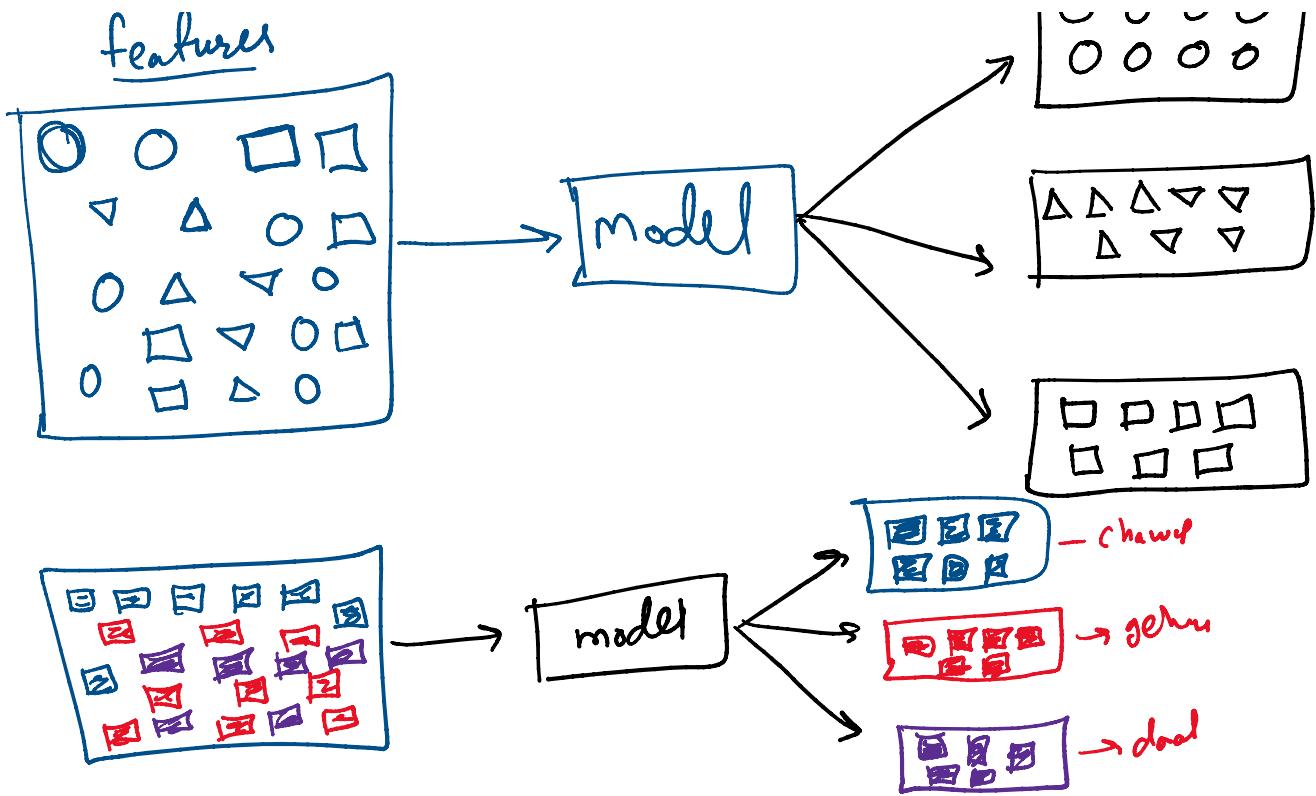


$\Rightarrow$  Unsupervised Machine Learning  $\rightarrow$

Training Dataset only contains features

