MACHINE LEARNING

Shoaib Farooq

Department of Computer Science

Disclaimer: This presentation includes contents available online including images copied from Google search and contents of presentations of other professors. I don't claim any image or text to be my own. All the credit goes to the original authors.

Instructor Information

Instructor	Shoaib Farooq	E-mail	m.shoaib1050@gmail.com
 GitHub	https://github.com/SHOAIB1050		
 YouTube	https://www.youtube.com/c/pakithub		
 LinkedIn	https://www.linkedin.com/in/shoaib-farooq-b5b190105/		

Let's Start .....

Lecture #2

Goals

This Lecture Will Cover:

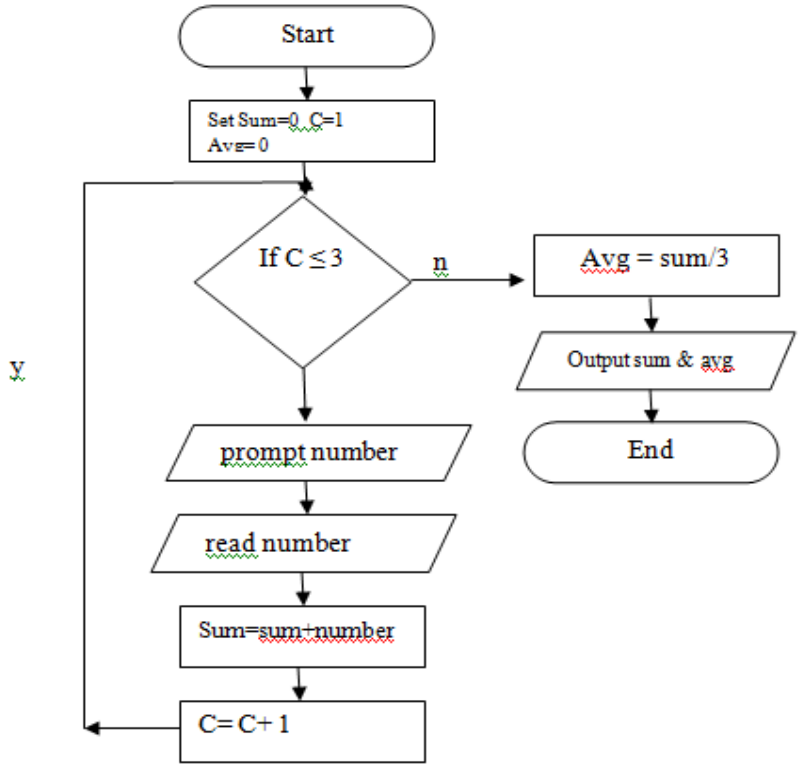
- What is Learning?
- What is Machine Learning?
- Types of Machine Learning
- Types of Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Data And Types
- Structured Data
- Unstructured Data
- Semi-Structured Data

WHAT IS LEARNING?

- How can we solve a specific problem?

  We write a program with a set of rules that are useful to solve the problem.

  Example: Find average of three numbers



WHAT IS LEARNING?

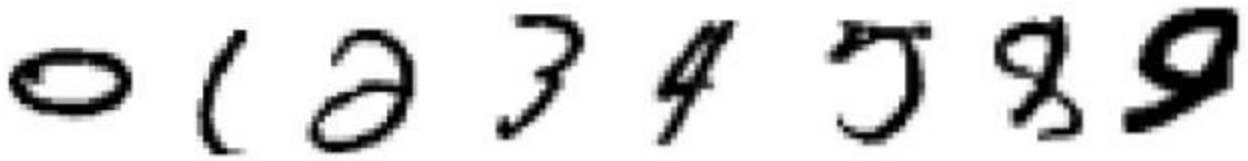
- In many situations it is very difficult to specify those rules to solve a problem.
- For example, given a picture determine whether there is a cat in the image

WHAT IS LEARNING?

- Find face of a specific person?

WHAT IS LEARNING?

- Any learning systems are not **directly programmed using conditions to solve a problem**
- Instead it should learn from **examples (data)**
- From **trial-and-error** experience trying to solve the problem



WHAT IS MACHINE LEARNING?

- Machine Learning is the **science (and art)** of programming computers so they can learn from data.

- **[Machine Learning is the]** field of study that gives computers the ability to **learn without being explicitly programmed**.

  (Arthur Samuel, 1959)

- Machine learning can be defined as **computational methods** using experience **to improve performance** or to make **accurate predictions**.

  Experience refers to the past information.

WHAT IS MACHINE LEARNING?

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**

Tom M. Mitchel

WHAT IS MACHINE LEARNING?

- **Task T:** playing checkers
- **Performance measure P:** percent of games won against opponents
- **Training experience E:** playing practice games against itself

SPAM TAGGING PROBLEM

- Your **spam filter** is a Machine Learning program
- **Binary Classification Problem:** spam emails or Nonspam
- To train a machine learning model, examples of emails that are spam and no spam should be presented to the model
- The examples that the model uses to learn are called **the training set**.
  - Training instance (or sample).

