

# **Introduzione al machine learning**

**Parte prima: algoritmi supervisionati, tecniche di validazione, metriche di valutazione**

**Advanced School on Artificial Intelligence, Bertinoro, 18 luglio 2023**

**Ciao!**  
**Il mio nome è**  
**Stefano**

<https://lozingaro.github.io>

# Quali argomenti tratteremo?

In circa 4 slot da 45 minuti ciascuno (con 10-15 minuti di pausa)

1. Diversi tipi di apprendimento (*learning*)
2. Tecniche di apprendimento (supervisionate, non-supervisionate, semi-supervisionate, di rinforzo)
3. Principali algoritmi di apprendimento supervisionato
4. Tecniche per la validazione dei modelli di apprendimento
5. Metriche per la valutazione dei modelli di apprendimento

## Consider a simple one player learning game.

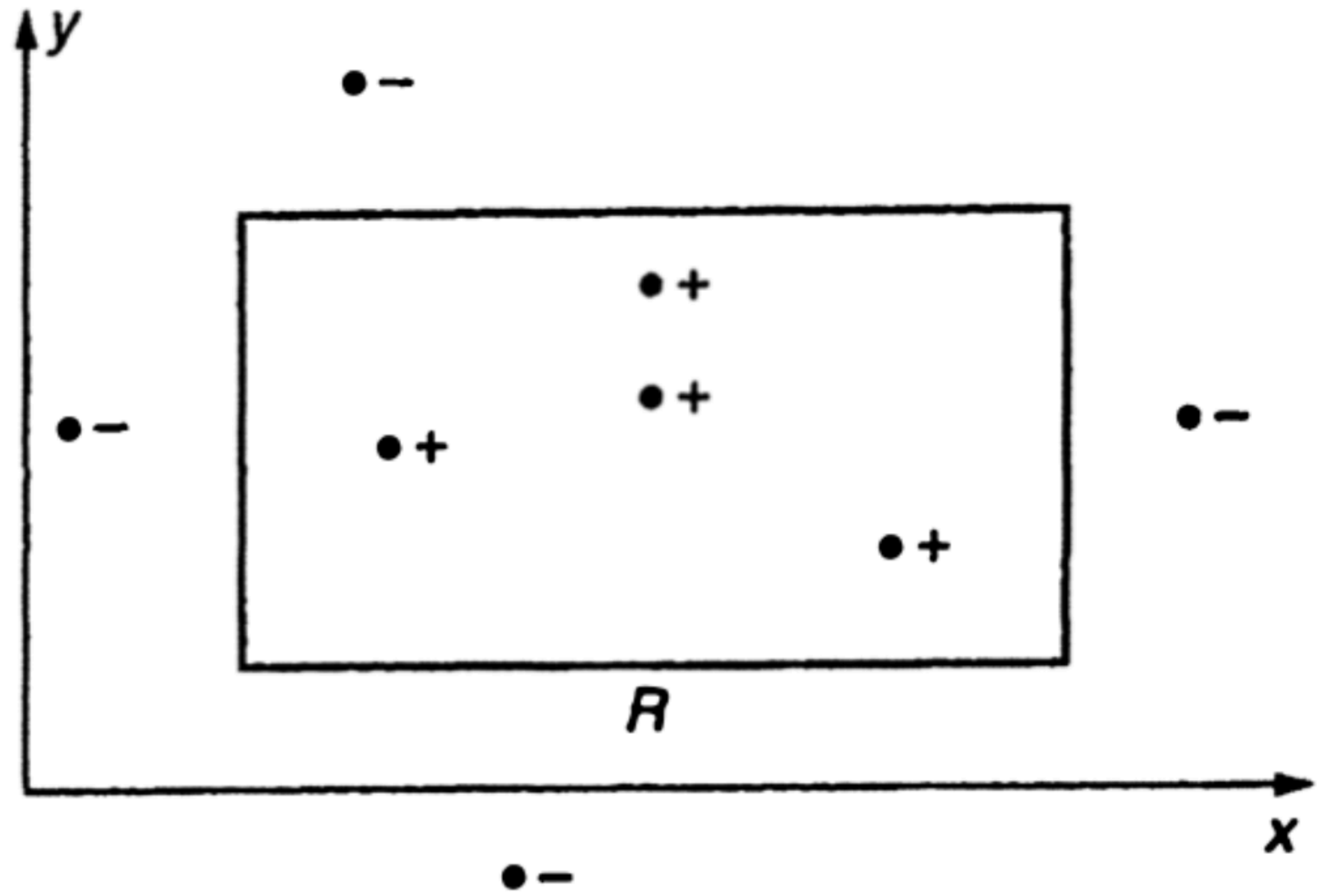
The object of the game is to learn an unknown axis-aligned rectangle  $R$ , that we will call target rectangle.

