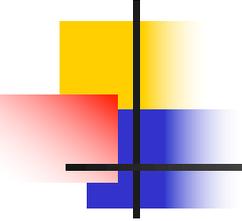


# ***Artificial Intelligence e Machine Learning: fini e confini***

---

Angelo Montanari  
Dipartimento di Scienze Matematiche,  
Informatiche e Fisiche  
Università degli Studi di Udine

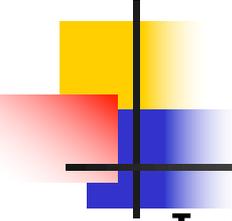
Roma, 3 dicembre, 2022



# Sommario

---

- Inquadramento
- **Intelligenza** artificiale simbolica (**basata sui modelli**)
- Limiti pratici e teorici dell'IA simbolica
- **Intelligenza** artificiale sub-simbolica (**guidata dai dati**): machine learning and deep learning
- Questioni pratiche e teoriche sull'IA guidata dai dati
- Alcune considerazioni finali



# Un breve inquadramento

---

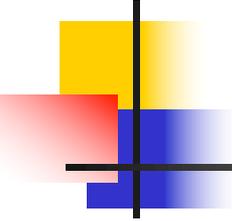
Intelligenza artificiale simbolica e sub-simbolica

**Intelligenza artificiale simbolica:** rappresentazione della conoscenza e ragionamento automatico

**Intelligenza artificiale sub-simbolica:** modelli e tecniche di machine learning e deep learning

Trade-off tra efficienza / efficacia and **interpretabilità**

Explainable artificial intelligence e **trustworthy artificial intelligence**



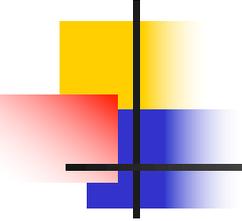
# IA simbolica: le origini

---

L'inizio dell'IA simbolica viene ricondotta al lavoro di Turing sulle relazioni tra sistemi computerizzati e intelligenza

**Turing** si interroga sulla possibilità di rendere intelligente un sistema artificiale, ma non delinea un percorso per raggiungere tale obiettivo, sebbene lo stretto legame che stabilisce tra intelligenza e linguaggio di fatto attribuisce un ruolo fondamentale alla dimensione simbolica umana

Più tardi, **Newell** e **Simon** formularono la cosiddetta "**physical symbol system hypothesis**" (la capacità di elaborare simboli / espressioni è sufficiente per rendere un sistema intelligente; manipolazioni simboliche dello stesso tipo sono alla base dell'intelligenza umana)



# La visione “macchinista”

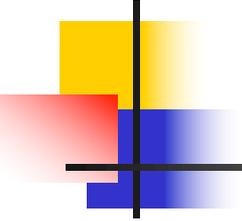
---

La visione macchinista è affermata in modo esplicito da Minsky (M. Minsky, *The Society of Mind*, Simon and Schuster, 1988):

“Non vi alcun motivo per credere che il **cervello** sia qualcosa di diverso da una **macchina** con un numero enorme di componenti che funzionano in perfetto accordo con le leggi della fisica”

# Marvin Minsky





# La mente come processo

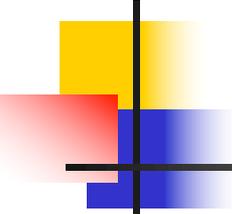
---

E' una declinazione particolare della posizione materialistica classica, che si contrappone ad ogni dualismo mente/corpo

Per Minsky la mente è semplicemente ciò che fa il cervello (la **mente come processo**)

Tale interpretazione della mente stabilisce una stretta analogia tra la relazione tra la mente e il cervello e quella che intercorre tra le nozioni di **processo** (un programma in esecuzione) e di **programma** in informatica: per Minsky la mente è semplicemente il