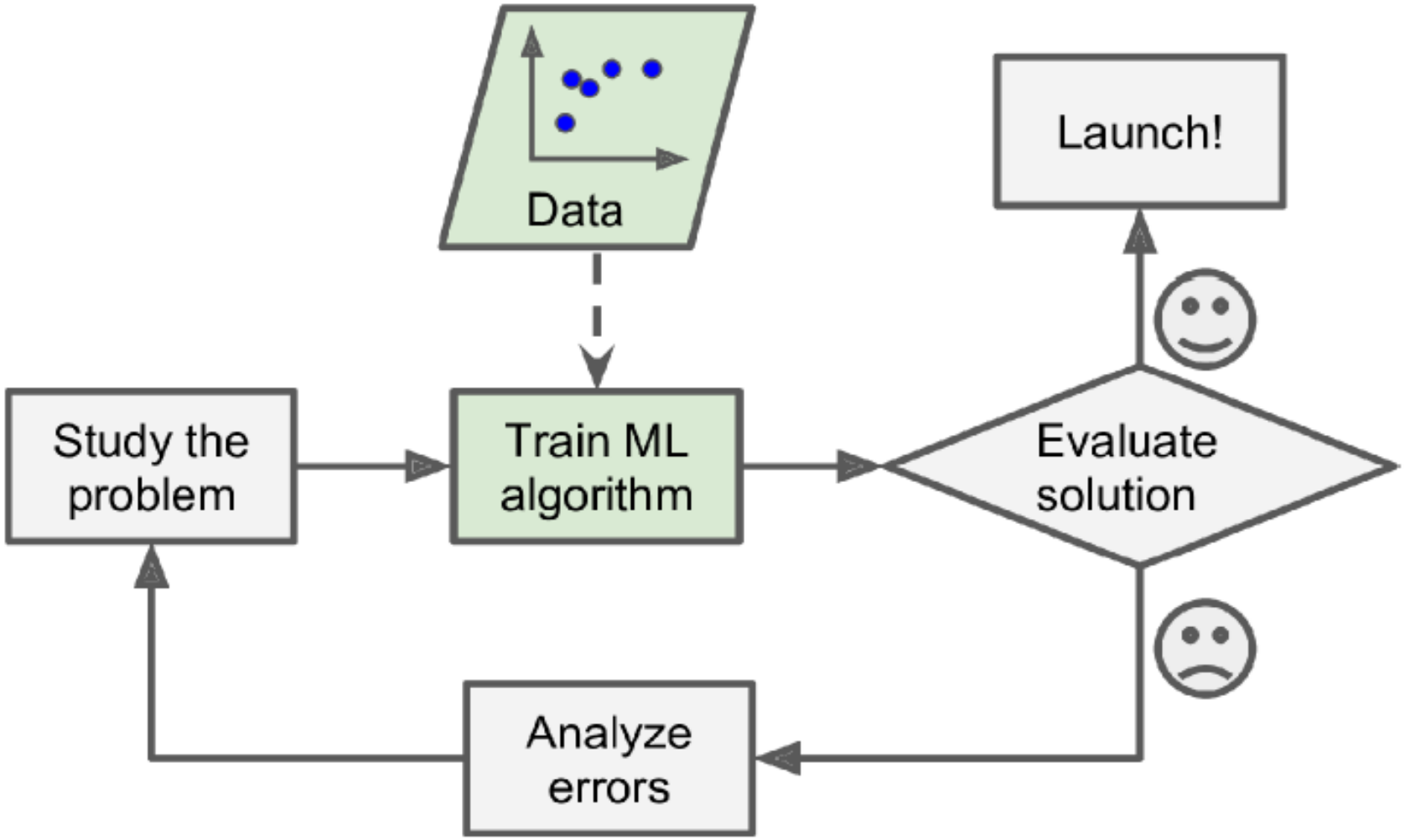
SPAM TAGGING PROBLEM

- Your **spam filter** is a Machine Learning program
- **Binary Classification Problem:** spam emails or Nonspam
- To train a machine learning model, examples of emails that are spam and no spam should be presented to the model
- The examples that the model uses to learn are called **the training set**.
  - Training instance (or sample).

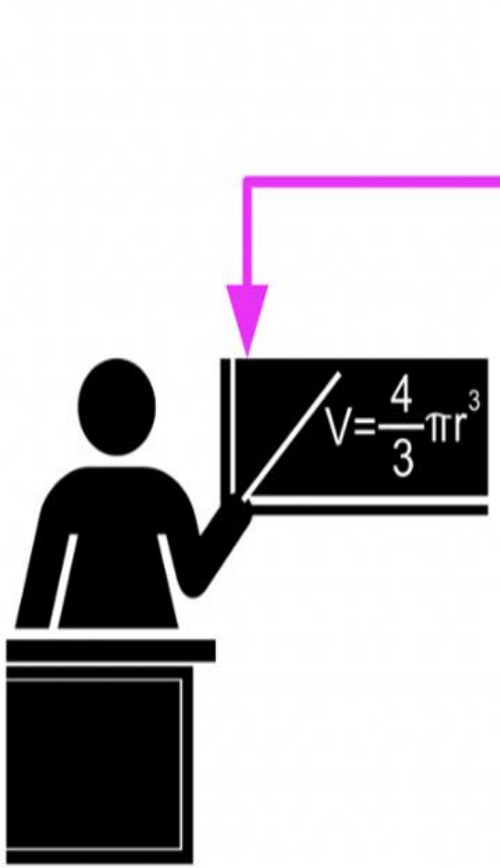
SPAM TAGGING PROBLEM

- For Spam classification:
  - The task T is to flag spam for new emails
  - The experience *E* is the *training data*,
  - The performance measure *P* needs to be defined;
    - Percentage of correctly classified emails (*accuracy)*

GENERAL FRAMEWORK FOR ML

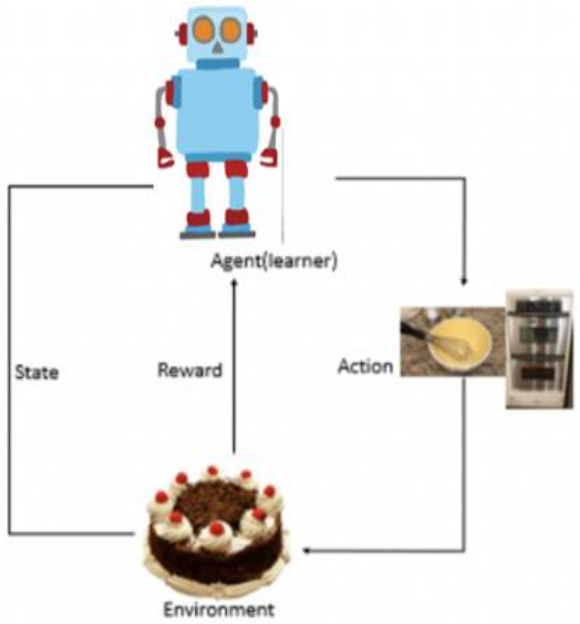


TYPES OF MACHINE LEARNING...



