

Free GATE-DA Course: L1

Introduction to ML

Special class



By: Vishvadeep Gothi

Data Science: Machine Learning



Today \Rightarrow Introductⁿ

28th \Rightarrow classificatⁿ algo $\left\{ \begin{array}{l} \rightarrow \checkmark \\ \rightarrow \checkmark \end{array} \right\} \leftarrow \text{Math + Theory}$

\Rightarrow Regression (Math heavy) \Rightarrow Math + Theory

\Rightarrow clustering \rightarrow Types \Rightarrow Math + Theory

\rightarrow Rest of the topics

▶ Learning

A process that leads to change, which occurs as a result of experience and increases the potential for improved performance and future learning.

Learning is the process of acquiring new understanding, knowledge, behaviors, skills, values, attitudes, and preferences.

▶ Learning

Learning is experience from past

Machine Learning

Learning is experience from past

A machine can be programmed to gather experience in the form of instances, facts, rules etc.

▶ Machine Learning

Learning is experience from past

A machine can be programmed to gather experience in the form of instances, facts, rules etc.

A machine with learning capability can predict about new situation (seen or unseen) using past experience.

► Machine Learning

As humans can tell a person's name seeing him/her second, third or fifth time, a machine can also do that.

As humans can recognize a person's voice even if not seeing person's face, a machine can also do the same.

Machine Learning

Tom Mitchell (1998)

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ”

Question

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

- a) Classifying emails as spam or not spam
- b) Watching you label emails as spam or not spam
- c) The number (or fraction) of emails correctly classified as spam/not spam
- d) None

Machine Learning

Tom Mitchell (1998)

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ”

T := Classifying emails as spam or not spam.

E := Watching you label emails as spam or not spam.

P := The number (or fraction) of emails correctly classified as spam/not spam.

Class Experiment: Training

Let:

- ▶ AA denote 5
- ▶ BB denote 6
- ▶ AAA denote 50
- ▶ BBB denote 60
- ▶ AAAA denote 500
- ▶ BBBB denote 600

Can you find out equivalent value of AAAAAA? 5000

AAAAAA 5000

BBBBBB 6000

Class Experiment: Training

Let:

- ▶ AA denote 5
- ▶ BB denote 6
- ▶ AAA denote 50
- ▶ BBB denote 60
- ▶ AAAA denote 500
- ▶ BBBB denote 600

you ✓

Can you find out equivalent value of AABB?

→ not trained

► Pronunciation Learning

Training:

- ▶ Cat (ae sound)
- ▶ Pot (aw sound)
- ▶ Pat (ae sound)
- ▶ Tap (ae sound)
- ▶ Cot (aw sound)

Testing:

How do you pronounce 'not' ?

Pronunciation Learning

Training:

- ▶ Cat (ae sound)
- ▶ Pot (aw sound)
- ▶ Pat (ae sound)
- ▶ Tap (ae sound)
- ▶ Cot (aw sound)

Testing:

How do you pronounce 'check' ?

↳ not trained

► Learning in Human vs Machine

Human:

Gain Experience from day-to-day activities and gain ability to predict

Machine:

Get trained with the numerical data (text, image, sound, rules, numbers) and be able to predict