The player receives information about  $R$  only through the following process:

1. Every so often, a random point  $p$  is chosen in the plane according to some fixed probability distribution  $D$ .
2. The player is given the point  $p$  together with a label indicating whether  $p$  is contained in  $R$  (a positive example) or not contained in  $R$  (a negative example).

The goal of the player is to use as few examples as possible, and as little computation as possible, to pick a hypothesis rectangle  $R'$  which is a close approximation to  $R$ .

We measure the error of  $R'$  as the probability that a randomly chosen point from  $D$  falls in the region  $R \Delta R' = (R - R') \cup (R' - R)$ .



**Apprendimento?**

**<https://youtu.be/PBb1CH18Smg>**

**The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories.**

(Bishop, Springer, 2006)

# Apprendimento associativo supervisionato

## Regressione e classificazione

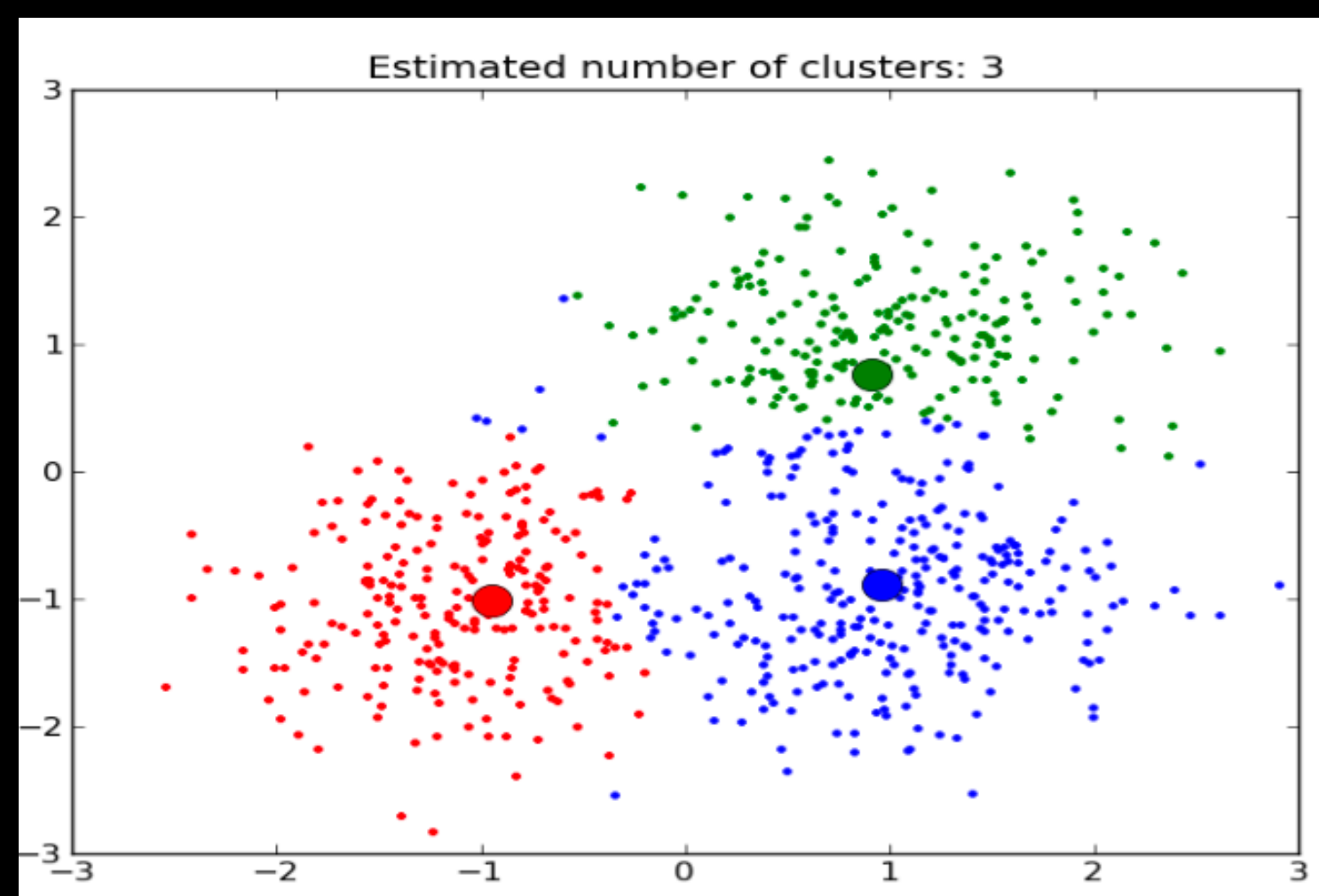
- Applications in which training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised learning problems.
- Cases such as the digit recognition example, in which the aim is to assign each input vector to one of a finite number of discrete categories, are called classification problems.
- If the desired output consists of one or more continuous variables, then the task is called regression.



