



# UNIVERSITÀ DEGLI STUDI DI PADOVA

## Introduzione all'IA

Stefano Ghidoni

[stefano.ghidoni@unipd.it](mailto:stefano.ghidoni@unipd.it)





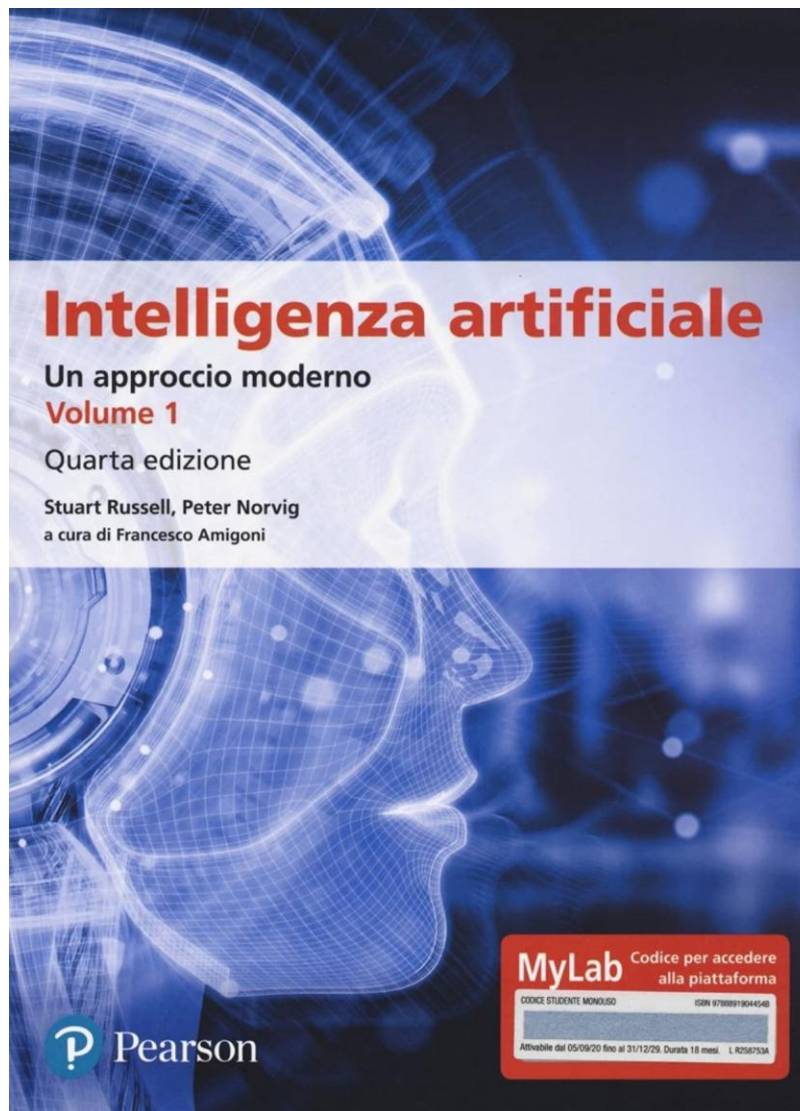
- Introduzione all'IA
- Vari approcci metodologici all'IA
- Una breve panoramica storica
- Agenti intelligenti
- Un esempio implementativo