► Why Machine Learning

Humans have limitations in terms of accessibility and computational efficiency

Types of Machine Learning

Supervised Learning

- Inputs & Outputs

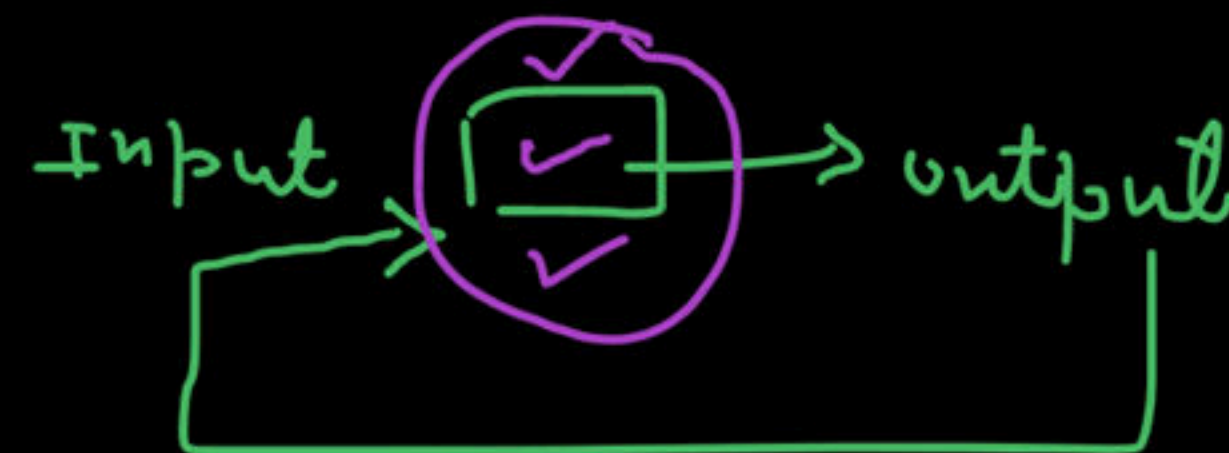
Banana
apple

Unsupervised Learning

- Inputs, but no specific Output

Reinforcement Learning

- Reward or Penalty

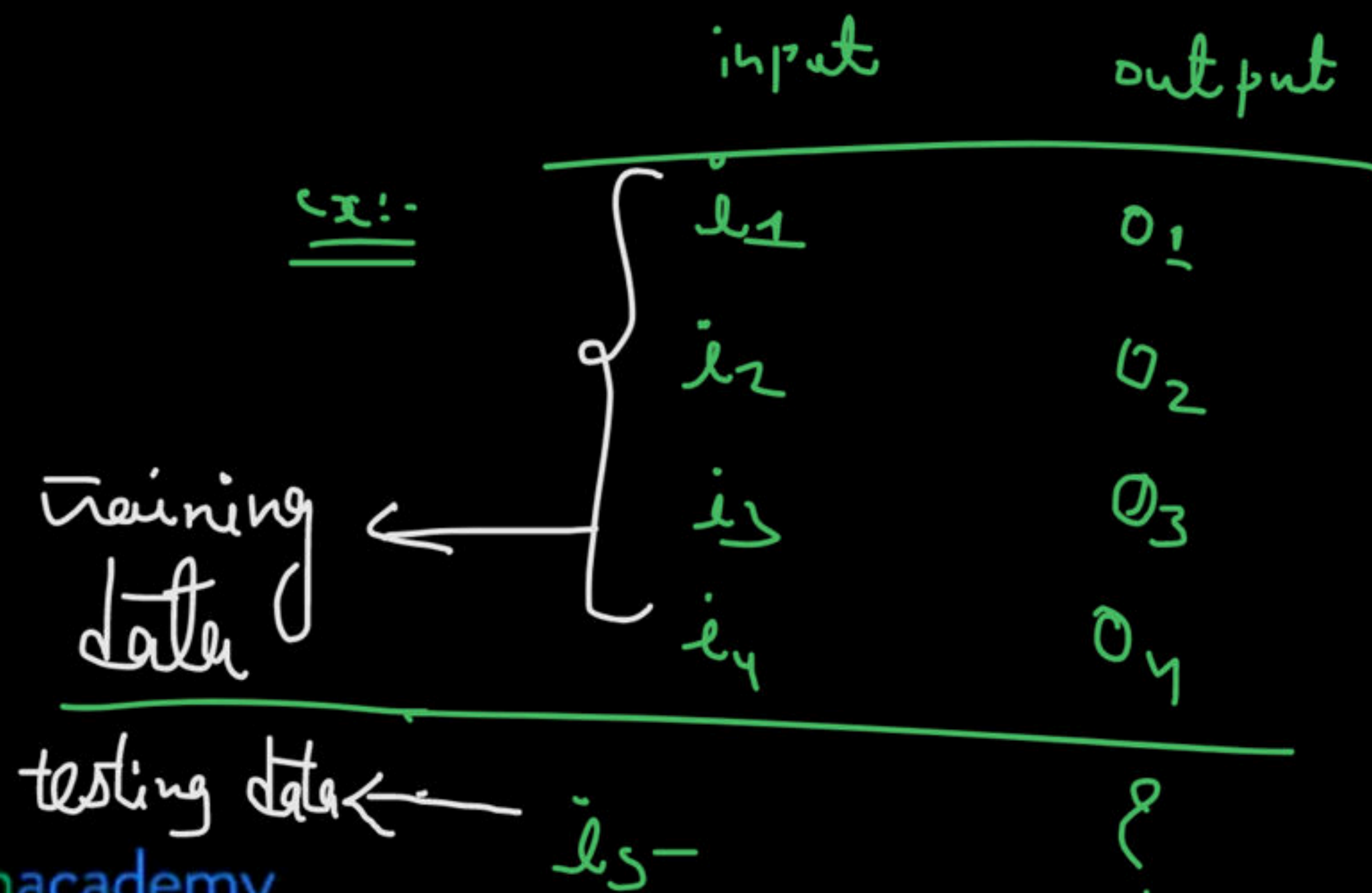


artificial neural network
(A.N.N.)

(deep learning)

Supervised Learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs



Supervised Learning

Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(X)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

$y = f(x_1, x_2, x_3, x_4, x_5)$

x_1	x_2	x_3	x_4	x_5	y
Gender	age group	yearly income	having car or not	family members	Buy or not
M	30-32	20 LPA	No	2	Buy
...					not buy

Supervised Learning: Types

1. Classification \Rightarrow result \Rightarrow one or more of known type
2. Regression \Rightarrow result \Rightarrow numerical value

Ex:-

Students					
Syllabus finish	Revision n-	no. of qs	Tests	ans	final C/AIE marks
✓	✓	✓	✓	✓	65
...					60
✓	0	1000	2	30	30
					70
					70

