

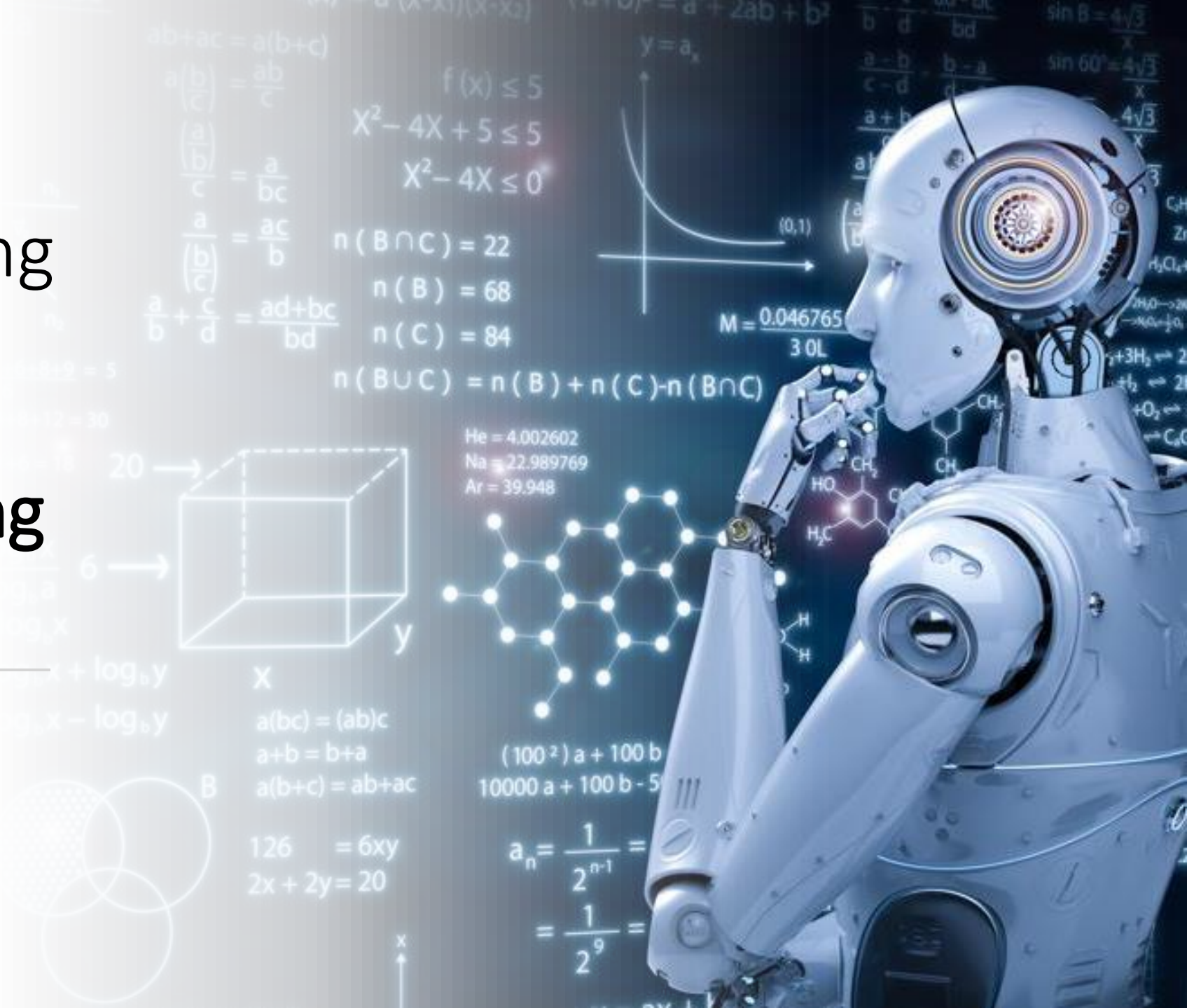
# Introduction to Machine Learning

## Lecture 1 Machine Learning

A. S. M. Sanwar Hosen

**Email:** sanwar@wsu.ac.kr

**Date:** 14 Sept. 2023



# Machine Learning

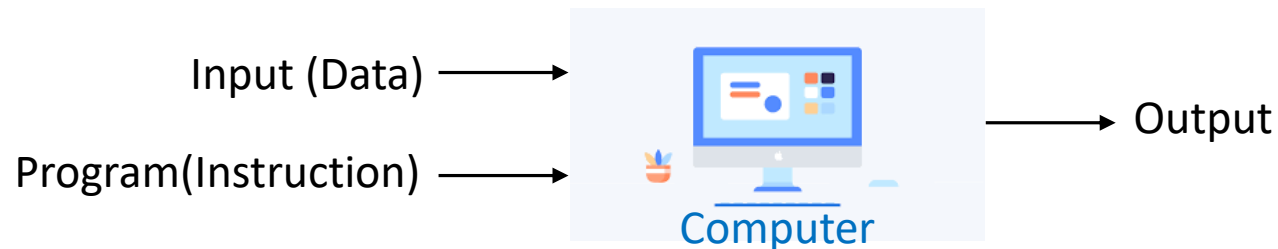
## ❑ What is Machine Learning (ML)?

“Learning is any process by which a system improves performance from experience”

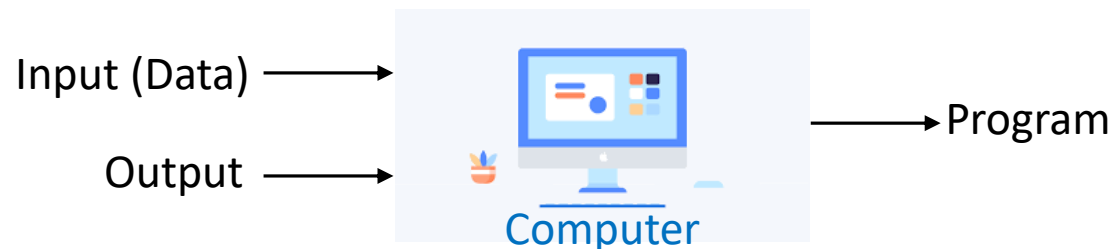
- Herbert Simon

## ❑ Traditional Programming vs. ML

**Traditional Programming:** It provides an output based on input and instruction given to a computer.



**Machine Learning (ML):** It provides a program based on input and output given to a computer.