features





Tomorrow

→ Reinforcement

→ process to create a ml program

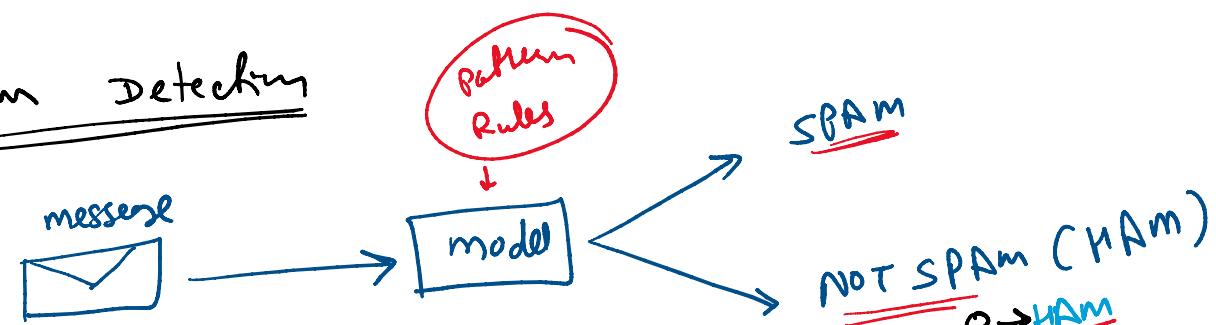
→ Chapter - 1

types of ml

steps

