Machine Learning

1

AI, ML, DL and GI – How it all fits together!

Data Analytics

Artificial Intelligence

Machine Learning

Deep Learning
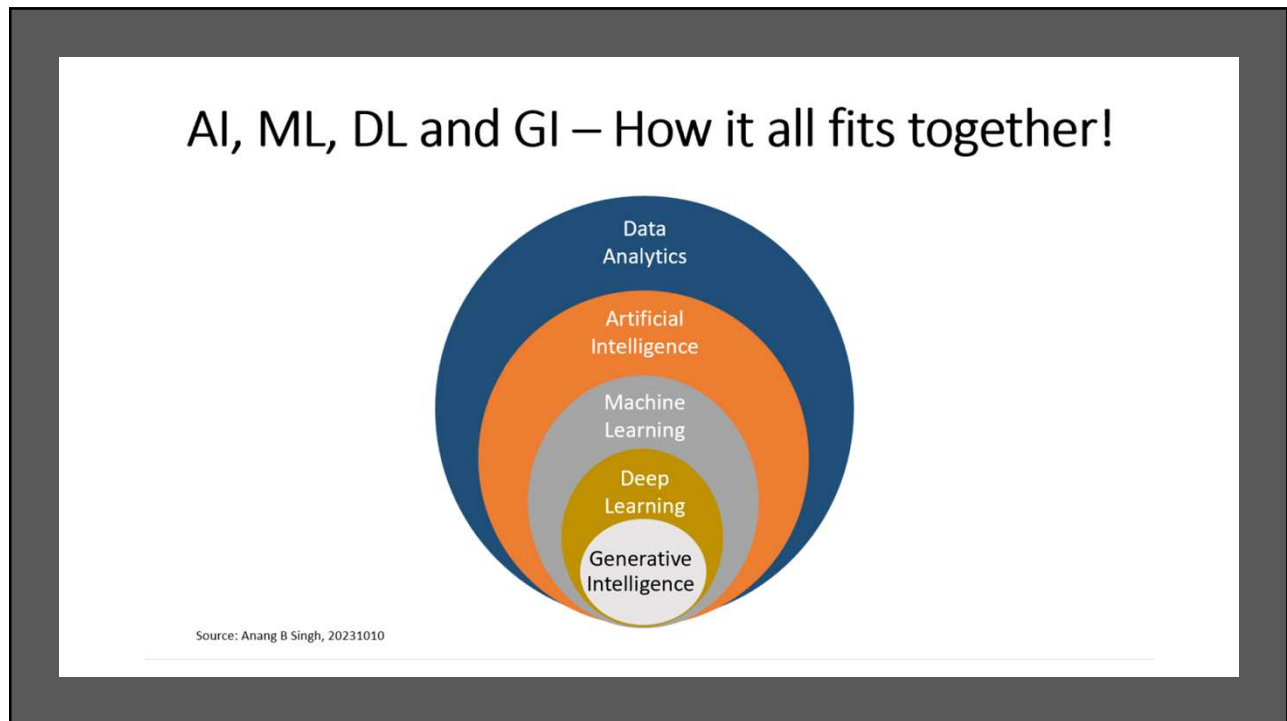
Generative Intelligence

Source: Anang B Singh, 20231010

2

Artificial Intelligence

John McCarthy, widely recognized as one of the godfathers of AI, defined it as **"the science and engineering of making intelligent machines."** (1956)
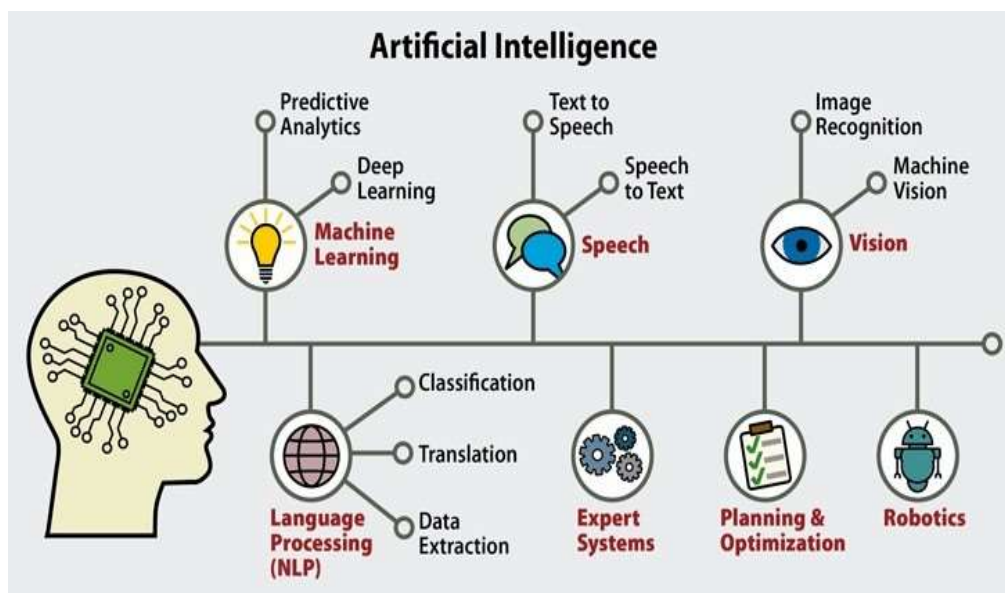
A branch of computer science dealing with the simulation of intelligent behavior in computers (vision – object detection and classification from images, scene understanding and many more….).

