# Machine Learning

## Introduction

**Dr. Pratyay Kuila**

Dept. of Computer Science & Engineering
NIT Sikkim, Ravangla-737139

# Introduction

- To solve a problem on a computer, we need an *Algorithm*.
- For some tasks, the outcome depends on the *experience* or *knowledge*.
- However, a conventional algorithm cannot be devised to answer or predict whether it will be raining today or not.
- This is due to lack of *knowledge*/*experience*.
- We can easily compile thousands of examples of weather parameters of past days and by observing today's weather we may answer it.
- We may not be able to predict accurately, but we believe we can construct *a good and useful approximation.*

# Machine Learning

- *Machine learning* is an sub-branch of artificial intelligence (AI).

> *Definition* 1: *Machine learning is the field of study that gives the computers ability to learn without being explicitly programmed.*

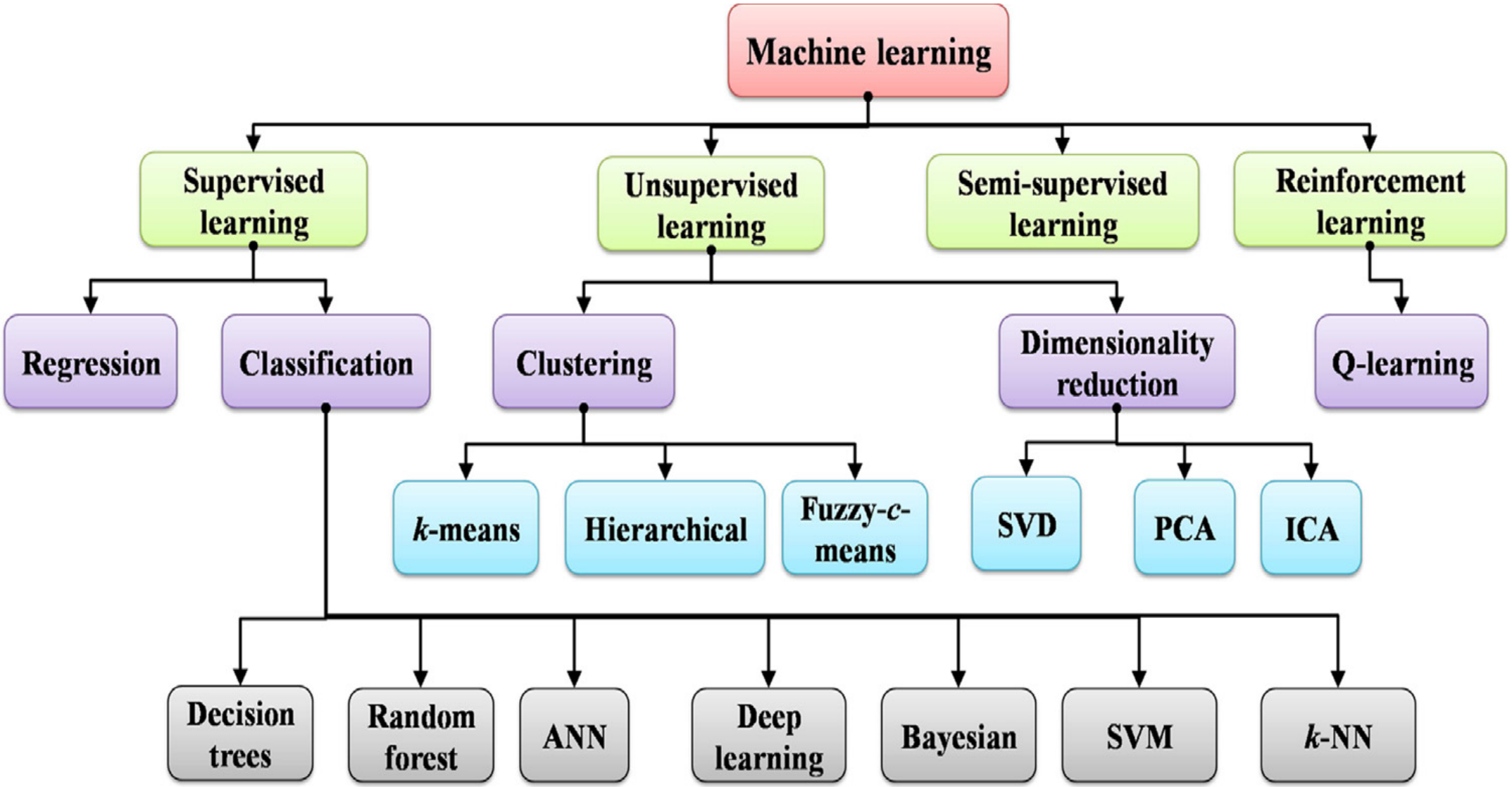
- The key to this definition is that the system should learn by itself without being explicit programming.
- We wish to program computers so that they can learn from input available to them.
- Roughly speaking, learning is the process of converting experience into expertise or knowledge.