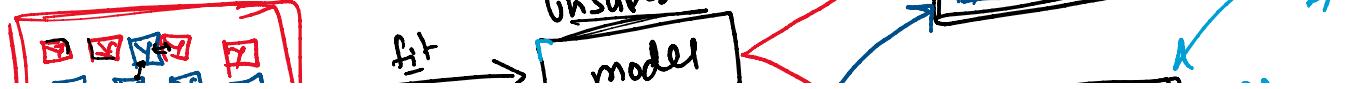
challenges

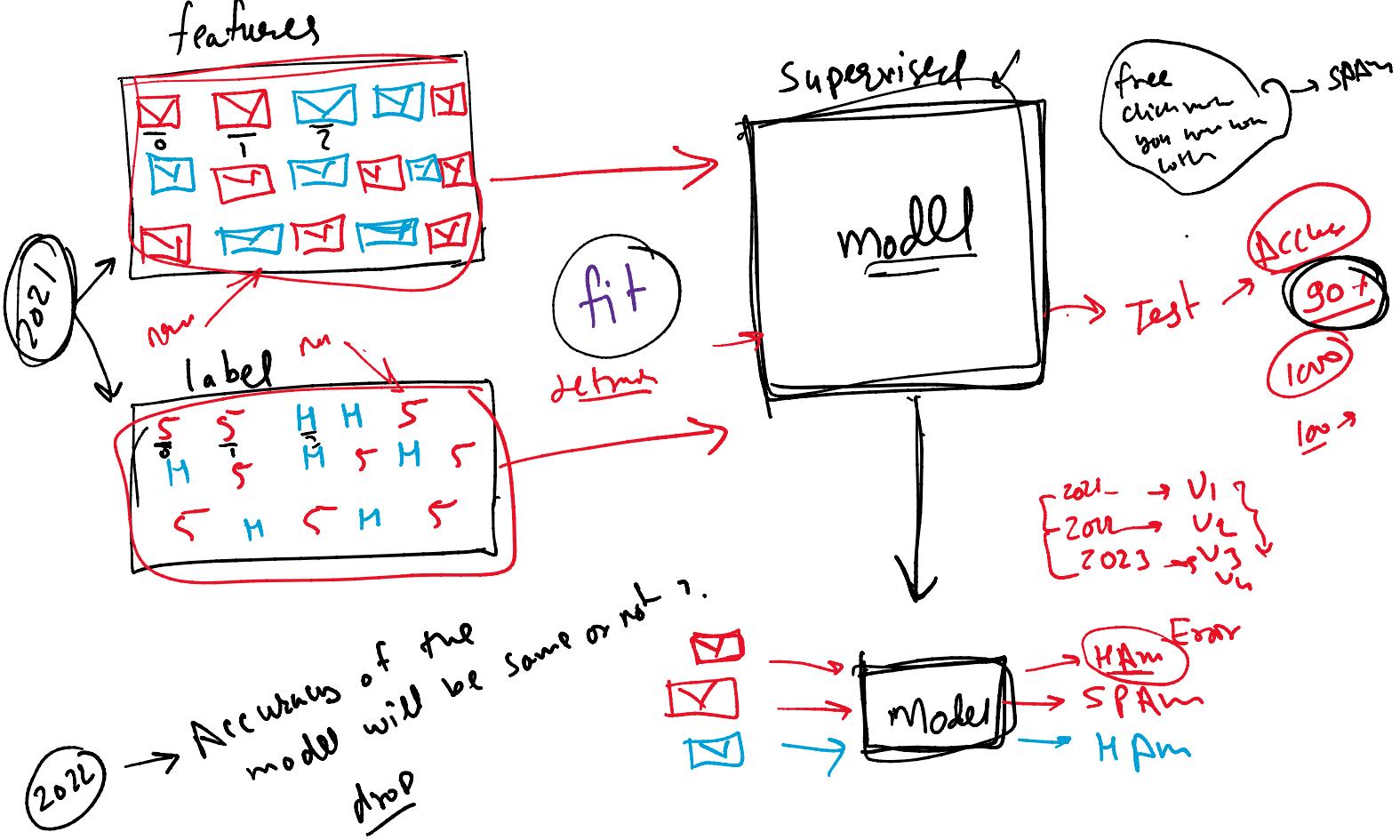
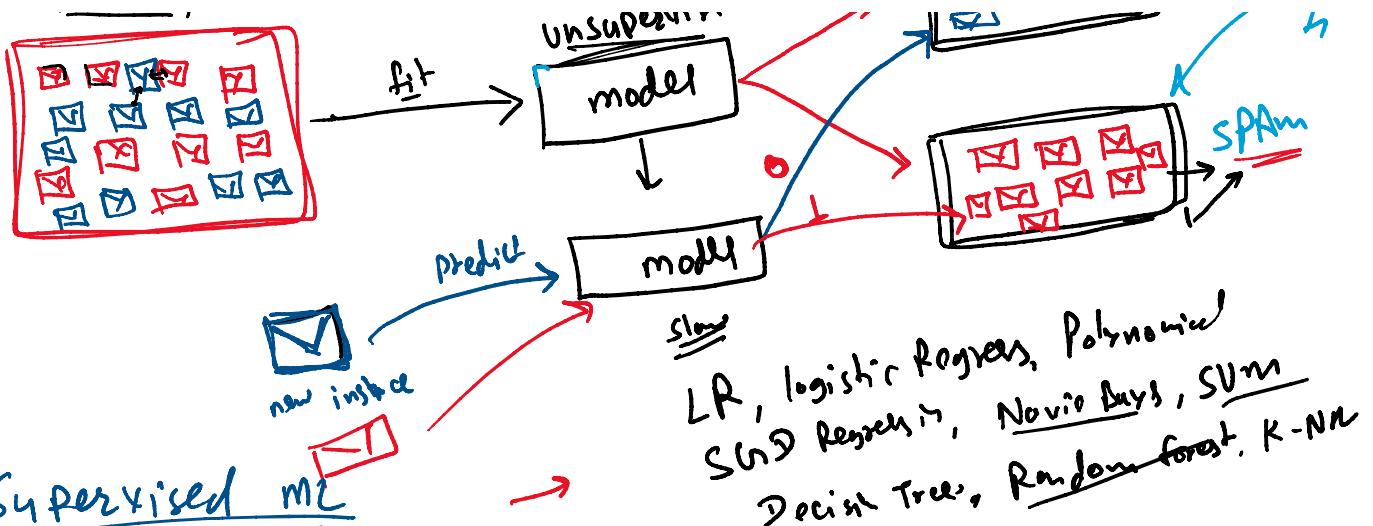
→ SPAM Detection