The capability of a machine to imitate intelligent human behavior (**Goal of General Purpose AI**).
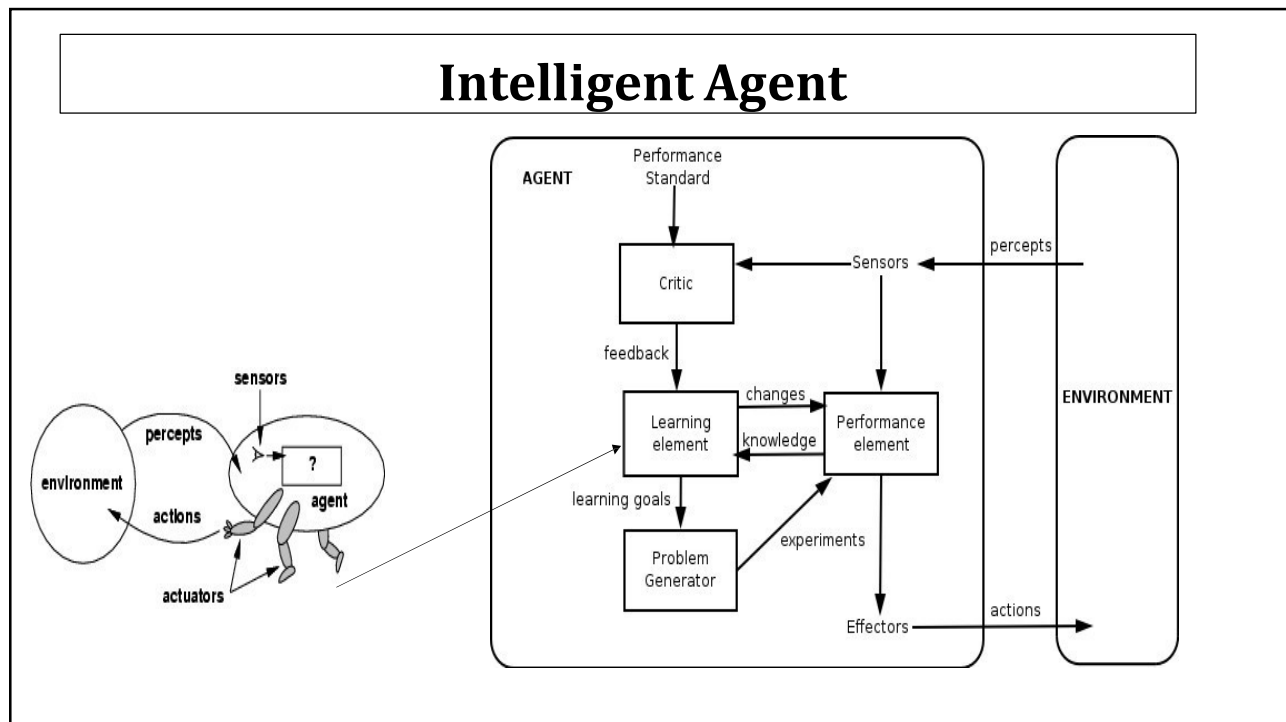
A computer system **able to perform tasks** that normally **require human intelligence**, such as visual perception, speech recognition, decision-making and translation between languages.
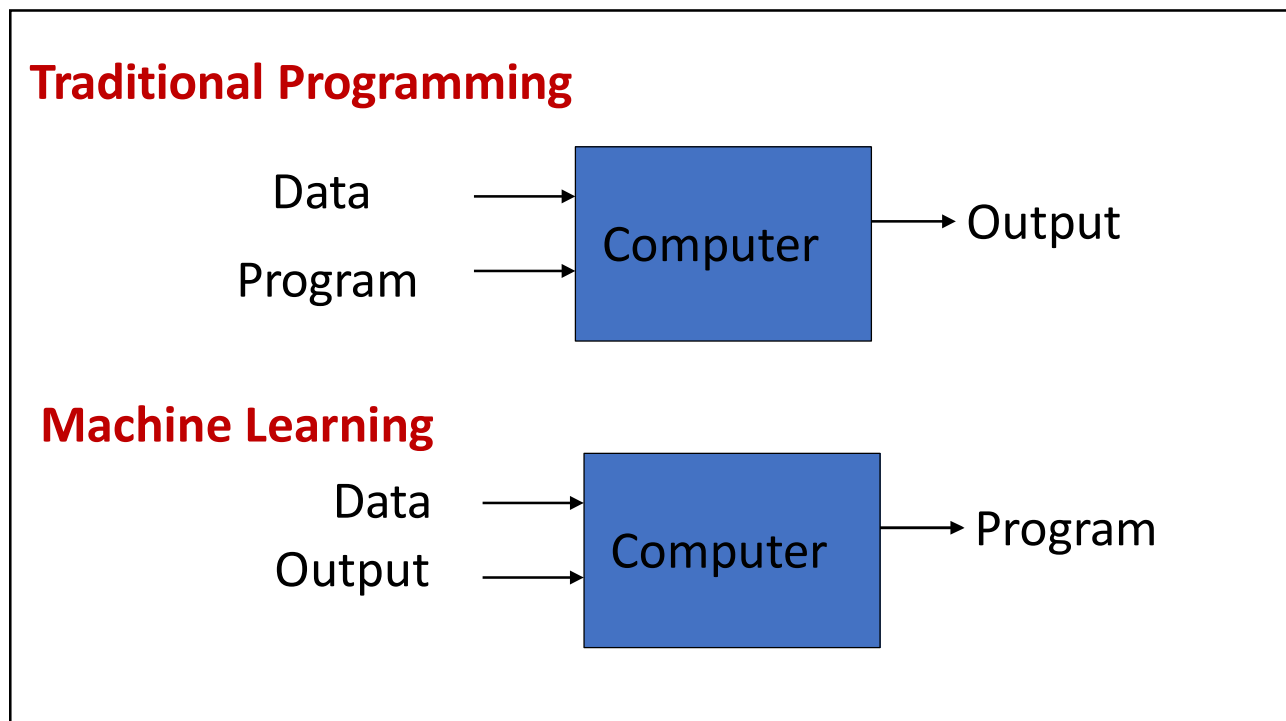
3

4

Intelligent Agent

AGENT

Performance Standard

percepts

Sensors

Critic

feedback

changes

knowledge

Learning element

Performance element

ENVIRONMENT

learning goals

experiments

Problem Generator

Effectors

actions

sensors

percepts

environment

actions

agent

?

actuators

5

Traditional Programming

Data

Program

Computer

Output

Machine Learning

Data

Output

Computer

Program

6

What is Machine Learning?

- Subset of Artificial Intelligence(AI)
- Gives "computers the **ability to learn without being explicitly programmed**" (Arthur Samuel, 1959)
- Focuses on the development of computer programs that can **access data** and use it to **learn for themselves**.
- The process of learning begins with data to **find patterns in the data** and **make better decisions** in the future (predictions) based on examples provided in the data.
- Aim is to allow computers to learn automatically **without human intervention** and adjust actions accordingly.

7

Machine Learning = Learning from Data

- A computer program is said to **learn from experience E** for 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 Mitchell, Book on Machine Learning*

  Spam Email Detection

  - T: Classify emails as spam or not spam
  - E: Watch the user label emails as spam or not spam
  - P: Percentage of correctly identified emails

- **Ability to modify itself when exposed to more data**
- **Dynamic/adaptive** and **does not require human intervention** to make certain changes.
- Less brittle – more flexible
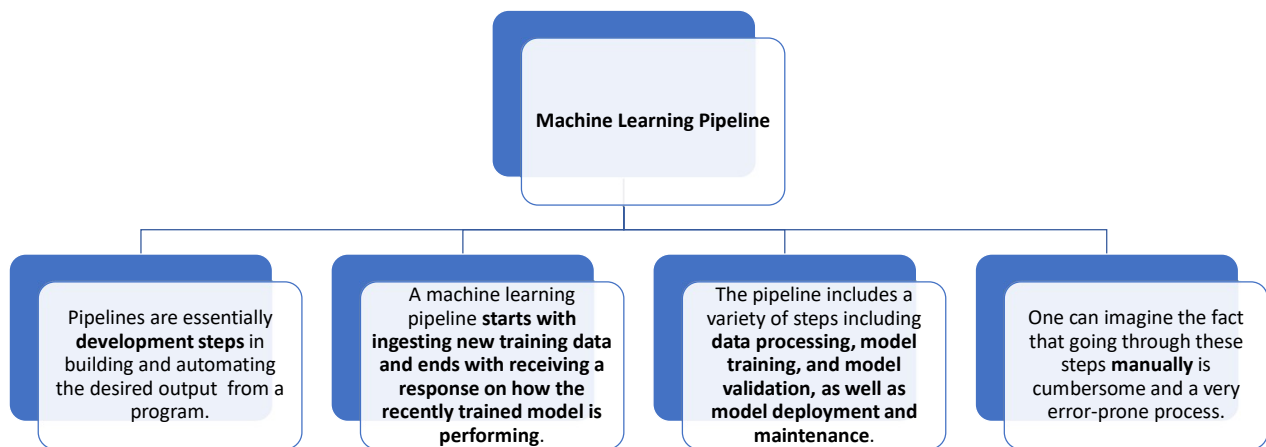- Less reliant on human experts – less biased

