Minería de Datos

U1 Introducción

Héctor Maravillo

Section 1

Learning and intelligence

Control Systems

- A **control system** is an configuration of components **connected** or related in a way that allows it to **manage**, **guide**, or **regulate** itself or another system.
- A control system receives an **input**, which consist of the **environment's influence** of the system, to produce a specified **response**

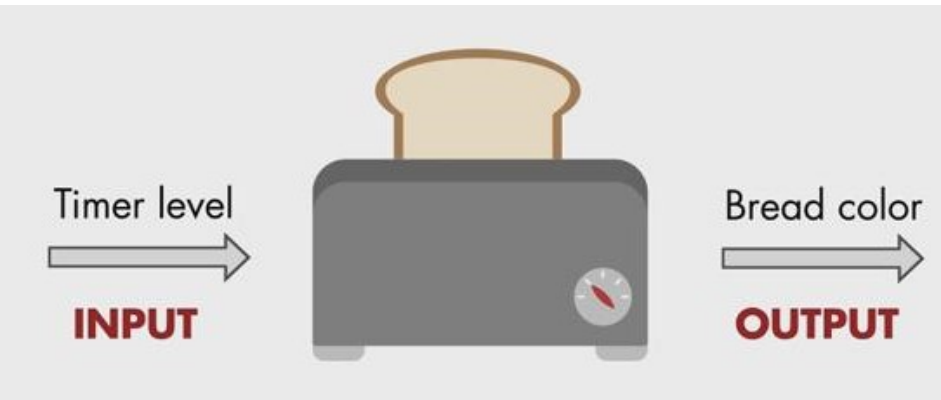


Figure: Example of a control system

Self-regulating control system

- In a **feedback control system**, the output is continuously measured and monitored. This information is used to **dynamically adjust** future actions, allowing the system to **adapt to changes** and **disturbances** and ensuring the outcome aligns with the **desired objectives**.

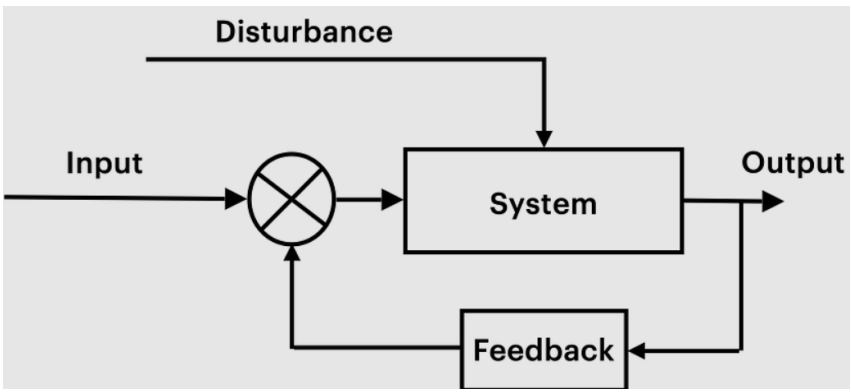


Figure: Feedback control system outline

Learning

What is learning?

Learning denotes **changes** *in the system that are* **adaptive** *in the sense that they enable the system to do the same task or tasks drawn from the same population more* **efficiently** *and more* **effectively** *the next time.*

- Herbert A. Simon

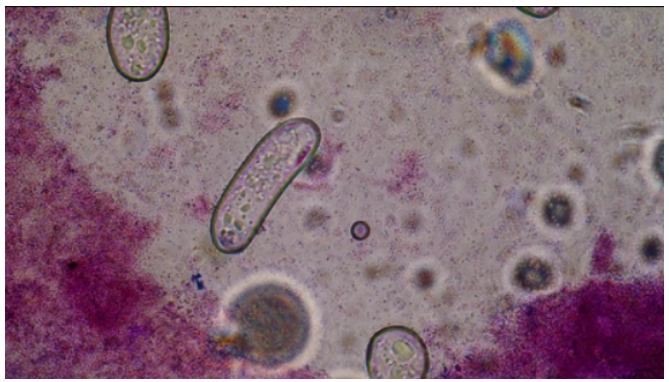
Learning

Learning can be defined as the process by which a **system** (the learner), through **practice** (i.e. by imitation or trial and error), to **perform** a given **task**.