# Machine Learning

## □ A Classic Example of ML

It is very hard to say what makes a 6!



Hard to say what number is it

Handwritten digits

# Machine Learning

## ☐ Solving Tasks Using ML

### ✓ Recognizing Patterns

- Facial identities or facial expression
- Handwritten or spoken words
- Medical image

### ✓ Generating Patterns

- Generating images or motion sequences

### ✓ Recognizing Anomalies

- Unusual credit card transactions
- Unusual patterns of sensor reading

### ✓ Predictions

- Future stock prices or currency exchange rates

# Machine Learning

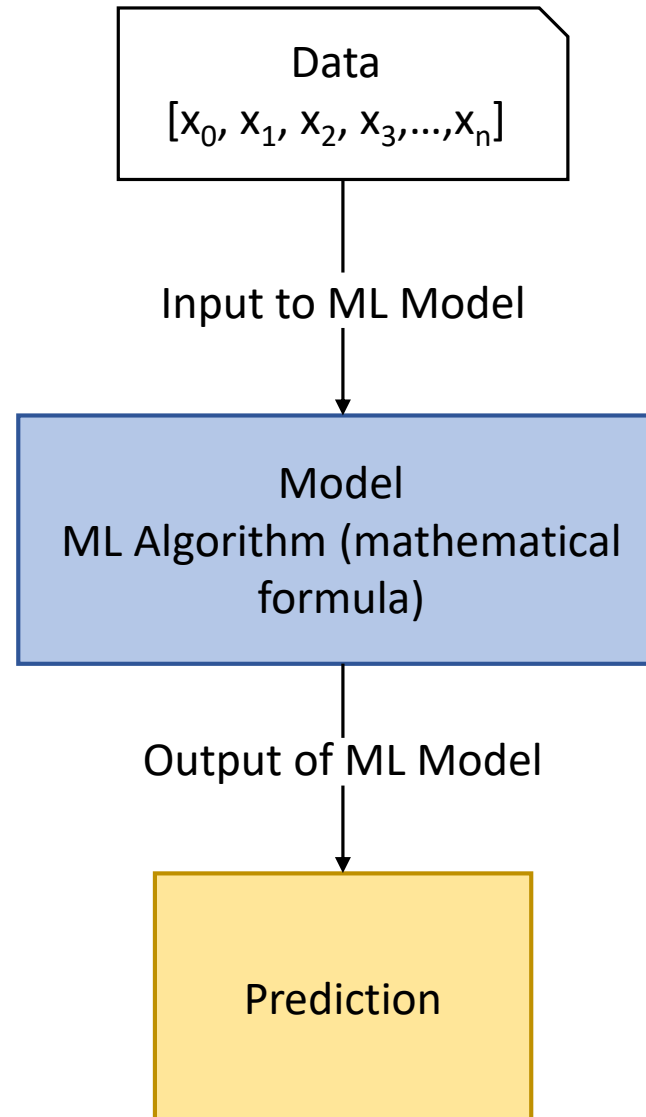
## ☐ Applications of ML (some of them)