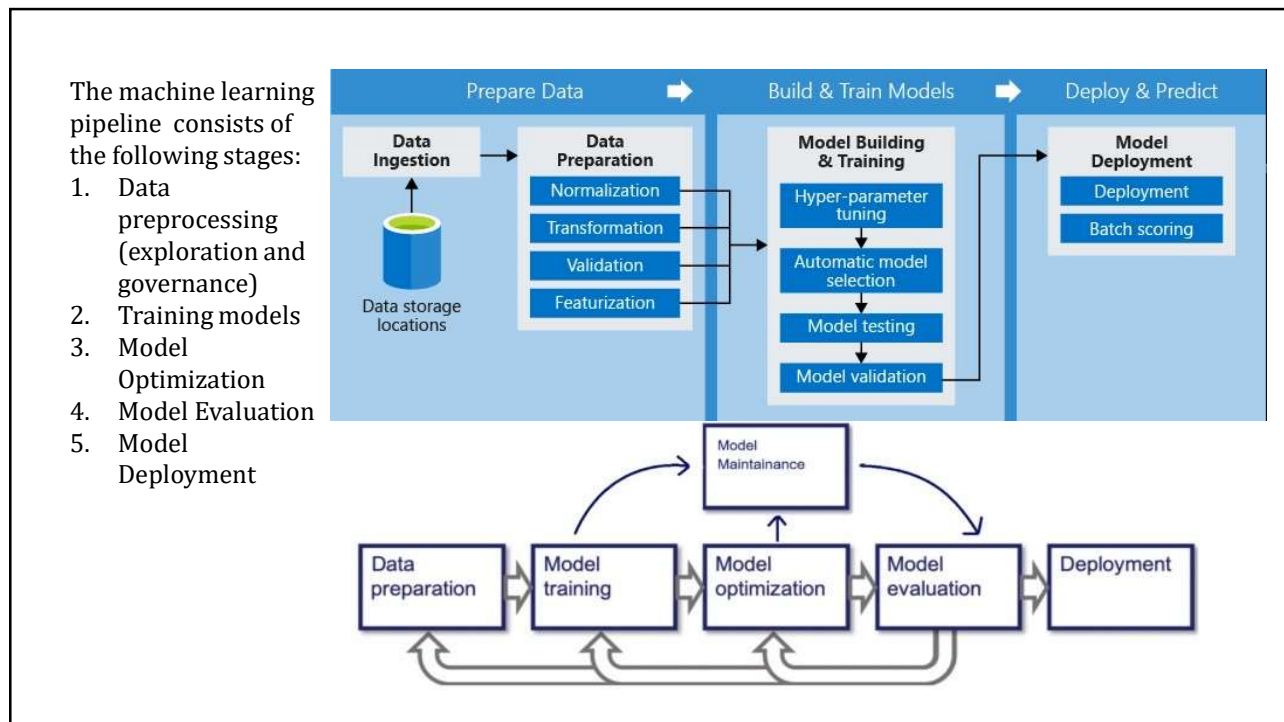For

8

4

Machine Learning

- **Learning from Data.**
- The "learning" part of machine learning means that ML algorithms attempt to **optimize** along a certain dimension; i.e. they usually try to **minimize error** or **maximize the likelihood** of their predictions being true.
- Optimizing an **error/loss/cost function**.
- Learning from a **mathematical equation representing the relationship between the inputs and output**.
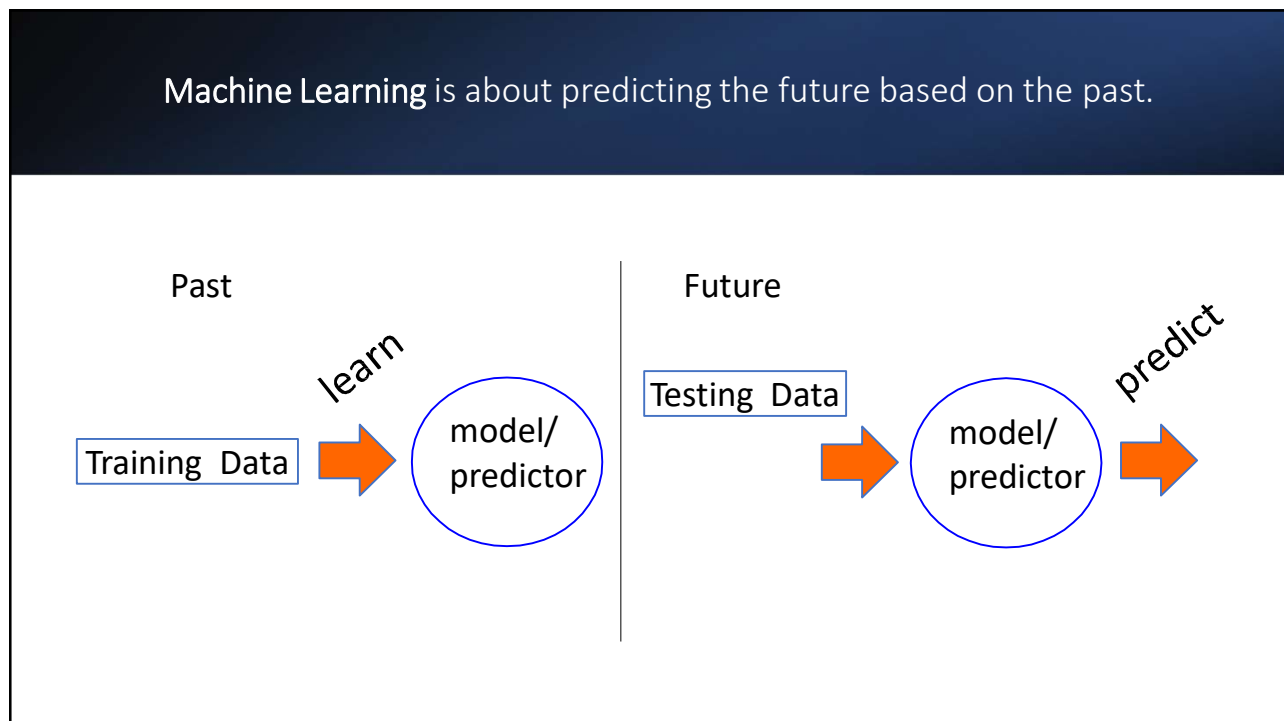
9

A machine learning pipeline encapsulates and automates the multiple Machine Learning processes such as performing data extractions and wrangling, creating training models, model validation, and model deployment for predictions.