Supervised Learning

Unsupervised Learning



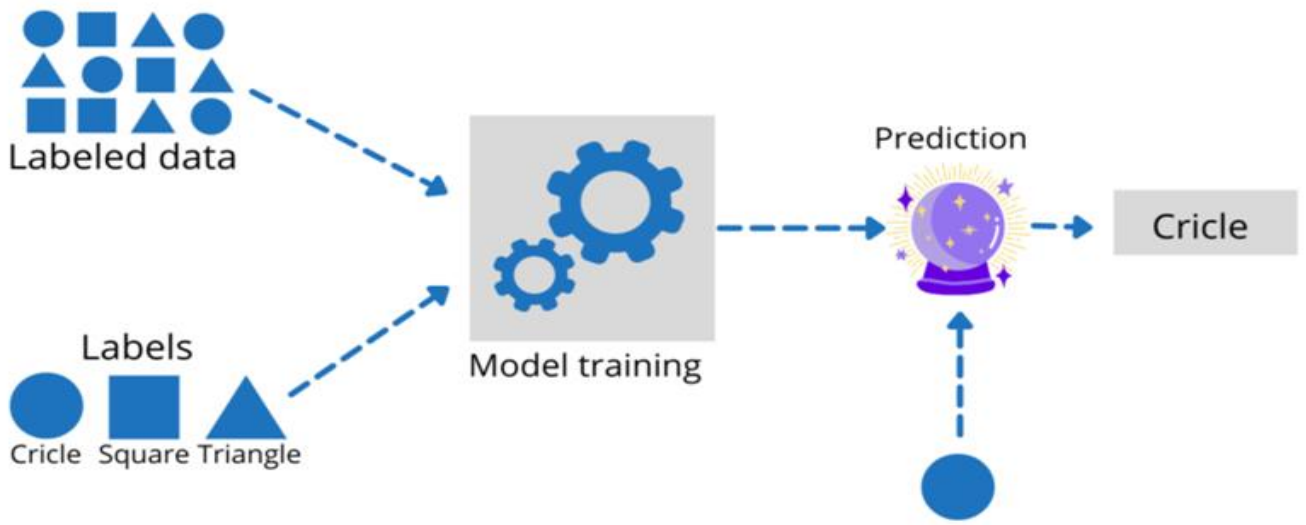
Reinforcement Learning

TYPES OF MACHINE LEARNING...

1. Supervised Learning (learning with **labeled data**)
2. Unsupervised Learning (discover patterns in **unlabeled data**)
3. Reinforcement learning (learn to act based on **feedback/rewards**)

SUPERVISED LEARNING

- For supervised learning, we provide **both data and labels** for training the algorithm.
- The algorithms learns from the **data and labels**
- After training, we can pass **test samples** to check if the **algorithm learned the data or not**
- **Most popular** in ML community



SUPERVISED LEARNING

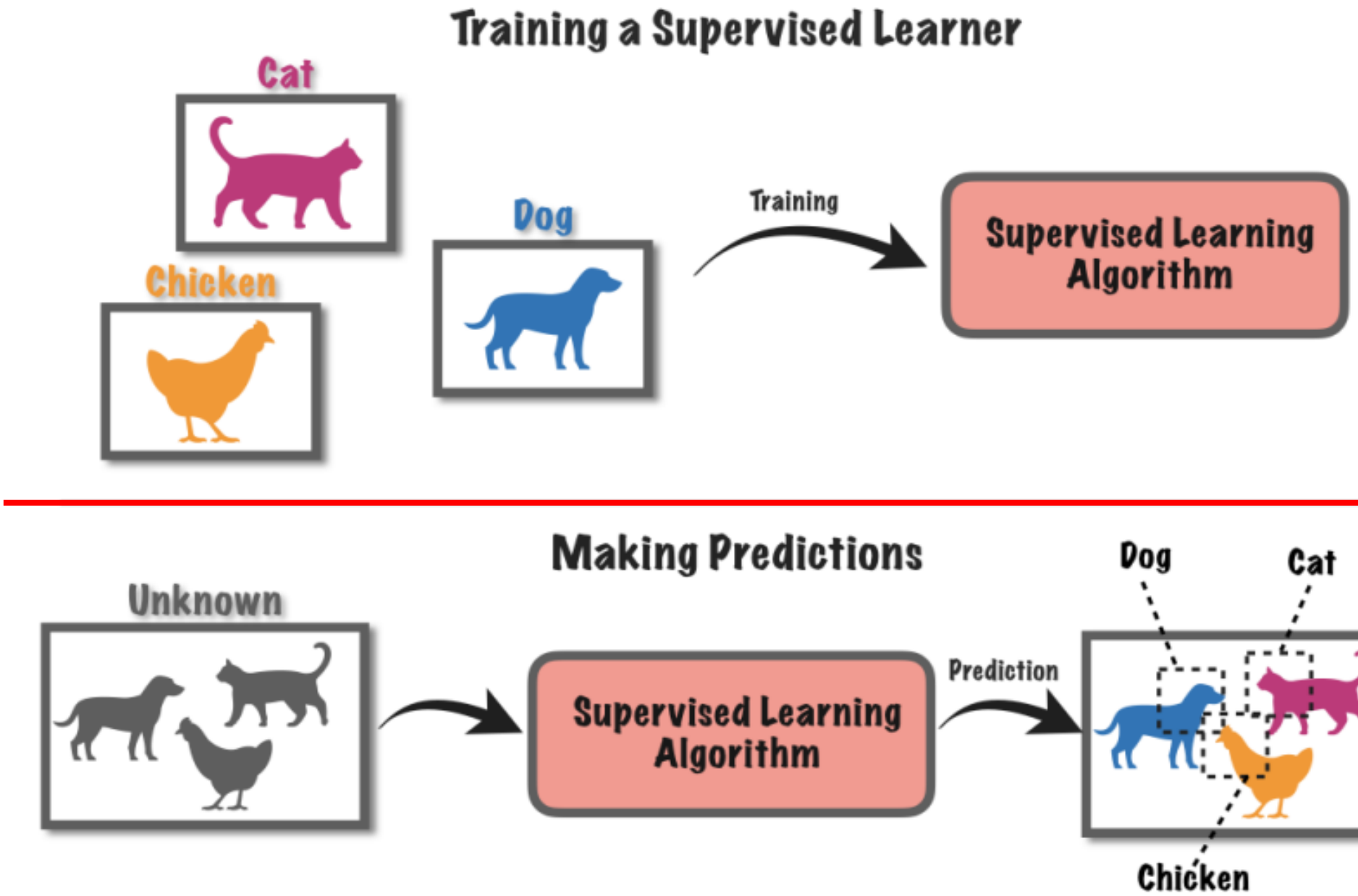
Data: $x = \{x_1, x_2, .., x_n\}$ ***n* examples**