- ✓ Web search (searching a topic you are interested in)
- ✓ Computational biology (i.e., DNA sequencing)
- ✓ E-commerce (i.e., market analysis and delivery management)
- ✓ Space exploration (i.e., finding a new planet in the universe)
- ✓ Robotics (i.e., thinking and active like a human)
- ✓ Information extraction (i.e., finding a key information you are interested in)
- ✓ Social networking (i.e., sentiment analysis of users)
- ✓ Debugging software (i.e., finding error in a program)
- ✓ Smart transportation (i.e., autonomous vehicle)
- ✓ Weather forecasting (i.e., predict weather conditions in near future)

# Machine Learning

## □ ML Model Diagram

- ✓ Data
- ✓ ML algorithm
- ✓ Prediction



# Machine Learning

## ❑ Data/Dataset in ML

A dataset in ML is collection of data pieces that can be treated by a computer as a single unit for analytic and prediction purposes.

## ❑ Feature Variables in a Dataset

A feature is a measurable property or characteristic of a phenomenon or an object that is used to analyze.

- ✓ **Input Feature:** It is a measurable property or characteristic of a phenomenon as input to the ML.
- ✓ **Output Feature/Label:** It is a measurable property or characteristic of a phenomenon as output to the ML.

**Dataset:** Student grading

Student ID	Subject	Score	Result
20111521	Physics	25.15	0 (F)
20192451	Physics	60.25	1 (P)
20221452	Physics	90.12	1 (P)
20192536	Physics	29.85	0 (F)

Input feature = (Student ID, Subject, Score)

Output feature/Label = Result



# Machine Learning

## ☐ Data Types in ML Dataset

- ✓ **Numerical:** It is information about quantities and represented as numbers. Examples are height, age, weight etc.
- ✓ **Categorical:** It is information that characterizes and approximates about a phenomena or an object. This data type is non-numeric in nature. Examples are color, test, feelings, names etc.

**Dataset:** Student grading

Student ID	Subject	Points	Result
20111521	Physics	25.15	0
20192451	Physics	60.25	1
20221452	Physics	90.12	1
20192536	Physics	29.85	0

Categorical data

Numeric data

# Machine Learning

## **ML Types**

- ✓ Supervised
- ✓ Unsupervised
- ✓ Semi-supervised
- ✓ Reinforcement

# Machine Learning

## ☐ Supervised ML

It refers to use of ML to identify patterns in datasets containing data points that are classified or labeled. A labelled dataset is one that has both input and output features.

**Dataset:** Purchasing trend of customers

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Input (Data)      Output (Data)

**Dataset:** Weather conditions

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Input (Data)      Output (Data)

# Machine Learning

## ☐ Supervised ML Types

It has two types:

- ✓ **Classification:** It is a ML process of categorizing a given set of data into classes. Labels in the dataset are defined.
- ✓ **Regression:** It is a ML technique to predict continuous values where labels are not defined. The goal of this model is to predict a value as much as closer to the actual output values the model can.