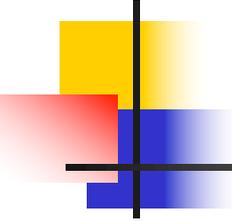
**“cervello in esecuzione”**



# Rapporto mente-cervello

---

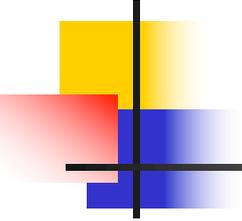
- Per spiegare la mente evitando la circolarità, occorre descrivere il modo in cui le menti sono costruite a partire da materia priva di mente, parti molto più piccole e più semplici di tutto ciò che può essere considerato intelligente
- **Questione:** una mente può essere associata solo ad un cervello o, invece, qualità tipiche della mente possono appartenere, in grado diverso, a tutte le cose?
- Per Minsky, il **cervello** può essere visto come una **società organizzata**, composta da una molteplicità di componenti strutturate in modo gerarchico, alcune delle quali operano in modo del tutto autonomo, la maggior parte in un rapporto alle volte di collaborazione, più spesso di competizione, con altre componenti



# La società della mente

---

- Intelligenza umana frutto dell'interazione di un numero enorme di componenti fortemente diverse fra loro, i cosiddetti **agenti della mente**, componenti elementari ("particelle") di una (teoria della) mente
- **Questione:** come può l'opera combinata di un insieme di agenti produrre un comportamento che ogni singolo agente, considerato separatamente, non è in grado di fornire?
- Si tratta di una questione ben nota: comportamento emergente in sistemi complessi, dove la **nozione di emergenza** fa riferimento a proprietà di un sistema non derivabili dalle proprietà note dei suoi componenti



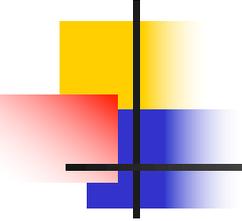
# La nozione di agenzia

---

- Per superare posizioni di riduzionismo ingenuo difficilmente sostenibili (Minsky contesta chi considera la fisica e la chimica modelli ideali di come dovrebbe essere la psicologia), Minsky introduce la **nozione di agenzia**

un'agenzia è un insieme di agenti collegati fra loro da un'opportuna rete di interconnessioni

- La **gerarchia delle agenzie**



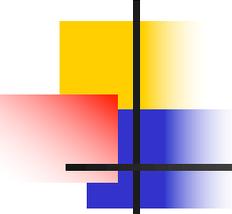
# La teoria degli agenti

---

La **teoria degli agenti** / i **sistemi multi-agente** in intelligenza artificiale possono essere visti come la controparte "applicativa" della società della mente di Minsky

**Agente (artificiale) intelligente:** un agente intelligente è un sistema in grado di decidere cosa deve fare e di intraprendere le azioni necessarie a realizzare quanto deciso

**Sistemi multi-agente:** sistemi costituiti da un insieme di agenti interagenti (interazione = scambio di messaggi tra agenti artificiali)



# Naturale e artificiale

---

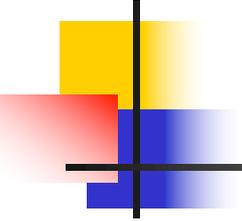
La prospettiva di Minsky non è l'unica possibile.

Punto di vista alternativo: le macchine sono ciò che vi è di più umano nella natura inanimata:

**l'artificiale** quale tratto distintivo dell'umano

Per il paleoantropologo Y. Coppens, la costruzione dei primi utensili (oggetti artificiali) segna l'inizio di una **storia culturale**, di tutto ciò che non è natura (*Storia dell'uomo e cambi di clima*, Jaca Book, 2006)

Oltre che nella tecnologia, tale storia si manifesta nelle dimensioni intellettuale, spirituale, morale ed estetica dell'uomo



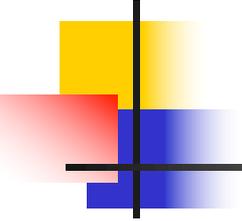
# L'uomo e le macchine

---

La questione fondamentale è quella del **rapporto** dell'**uomo** con le **macchine**, a fronte della crescita della **complessità** e del grado di autonomia di queste ultime

La distinzione tra chi progetta e costruisce (**progettista / costruttore**) una macchina e chi la utilizza (**utilizzatore**)

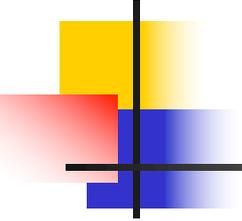
Le diverse modalità di progettazione, sviluppo e realizzazione di una macchina rispetto al passato: una singola persona non è in grado di controllare l'intero processo (conoscenza **distribuita**)