→ Instance Based ML Algo

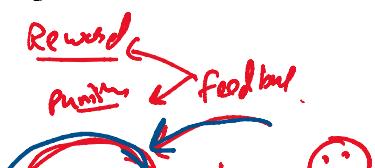
Features



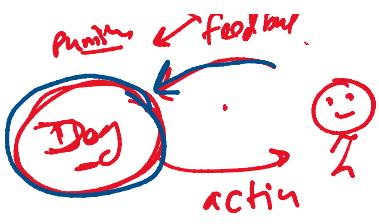


## → Reinforcement Learning

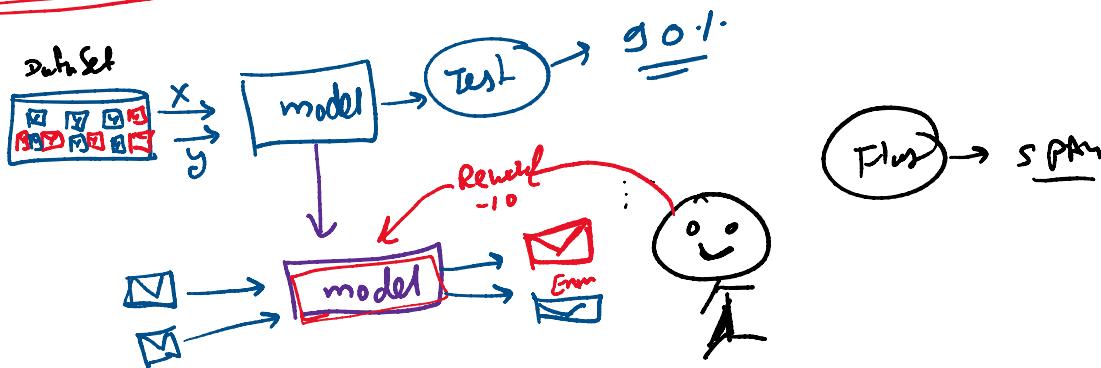
with time model updates (return) it self to improve its performance.



self to improve its performance.



## Reward & punishment →



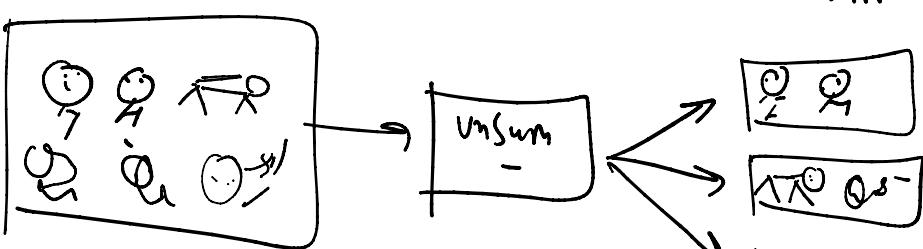
## # Supervised ML



① Regression → when your target variable is a continuous value. e.g. price, height, stock price, temp -

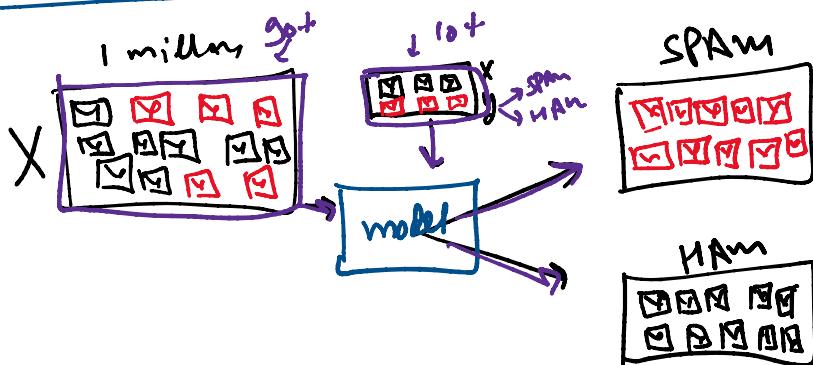
② Classification → when your target variable is a categorical value or represents a group.

