



Per imparare bisogna sbagliare

Alla scoperta del Reinforcement Learning



UNIVERSITÀ
DI TRENTO

Dipartimento di Matematica

Laboratorio di comunicazione storia e filosofia della matematica

MuSe

Cosa hanno in comune? In che modo *imparano*?



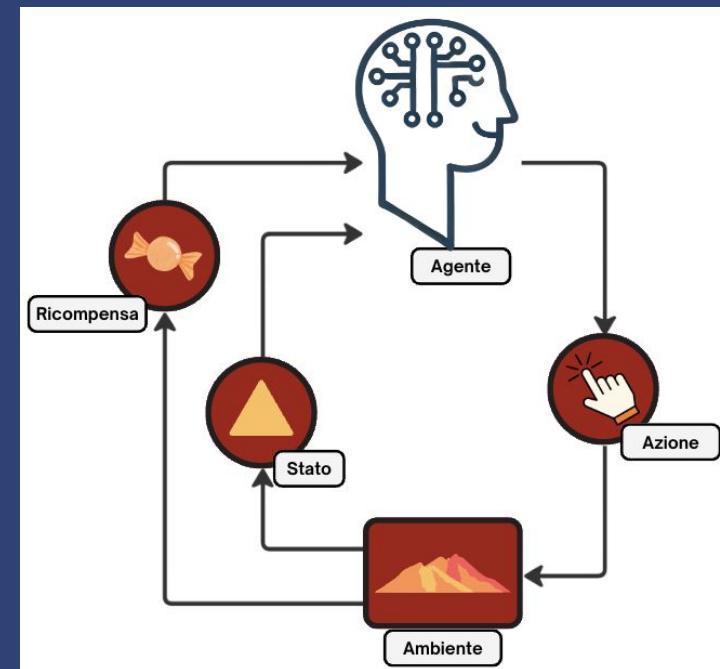
Cosa hanno in comune? In che modo *imparano*?



- interagiscono con l'ambiente circostante, modificandolo
- imparano dalle proprie esperienze e dagli errori
- cercano di capire le azioni più efficienti
- ...

Questo si chiama apprendimento per rinforzo o reinforcement learning

Si ha un **agente** che opera e interagisce con un **ambiente** attraverso una scelta di **azioni** che permettono di modificare il proprio **stato** e ottenere delle **ricompense**, cioè dei feedback positivi o negativi.