# Intenzionalità e macchine

---

- Il legame tra **intenzionalità** (umana) e capacità di creare artefatti: la creazione di artefatti è per l'uomo un modo per estendere la propria intenzionalità
- “i nostri strumenti – osserva Searle -- sono estensioni dei nostri scopi/intenzioni”
- L'intenzionalità trova modo di esprimersi nella creazione di artefatti, dei quali costituisce, in un certo senso, la causa finale (**intenzionalità derivata**), e, successivamente e in modo diverso, nel loro utilizzo, ma non si trasferisce ad essi



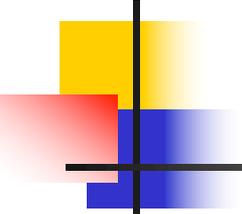
# L'intelligibilità delle macchine - 1

---

Il legame tra intenzionalità e intelligibilità delle macchine

**L'intelligibilità delle macchine**, ossia la possibilità di descriverne in modo comprensibile le caratteristiche strutturali e funzionali e le tecniche di costruzione, è condizione essenziale per il loro sviluppo e il loro utilizzo

Solo l'esistenza di una **spiegazione adeguata** (razionale) del funzionamento di una macchina complessa consente, infatti, di predirne, nei limiti del possibile, il comportamento e di diagnosticarne eventuali guasti e malfunzionamenti

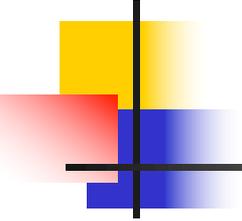


# L'intelligibilità delle macchine - 2

---

La spiegazione mediante il **paradigma riduzionista**: l'analisi del sistema nel suo complesso viene ridotta all'analisi separata delle sue componenti elementari e delle loro interazioni

Efficace nel caso di macchine relativamente semplici, tale approccio diventa problematico in presenza di **meccanismi di controllo** (meccanismi di anticipazione e meccanismi di retroazione). Tali meccanismi possono essere visti come il tentativo di introdurre nella macchina un'opportuna **rappresentazione** dell'obiettivo (causa finale) per il quale la macchina è stata costruita



# Alan Turing

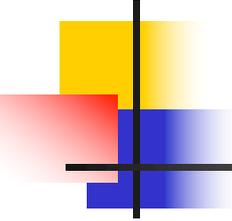
---



# Intelligenza e macchine

Il **test di Turing** (o gioco dell'imitazione): una macchina può essere definita intelligente se riesce a convincere una persona che il suo comportamento, dal punto di vista intellettuale, non è diverso da quello di un essere umano medio



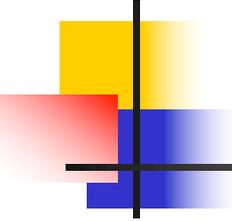


# Il test di Turing: dettagli - 1

---

Il test si svolge in 3 stanze separate. Nella prima si trova l'esaminatore umano (A); nelle altre due vi sono rispettivamente un'altra persona e il computer che si sottopone al test. Dei due A conosce i nomi (B e C), ma ignora chi sia la persona e chi il computer.

Sia B che C si relazionano separatamente con A attraverso un computer. Via computer A può porre domande a B e C e leggere le loro risposte. Compito di A è scoprire l'identità di B e C (**chi è la persona, chi è la macchina?**) entro un limite di tempo prefissato.

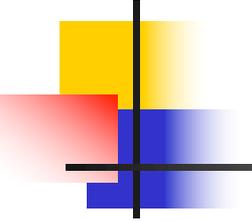


# Il test di Turing: dettagli - 2

---

A può effettuare qualunque tipo di domanda; il computer ovviamente cercherà di rispondere in modo tale da celare la propria identità.

La **macchina supera il test** se A non riesce a identificarla nel tempo prefissato. Il test verrà ripetuto più volte, coinvolgendo anche esaminatori diversi, in modo da ridurre i margini di soggettività.