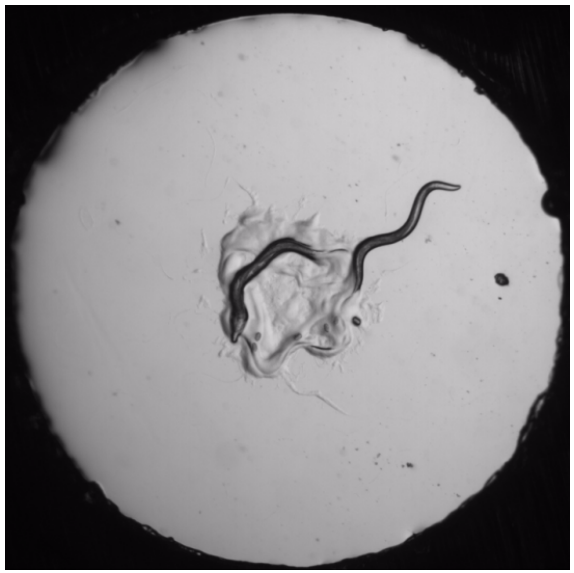
The learner has learned to perform the task, if he or she can **repeat** the task equally well **without relearning**.

Figure: A classic example of learning

Learning vs. adaptation



(a) Bacteria react quickly and accurately to changes in the environment

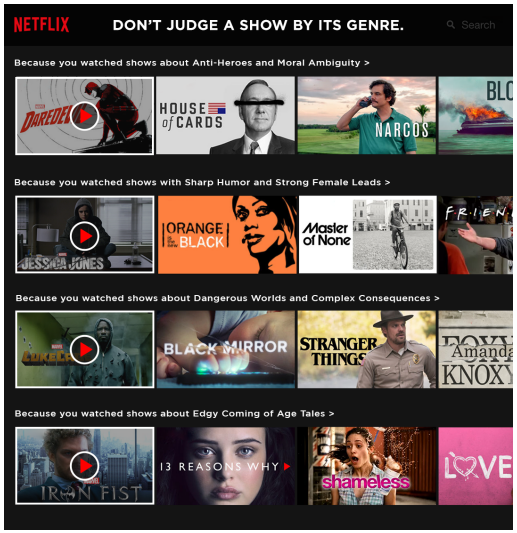


(b) Earthworms, with only 302 neurons, have the ability to make goal-directed decisions

Types of learning

- Learning by a living system is called **natural learning**.
- If the learning is a machine - a computer - it is called **machine learning**.

(a) Natural learning



(b) Machine learning

Figure: Types of learning

Natural learning process

- **Imprinting**: A certain knowledge is imprinted in the learner's **memory** and after that the knowledge **does not change** any more.
- **Conditioning** and **associating**: With conditioning a conditional **reflex emerges** (besides innate reflexes which are unconditional). It appears as an association between various mechanical or mental **sensations** or **stimuli**.

(a) Imprinting

(b) Conditioning

- **Memorizing** :Simplistic memorizing without any understanding of the meaning does not require additional mental activities. However, this kind of learning is hard just because during the learning **no associations** with known concepts are created.
- **Learning by trial and error**: The learner starts from an initial state and in each turn **chooses** from a set of alternatives (actions, directions, moves, decisions) and changes the current state to achieve the goal state. By **practicing**, the learner acquires the ability to select more promising alternatives. Memory has a crucial role in this process.