Machine Learning Pipeline

Pipelines are essentially **development steps** in building and automating the desired output from a program.

A machine learning pipeline **starts with ingesting new training data and ends with receiving a response on how the recently trained model is performing**.

The pipeline includes a variety of steps including **data processing, model training, and model validation, as well as model deployment and maintenance**.

One can imagine the fact that going through these steps **manually** is cumbersome and a very error-prone process.

10

The machine learning pipeline consists of the following stages:
1. Data preprocessing (exploration and governance)
2. Training models
3. Model Optimization
4. Model Evaluation
5. Model Deployment

11

Machine Learning is about predicting the future based on the past.

Past

Future

learn

predict

Training Data ➡ model/predictor

Testing Data ➡ model/predictor ➡

12

Training and Testing

•**Training** is the process of **making the system able to learn**.

•No free lunch rule:
  •Training set and Testing set come from the **same distribution**
  •Need to make **some assumptions or bias**
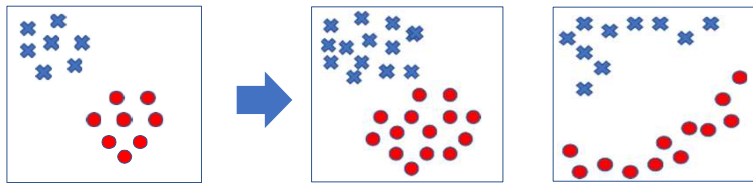
13

"No Free Lunch" :(

D. H. Wolpert. The supervised learning no-free-lunch theorems. In Soft Computing and Industry, pages 25–42. Springer, 2002.

Our model is a simplification of reality

⇩

Simplification is based on assumptions (model bias)

⇩

Assumptions fail in certain situations

Roughly speaking:

"No one model works best for all possible situations."

14

15

Train
Evaluate
Predict

16

Types of Learning

17

Supervised learning (classification)