$d_i = < \mathbf{x}_i, y_i >$

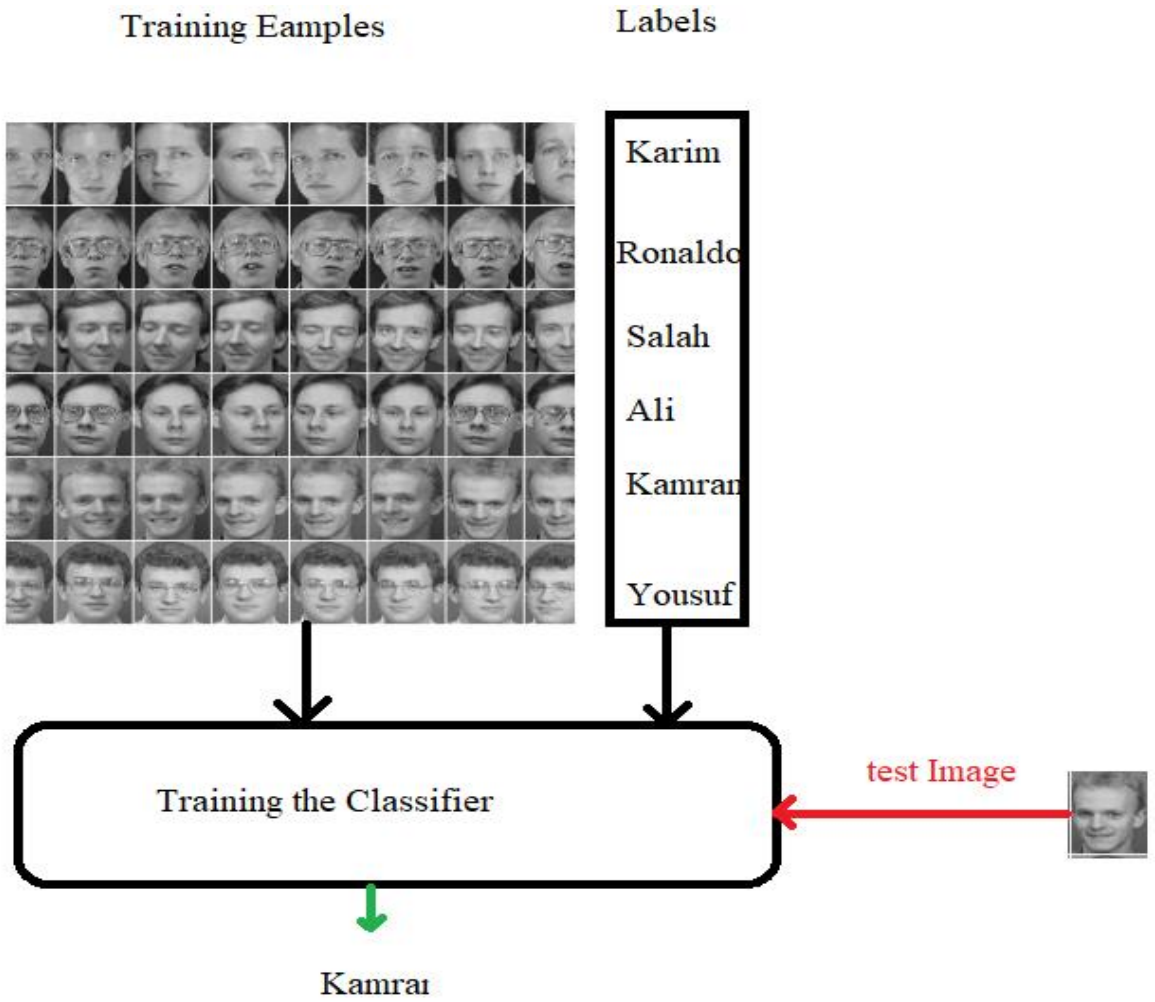
$\mathbf{x}_i$ is input vector, and y is desired output (given by a teacher)

Objective: learn the mapping $f : X \rightarrow Y$

s.t. $y_i \approx f(x_i)$ for all $i = 1, .., n$


Training a Supervised Learner
Cat
Dog
Chicken
Training
Supervised Learning Algorithm
Making Predictions
Unknown
Supervised Learning Algorithm
Prediction
Dog
Cat
Chicken

SUPERVISED LEARNING EXAMPLE



TYPES OF SUPERVISED LEARNING

Two types of problems:

- **Regression:** X discrete or continuous → Y is **continuous**
- **Classification:** X discrete or continuous → Y is **discrete**

TYPES OF SUPERVISED LEARNING

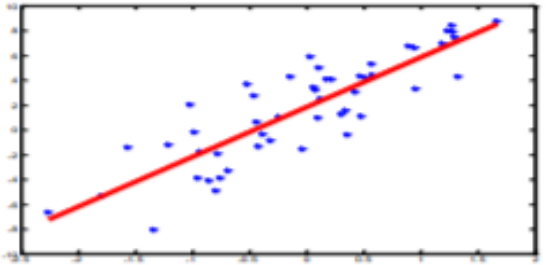
- Can regression algorithms be used for classification and vice versa?
  - **Yes, some algorithms can be used.**
- **Logistic Regression** is commonly used for classification
  - Predicts probability belonging to a class

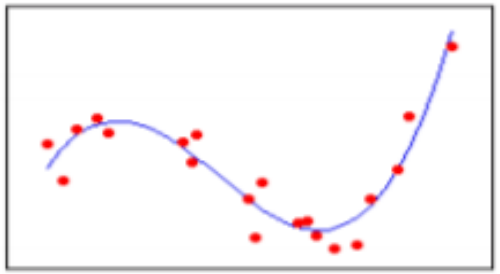
TYPES OF SUPERVISED LEARNING

- **Regression:** Y is **continuous**

Debt/equity
Earnings
Future product orders

→ Stock price





Data:

Debt/equity	Earnings	Future prod orders	Stock price
20	115	20	123.45
18	120	31	140.56
….			