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

# Fonte principale

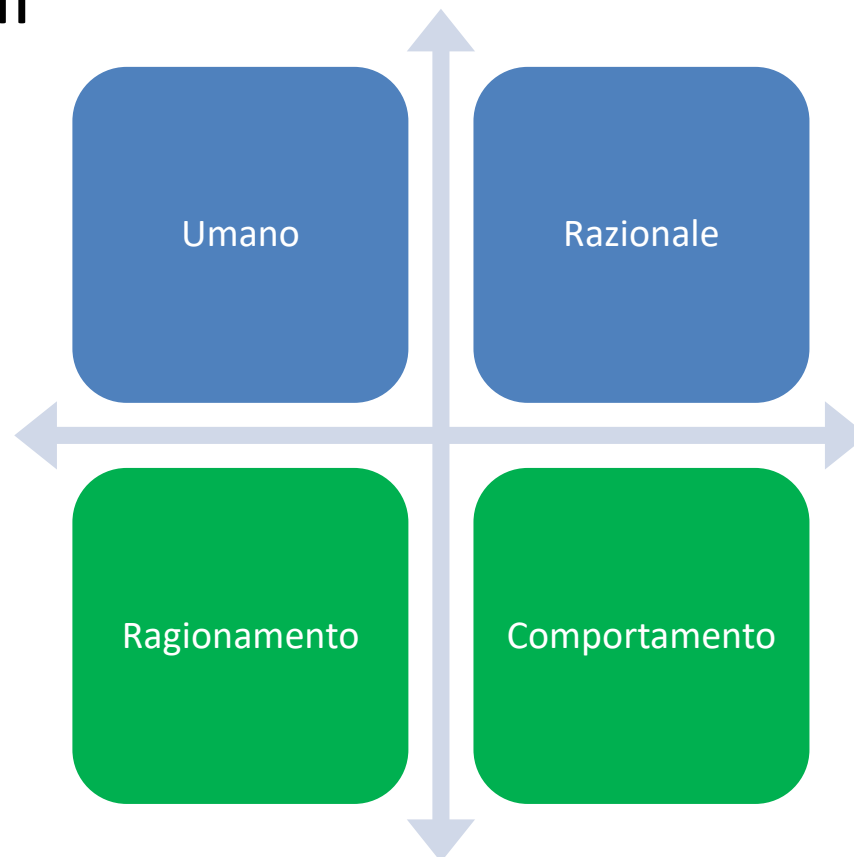
IAS-LAB



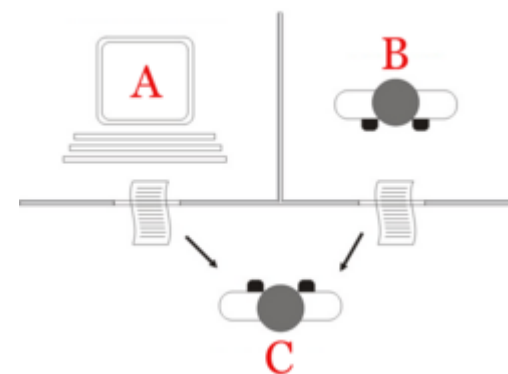


- La definizione di IA non è univoca
  - Fedeltà alla prestazione umana
    - Ma gli esseri umani possono prendere decisioni non ottime
  - Razionalità, intesa come "fare la cosa giusta"
- L'IA può essere una proprietà di
  - Processi del pensiero e del ragionamento
  - Comportamento

- Abbiamo quindi una definizione in due dimensioni



- Test di Turing
  - Una macchina supera il test se un esaminatore umano che può formulare domande non è in grado di capire se le risposte sono state fornite da una macchina o da un umano
  - Non contempla la parte fisica
- Test di Turing totale
  - Include l'interazione col mondo fisico





- Per superare il test di Turing una macchina necessita di:
  - Interpretazione del linguaggio naturale (NLP)
  - Rappresentazione della conoscenza
  - Ragionamento automatico
  - Apprendimento automatico (ML)
- Per il test totale sono inoltre necessari:
  - Visione artificiale
  - Robotica

- Queste discipline comprendono la maggior parte dell'IA
- Dilemma tra i ricercatori IA:
  - Sviluppare sistemi in grado di superare il test di Turing
  - Concentrarsi sui principi alla base dell'IA
- Imitare gli umani vs sintetizzare sistemi intelligenti

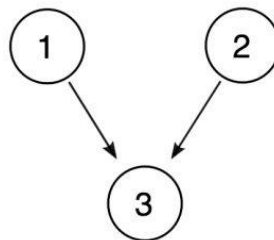




- Imitazione dell'umano:
  - Determinare i percorsi del pensiero umano
  - Introspezione (cattura al volo dei pensieri mentre scorrono)
  - Sperimentazione psicologica
  - Imaging cerebrale
- Forte interdisciplinarietà con le scienze cognitive