**Dataset:** Purchasing trend of customers

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

An example of classification dataset: predicting whether a customer will purchase a particular product “yes” (1) or “no” (0) based on different parameters- his/her gender, age and salary

**Dataset:** Weather conditions

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

An example of regression dataset: predicting wind speed based on different parameters- temperature, pressure, humidity, wind direction

# Machine Learning

## ❑ Classification

It has mainly two types:

- ✓ **Binary Classification:** The ML model predicts one class either 0 or 1 (yes or no).
- ✓ **Multi-class Classification:** The ML model predicts more than one class.

**Dataset:** Student grading

Student ID	Subject	Points	Result
20111521	Physics	25.15	0 (F)
20192451	Physics	60.25	1 (P)
20221452	Physics	90.12	1 (P)
20192536	Physics	29.85	0 (F)

An example of binary classification dataset

Output belong to either 0 or 1  
(one class)

**Dataset:** Student grading

Student ID	Subject	Points	Grade
20111521	Physics	25.15	D
20192451	Physics	60.25	B
20221452	Physics	90.12	A
20192536	Physics	29.85	D

An example of binary multi-class classification dataset

Output belong to  
more than one class

# Machine Learning

## ☐ Supervised ML Algorithms

- ✓ K-Nearest Neighbor (KNN)
- ✓ Support Vector Machine (SVM)
- ✓ Decision Tree (DT)
- ✓ Random Forest (RF)
- ✓ Gaussian Naïve Bayes (GN)
- ✓ Linear Regression (LR)

# Machine Learning

## ❑ Unsupervised ML

It is an ML technique in which the users do not need to supervise the model. Instead, it allows to model to work on its own discover patterns and information that was previously undetected. It mainly deals with the unlabeled data.

**Dataset:** Student information

ID	Name	Subject	Grade
1	John	Physics	D
2	Deny	Math	B+
3	Jenifer	Physics	A+
4	Wafiya	Business	A
5	Jullia	Math	B
6	Angel	Physics	A
7	Mark	Business	A+

**Dataset:** Average income of people

Age	Gender	Income	Profession	Tenure	City
35	M	60,000	IT	12	KRK
23	F	90,000	Sales	3	WAW
18	M	12,000	Student	1	KRK
42	F	128,000	Doctor	13	KRK
34	M	63,000	Manager	8	WAW
56	M	82,000	Teacher	30	WAW

No target value

No target value

# Machine Learning

## **Unsupervised ML Types**

It has mainly three types:

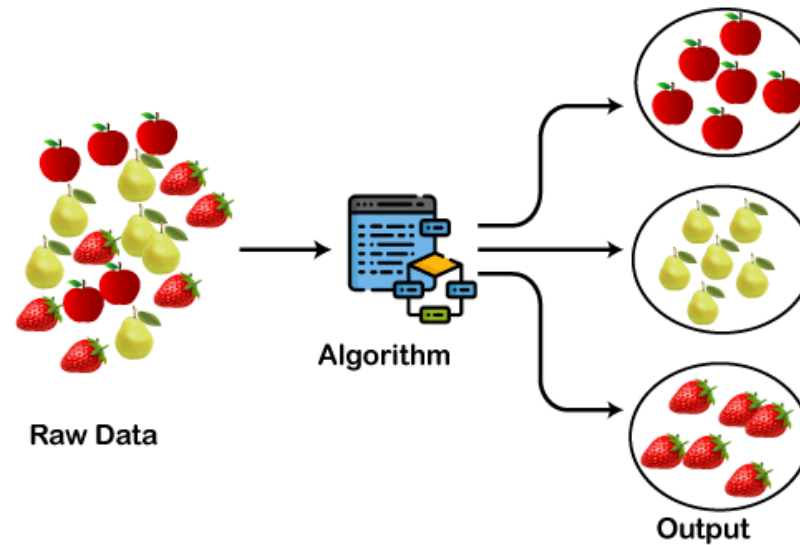
- ✓ Clustering
- ✓ Association
- ✓ Dimensionality Reduction



# Machine Learning

## ☐ Types of Unsupervised ML

- ✓ **Clustering:** It is an ML technique which groups unlabeled data based on their similarities or differences.