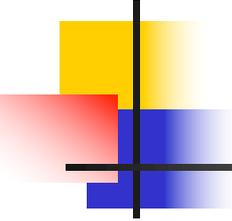


# Turing e il comportamentismo

---

E' evidente l'influenza del **comportamentismo** in auge nella prima metà del novecento sul gioco dell'imitazione proposto da Turing per stabilire se una macchina possa essere definita intelligente

Per il comportamentismo (metodologico), l'introspezione non è uno strumento adatto allo studio della mente in quanto non può fornire alcun dato affidabile. L'unica alternativa praticabile è lo studio delle misurazioni degli stimoli/percezioni forniti ad un uomo/animale e delle risposte/azioni risultanti



# Intelligenza e corporeità

---

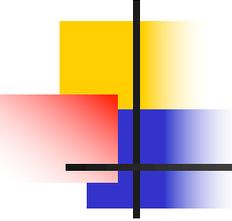
Il modello di intelligenza sotteso al Test di Turing è un modello astratto/disincarnato dell'intelligenza.

Una delle acquisizioni più importanti della ricerca in IA degli ultimi decenni è la consapevolezza del ruolo cruciale che gli organi di senso svolgono nell'interazione dell'uomo col mondo e della conseguente impossibilità di un'intelligenza (artificiale) priva di "corporeità". Ciò ha portato allo sviluppo di un rapporto sempre più stretto tra IA ("cervello senza corpo") e robotica ("corpo senza cervello").

Per paradossale che possa suonare, per avvicinarsi all'intelligenza umana l'IA deve diventare un **intelligenza incarnata**.

# John Searle





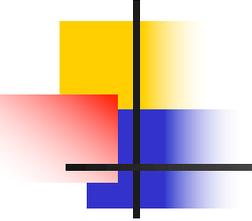
# La stanza cinese di Searle

---

Tesi fondamentale: impossibilità per una macchina di manifestare l'**intenzionalità** che caratterizza gli esseri umani e, sia pure in forme diverse, gli animali

**Intenzionalità:** caratteristica che contraddistingue certi stati mentali, quali le credenze, i desideri e le intenzioni, diretti verso oggetti e situazioni del mondo

Per Searle, l'intenzionalità è un dato di fatto empirico circa le effettive relazioni causali tra mente e cervello, che consente (unicamente) di affermare che certi processi cerebrali sono sufficienti per l'intenzionalità



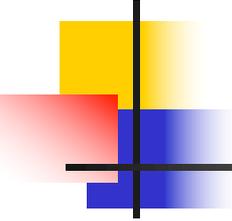
# Intenzionalità e programmi

---

Per Searle, l'esecuzione di un programma su un dato input (**processo** nel linguaggio informatico comune) non è mai di per se stessa una condizione sufficiente per l'intenzionalità

## La "dimostrazione"

Searle immagina di sostituire un agente umano al calcolatore nel ruolo di esecutore di una specifica istanza di un programma e mostra come tale esecuzione possa avvenire alcuna forma significativa di intenzionalità



# L'esperimento mentale di Searle

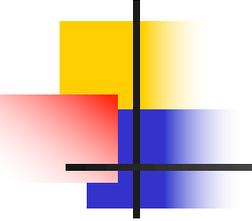
---

La struttura dell'**esperimento mentale**: una teoria della mente può essere confermata/falsificata immaginando che la propria mente operi secondo i principi di tale teoria e verificando la validità o meno delle affermazioni/previsioni della teoria

## L'**esperimento mentale di Searle**:

Searle prende in esame i lavori sulla simulazione della capacità umana di **comprendere narrazioni**

Caratteristica distintiva di tale abilità: la capacità di rispondere a domande che coinvolgono informazioni non fornite in modo esplicito dalla narrazione, ma desumibili da essa sfruttando conoscenze di natura generale

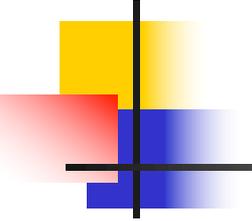


# L'esperimento in dettaglio - 1

---

Searle immagina che una persona venga chiusa in una stanza e riceva **3 gruppi di testi** scritti in una lingua a lei sconosciuta (**cinese**), interpretabili (da chi fornisce i testi) rispettivamente come il testo di una narrazione, un insieme di conoscenze di senso comune sul domino della narrazione e un insieme di domande relative alla narrazione

Immagina, inoltre, che tale persona riceva un **insieme di regole**, espresse nella propria lingua (**inglese**), che consentano di collegare in modo preciso i simboli formali che compaiono nel primo gruppo di testi a quelli che compaiono nel secondo e **un altro insieme di regole**, anch'esse scritte in una lingua a lei nota, che permettano di collegare i simboli formali che compaiono nel terzo gruppo di testi a quelli degli altri due e che rendano possibile la produzione di opportuni simboli formali in corrispondenza di certi simboli presenti nel terzo gruppo di testi