Figure: Trial and error

- **Imitation**: The learner first **observes** someone who is solving the same or a similar task, and then imitates him/her in order to solve the given task.
- **Learning by understanding and insight**: It requires **memorizing**, **abstract (symbolic)** thinking, **logical reasoning**, and causal integration which leads to problem understanding. The insight comes when the learner discovers the solution by **integrating** the relationships in each problem situation.

(a) Imitation

(b) Learning by understanding

Knowledge

- The result of learning is **knowledge** which the system can use for solving new problems.
- **Knowledge** is defined as an **interpretation** of the **information** contained in **data**.

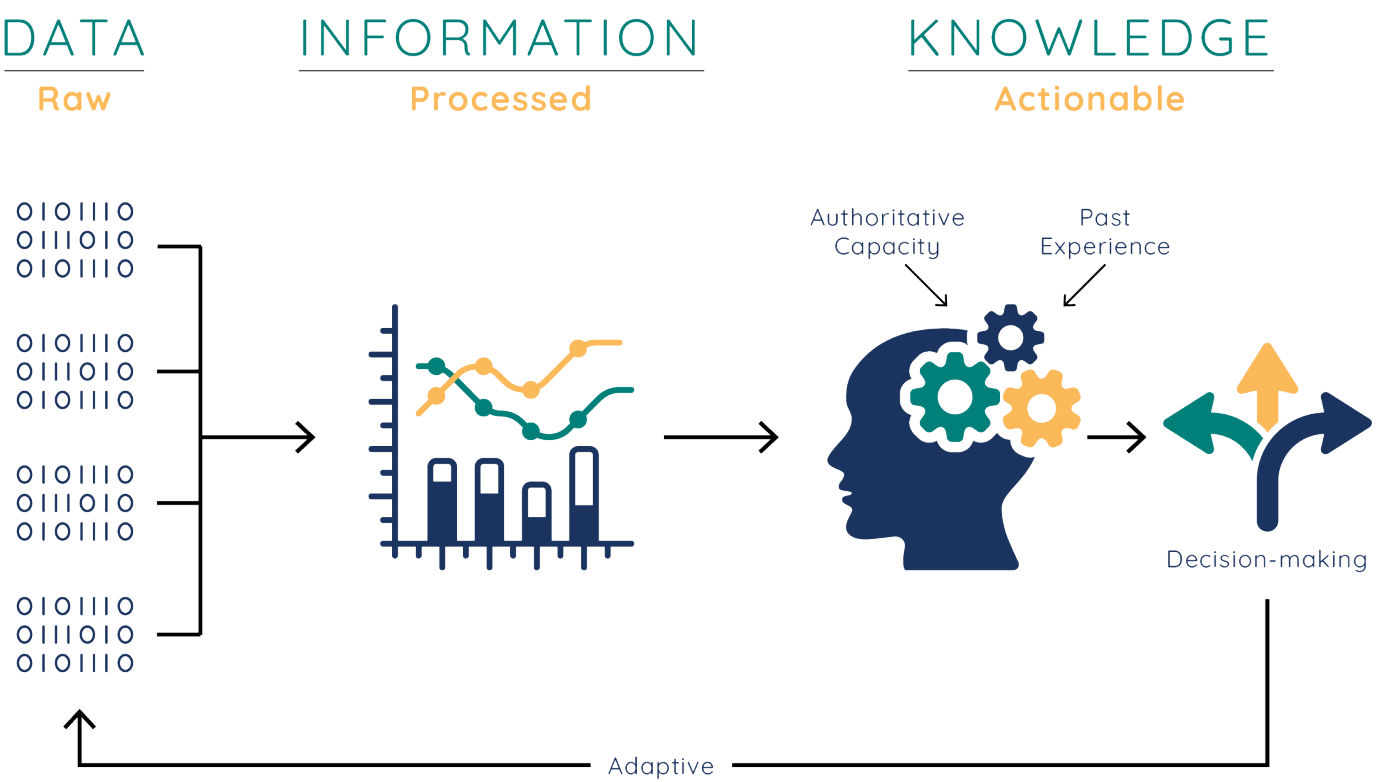


Figure: The flow from data to knowledge

Knowledge

- Knowledge can be either **given in advance** (for example inherited or pre-coded), or is the **result of learning**.