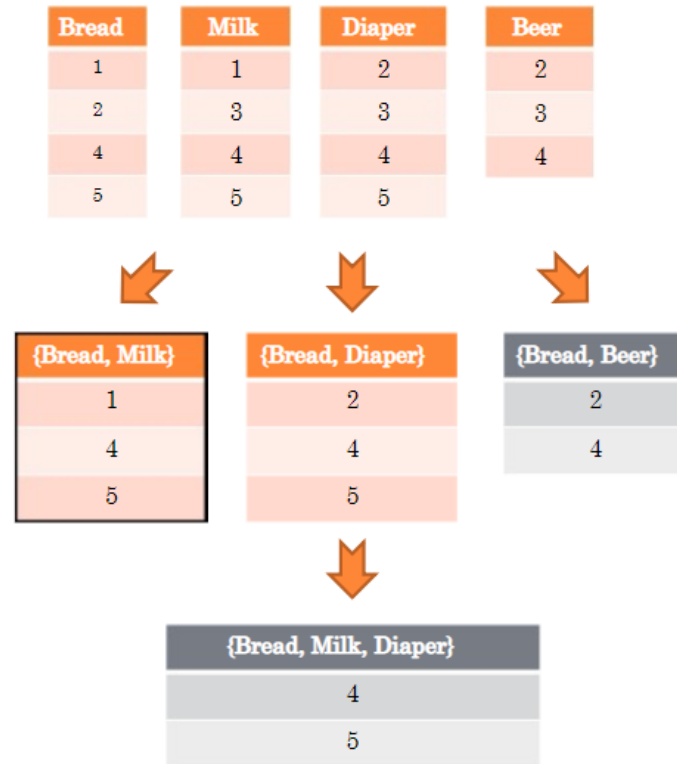


An example of clustering

# Machine Learning

## ☐ Types of Unsupervised ML (cont..)

- ✓ **Association:** It establishes association amongst data objects inside large dataset. This technique is used to discover interesting relationships between variables in large dataset. For example, a customer who buy a Bread most likely to buy Milk and Diaper.




An example of association

# Machine Learning

## ❑ Types of Unsupervised ML (cont..)

- ✓ **Dimensionality Reduction:** It transforms data from high-dimensional space into low-dimensional space. For example, data compression.

Age	Weight	Height
20	55	5.6
25	60	5.3
15	47	4.9
35	75	5.7



3.5	2.5
4.7	3.4
5.3	1.5

An example of data dimensionality reduction:  
(4 × 3) size data into (3 × 2) size compressed data