# Machine Learning

*Definition 2: A computer program is said to learn from experience E, with respect to task T and some performance measure P, if its performance on T measured by P improves with experience E.*

- The important component of this definition are:
  - Experience, *E*
  - Task, *T* and
  - Performance measure, *P*.

# Categorization of Machine Learning



# 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.

- In *supervised learning*, the computer is provided with example inputs that are labeled with their desired outputs.

# Supervised Learning

- As example, suppose there is a basket which is filled with some fresh fruits, apple, banana, cherry and grape. *The task is to arrange the same type of fruits at one place.*
- If one already knows from their experience the shape of every fruit, it is easy for them to recognize and arrange them.

| Sl. No. | Size | Colour | Shape | Fruit Name |
|---|---|---|---|---|
| 1 | Big | Red | Rounded with depression at top | Apple |
| 2 | Small | Red | Heart shaped and nearly globular | Cherry |
| 3 | Big | Green | Long curving cylinder | Banana |
| 4 | Small | Green | Round | Grape |

- Here, the previous work is called as **training data**. It learns the things from the training data.
- This type of learning is called **Supervised Learning**.

# Un-supervised Learning

- *Unsupervised learning* is where you only have input data ($X$) and no corresponding output variables.
- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.
- Here, data is unlabeled. The learning algorithm is left to find commonalities among its input data.

# Un-supervised Learning

- Let us take the previous example. This time there is no information about the fruits beforehand. It is the first time that the fruits are being seen or discovered.

- First, any physical characteristic is selected. Suppose *color*.
  RED COLOR GROUP: apples & cherry fruits.
  GREEN COLOR GROUP: bananas & grapes.

- Now, take another character say, *size*.
  RED COLOR AND BIG SIZE: apple.
  RED COLOR AND SMALL SIZE: cherry fruits.
  GREEN COLOR AND BIG SIZE: bananas.
  GREEN COLOR AND SMALL SIZE: grapes.

- This type of learning is known as **Unsupervised Learning**.

# Semi-supervised Learning

- Problems where you have a large amount of input data (*X*) and only some of the data is labeled (*Y*) are called *semi-supervised* learning problems.

- These problems sit in between both *supervised* and *unsupervised* learning.

- A good example is a photo archive where only some of the images are labeled, (e.g. dog, cat, person) and the majority are unlabeled.

# Reinforcement Learning

- Learning through rewards.

- It is about learning the optimal behavior in an environment to obtain maximum reward.

- This optimal behavior is learned through interactions with the environment and observations of how it responds.

- In the absence of a supervisor, the learner must independently discover the sequence of actions that maximize the reward.

# Issues in Machine Learning

- What algorithms exist for learning general target functions from specific training examples?
- In what settings will particular algorithms converge to the desired function, given sufficient training data?
- Which algorithms perform best for which types of problems and representations?
- How much training data is sufficient? What general bounds can be found to relate the confidence in learned hypotheses to the amount of training experience and the character of the learner's hypothesis space?

# Issues in Machine Learning

- What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy alter the complexity of the learning problem?

- What is the best way to reduce the learning task to one or more function approximation problems? Put another way, what specific functions should the system attempt to learn? Can this process itself be automated?

- How can the learner *automatically alter its representation to improve its ability to represent and learn the target function*?

# Machine Learning and Data Mining

- Application of machine learning methods to large databases is called *data mining*.
  - The analogy is that a large volume of earth and raw material is extracted from a mine, which when processed leads to a small amount of very precious material. Similarly, *in data mining, a large volume of data is processed to construct a simple model with valuable use*, for example, having high predictive accuracy.
  - But *machine learning is not just a database problem*. To be intelligent, *a system that is in a changing environment should have the ability to learn*.

**Books:**

1. Introduction to Machine Learning by Ethem Alpaydin, MIT Press, PHI, 3rd Edition 2014.
2. Machine Learning by Tom Mitchell, McGraw Hill.
3. Applied Machine Learning by M. Gopal, McGraw Hill.