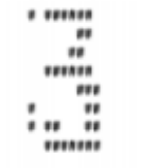
TYPES OF SUPERVISED LEARNING

- **Classification:** Y is **discrete**

Label "3"

Handwritten digit (array of 0,1s)

Data:



image

digit

3

7

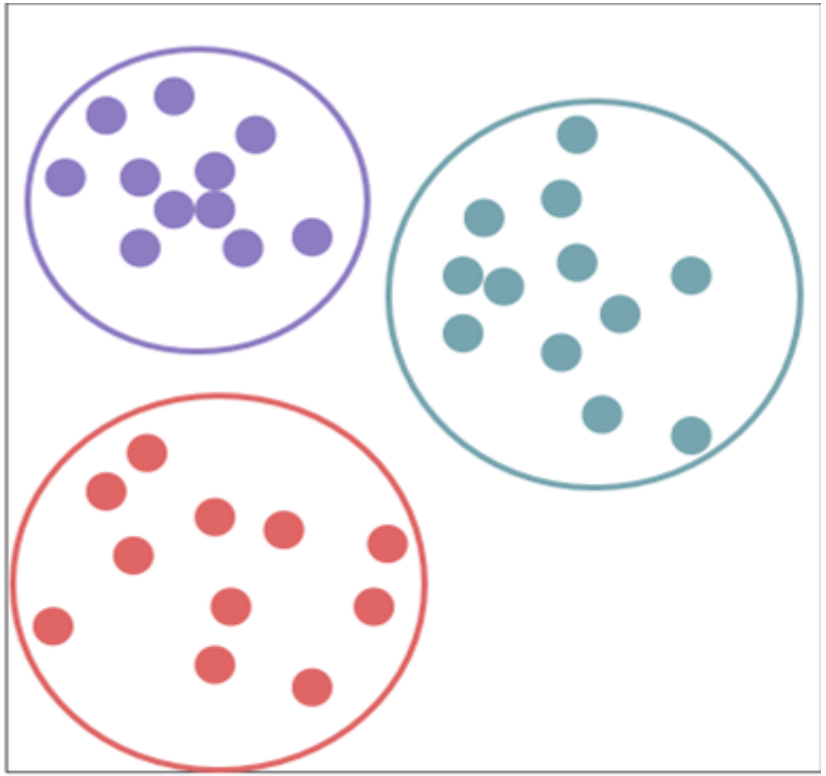
5

ALGORITHMS OF SUPERVISED LEARNING

- **Some widely used supervised ML algorithms:**
  - Linear Regression
  - Logistic Regression
  - Support Vector Machines (SVMs)
  - Decision Trees and Random Forests
  - Neural networks
  - k-Nearest Neighbors

UNSUPERVISED LEARNING

- For unsupervised learning, we provide **data but NOT labels** for training the algorithm
- The system tries to learn **without a teacher**.
- Learns **relations among data** by itself
- Then put the data into different **groups/clusters**



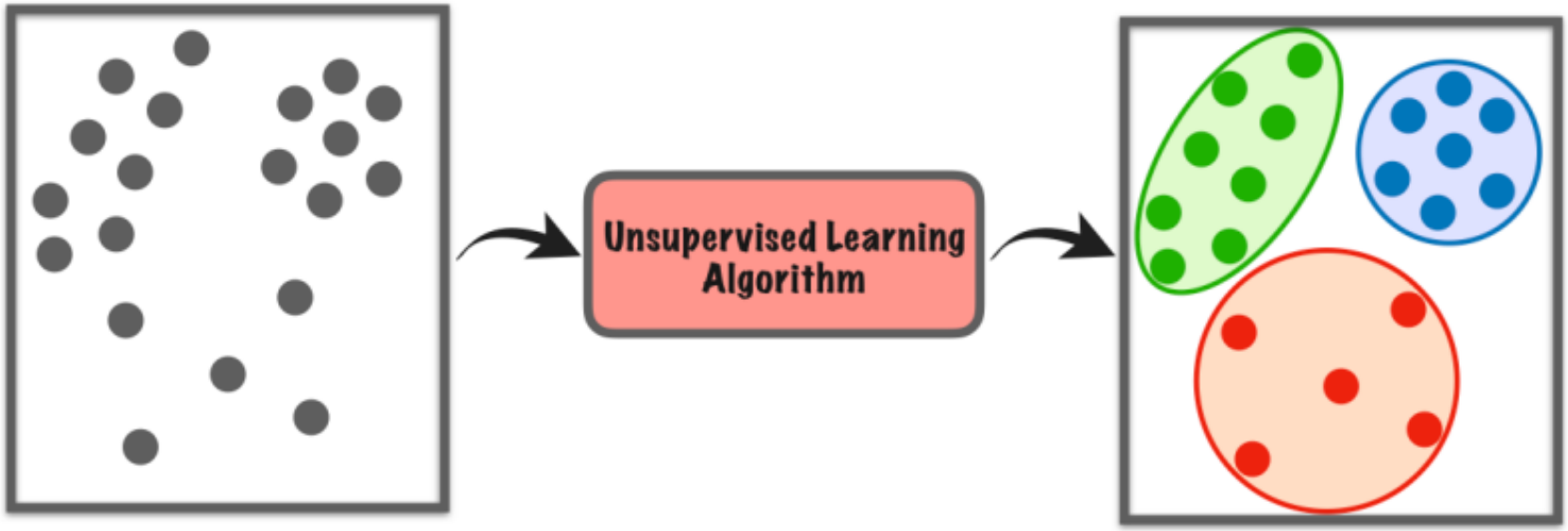
- **Data:** $x = \{x_1, x_2, .., x_n\}$ vector of values