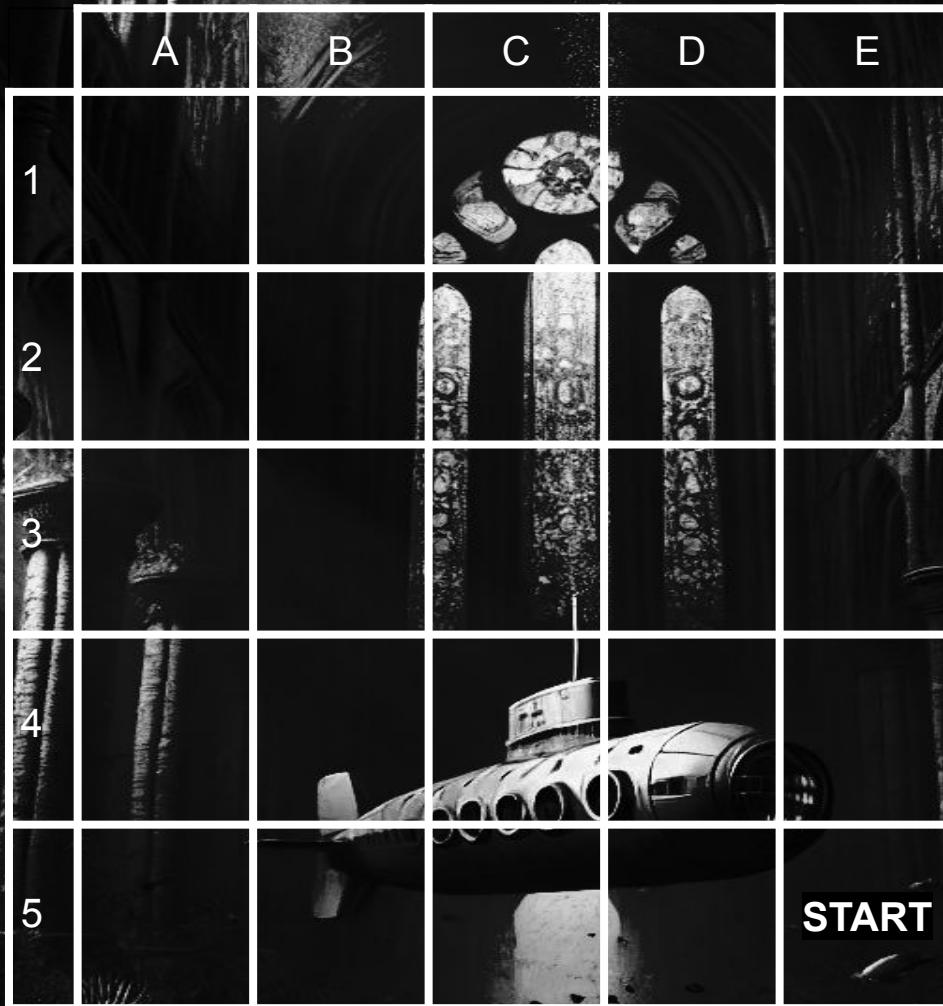


È il 3024 e la città di Genova è sepolta dalle acque.

Nei fondali marini si trova un tesoro,
la Margherita di Bramante di Giovanni Pisano
nascosto in una vecchia chiesa gotica,
la più grande della città.
Aiutateci a ritrovarla!



Ambiente



Regole

1

Dividetevi in squadre (massimo 6 persone)

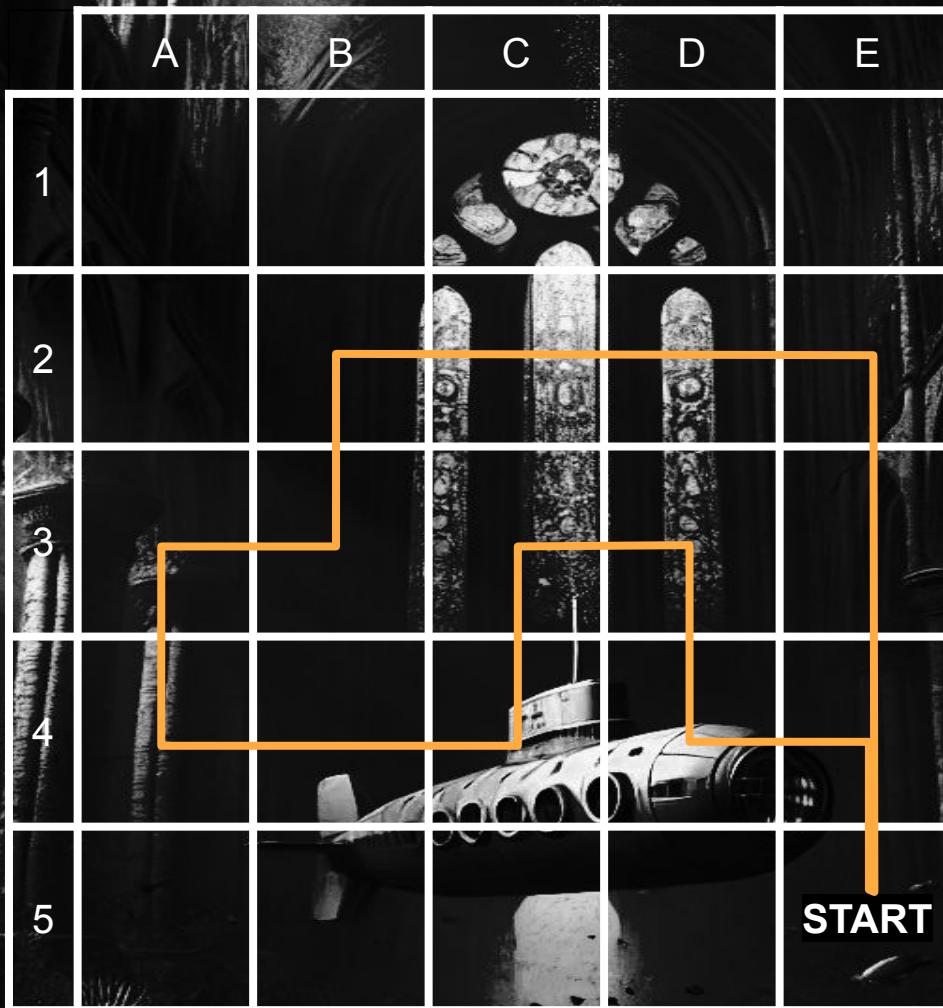
2

Il gioco è composto da più episodi:
in ogni episodio, un giocatore per squadra
si muoverà sulla griglia

3

Ci si può muovere a destra, sinistra, avanti e
indietro, senza mai uscire dalla griglia