# Machine Learning

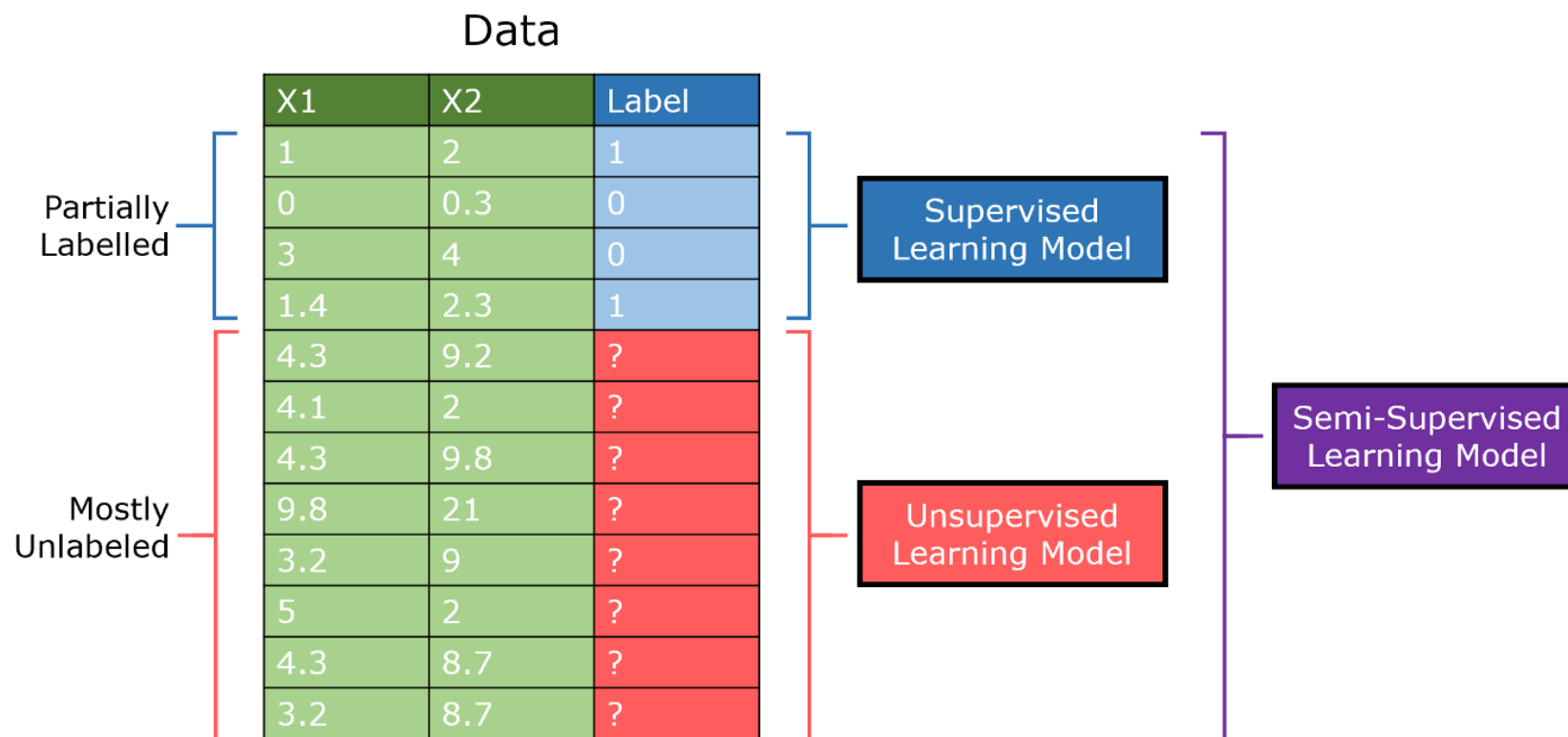
## ☐ **Unsupervised ML Algorithms**

- ✓ K-Nearest Neighbor (KNN) for clustering
- ✓ Principle Component Analysis (PCA) for association
- ✓ Autoencoders for dimensionality reduction

# Machine Learning

## Semi-supervised ML

It is a middle ground between supervised and unsupervised learning. In this learning model, a small portion of data are labeled and remains data are unlabeled.