  No target value (output) y

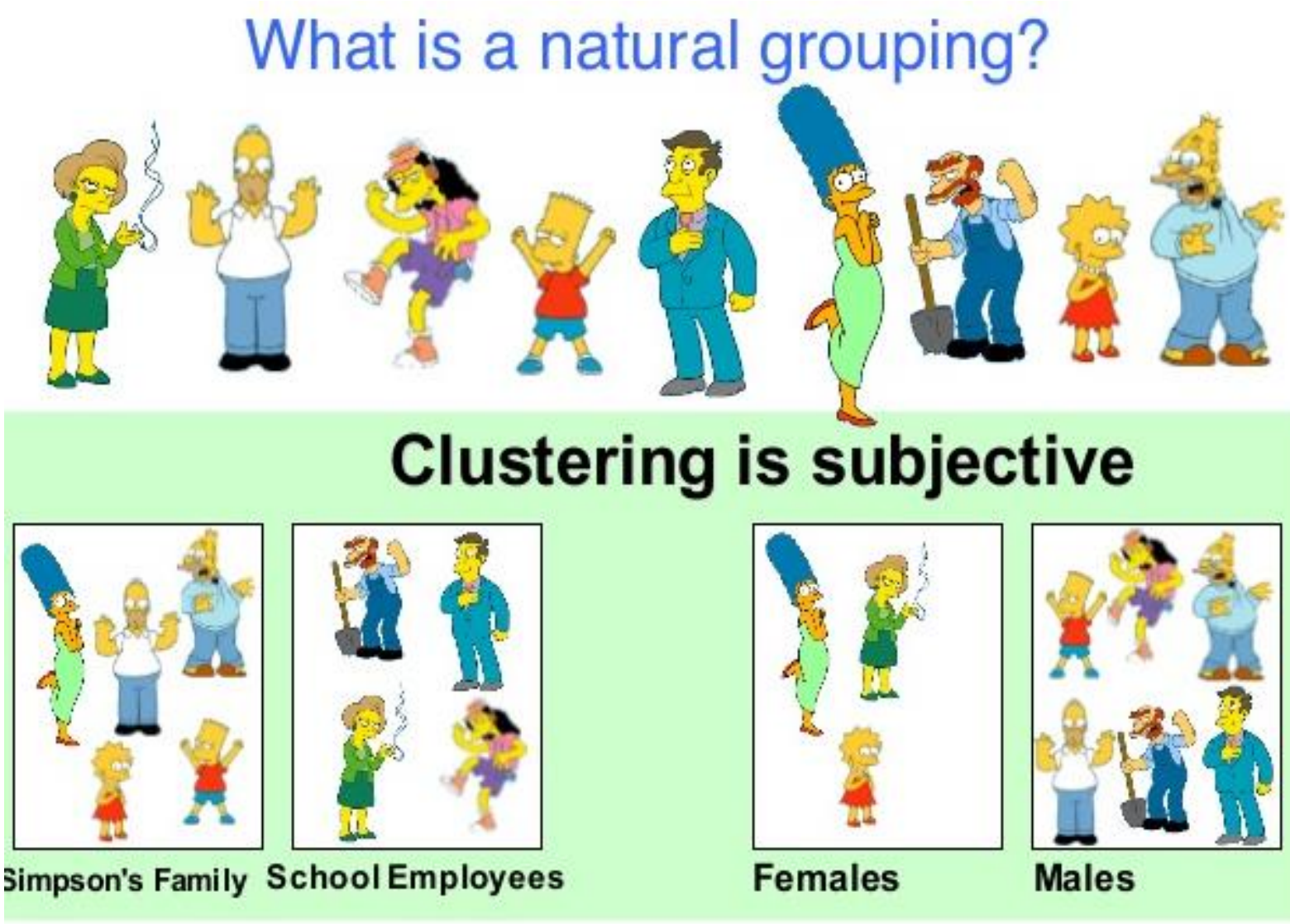
- **Objective:**
  - learn relations between samples, components of samples

UNSUPERVISED LEARNING

- Clustering
- Dimensionality Reduction
- Anomaly Detection



EXAMPLES OF UNSUPERVISED LEARNING



ALGORITHUMS OF UNSUPERVISED LEARNING

- **Some widely used unsupervised learning algorithms:**
  - **K-Means**
  - **Principal Component Analysis (PCA)**
  - **Apriori**
  - **Hierarchical Cluster Analysis (HCA)**
  - **One-class SVM**

USE OF UNSUPERVISED LEARNING

- Data visualization
- Dimensionality reduction
- Clustering
- Anomaly detection
- Products Segmentation
- Customer Segmentation
- Similarity Detection
- Recommendation Systems
- Labelling unlabeled datasets

SUPERVISED LEARNING VS UNSUPERVISED LEARNING

Supervised learning	Unsupervised learning
Input data is labelled	Input data is unlabelled
There is a training phase	There is no training phase
Data is modelled based on training dataset	Uses properties of given data for classification
Divided into two types: Classification and Regression	Most popular types: Clustering and Dimensionality reduction
Known number of classes (for classification)	Unknown number of classes

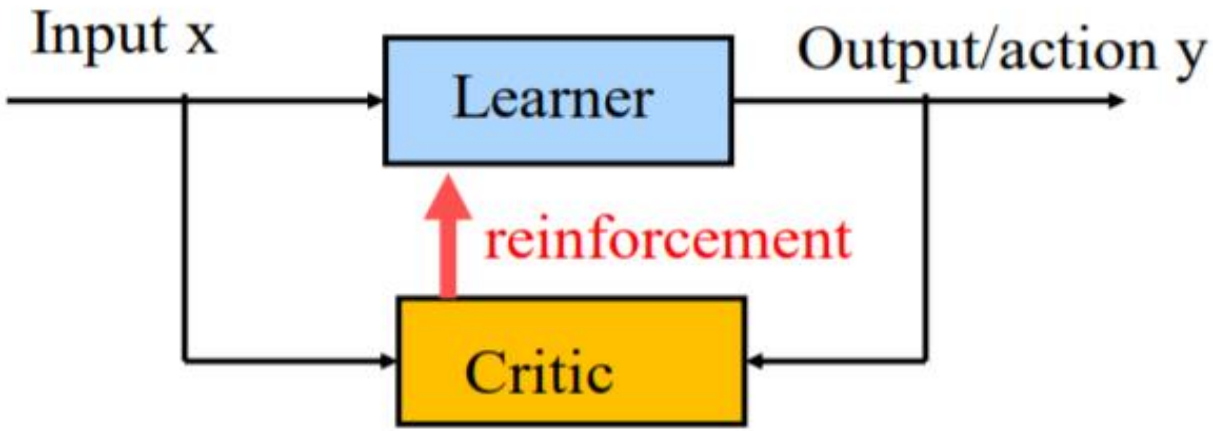
REINFORCEMENT LEARNING

- The learning system, called an **agent**, can **observe the environment, select and perform actions**:
  - Get **positive rewards** for good actions
  - Get **negative rewards** for wrong action
- Reinforcement **learning** **refers to goal-oriented algorithms**, which learn how to attain a complex **objective (goal) or maximize**