Supervised Learning: Types

Classification

- Learning a discrete valued function

Regression

- Learning a continuous valued function

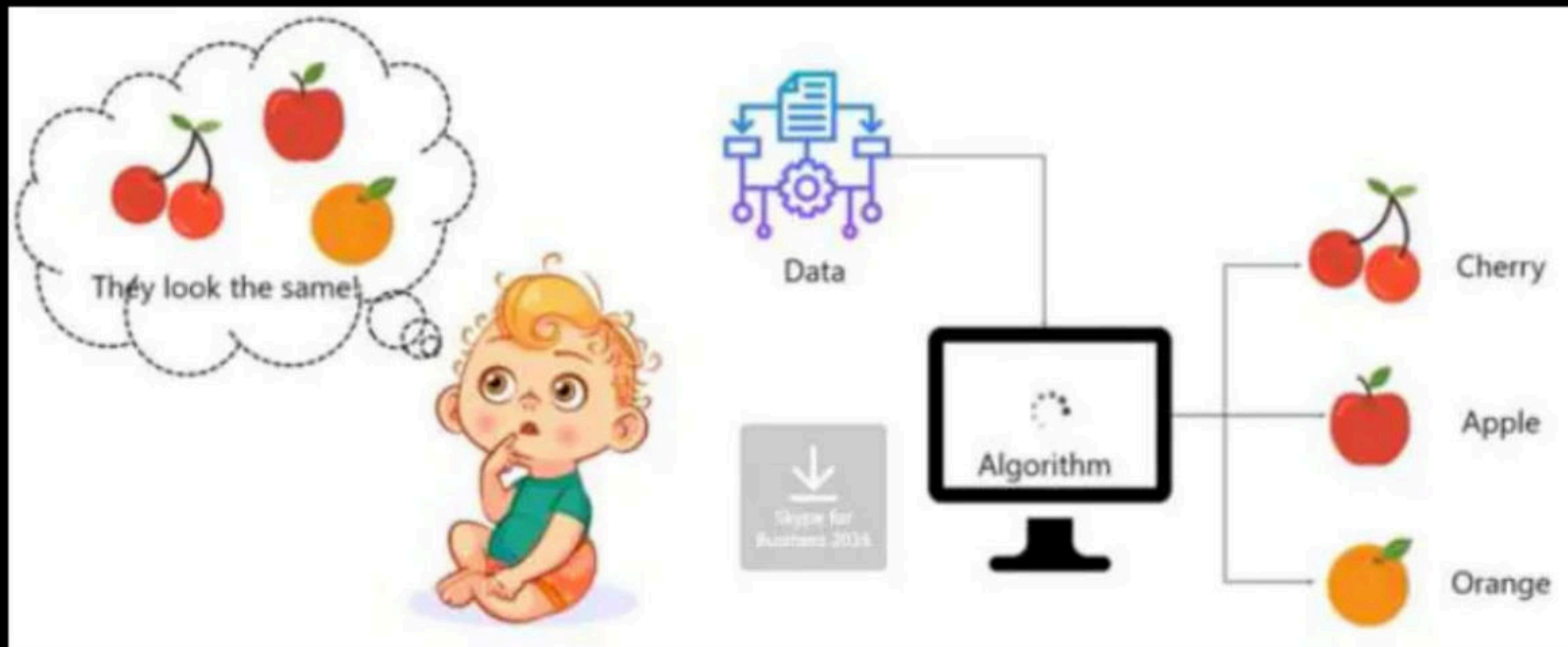
► Classification

A predictive modeling problem where a class label is predicted for a given example of input data.