e.g. → Corona → pos / neg, Test → pos / fail, product → good / bad, work → excellent / good / avg / worst

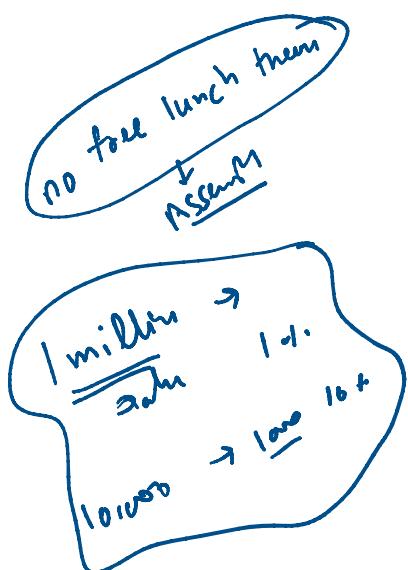




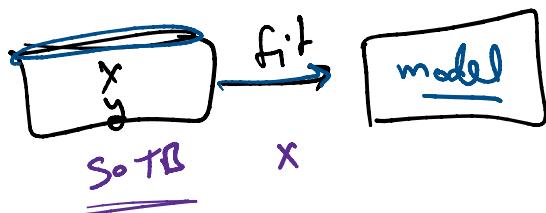
$\rightarrow$  Semi-Supervised  $\times$  ML  $\rightarrow$