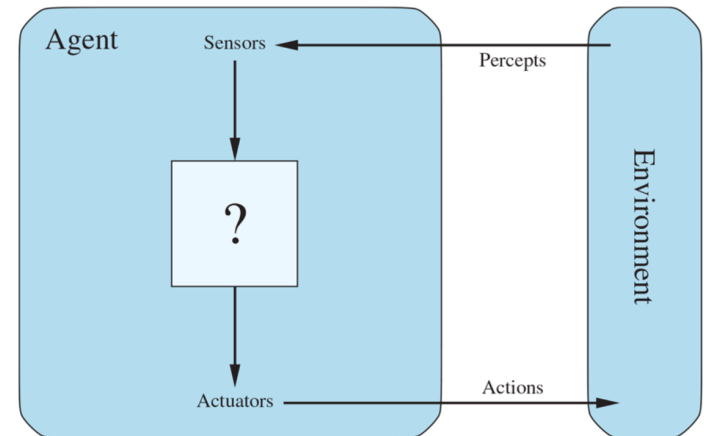
- Come codificare il pensiero corretto?
- Punto di partenza: sillogismo
  - Schema di deduzione che porta a conclusioni corrette quando le premesse sono corrette
    - Da "Socrate è un uomo" e
    - "Tutti gli uomini sono mortali"
    - si deduce per sillogismo "Socrate è mortale"





- Questi studi hanno dato origine alla disciplina detta **logica**
- La visione dell'IA basata sulla logica prende il nome di *tradizione logicista*
- La logica non contempla situazioni che non partono da assunti certi
  - Il mondo è popolato di eventi non certi!
- Si adotta quindi la teoria della probabilità
  - Consente di sviluppare ragionamenti corretti in presenza di eventi incerti

- Approccio degli agenti razionali
- Si definisce agente qualcosa che agisce
- Gli agenti artificiali:
  - Percepiscono l'ambiente
  - Possono persistere in un'attività
  - Si adattano ai cambiamenti
  - Perseguono obiettivi



- Un agente razionale agisce per ottenere:
  - In un mondo certo: il miglior risultato
  - In un mondo incerto: il miglior risultato *atteso*





- Agire razionalmente va oltre il pensare razionalmente
- Esistono comportamenti razionali che non dipendono da un ragionamento
  - Molti comportamenti istintivi
    - Ritrarre la mano a contatto con una superficie rovente



- L'approccio all'IA basato su agenti intelligente è stato spesso prevalente
  - È dimostrabile – l'approccio sull'imitazione umana lo è molto meno
- Lo studio e la realizzazione di agenti che "fanno la cosa giusta" è chiamato **modello standard**

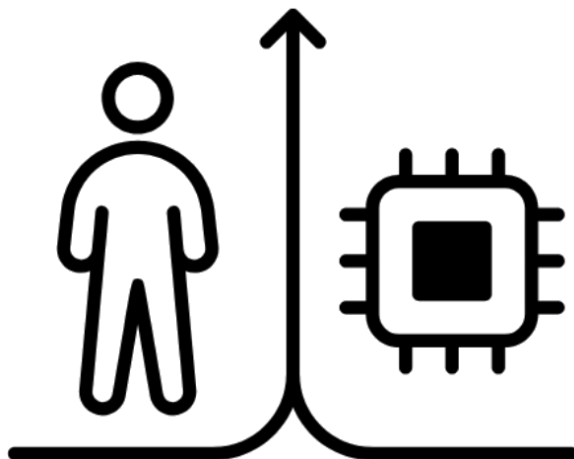
- Cosa significa "fare la cosa giusta"?
- Come possiamo definire "la cosa giusta"?





- Qualche controesempio:
- Qual è l'obiettivo che possiamo definire per un'auto a guida autonoma?
  - "Raggiungere la destinazione garantendo la sicurezza"
- Ma non è razionalmente possibile garantire la sicurezza
  - Quindi un agente potrebbe rifiutarsi di intraprendere il viaggio

- Problema di allineamento dei valori
  - I valori e gli obiettivi affidati alla macchina devono essere allineati a quelli dell'umano
- La definizione degli obiettivi è tanto più critica quanto più l'agente è intelligente





- Esempio: una IA che gioca a scacchi e che ha come unico obiettivo vincere, potrebbe
  - Ipnotizzare o ricattare l'avversario
  - Infastidire l'avversario
    - Ruy Lopez scrisse in uno dei primi libri sugli scacchi (1561): "Posiziona sempre la scacchiera in modo che il tuo avversario abbia il sole negli occhi"
- Questo comportamento è scorretto?