Figure: Example of knowledge through learning and observation

Knowledge

- Knowledge can be **correct or wrong**, **correct but useless**, incomplete, etc. However, in practice only useful knowledge is interesting.

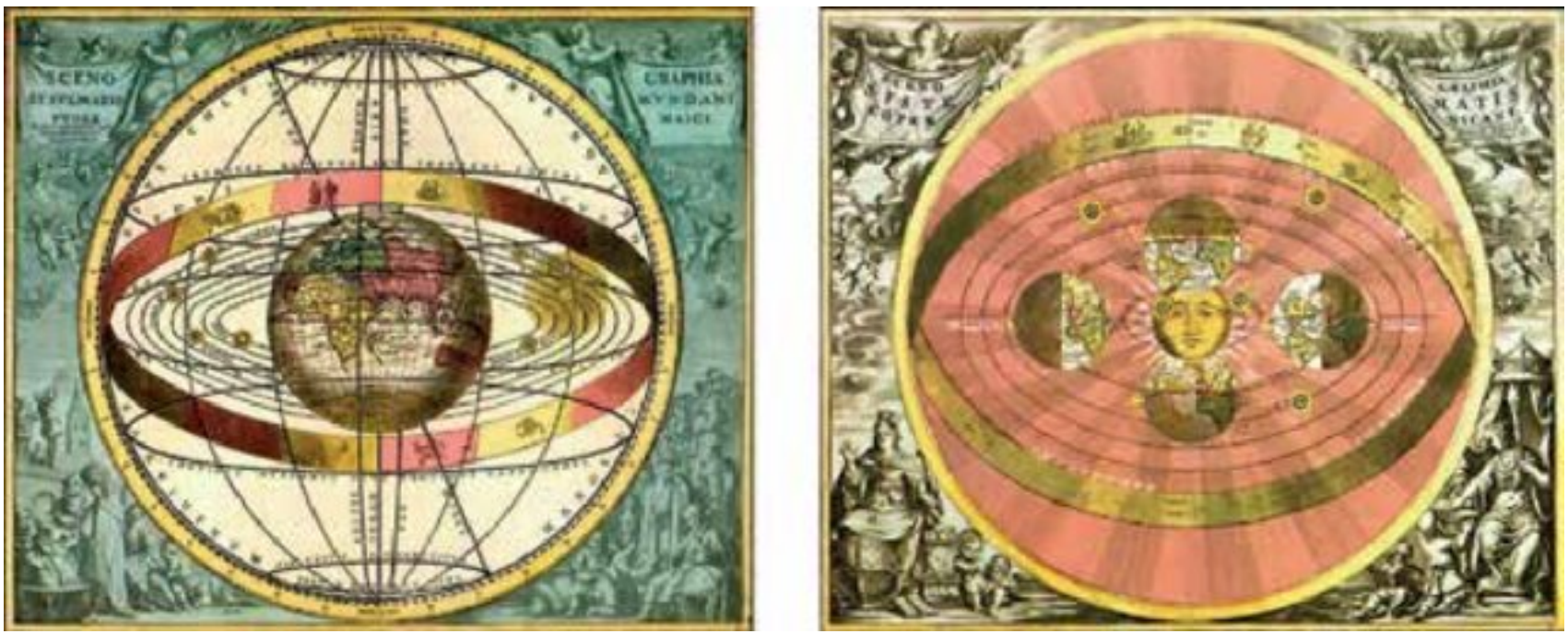


Figure: Models of the solar system: Geocentric by Ptolemy (left) and Heliocentric by Copernicus (right)

Inductive bias

- The stronger prior knowledge (or **prior assumptions**) facilitate **easier learning** from new examples.
- However, the stronger these prior assumptions **reduce the flexibility** of learning, as they constraint the process based on those initial assumptions.

