

# Introduction to ML

ML Instruction Team, Fall 2022

CE Department  
Sharif University of Technology

# Machine Learning: An Overview

## ■ What is Machine Learning?

Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.

## ■ Applications of Machine Learning

- ▶ This Person Does not Exist!
- ▶ Github Copilot
- ▶ Imagen
- ▶ Dall-E Open AI
- ▶ DocQuery
- ▶ Zero Shot Object Detection!

# Machine Learning Categories

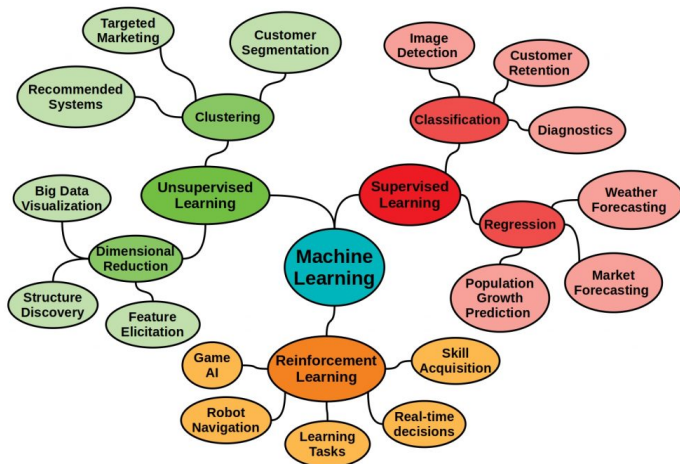


Figure: Classical Machine Learning Paradigm, Source

# Machine Learning Categories

■ The three broad categories of ML are summarized in:

- ▶ **Supervised Learning**
- ▶ **Unsupervised Learning**
- ▶ **Reinforcement Learning**

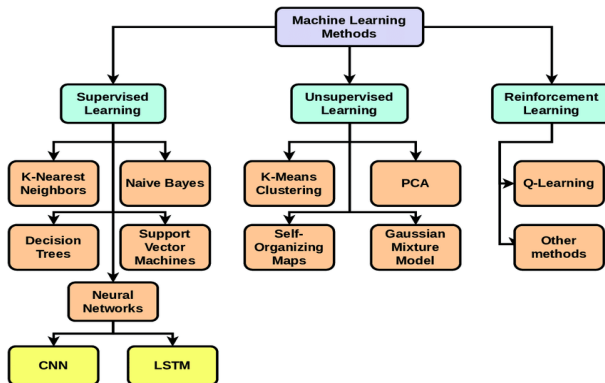


Figure: Categories of ML, [Source](#)

# Supervised Learning

## ■ Whats is Supervised Learning?

**Supervised Learning** is the subcategory of machine learning that focuses on learning from labeled training data, which can be divided to two main categories:

- ▶ **Classification**: Predicting the discrete values such as male/female, etc.
- ▶ **Regression**: Predicting the continuous values such as price, age, etc.

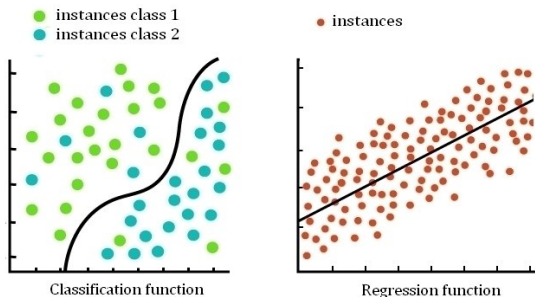


Figure: Classification vs Regression, [Source](#)

# Unsupervised Learning

## ■ What is Unsupervised Learning?

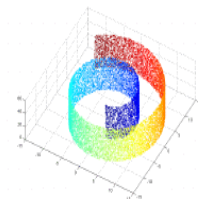
**Unsupervised Learning**, in contrast to supervised learning, is concerned with unlabeled data.

## ■ Common tasks in unsupervised learning are:

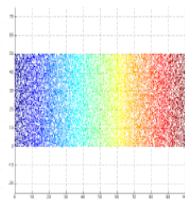
- ▶ Clustering
- ▶ Dimensionality Reduction



(a) Clustering, [Source](#)



(b) Dimensionality Reduction, [Source](#)



**Figure:** Clustering vs Dimensionality Reduction

# Reinforcement Learning

- Reinforcement is the process of learning from rewards while performing a series of actions.
- An agent in this context is a learning system that observes the environment, selects and performs actions, and receives rewards.
- As time goes on, it must learn how to get the most rewards using the best strategy, called a policy.



Figure: Reinforcement Learning, [Source](#).

# ML Categorization Schemes

## ■ Eager vs Lazy:

- ▶ **Eager** learners are algorithms that process training data immediately.
- ▶ **Lazy** learners, however, defer the processing step until the prediction.

## ■ Batch vs Online:

- ▶ **Batch** learning refers to the fact that the model is learned on the entire set of training examples.
- ▶ **Online** learners, in contrast, learn from one training example at the time.

## ■ Generative vs Discriminative:

- ▶ **Generative** models (classically) describe methods that model the joint distribution  $\mathbb{P}(X, Y) = \mathbb{P}(Y)\mathbb{P}(X|Y) = \mathbb{P}(X)\mathbb{P}(Y|X)$  for training pairs  $(x_i, y_i)$ .
- ▶ **Discriminative** Discriminative models are taking a more "direct" approach, modeling  $\mathbb{P}(Y|X)$  directly.



# ML Categorization Schemes

## ■ Instance-based vs Model-Based:

- ▶ **Instance-based** learners learn the training examples by heart and then generalizes to new instances based on some similarity measure.
  - Here, the algorithm looks at a set of training data and tries to find a pattern that can be generalized to new data.
  - This pattern is then used to make predictions on new data.
- ▶ **Model-Based** learners, on the other hand, learn from a model that is created from the training data.
  - This model can be thought of as a mathematical representation of the training data.
  - The model is then used to make predictions on new data.

## ■ Parametric vs Non-parametric:

- ▶ **Parametric** models have "fixed" number of parameters.
- ▶ **Non-parametric** models are more "flexible" and do not have a pre-specified number of parameters.

# How to Solve A Machine Learning Problem

- Collect data.
- Preprocess the data.
- Select a suitable model and train it.
- Evaluate the generalization error on the test dataset.
- Improve the model using various techniques.

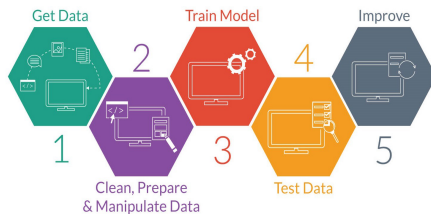


Figure: Required steps to solve an ML problem, [Source](#)

**Thank You!**

**Any Question?**