- Esempio: una IA che gioca a scacchi e che ha come unico obiettivo vincere, potrebbe
  - Ipnotizzare o ricattare l'avversario
  - Infastidire l'avversario
    - Ruy Lopez scrisse in uno dei primi libri sugli scacchi (1561): "Posiziona sempre la scacchiera in modo che il tuo avversario abbia il sole negli occhi"
- Questo comportamento è scorretto?
  - È solo la conseguenza dell'obiettivo fissato

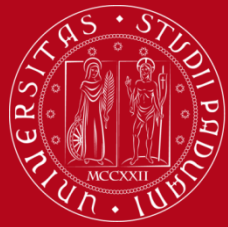


- Trasferire i nostri obiettivi alla macchina può essere molto difficile, soprattutto quando
  - La macchina è complessa e intelligente
  - L'ambiente è complesso



- Possiamo quindi andare verso una nuova formulazione
  - La macchina persegue i nostri obiettivi, ma è *incerta* su quali siano
- Questo rende il comportamento più votato alla cautela
  - La macchina non sa quali siano realmente gli obiettivi!

# Breve storia dell'IA



- 1943: primo lavoro McCulloch e Pitts
- Si rifanno a:
  - Conoscenze di fisiologia dei neuroni umani
  - Analisi formale della logica proposizionale (Russel, Whitehead)
  - Teoria della computazione (Turing)
- Dimostrano che tutti gli operatori logici possono essere implementati con strutture a rete

- 1950: Minsky e Edmonds costruiscono il primo computer basato su reti neurali, Snark
- 1950: Test di Turing
- 1957: GPS – General Problem Solver, primo esempio del "pensare umanamente"
  - Solutore di problemi che aveva come obiettivo percorrere gli stessi passaggi logici dell'umano



Newell & Simon



- 1963: SAINT risolve problemi di integrali chiusi del I anno di un corso universitario
- Questi successi, brillanti da un lato, furono ottenuti in domini molto semplificati



- L'IA iniziò a scontrarsi con la realtà verso la seconda metà degli anni '60
  - Problemi difficili da scalare, spesso intrattabili
  - Troppa fiducia nel progresso dell'hardware, su cui era scaricato l'onere per scalare le soluzioni
- I metodi con difficoltà a scalare presero il nome di "metodi deboli"

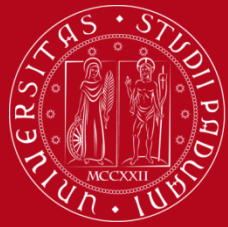
- Andare oltre ai metodi deboli
  - Aumentare la conoscenza
    - Del problema
    - Del dominio
- Porta allo sviluppo dei sistemi esperti
  - Dendral, 1969, riconoscimento della struttura molecolare dai dati di spettrometria di massa
    - Sistema a conoscenza intensiva





- Inverno AI: verso fine anni '80, dopo decenni di fiducia, molte azienda falliscono per l'impossibilità di applicare al mondo reale i loro sistemi





- Anni '80
  - Hidden Markov Model (HMM)
  - Reti bayesiane
- Modelli generali adattabili a molti problemi diversi mediante training
- Paradigma dell'apprendimento per rinforzo
  - Markov decision process (MDP)



- Deep learning: estendere la profondità di una rete
  - Alcuni accenni già negli anni '70
  - Efficaci solo con grandi dataset e training onerosi

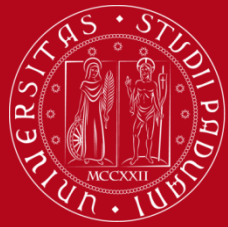
# Agenti intelligenti



- Agente: qualsiasi cosa che
  - Percepisce un ambiente (tramite sensori)
  - Agisce su un ambiente (tramite attuatori)
- Simile al concetto di robot
- Problema: dobbiamo definire l'ambiente