Intelligence

What is intelligence?

Intelligence is the ability to **adapt** *to the environment and to* **solve** *problems.*

[Knonenkon & Kukar, 2007]

Intelligence

There is no intelligence without learning, but learning alone is not enough.

In order to learn, a system must have some capabilities:

- Sufficient **memory** capacity
- Ability to **reason** (processor)
- Ability to **perceive** (input and output)

These abilities do not suffice if they are not appropriately **integrated** or if they lack appropriate learning algorithms.

Besides, for efficient learning one needs also some **initial knowledge**.

Artificial intelligence

Artificial intelligence (AI) is technology that enables computers and machines to **simulate** human intelligence and problem-solving capabilities [IBM].

- **Weak AI** is AI trained and focused to perform **specific task**. Weak AI drives most of the AI that surrounds us today.
- **Strong AI** is a theoretical form of AI where a machine would have an intelligence equal to or **surpass** that humans.

Artificial intelligence

Some areas of artificial intelligence:

- Machine learning
- Robotics
- Natural language processing
- Expert systems
- Automatic reasoning and theorem proving
- Computer vision and speech recognition

Learning task

Definition (Mitchell, 1997)

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.

Example: A handwriting recognition learning problem:

- Task: Recognizing and classifying handwritten words within images.
- Performance measure: Percent of words correctly classified.
- Training experience: A database of handwritten words with given classification.

Learning task

Exercise. Identify the three elements of the following learning tasks:

- Email spam detection.
- House price prediction.
- Speech-to-Text conversion.
- Medical diagnosis.
- Autonomous driving.
- Facial recognition in surveillance cameras.
- Customer segmentation for marketing campaigns.

Machine Learning

Machine learning (ML) is a branch of **artificial intelligence** and **computer science** that focuses on the using **data and algorithms** to enable AI to imitate the way that humans **learn**, gradually improving its accuracy [IBM].

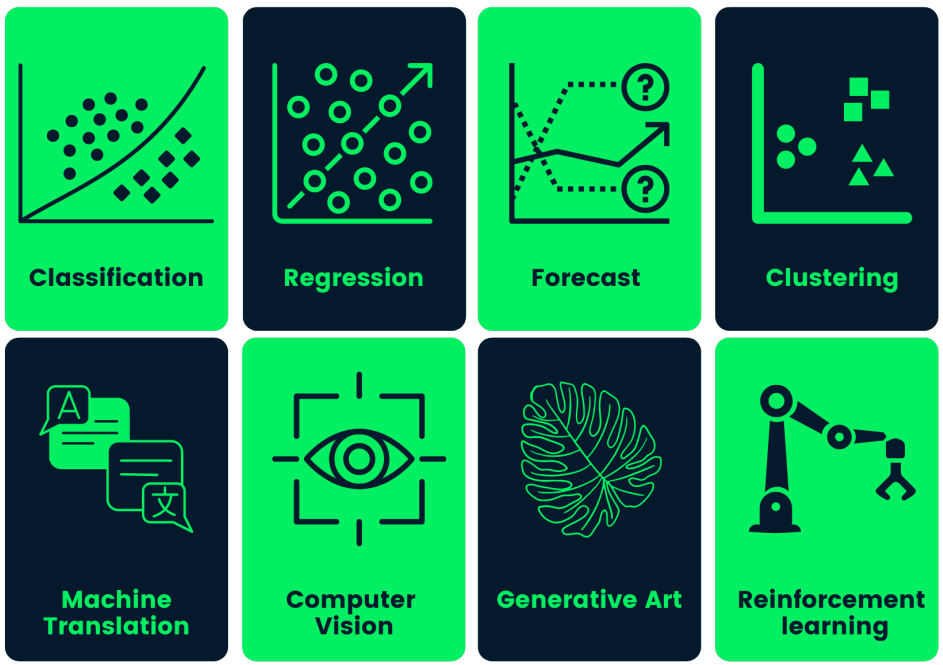


Figure: Machine learning tasks

Machine Learning components

An machine learning algorithm consists of three parts:

- A **decision process**: Based on some input data the algorithm will produce an **estimate** about a **pattern** in the data.
- An **error function**: An error function **evaluates** the model's prediction. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
- A **model optimization process**: Adjustments are made to reduce the discrepancy between known examples and the model estimate.