Supervised learning (regression)
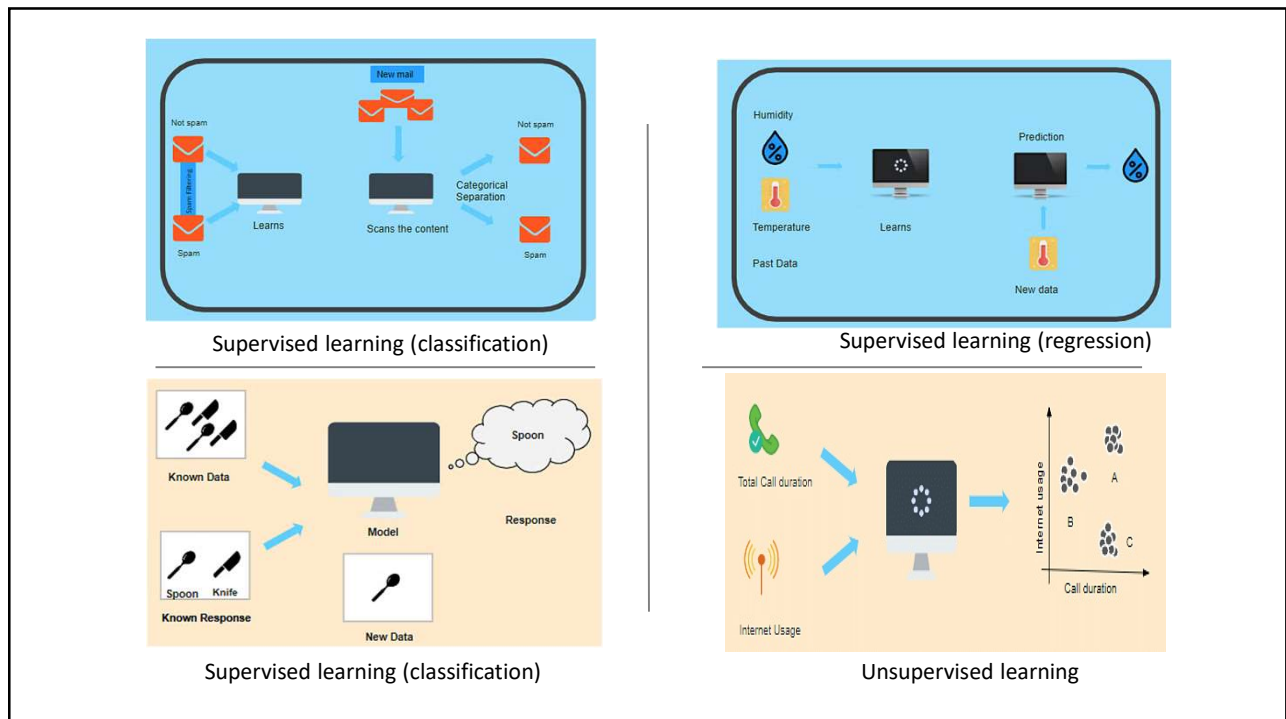
Supervised learning (classification)
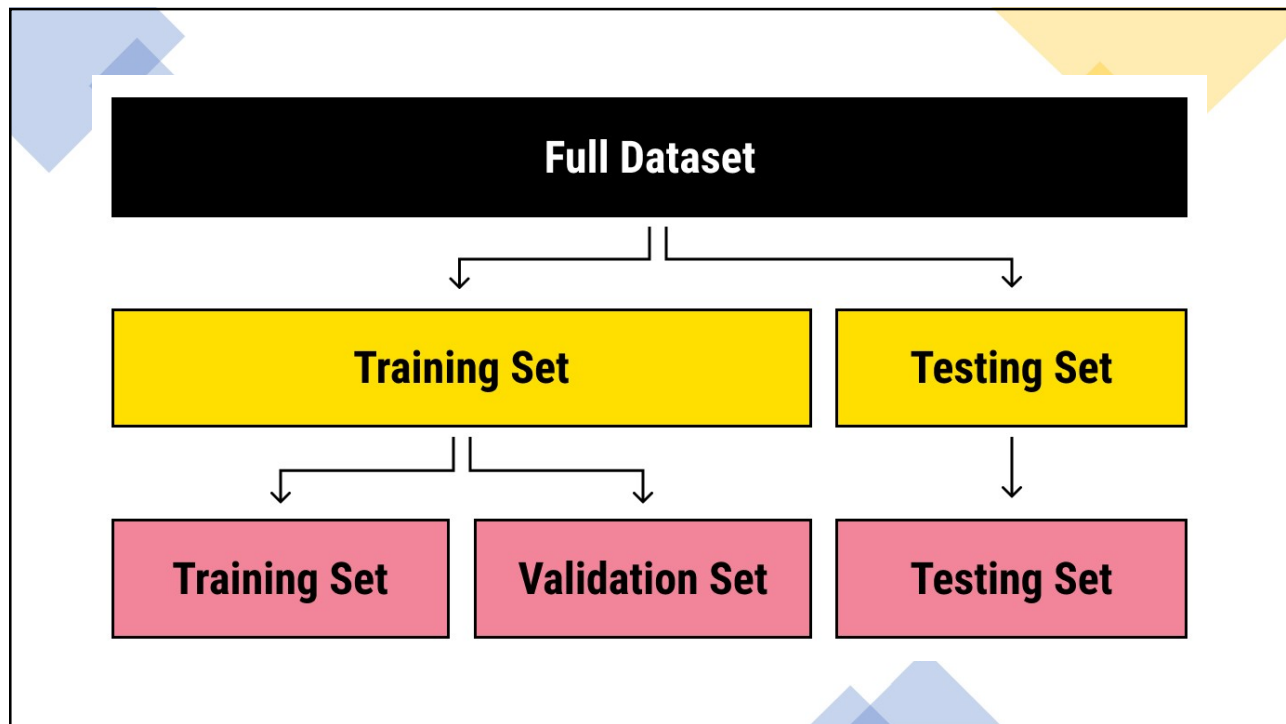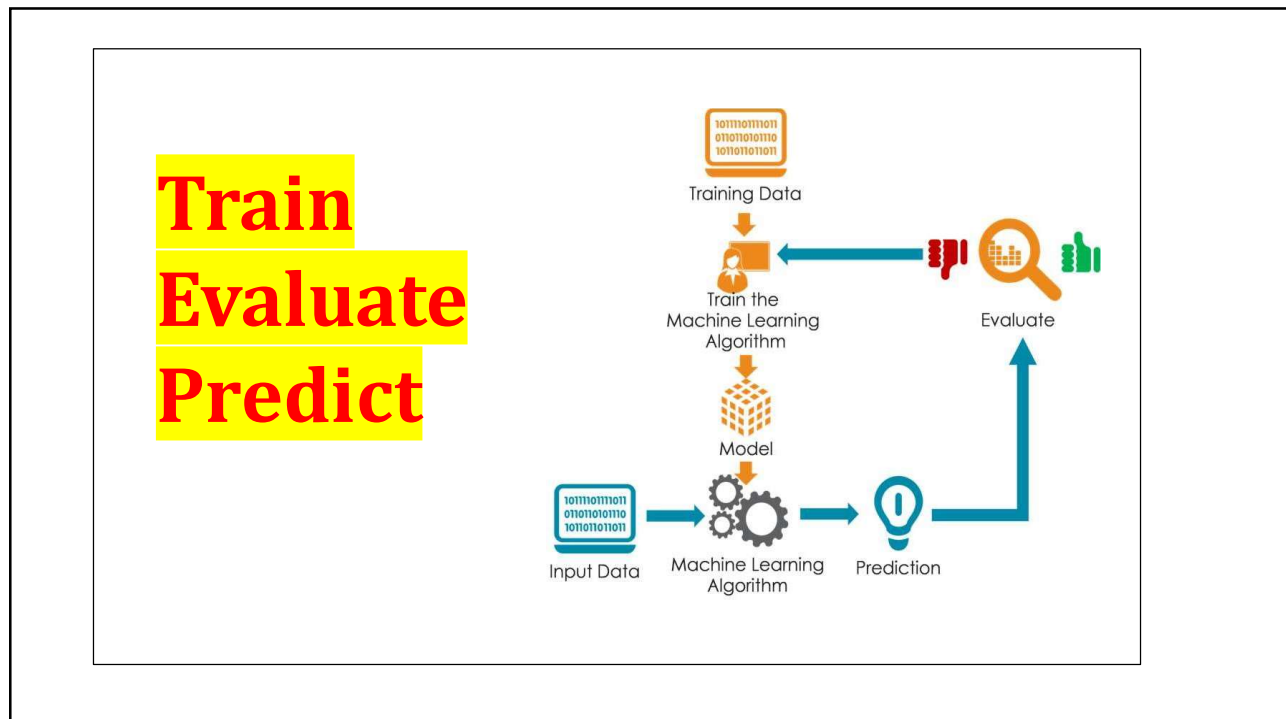
Unsupervised learning
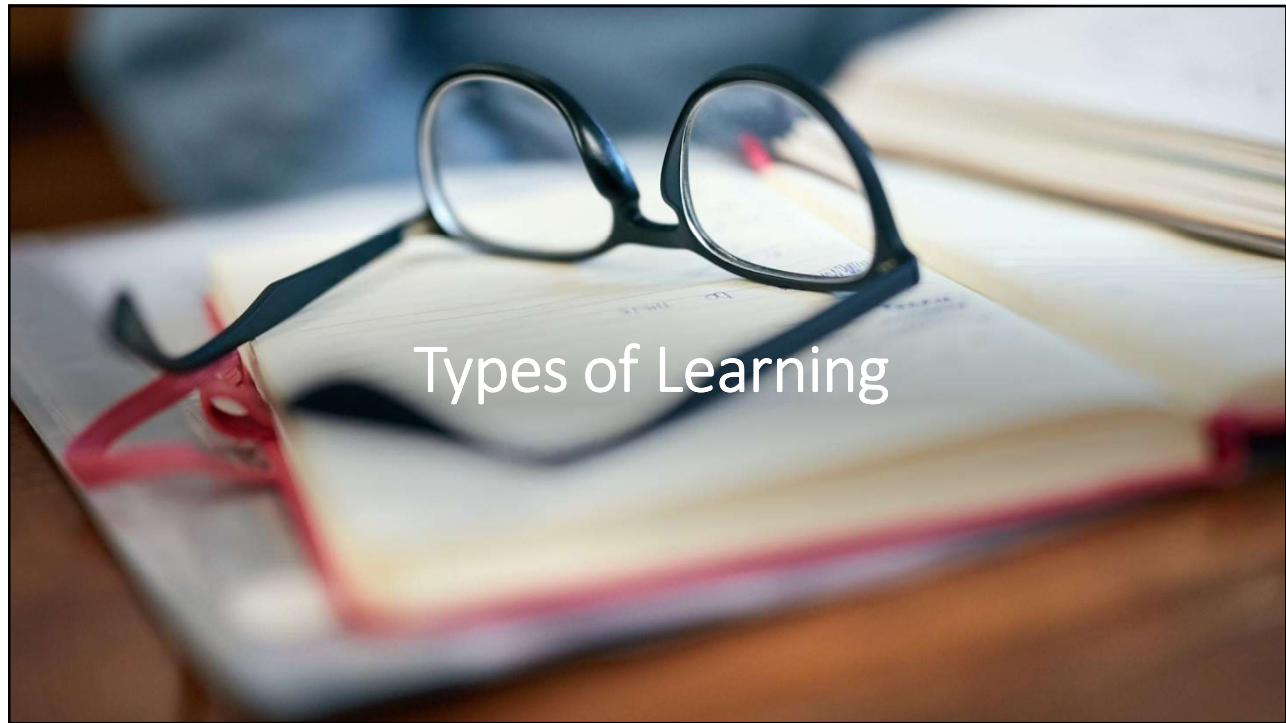
18