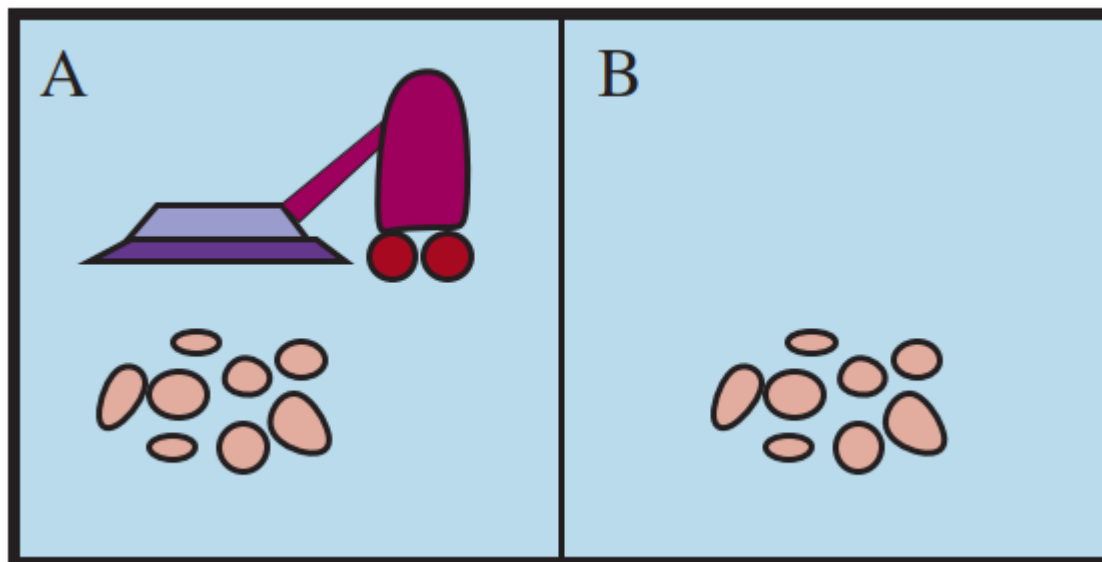


- Percezione: il dato inviato dal sensore
- Sequenza percettiva: la storia delle percezioni
- L'azione intrapresa dall'agente dipende da
  - Conoscenza integrata nell'agente
  - Sequenza percettiva

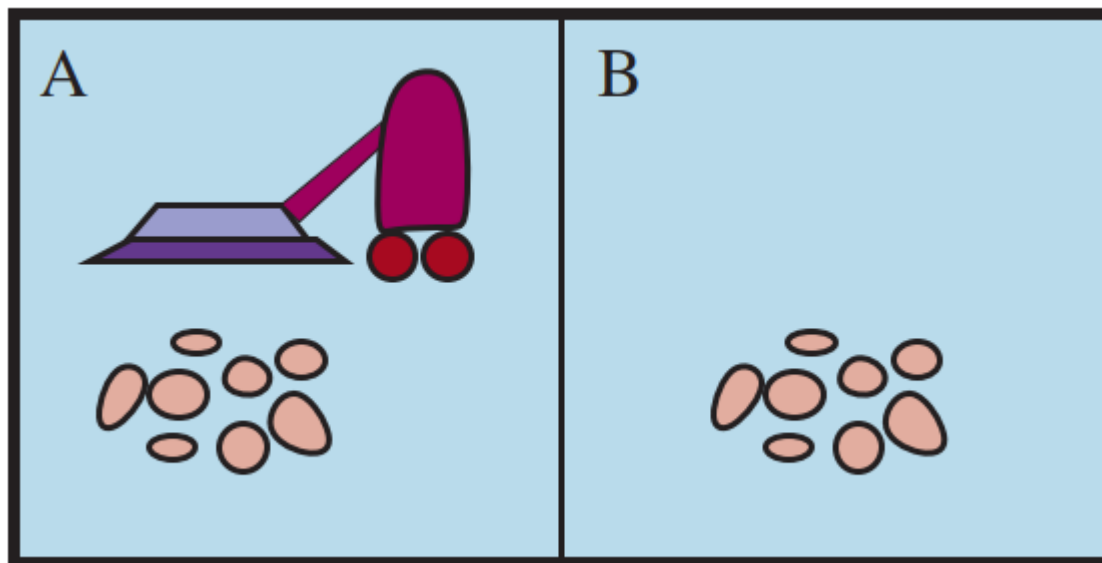


- Azione intrapresa dall'agente
  - Formalizzata da una **funzione agente**
    - Azione da intraprendere per ogni sequenza percettiva
    - Può essere una descrizione molto onerosa
  - Implementata da un **programma agente**