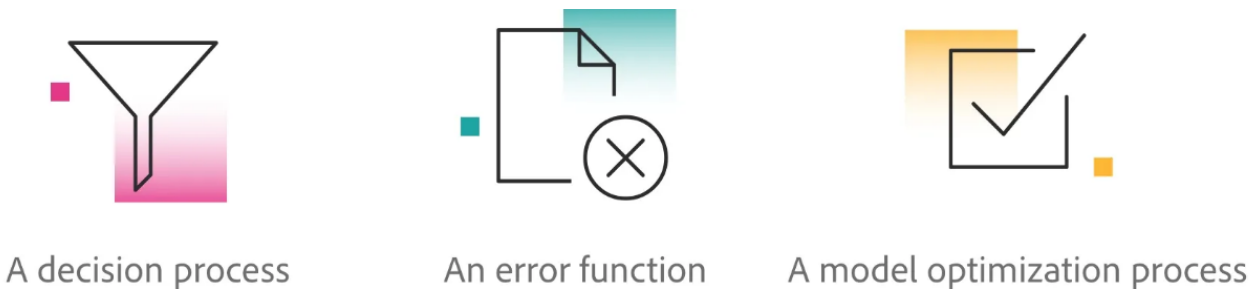


Figure: Machine learning components

Designing a learning system

To design a learning system, it is necessary to make the following decisions:

- Choosing the **training experience**.
- Determine **target function**.
- Determine **representation** of learned function.
- Determine **learning algorithm**.
- Design the performance or **evaluation** system.