# Apprendimento associativo non-supervisionato

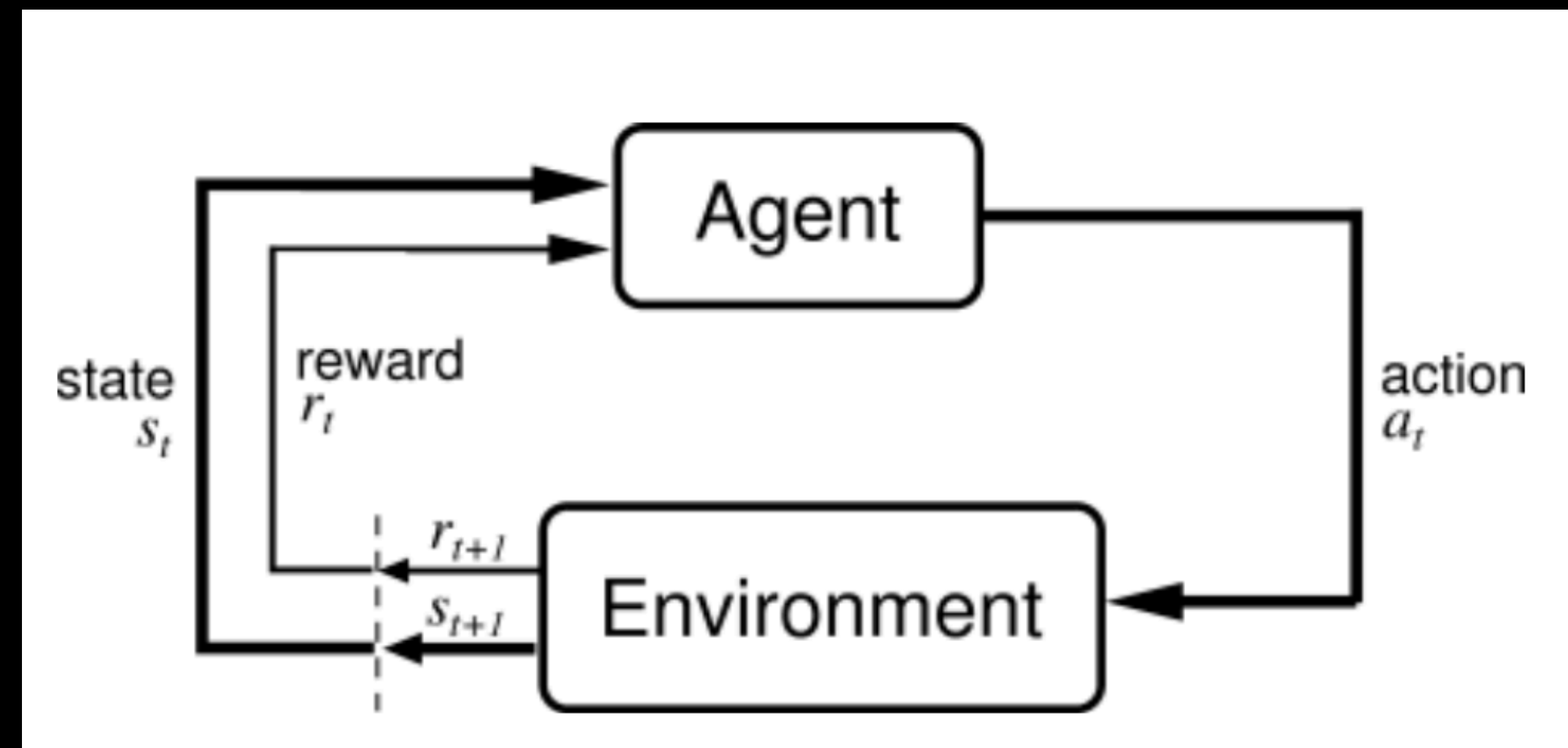
## Clusterizzazione

- In some pattern recognition problem the training data consists of a set of input vectors  $x$  without any corresponding target values.
- The goal in such unsupervised learning problems may be to discover groups of similar examples within the data, where it is called clustering, or to determine the distribution of data within the input space, known as density estimation.



# Apprendimento associativo di rinforzo

- The agent learns a strategy—a series of moves—that maximizes the reward. A general feature of reinforcement learning is the trade-off between exploration, in which the system tries out new kinds of actions to see how effective they are, and exploitation, in which the system makes use of actions that known to yield a high reward.



**> l'intelligenza < l'errore**

# Riferimenti utili

## Libri, pagine web, articoli

- [Pattern recognition and machine learning](#), Bishop, 2006: manuale di ML
- <https://github.com/lozingaro/asai-er-ml>: materiale di laboratorio
- <https://anatomyof.ai>: Kate Crawford's anatomy of AI
- <https://2023.ecmlpkdd.org/>: next European conference on ML
- <https://atcold.github.io/>: Alfredo's website and didactics (rif. energy-based)
- [https://www.youtube.com/watch?v=xnf8i\\_IRCcw](https://www.youtube.com/watch?v=xnf8i_IRCcw): The office, Pavlov exp
-