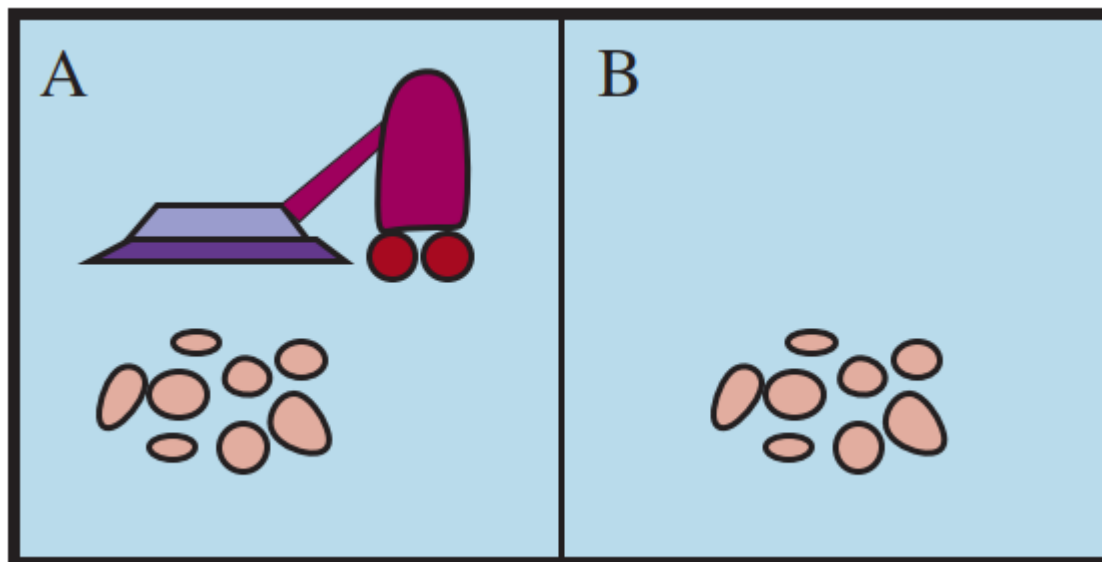
- Esempio: robot aspirapolvere
  - Come descrivereste l'ambiente?



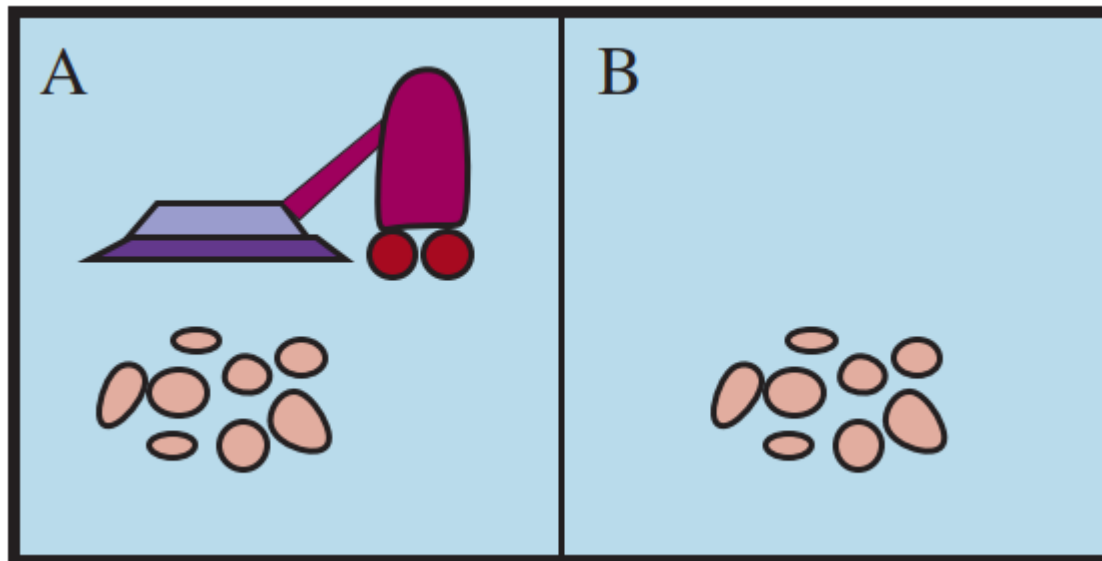
- Esempio: robot aspirapolvere
  - Che sensori sarebbero utili?
  - Come descrivereste l'insieme delle possibili azioni?