Ambiente



Regole

4

Ogni episodio, si inizia con **20 punti energia**

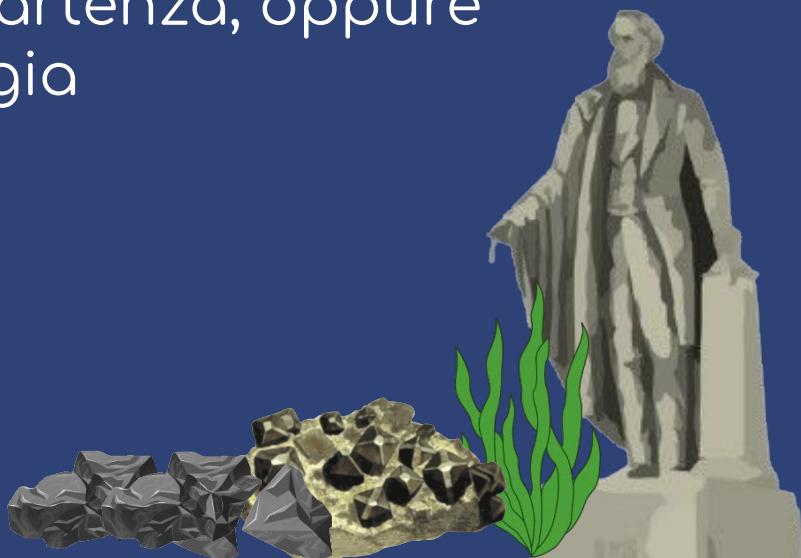
5

Ad ogni passo del vostro cammino
perderete punti energia: il numero di punti
persi dipenderà dagli ostacoli che incontrerete

6

L'episodio termina quando:

1. tornate al punto di partenza, oppure
2. esaurite i punti energia



Ostacoli



Acqua
- 1 pt. energia

Alghe
- 2 pt.

Ostacoli



Rovine
- 2 pt.

e inoltre non si possono attraversare
cioè si torna alla casella precedente

Il tesoro



Angelo DX



Regina



Angelo SX

-1 pt. energia

Feedback

Alla fine dell'episodio,
il punteggio ottenuto è calcolato con la formula:

pt. energia rimasti

+

bonus statue trovate

se non siete riusciti a tornare al punto di partenza
c'è un ulteriore malus di -20 pt.