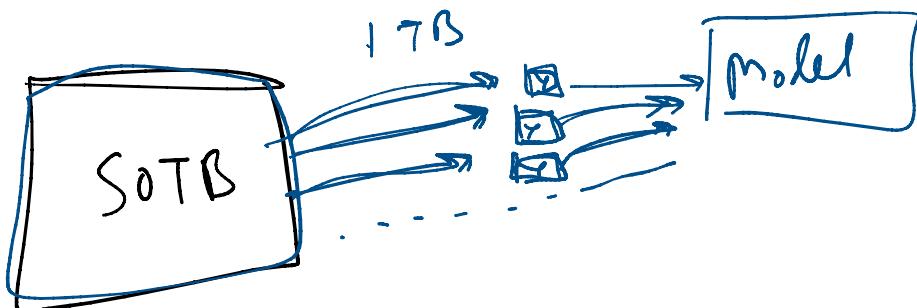
$\rightarrow$  Batch vs online learners



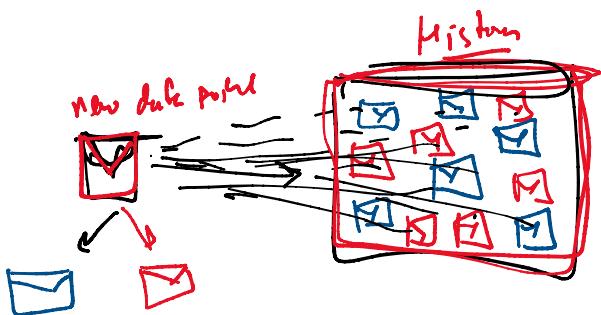
Batch learners



Online learning / on the fly / out of core learning



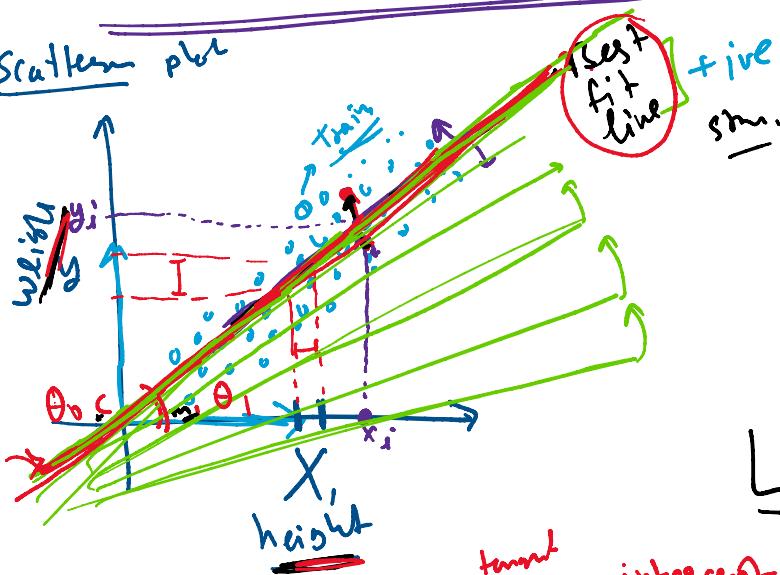
## → Instance Based or Model Based Learning