# Machine Learning

## ☐ **Semi-supervised ML Algorithms**

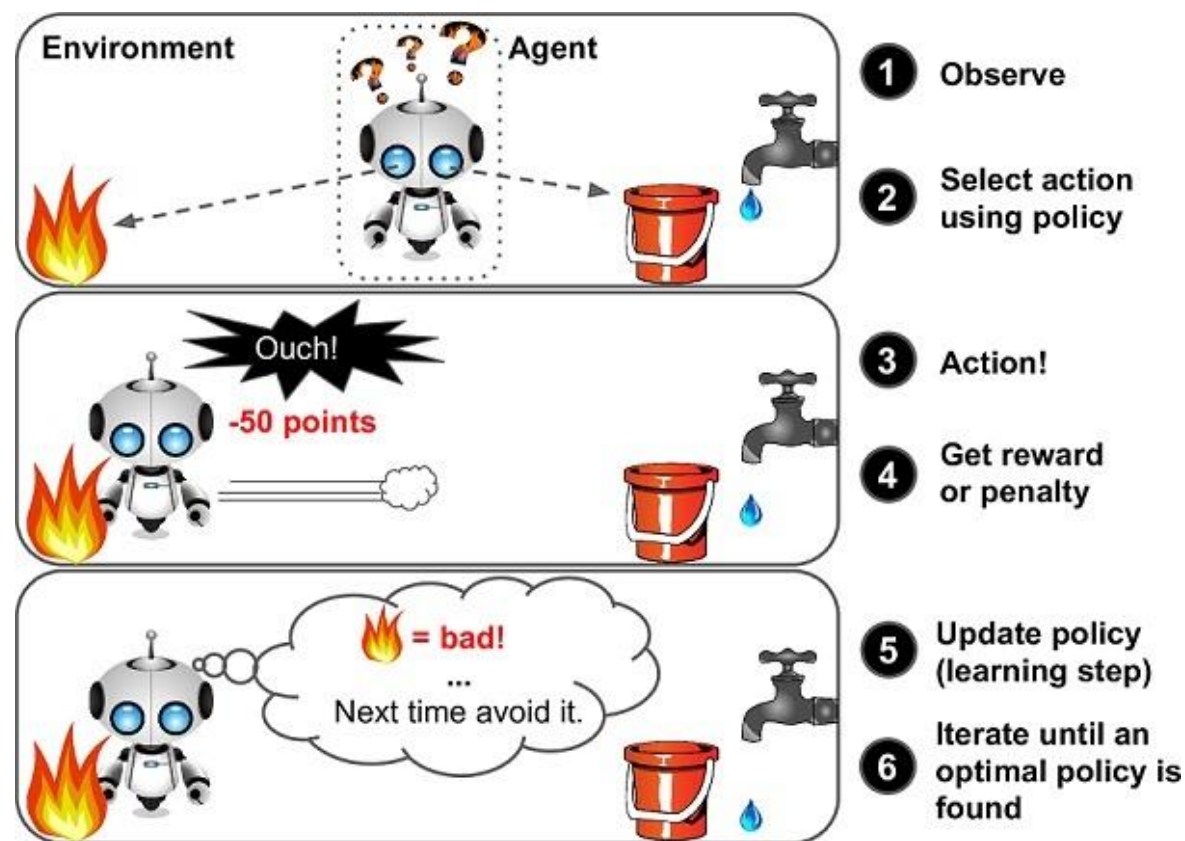
- ✓ Self-training
- ✓ Co-training
- ✓ Multi-view Learning
- ✓ Label Propagation
- ✓ Graph-based Method
- ✓ Semi-supervised Support Vector Machine (S3VM)
- ✓ Transductive Support Vector Machine (TSVM)
- ✓ Generative Models

# Machine Learning

## ☐ Reinforcement ML

It learns to achieve a goal in an uncertain, potential complex environment. It allows to make sequence of decisions based on reward and punishment. The components of reinforcement learning are:

- 1) Intelligent Agent
- 2) Environment
- 3) Actions



# Machine Learning

## ☐ Reinforcement ML Algorithms

- ✓ Q-Learning
- ✓ Deep Q-Network (DQN)
- ✓ Policy Gradient Methods (PG)
- ✓ Trust Region Policy Optimization (TROP)
- ✓ Soft Actor-Critic (SAC)
- ✓ Deep Deterministic Policy Gradients (DDPG)
- ✓ Proximal Policy Optimization (PPO)
- ✓ Monte Carlo Tree Search (MCTS)
- ✓ Twin Delayed Deep Deterministic Policy Gradient (TD3)



# Machine Learning

## ❑ Difference between Supervised and Unsupervised ML

### ✓ Supervised ML

- 1) It uses labeled data.
- 2) It predicts future outcomes or assign data to specific categories based on the regression or classification problem.
- 3) It requires human intervention to label the data appropriately.
- 4) The computational complexity is less as the data are labeled.

### ✓ Unsupervised ML

- 1) It uses unlabeled data.
- 2) It predicts future outcomes or assign data to categories based on clustering or association problem.
- 3) It does not require human intervention to label the data.
- 4) The computational complexity is higher than supervised ML.

# Machine Learning

## ☐ Lecture Overview

- ✓ What is ML?
- ✓ ML Applications
- ✓ Components of ML Model
- ✓ What is Data/Dataset in ML?
- ✓ Data Types in ML Dataset
  - Numerical
  - Categorical
- ✓ ML types
  - Supervised ML (Classification and Regression)
  - Unsupervised ML (Clustering, Association and Dimensionality Reduction)
  - Semi-supervised ML
  - Reinforcement ML
- ✓ Difference between Supervised and Unsupervised ML