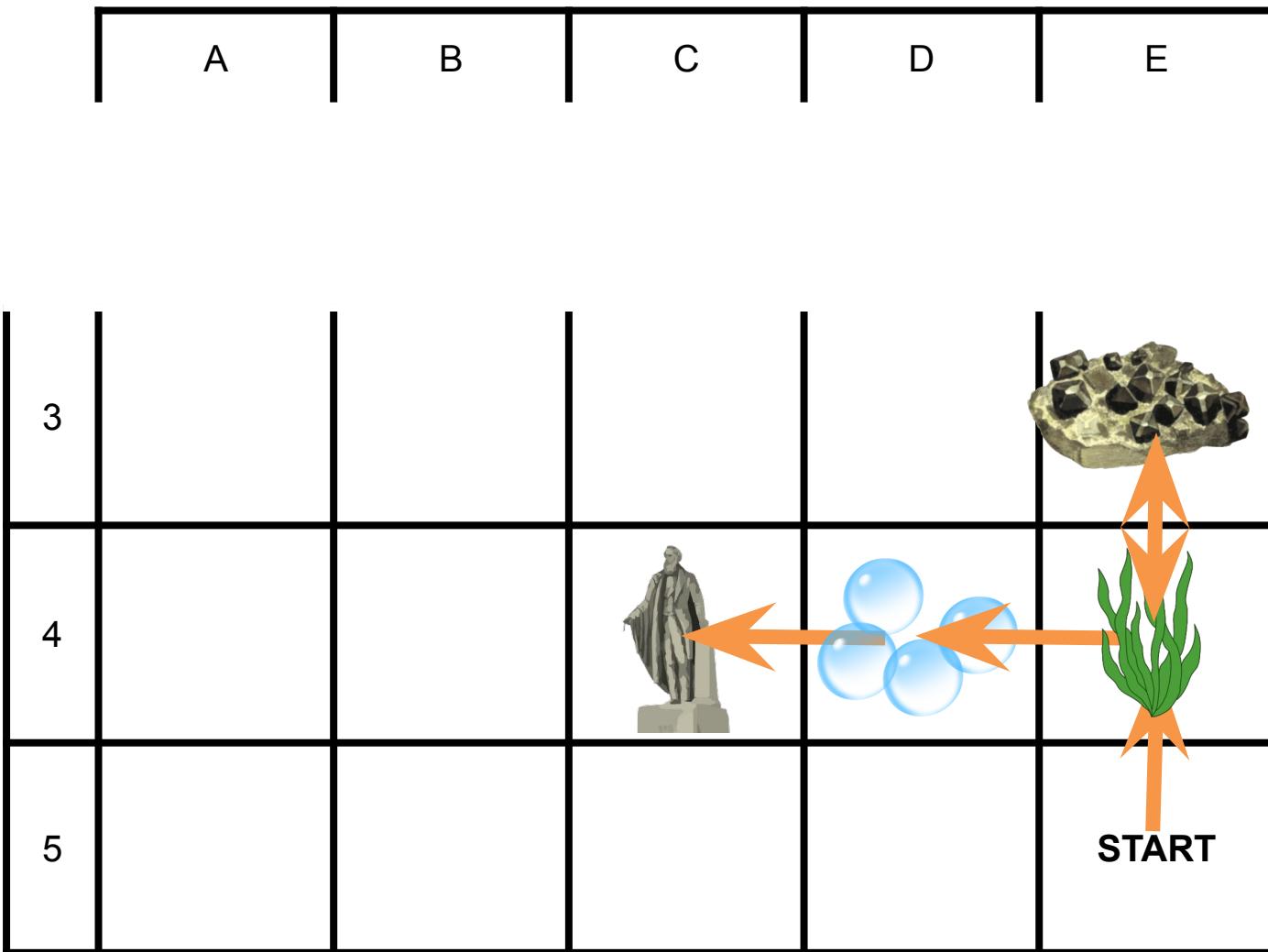
Il tesoro



Angelo DX
10 pt. bonus

Regina
20 pt. bonus

Angelo SX
15 pt. bonus



percorso: E1 - E4 - E3 - D4 - C4 - ...
energia: 20 - 18 - 16 - 15 - 14 - ...



Buona ricerca!

Come è andata?

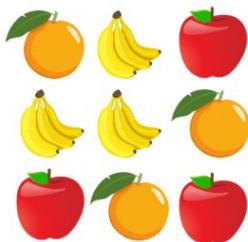
- Qual è stato il percorso migliore che avete trovato?
- Dopo quanti episodi siete riusciti ad arrivare al risultato migliore?
- Come avete ragionato per arrivare all'obiettivo?

	A	B	C	D	E
1	rovine -2	<i>angelo dx</i> -1	acqua -1	acqua -1	rovine -2
2	rovine -2	acqua -1	rovine -2	<i>angelo sx</i> -1	rovine -2
3	algne -2	acqua -1	acqua -1	algne -2	acqua -1
4	acqua -1	rovine -2	acqua -1	acqua -1	rovine -2
5	<i>regina</i> -1	algne -2	algne -2	acqua -1	START

Come imparano le macchine?

Apprendimento supervisionato

Input data set



**Modello di
supervised learning**



Etichette



Arancia



Banana

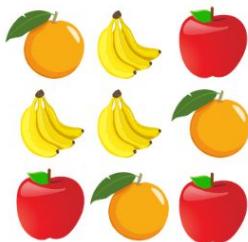


Mela

Come imparano le macchine?

Apprendimento supervisionato

Input data set



Modello di
supervised learning



Predizione



è una mela

Etichette



Arancia



Banana



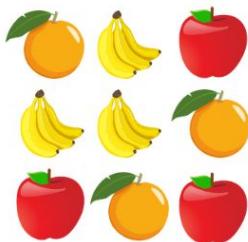
Mela



Come imparano le macchine?