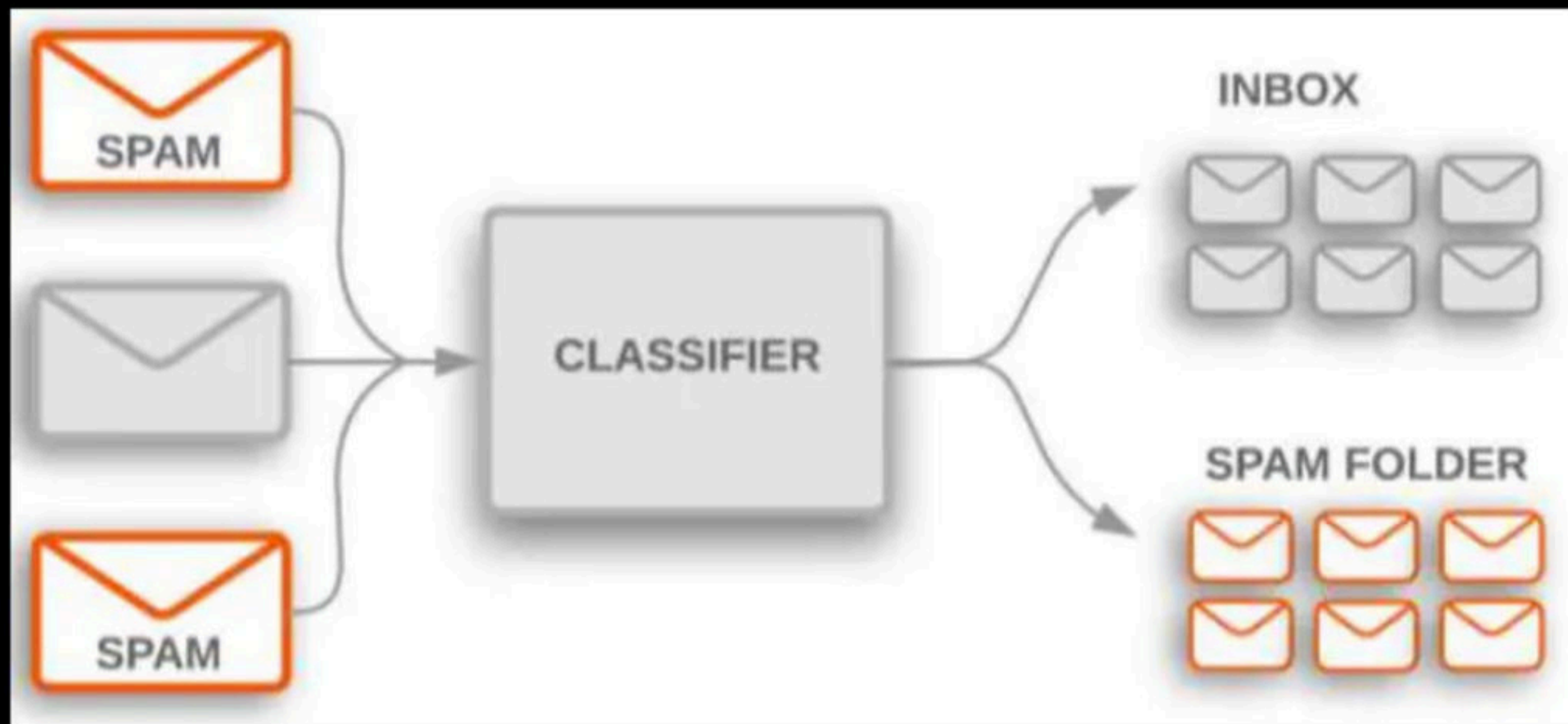
Classification Example



Classification Example



Classification Example



outlier => data which does not follow usual pattern

Classification Types

Binary Classification

- Spam or not
- Buy or not

↓↓
only 2 classes

Multi-Class Classification

- Face Classification
- Fruit Recognition

↓↓
multiple classes but input is classified into only one class

Multi-Label Classification

- Document Classification

↓
multiple classes but one input can be labeled as multiple classes

Imbalanced Classification

- Fraud detection
- Outlier detection
- Medical diagnostic tests

Regression

A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”.

Regression Example

Suppose there is a marketing company A, who does various advertisement every year and get sales on that. The below list shows the advertisement made by the company in the last 5 years and the corresponding sales

in lacs

Advertisement (INR)	Sales (INR)
900	10000
1200	13000
1500	18000
1000	12000
1300	13700
2000	??

Regression Example

- ▶ Determining Market trends
- ▶ House Price Prediction



► Unsupervised Learning

Determining Unsupervised Learning is a machine learning technique in which the users do not need to supervise the model. Instead, it allows the model to work on its own to discover patterns and information that was previously undetected.