REINFORCEMENT LEARNING

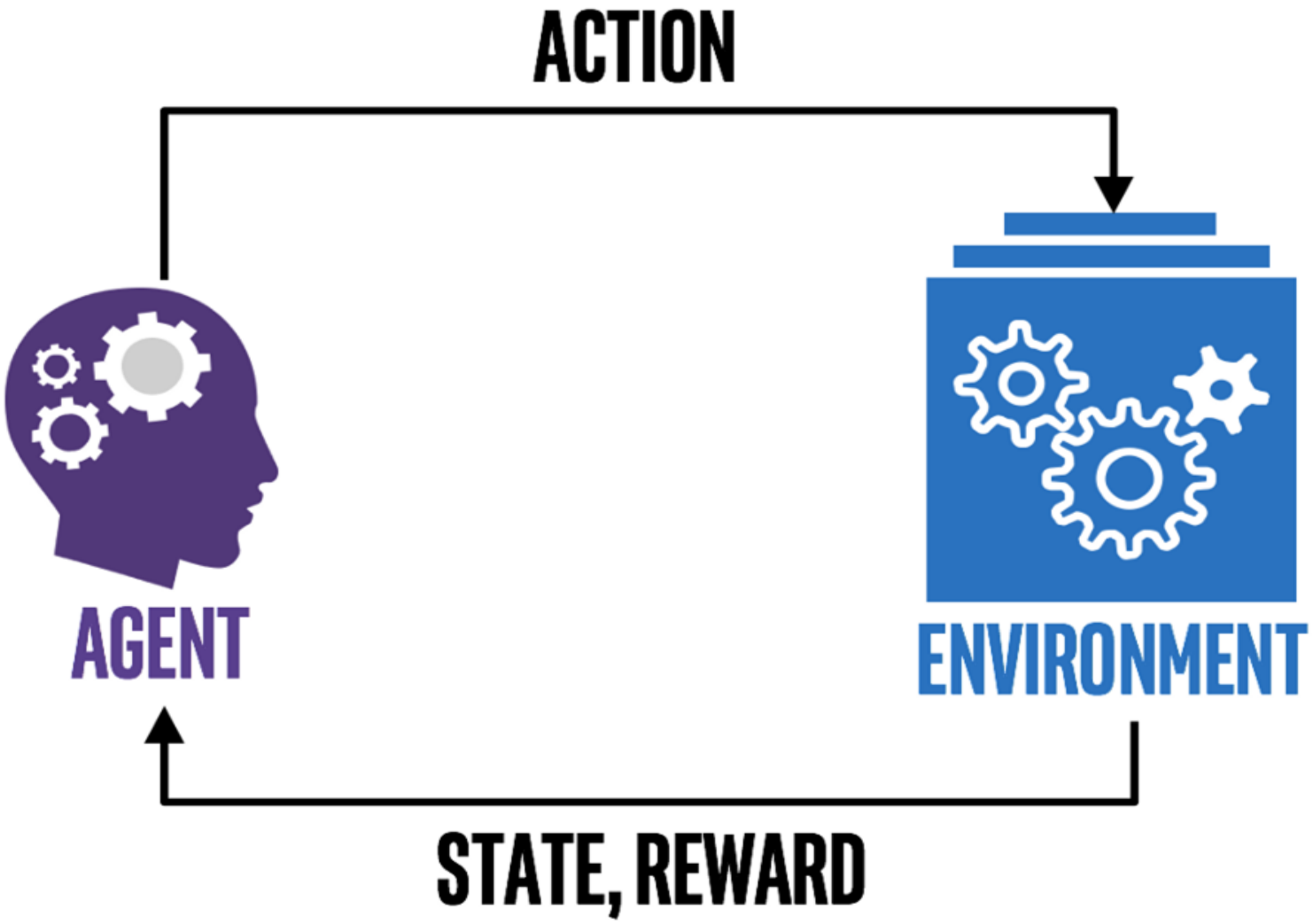
We want to learn: $f : X \rightarrow Y$

- We see examples of inputs **x** but not y
- We select y for observed x from available choices
- We get a feedback (reinforcement) from a **critic** about how good our choice of y was



- The goal is to select outputs that lead to the best reinforcement

REINFORCEMENT LEARNING



REINFORCEMENT LEARNING

- It must then learn by itself what is the best strategy
  - **Policy**: best strategy
- A policy defines what **action the agent should choose** when it is in a given situation.
- Example:
  - Playing games, Robotics
  - Robots learn how to walk.
  - DeepMind's AlphaGo