Types of Learning

- **Supervised Learning: Labelled Data** $\{x_n \in R^d, y_n \in R\}_{n=1}^N$
  - Prediction
  - Classification (discrete labels), Regression (real values)

- **Unsupervised Learning: Un-labelled Data** $\{x_n \in R^d\}_{n=1}^N$
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction

- **Semi-Supervised Learning: Only a part of the data is labeled**
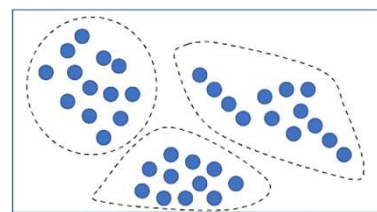
- **Reinforcement Learning: Rewards from a sequence of actions**
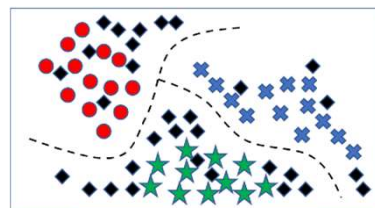 - Decision making (robot, chess machine)

19

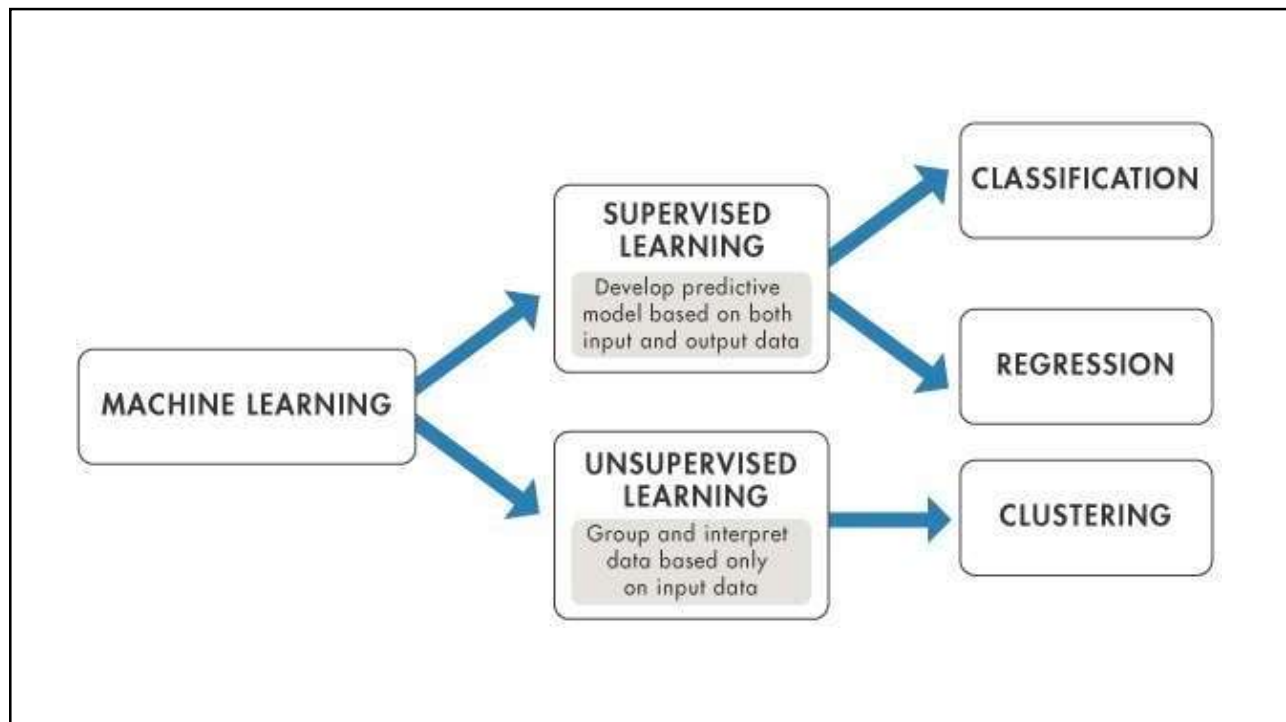## Algorithms: Types of Learning
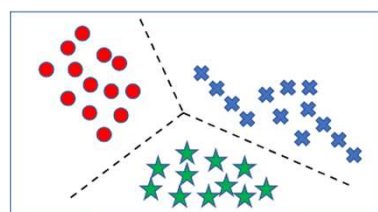
Supervised learning