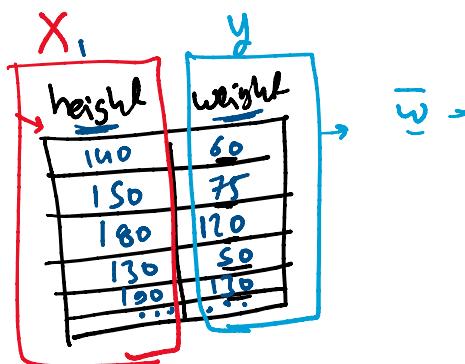
## Instance Based M2

## → model Based Learning →

Scatter plot



No free Lunch theorem



## Linear Regression

$$y = b_0 + b_1 x_1 + \dots + b_n x_n$$

↑ bias  
↓ intercept

$$y = w x + b$$

↓ coefficients  
↓ weights

$$y = B_0 + B_1 x_1$$

$$y = D_0 + \theta_1 x_1$$

hypothesis

$$y = m x_1 + c$$

↑ target  
↓ intercept

$$\hat{y} = \theta_1 \cdot \text{height} + \theta_0$$

↑ m  
↓ c

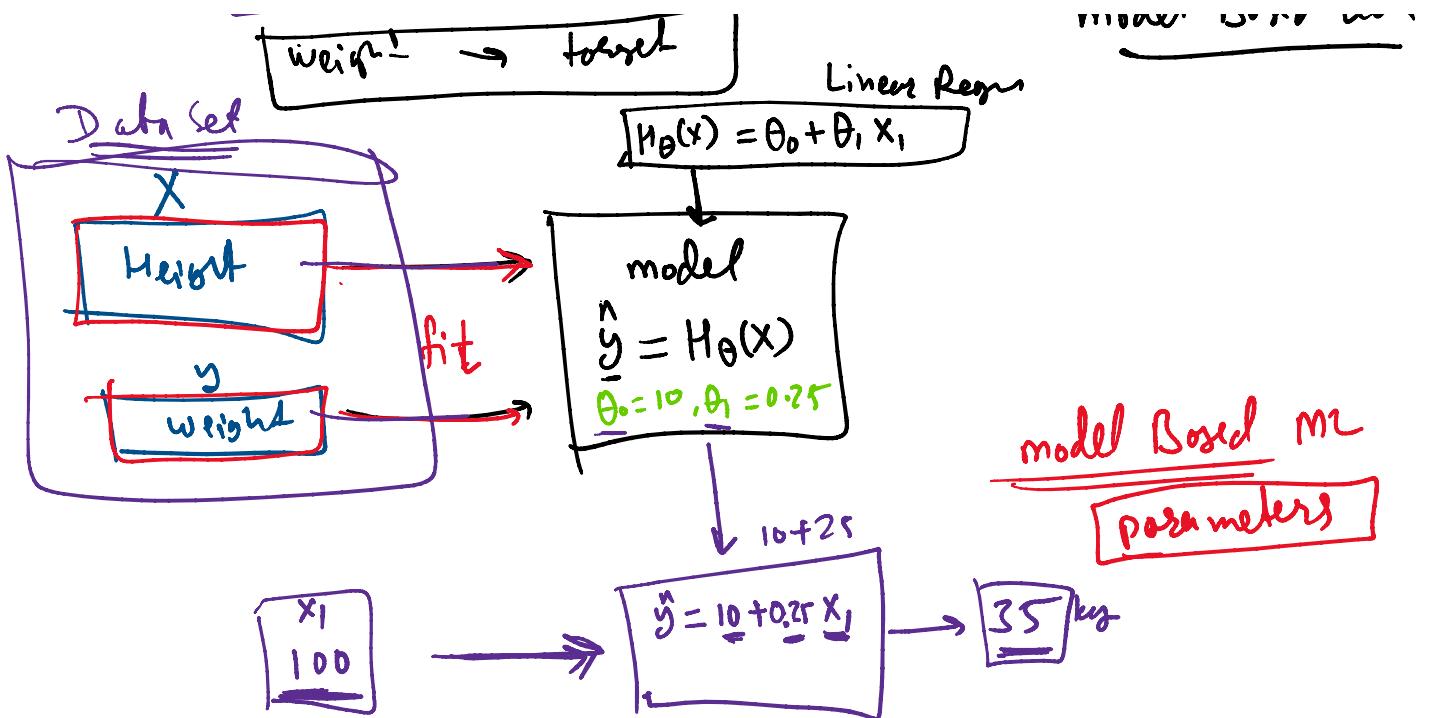
height → feature / Attribute

$\theta_0, \theta_1 \rightarrow$  parameter

weight → target

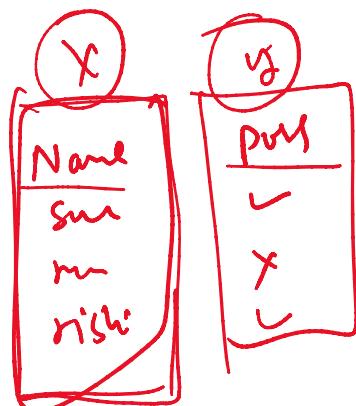
→ linear Regression  
↓  
model Based Learn