- Supponiamo che il robot percepisca
  - In quale riquadro si trova
  - Se questo è pulito o sporco



- Supponiamo che il robot possa svolgere le azioni
  - Spostarsi a sinistra o a destra
  - Aspirare





- Una possibile funzione agente

Sequenza percettiva	Azione
[A, pulito]	Destra
[A, sporco]	Aspira
[B, pulito]	Sinistra
[B, sporco]	Aspira
[A, pulito], [A, pulito]	Destra
[A, pulito], [A, sporco]	Aspira
...	...
[A, pulito], [A, pulito], [A, pulito]	Destra
[A, pulito], [A, pulito], [A, sporco]	Aspira
...	...



- L'agente deve essere progettato in modo da fare "la cosa giusta"
- Questa è espressa da una misura di prestazione
  - Valuta quanto è desiderabile la conseguenza dell'azione dell'agente
    - Conseguenzialismo: valutazione della desiderabilità in base alle conseguenze di un'azione



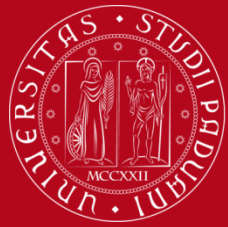
- Che misura di prestazione definireste?



- Opzione 1 – definire il comportamento dell'agente
- Misura di prestazione
  - Quantità di sporco aspirato in un lasso di tempo
- Possibile comportamento
  - Aspirapolvere aspira e svuota il cestino, poi aspira, poi svuota il cestino, ...
- **Questo massimizza la misura definita!**



- Opzione 2 – definire i risultati desiderati
- Misura di prestazione
  - Livello di pulizia dei riquadri
- Possibile comportamento
  - Pulizia frequente e superficiale
  - Pulizia approfondita e sporadica



- Come si definisce la razionalità?
- Consideriamo un esempio:
  - Vedo un amico dall'altro lato della strada
  - Controllo la strada, non ci sono veicoli
  - Decido di attraversare per salutarlo
  - Un aereo in volo perde un motore, mi uccide
- È stata una scelta irrazionale?



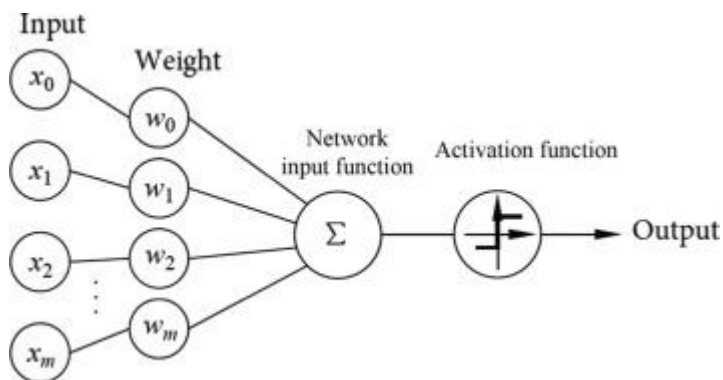
- La **razionalità** ha massimizzato il risultato atteso
  - In base alle osservazioni disponibili
- La **perfezione** massimizza il risultato reale
  - Possibile se conosciamo tutto del mondo, tramite l'**onniscienza**



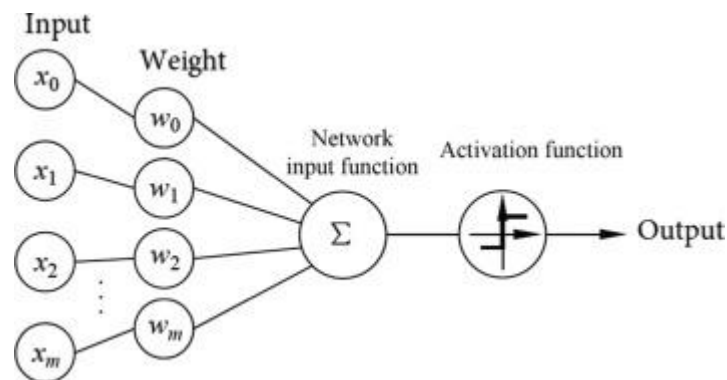
- Le azioni razionali devono tener conto delle osservazioni necessarie
  - Un agente senza sensori per guardare la strada potrebbe attraversare?