Apprendimento supervisionato

Input data set



Modello di
supervised learning



Predizione



è una mela

Etichette



Arancia



Banana

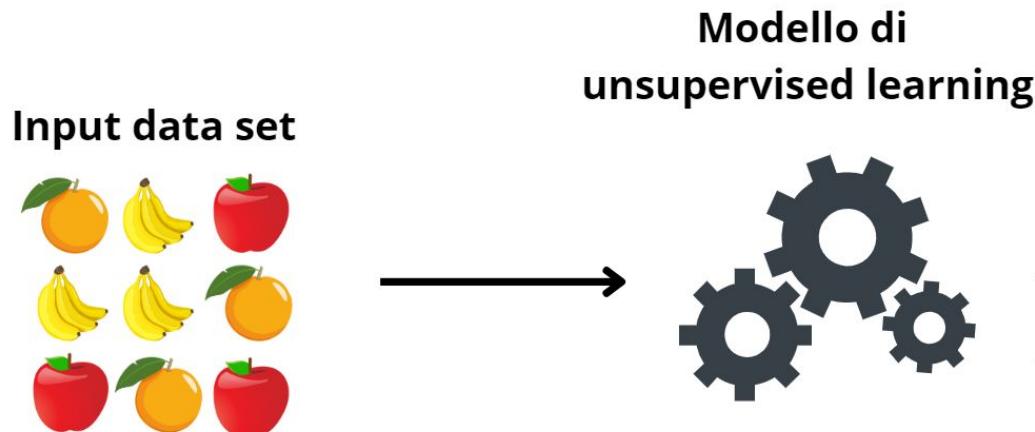


Mela



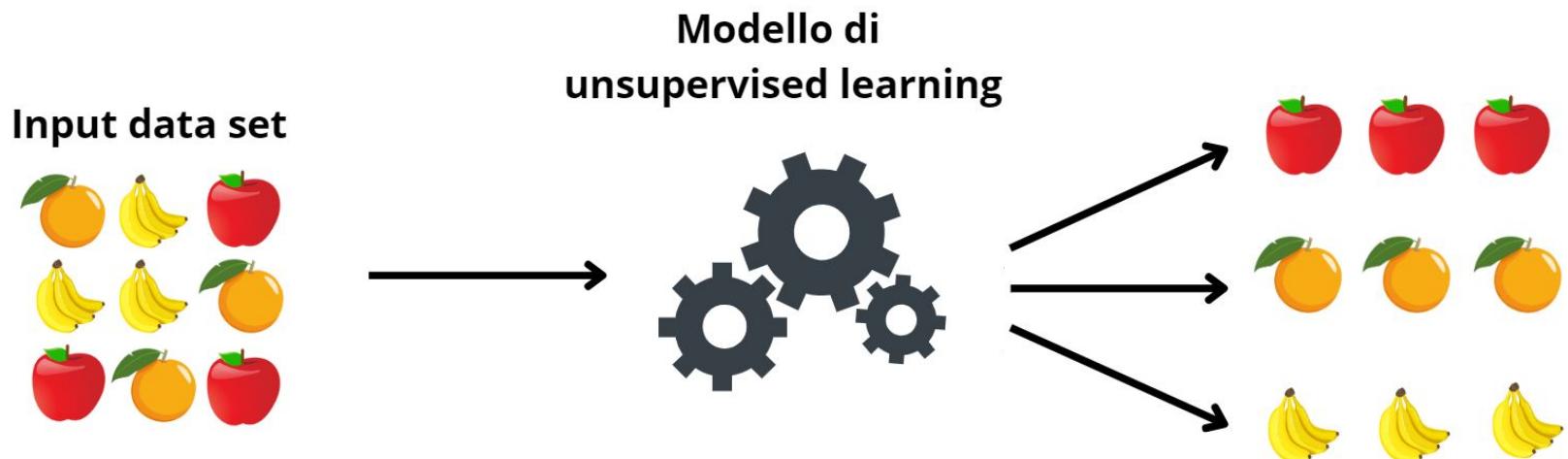
Come imparano le macchine?

Apprendimento non supervisionato



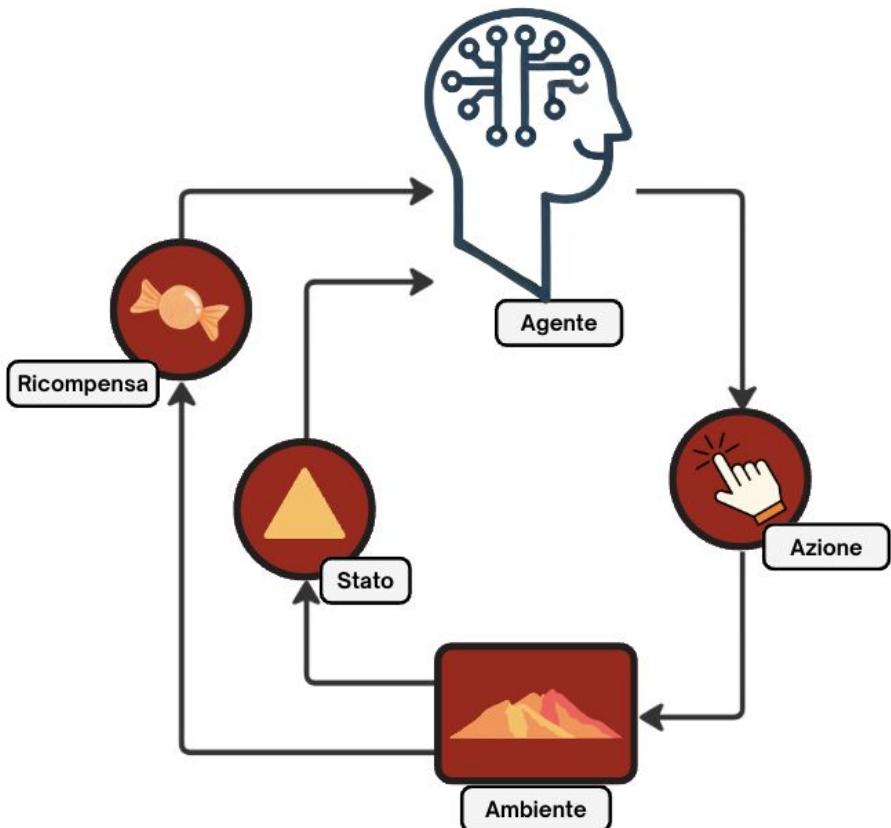
Come imparano le macchine?