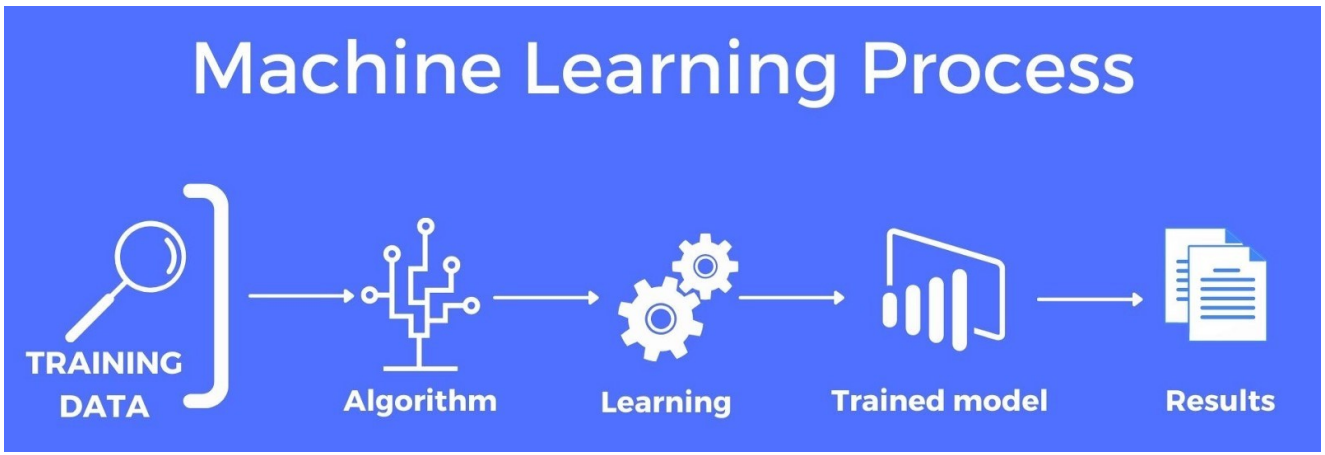


Figure: The process of machine learning

Properties of machine learning

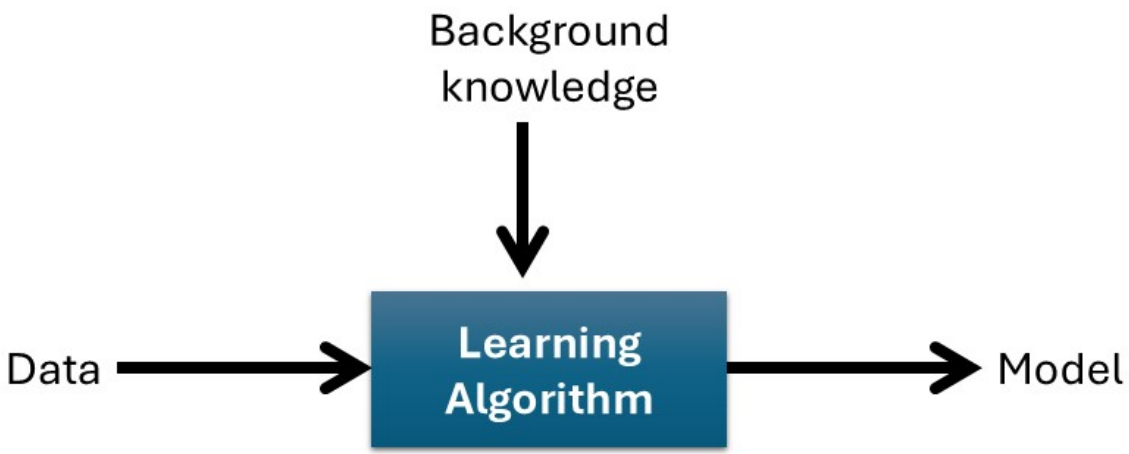
One of the most relevant features of machine learning algorithms is their **capability of learning** from the presentation of **samples (patterns)**, which expresses the system behavior.

Hence, after the algorithm has learned the **relationship** between inputs and outputs, it can **generalize solutions**, meaning that the algorithm can produce an output which is close to the expected (or desired) output of any given input values.

Machine Learning algorithm

In a machine learning system we will often distinguish between:

- **Learning algorithms** generates new knowledge (or modifies existing knowledge) from the set of learning data and background knowledge (if any).
- **Execution algorithms** uses the generated knowledge for solving new problems. The generated knowledge is often called the **model**.