Unsupervised learning

Semi-supervised learning
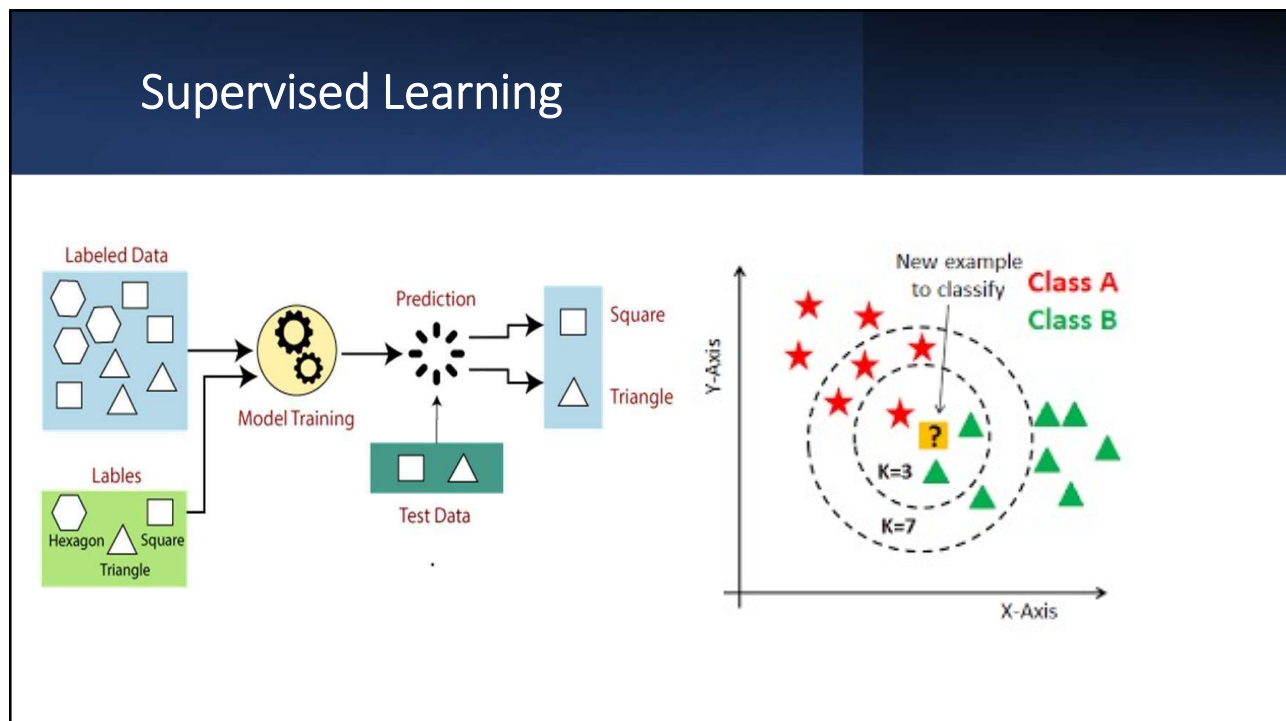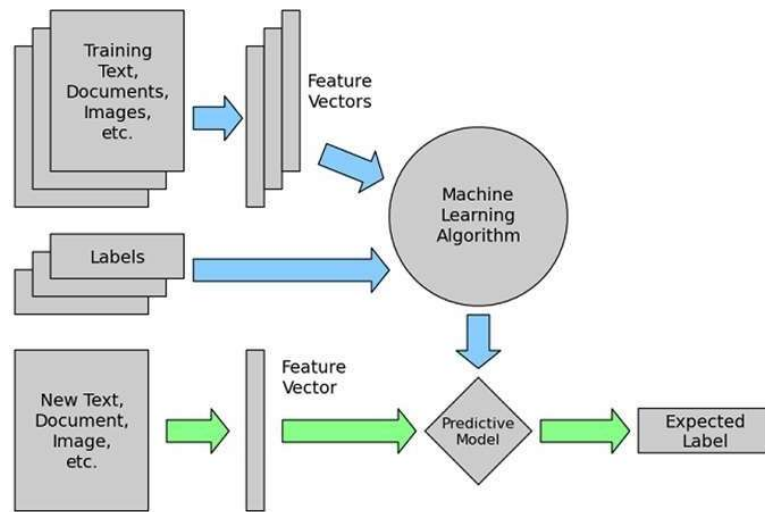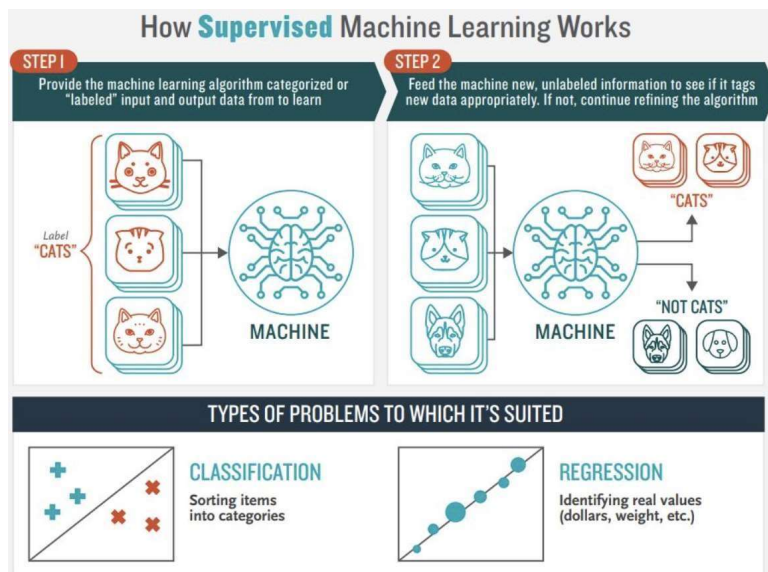
20