It mainly deals with the unlabeled data trends

► Unsupervised Learning Types

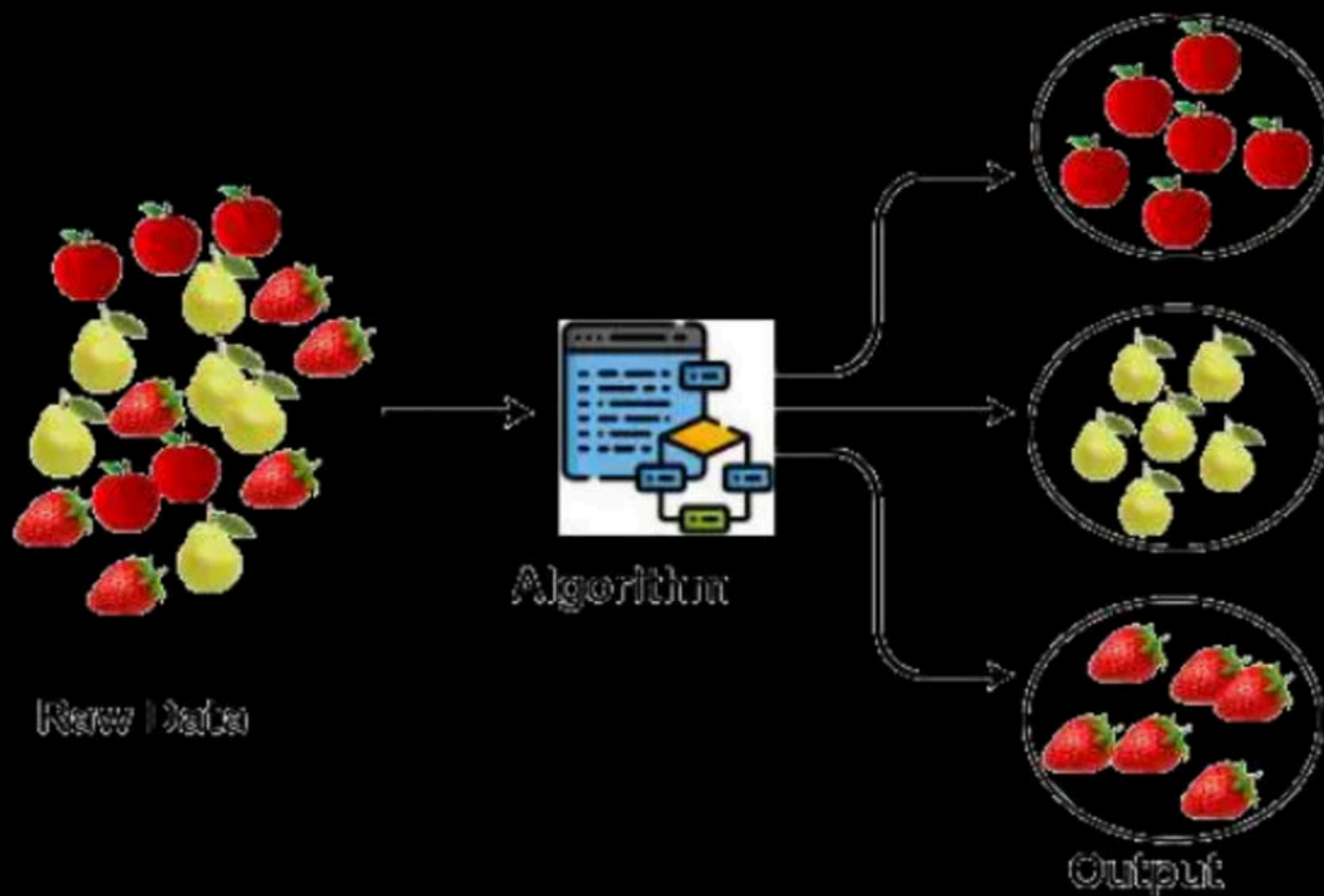
Clustering:

► Unsupervised Learning Types

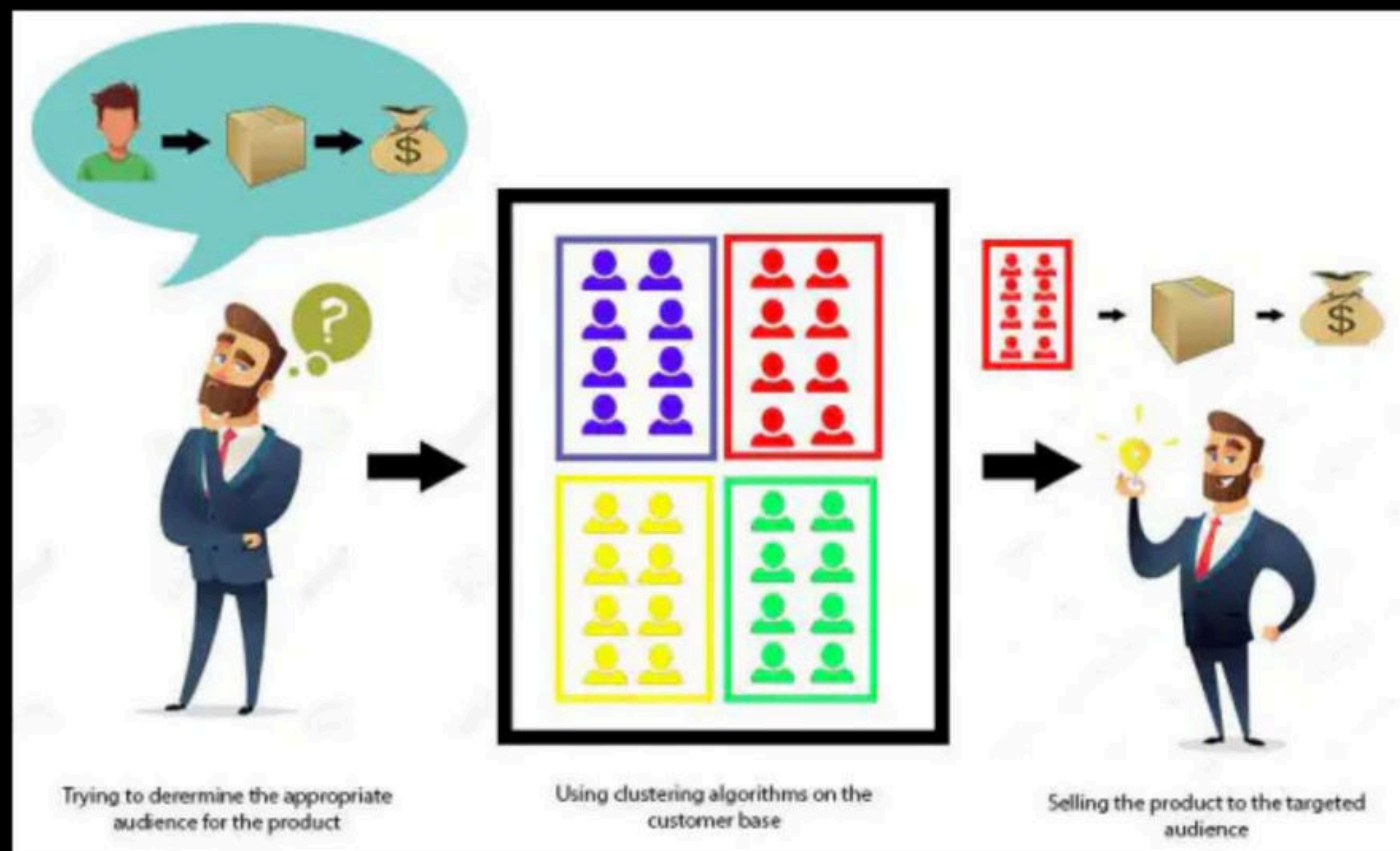
Clustering:

A way of grouping the data points into different clusters, consisting of similar data points. The objects with the possible similarities remain in a group that has less or no similarities with another group

Clustering



Clustering



Types of Clustering

1. Partitioning Clustering
2. Density-Based Clustering
3. Distribution Model-Based Clustering
4. Hierarchical Clustering
5. Fuzzy Clustering