Linear Regres



## Challenges

### ① Data $\rightarrow$



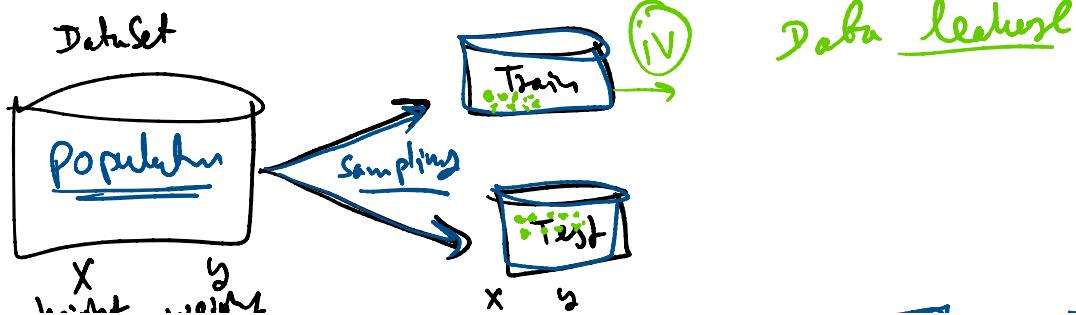
(i) insufficient Data

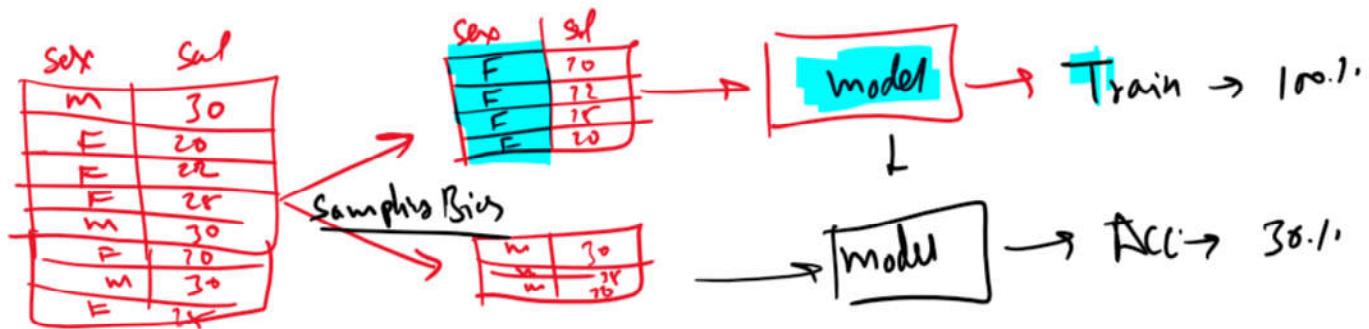
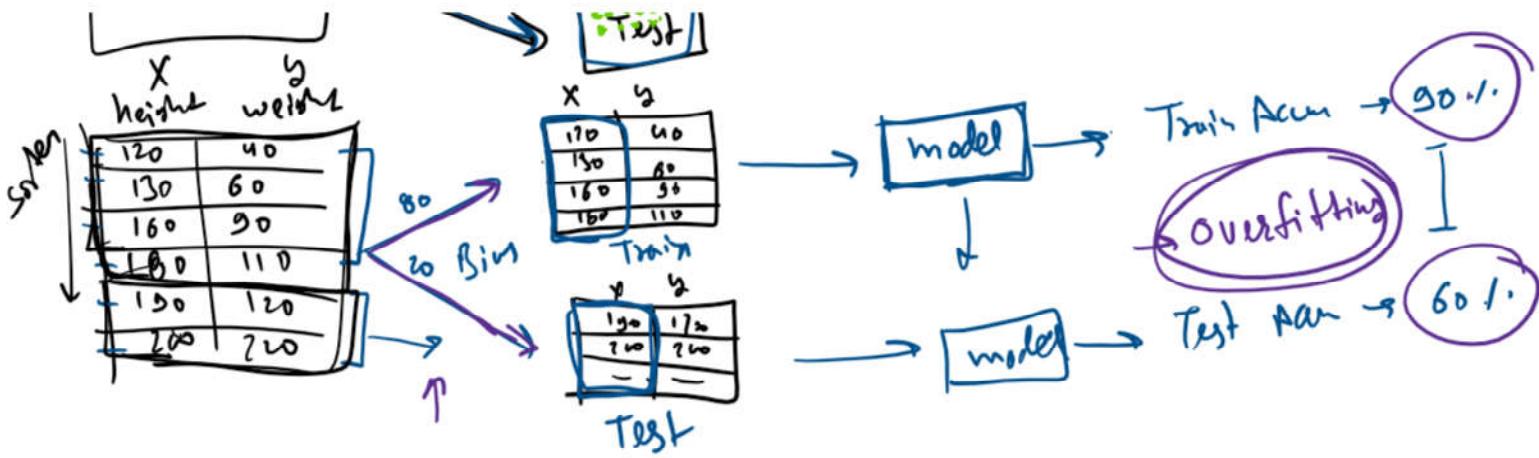
The Unreasonable Effectiveness of data

(ii) Non-Representative Dataset / Quality

(iii) Sampling Bias

Data leakage





→ Doubt

Tomorrow

→ Regression + Steps to create a project