Apprendimento non supervisionato



Come imparano le macchine?

Apprendimento per rinforzo

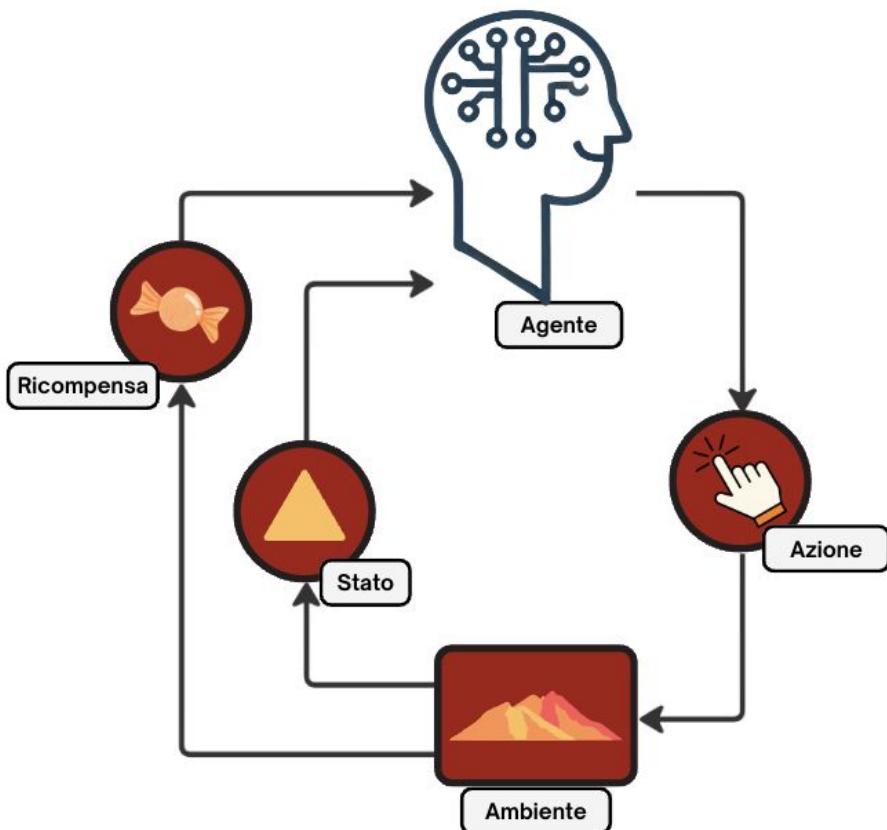


Branca dell'intelligenza artificiale e del machine learning che si ispira al modo in cui gli esseri umani e gli animali imparano.

In particolare, a differenza degli altri metodi, impara dall'esperienza e dagli errori.

Come imparano le macchine?

Apprendimento per rinforzo



Le parole chiave:

agente

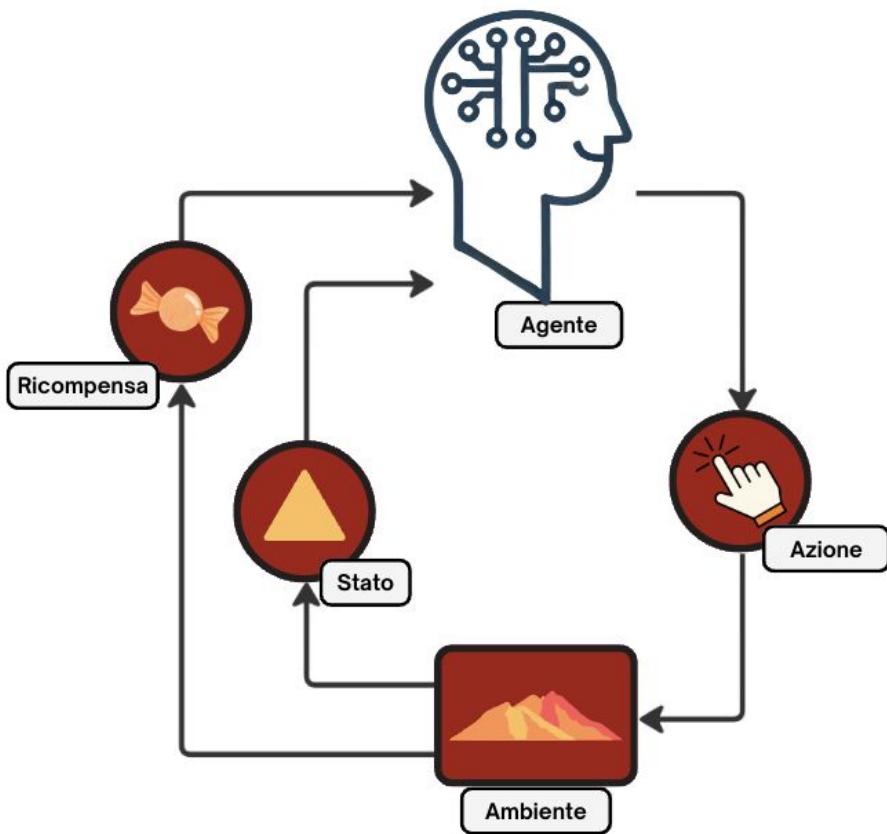
Un'entità (come un robot o un programma) che prende decisioni