21

Supervised Learning
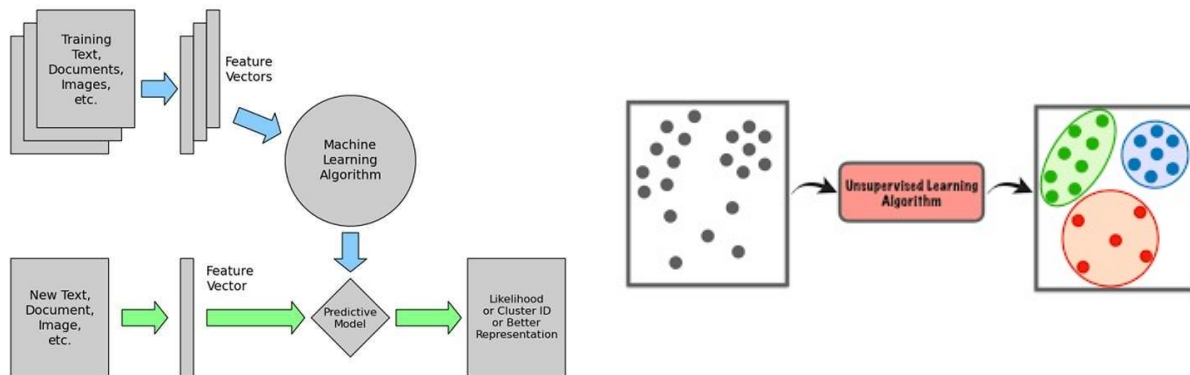

22

Supervised Learning Model

23

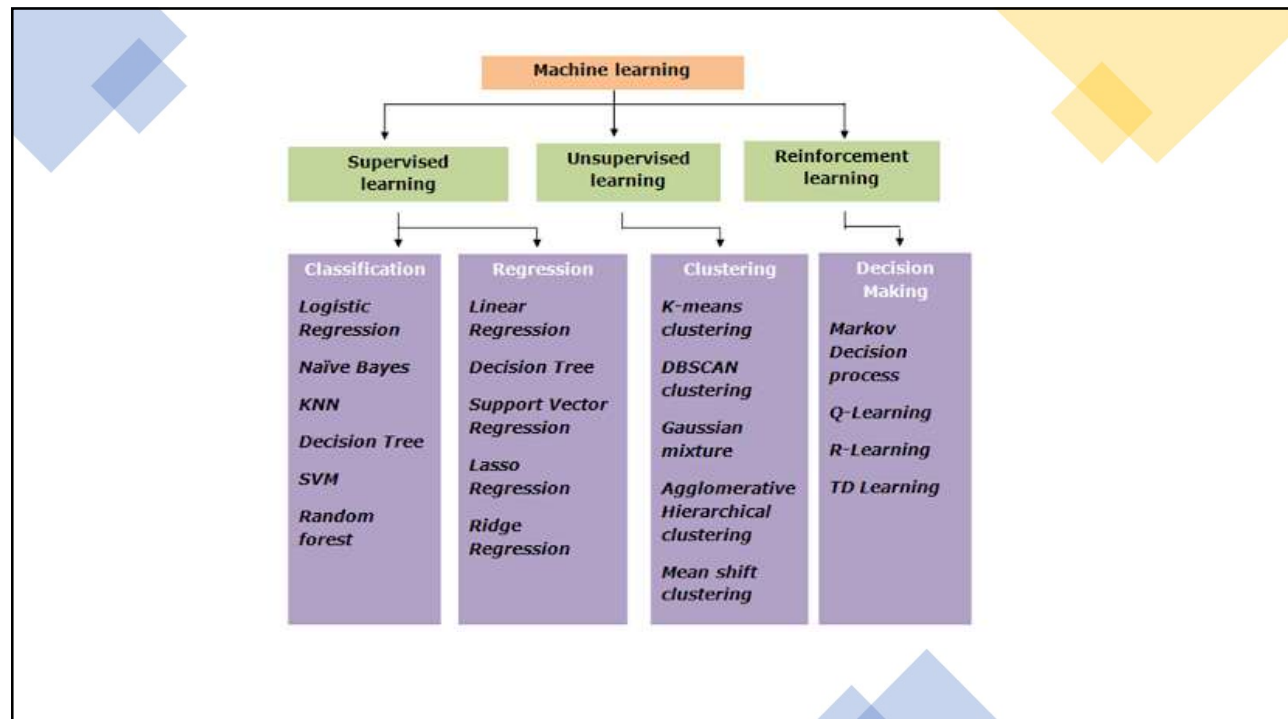## Example Supervised Learning
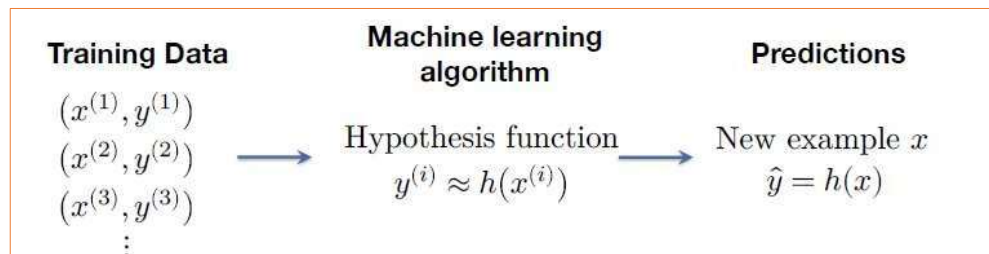

24

Unsupervised Learning Model

25

26

Machine Learning

Basic Idea: In many domains, it is difficult to hand-build a predictive model, but easy to collect lots of data; machine learning provides a way to automatically infer the predictive model from data.

Supervised Learning:

Training Data	Machine learning algorithm	Predictions
$(x^{(1)}, y^{(1)})$		
$(x^{(2)}, y^{(2)})$	Hypothesis function	New example x
$(x^{(3)}, y^{(3)})$	$y^{(i)} \approx h(x^{(i)})$	$\hat{y} = h(x)$
$\vdots$		

27

Terminology

Input features: $x^{(i)} \in \mathbb{R}^n, i = 1, \dots, m$

$$\text{E. g.: } x^{(i)} = \begin{bmatrix} \text{High\_Temperature}^{(i)} \\ \text{Is\_Weekday}^{(i)} \\ 1 \end{bmatrix}$$

Outputs: $y^{(i)} \in \mathcal{Y}, i = 1, \dots, m$

$$\text{E. g.: } y^{(i)} \in \mathbb{R} = \text{ Peak\_Demand}^{(i)}$$

Model parameters: $\theta \in \mathbb{R}^n$

Hypothesis function: $h_\theta \colon \mathbb{R}^n \to \mathcal{Y}$, predicts output given input

$$\text{E. g.: } h_\theta(x) = \sum_{j=1}^{n} \theta_j \cdot x_j$$

28

Terminology

Loss function: $\ell: \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$, measures the difference between a prediction and an actual output

$$\text{E. g.:} \quad \ell(\hat{y}, y) = (\hat{y} - y)^2$$

The machine learning optimization problem:

$$\underset{\theta}{\text{minimize}} \sum_{i=1}^{m} \ell(h_\theta(x^{(i)}), y^{(i)})$$

Virtually every machine learning algorithm has this form, just specify

- What is the hypothesis function?
- What is the loss function?
- How do we solve the optimization problem?

29

Example of ML Algorithms

- **Least Squares:** {linear hypothesis, squared loss, (usually) analytical solution}
- **Linear Regression:** {linear hypothesis}
- **Support Vector Machine:** {linear or kernel hypothesis, hinge loss}
- **Neural Network:** {Composed non-linear function,   (usually) gradient descent)
- **Decision Tree:** {Hierarchical   axis-aligned   halfplanes, greedy optimization}
- **Naïve Bayes:** {Linear hypothesis, joint probability under certain independence assumptions, analytical solution}

30

Loss function vs. Cost Function

- The loss function computes the error for a single training example, while the cost function is the average of the loss functions of the entire training set.
  - If we have *m* training data like this $\{(x_1,y_1),(x_2,y_2), \ldots (x_m,y_m)\}$.
  - $\hat{y}_i$ = output of the model for training example x_i *(prediction)*
  - y_i = expected output/true value for training example x_i *(reference/ground-truth)*
- The **loss function $L(\hat{y}_i, y_i)$** to defines the error/difference between $\hat{y}_i$ and y_i for the single training example x_i.
- This means **loss** refers to an **error in model output for an individual sample**.

If we want to find loss over **all the training examples present in a training-set**, we refer to it as the **cost function** (i.e. total or average loss over all training examples). This is the estimate of the **total error** computed for the whole training set.

31

Key Differences

Feature	Loss Function	Cost Function
Definition	Measures error for a single data point	Measures average error over the entire dataset
Scope	Individual sample	Whole dataset
Usage	Evaluates single instance performance	Used for optimization and parameter updates
Example	MSE for one sample	MSE for all samples

32

Loss/Error/Cost – Objective Function

The more general scenario is to define an objective function first, which we want to optimize. This objective function could be to

1. minimize a mean squared error cost (or loss) function (CART, Decision Tree regression, Linear Regression)
2. maximize the posterior probabilities (e.g., Naive Bayes)
3. maximize a fitness function (Genetic Programming)
4. maximize the total reward/value function (Reinforcement Learning)
5. maximize information gain/minimize child node impurities (CART Decision Tree classification)
6. maximize log-likelihood or minimize cross-entropy loss (or cost) function (ANN), minimize hinge loss (Support Vector Machine)

33

Some General Applications of ML

Classification/ Regression/ Clustering/Association

- **Image Processing**
- **Robotics**
- **Data Mining**
- **Recommender Systems**
- **Game Playing**
- **Text processing**
- **Health care**

34

References

- Trevor Hastie, Robert Tibshirani, Jerome Friedman; "*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*"; Springer.

- Tom Mitchell; "*Machine Learning*", McGraw Hill

- Stuart Russell, Peter Norvig; "*Artificial Intelligence: A Modern Approach*"; Prentice Hall.

35