Machine learning as optimization problem

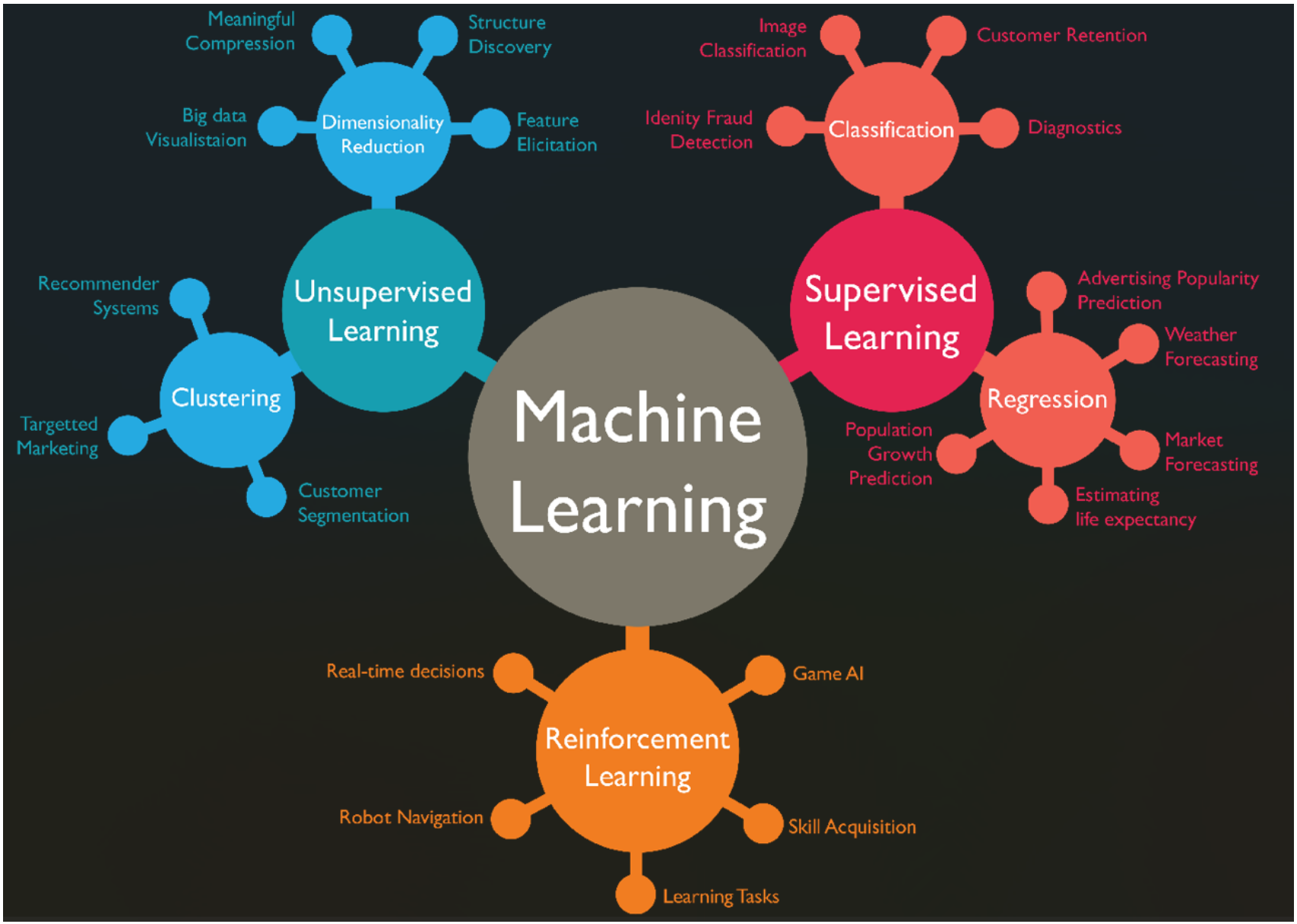
Machine learning can therefore be treated as an **optimization problem**.

Given a solution (model) space, and an **optimality criterion**, the model satisfying this criterion is sought. The criterion value depends upon the current model, background knowledge, and learning (modeled) data.

Relation to other fields

- As a subfield of **computer science** and **AI**, with goal is to program machine so that they will learn.
- In contrast with traditional AI, machine learning is not trying to build **automated imitation** of intelligent behavior, but rather to use the strengths and special abilities of computers to **complement** human intelligence.
- In contrast with traditional statistics, in ML **algorithmic** considerations play a major role.

Machine Learning methods



References

- DiStefano, J., Stubberud, A. & Williams, I., *Theory and Problems of Feedback and Control Systems*, McGraw-Hill, 1995.
- Kononenko, I. & Kukar, M. *Machine Learning and Data Mining: Introduction to Principles and Algorithms*, Horwood Publishing, 2007.
- Han, J. & Kamber, M. *Data Mining: Concepts and Techniques*, 2nd ed, Morgan Kaufmann Publishers, 2006.
- IBM. *What is artificial intelligence (AI)?*, https://www.ibm.com/topics/artificial-intelligence.
- Mitchell, T., *Machine Learning*, McGraw-Hill Science, 1997.
- Russell, S. & Norving, P., *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall, 2010.
- Shalev-Shwartz, S. & Ben-David, S., *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.