ambiente

Il mondo in cui l'agente opera e con il quale interagisce

Come imparano le macchine?

Apprendimento per rinforzo



Le parole chiave:

azioni

le scelte che l'agente può fare.

stati

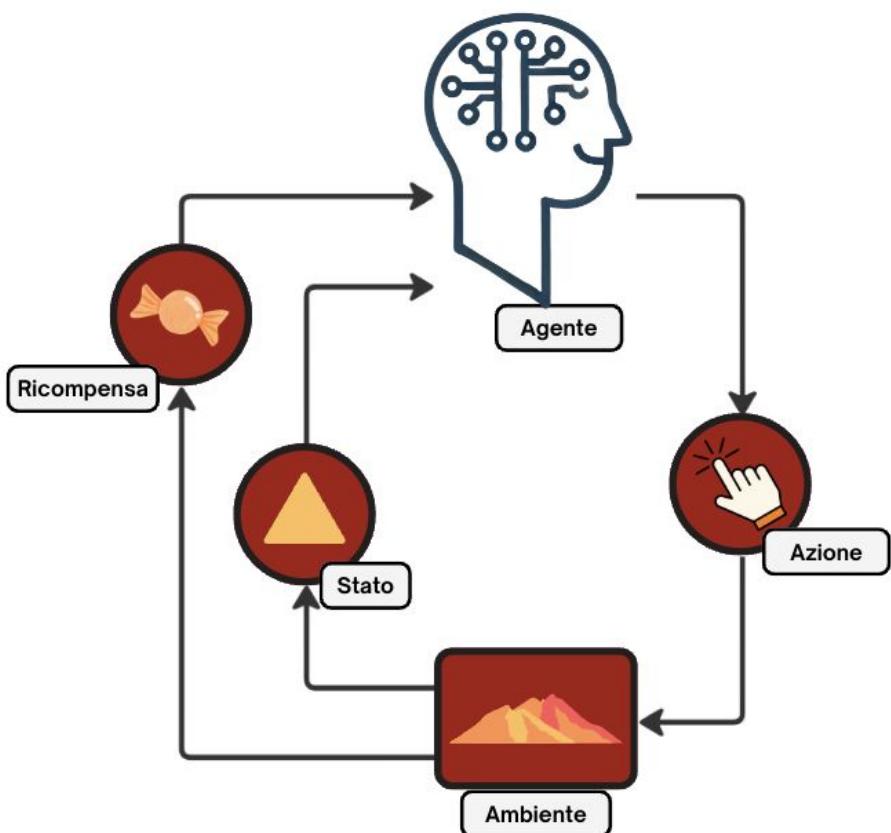
le situazioni in cui l'agente si trova.

ricompense

feedback positivi o negativi che l'agente riceve dopo ogni azione

Come imparano le macchine?

Apprendimento per rinforzo



L'agente impara a prendere le decisioni migliori attraverso un processo di prova ed errore, cercando di massimizzare le ricompense nel lungo termine.

Un ruolo cruciale è svolto dalla **funzione ricompensa**. Questa è scelta dal progettista e deve essere tale da stimolare l'agente ad arrivare all'obiettivo voluto.

Il ruolo del progettista (cioè, dove sta la matematica)

Chi progetta un sistema di reinforcement learning ha un ruolo cruciale, soprattutto nella definizione della funzione di ricompensa.