# Perceptron

- Implementazione di reti di apprendimento automatico
- Perceptron (Rosenblatt, 1958)



- Il perceptron è composto da
  - Input
  - Pesi ( $w$ ), modificabili in fase di training
    - Incluso un bias che non dipende dagli input
  - Una funzione di attivazione

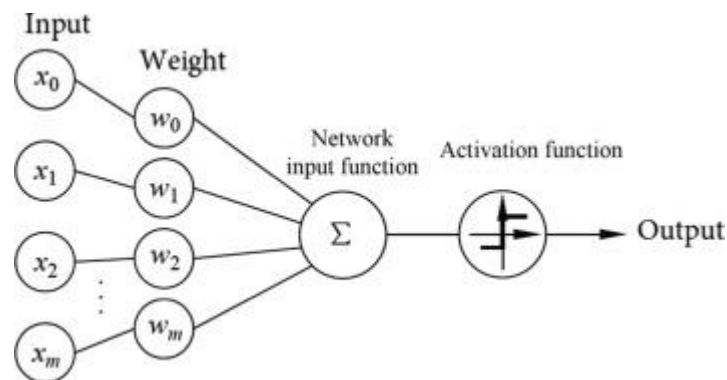


- Network input function

$$v(n) = \sum_{i=0}^n w_i(n) x_i(n)$$

- Output:

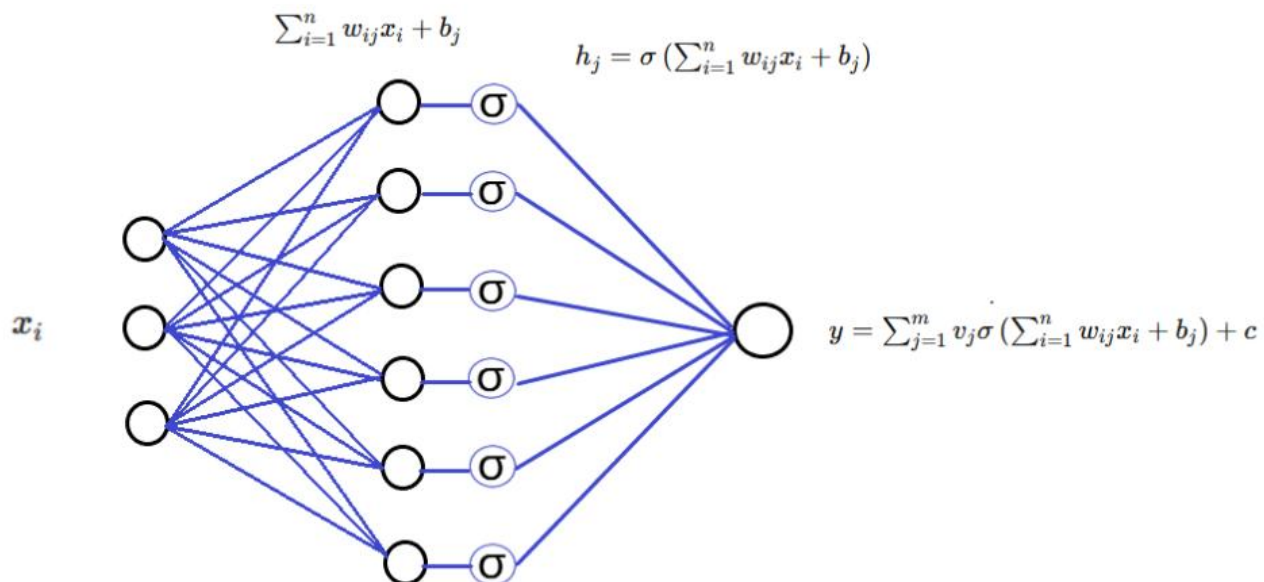
$$O_i(n) = f_{att} (v(n))$$





- Più perceptron possono essere combinati insieme
- Possibile struttura: gruppo di perceptron
  - Analizzano gli stessi dati
  - Contribuiscono all'output

- Single-layer perceptron



$y$  è l'output della rete.

$v_j$  è il peso della connessione tra il neurone nascosto  $j$  e il neurone di output.

$h_j$  è l'output del neurone nascosto  $j$  (già calcolato come sopra).

$c$  è il bias del neurone di output.



- Teorema dell'approssimazione universale  
(teorema di Cybenko)

*Un single-layer perceptron con funzione di attivazione continua e non costante può approssimare una qualunque funzione continua su un intervallo chiuso e limitato, purché vi sia un sufficiente numero di neuroni nello strato nascosto*

- Quindi il problema dell'apprendimento è risolto?



# UNIVERSITÀ DEGLI STUDI DI PADOVA

## Introduzione all'IA

Stefano Ghidoni