► Supervised vs Unsupervised

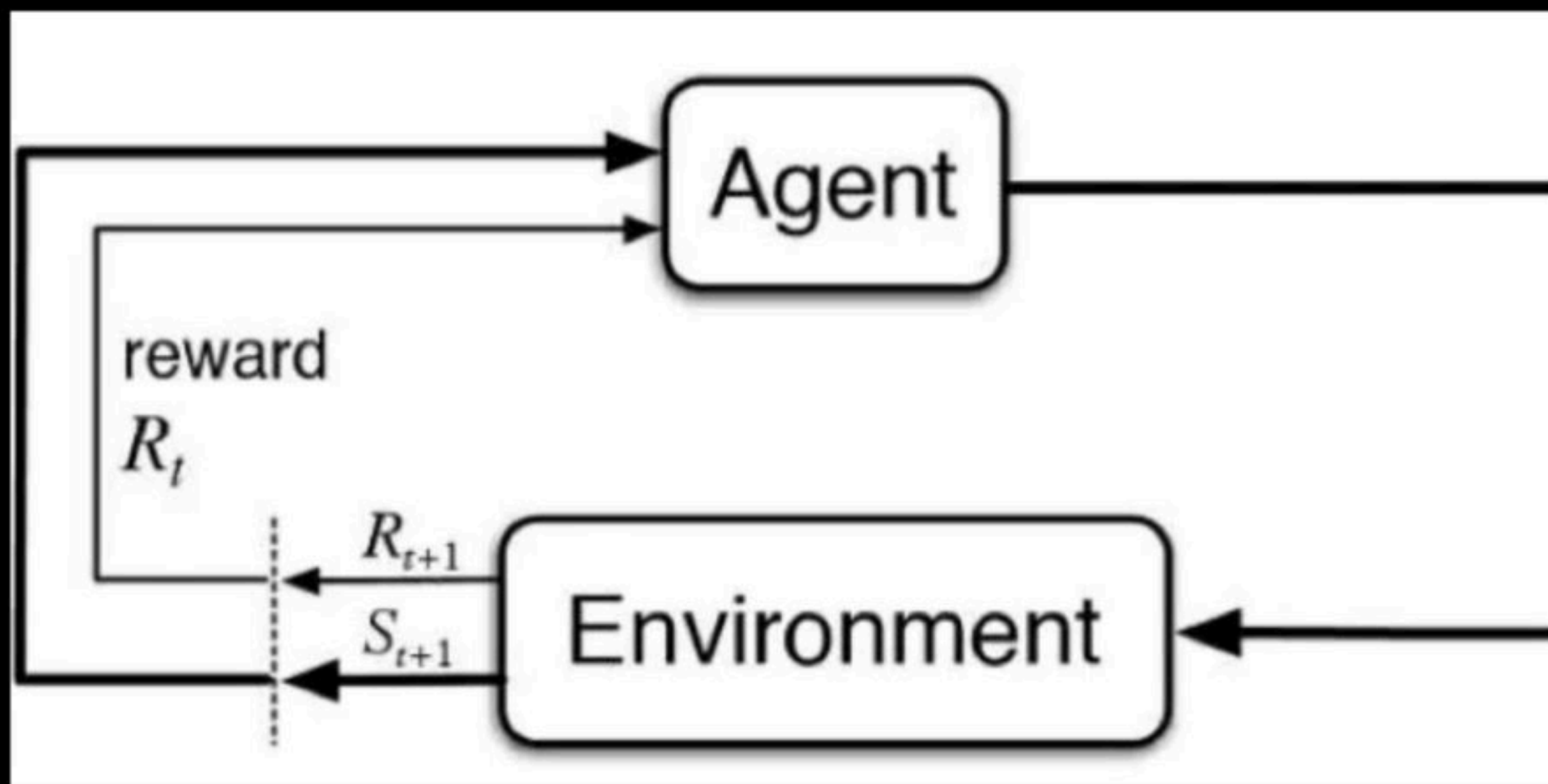
Supervised: (Input, Correct Output)

Unsupervised: : (Input, ?)

► Reinforcement Learning

Reinforcement Learning: (Input, *Some* Output, **We have to grade it**)

Reinforcement Learning



Monday \Rightarrow classificatⁿ

Happy Learning.!