1. **Definizione degli obiettivi.** Deve tradurre gli obiettivi desiderati in una funzione di ricompensa appropriata.
2. **Bilanciamento delle ricompense.** Deve decidere il peso relativo di diverse azioni o risultati.
3. **Prevenzione di comportamenti indesiderati.** Una funzione di ricompensa mal progettata potrebbe portare l'agente a trovare "scorciatoie" non previste per massimizzare la ricompensa.
4. **Considerazioni etiche.**

Esempio: Per un'auto a guida autonoma, il modello dovrebbe bilanciare fattori come velocità, sicurezza e comfort dei passeggeri.

Alcune applicazioni ed esperimenti

- Giochi -



SuperMario Bros



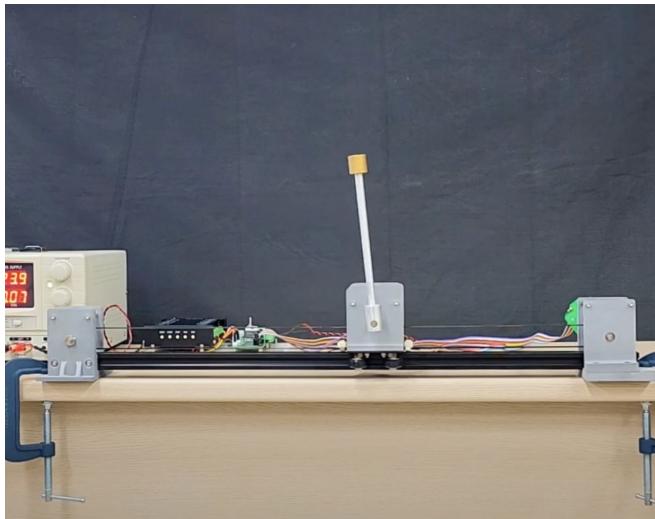
Nascondino



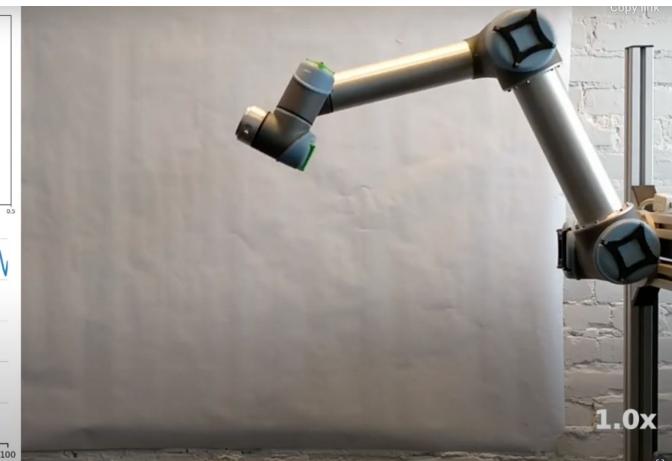
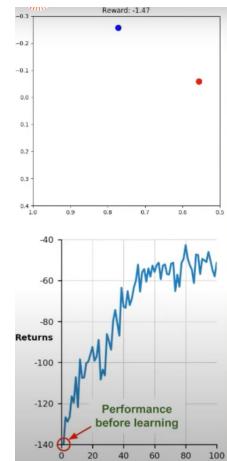
AlphaGo

Alcune applicazioni ed esperimenti

- Robotica -



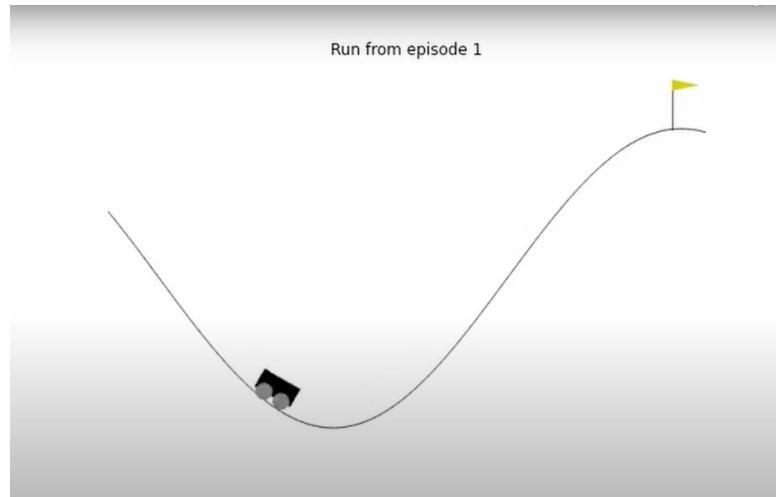
Pendolo invertito



Braccia robotiche

Alcune applicazioni ed esperimenti

- Guida automatica -



Il problema della
macchina di montagna



Parcheggio
automatico