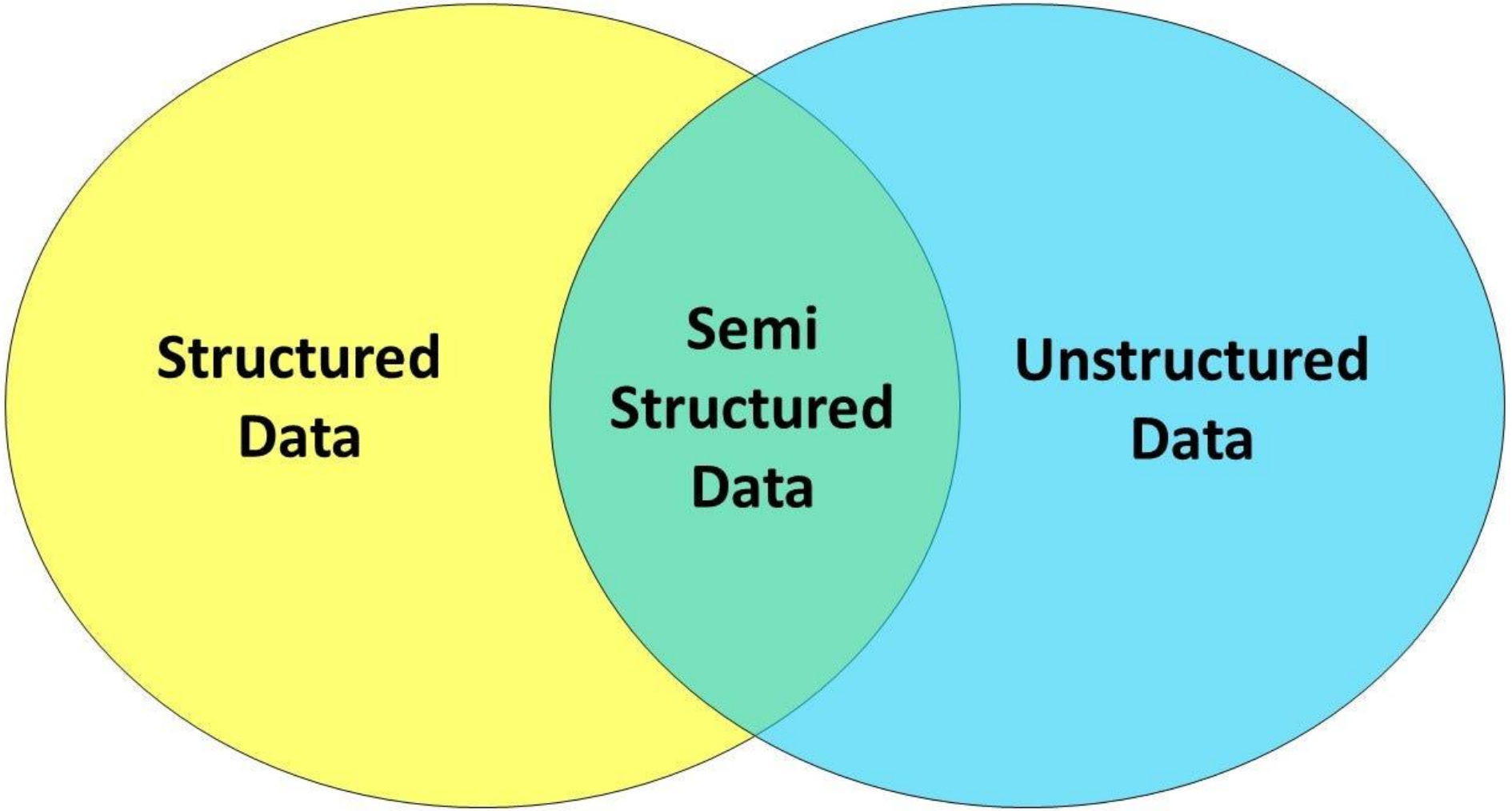
The Machine Learning Process

DATA AND TYPES

the term "data" refers to raw facts, observations, measurements, or information that can be collected and analyzed. Data is the foundation of any data-driven decision-making process and is essential for generating insights, making predictions, and uncovering patterns or trends.

- Structured Data
- Unstructured Data
- Semi-Structured Data

DATA AND TYPES



STRUCTURED DATA

- Structured data is organized and follows a **predefined format**, usually stored in **databases** or **spreadsheets**.
- **Example:** A customer database with columns for name, age, email, and purchase history.

  - Ordinal data
  - Nominal data
  - Numerical data

STRUCTURED DATA

STRUCTURED DATA

- **Nominal Data:**
  - Nominal data represents categories or labels with no inherent order or ranking among them.
  - **Examples** include gender, color, or types of fruits.

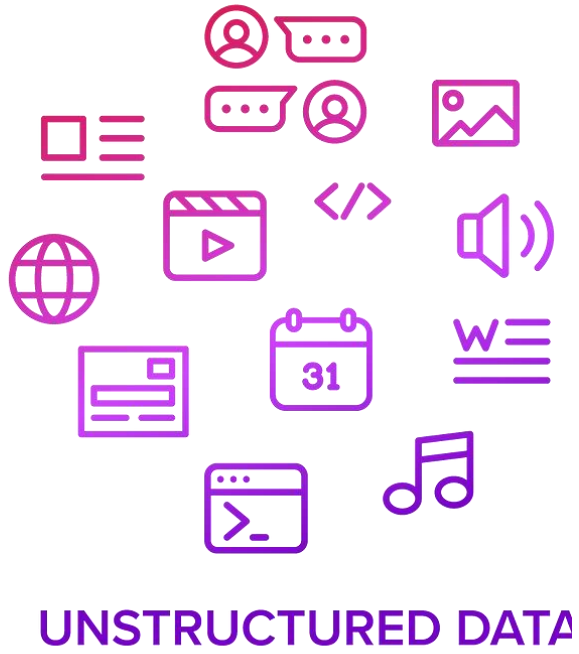
- **Ordinal Data:**
  - Ordinal data represents categories with a meaningful order or ranking.
  - The intervals between the categories are not uniform or measurable.
  - Examples include education levels (e.g., high school, college, graduate school) or customer satisfaction ratings (e.g., "low," "medium," "high").

STRUCTURED DATA

- **Numerical Data:**
  - Numerical data, also known as quantitative data, consists of numerical values that represent measurable quantities.
  - Numerical data can be further divided into two subtypes: **discrete and continuous.**
  - **Discrete numerical data** consists of separate, distinct values, often counted in whole numbers (e.g., the number of cars in a parking lot).
  - **Continuous numerical data** can take any value within a given range and can have decimal or fractional parts (e.g., height, weight, temperature).
  - Numerical data allows for mathematical operations such as addition, subtraction, multiplication, and division.

UNSTRUCTURED DATA

- Unstructured data is **not organized** and lacks a predefined format, often in the form of **text, images, audio, or video**.
- **Example:** Social media posts, customer reviews, or images from a surveillance camera.



Semi-Structured Data

- Semi-structured data has some organization but does not adhere to a **strict schema**, often **containing tags** or **labels.**
- **Example:** Emails, JSON files that contain data with tags or key-value pairs.

Structured Data

Often numbers or labels, stored in a structured framework of columns and rows relating to pre-set parameters.

- ID CODES IN DATABASES
- NUMERICAL DATA GOOGLE SHEETS
- STAR RATINGS

Semi-unstructured Data

Loosely organized into categories using meta tags

- EMAILS BY INBOX, SENT, DRAFT
- TWEETS ORGANIZED BY HASHTAGS
- FOLDERS ORGANIZED BY TOPIC

Unstructured Data

Text-heavy information that's not organized in a clearly defined framework or model.

- MEDIA POSTS, EMAILS, ONLINE REVIEWS
- VIDEOS, IMAGES
- SPEECH, SOUNDS

Thank You ☺