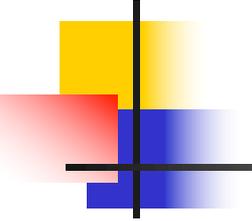


# L'esperimento in dettaglio - 2

---

Le **regole** vengono interpretate (da chi le fornisce) come un **programma** e i **simboli prodotti** come **risposte** alle domande poste attraverso il terzo gruppo di testi. Quanto più il programma è ben scritto e l'esecuzione delle regole spedita, tanto più il comportamento della persona sarà assimilabile a quello di un parlante nativo (un cinese)

Immaginiamo ora uno scenario in cui la persona riceva il testo narrativo e le domande ad esso relative nella propria lingua (**inglese**) e fornisca le risposte in tale lingua, sfruttando la propria conoscenza di senso comune. Tali risposte saranno indistinguibili da quelle di un qualunque altro parlante nativo, in quanto la persona è un parlante nativo



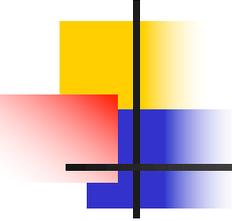
# L'esperimento in dettaglio - 3

---

Dal punto di vista esterno, le risposte fornite in lingua cinese e quelle fornite in lingua inglese saranno egualmente buone; il modo in cui vengono prodotte è, però, radicalmente diverso.

A differenza del secondo caso, nel primo caso le risposte vengono ottenute attraverso un'opportuna manipolazione algoritmica di simboli formali ai quali la persona non associa alcun significato (simboli non interpretati).

**Il comportamento della persona è, in questo caso, del tutto assimilabile all'esecuzione di un programma** su una specifica istanza da parte di un sistema artificiale.



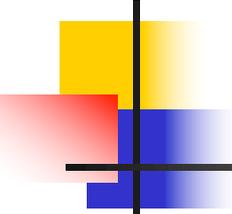
# Esito dell'esperimento

---

**Esito dell'esperimento:** la capacità (di un uomo/una macchina) di manipolare le informazioni ricevute secondo regole formali ben definite non è sufficiente a spiegare il processo di comprensione (non vi è nemmeno alcuna evidenza che essa debba essere una condizione necessaria) – “carattere non intenzionale, e, quindi, semanticamente vuoto, dei simboli elaborati da un sistema artificiale” (Diego Marconi)

**Conclusioni:** i processi mentali non possono essere ridotti a processi di natura computazionale che operano su elementi formalmente definiti

**Osservazione:** confutazione della validità del cosiddetto test di Turing



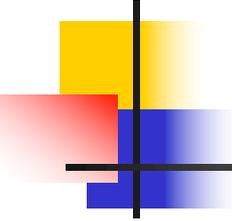
# Conseguenze

---

L'affermazione dell'**irriducibilità** dell'intenzionalità all'esecuzione di programmi su particolari input ha alcune importanti conseguenze:

- impossibilità di spiegare le modalità con le quali il cervello produce l'intenzionalità attraverso il meccanismo della istanziazione di programmi
- ogni meccanismo in grado di produrre intenzionalità deve avere abilità di tipo causale pari a quelle del cervello

Problemi (irrisolti): cosa differenzia il caso in cui la persona comprende il testo (inglese) da quello in cui non vi è alcuna comprensione (cinese)? Questo qualcosa può (se sì, come) essere trasferito ad un macchina?



# IA basata sui dati

---

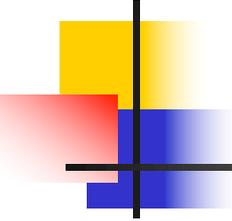
Negli ultimi anni, l'IA **basata sui dati** sta assumendo un ruolo dominante

In verità, il machine learning è presente in IA fin dagli albori

Ragioni della sua esplosione e pervasività:

- forte crescita del **potere computazionale** dei sistemi di calcolo
- disponibilità di **enormi quantità di informazioni** in rete (testi, immagini, registrazioni sonore, video)

La capacità di analizzare in modo sistematico grandissime collezioni di dati (big data) per scoprire significative similarità e di sfruttare tali similarità per **identificare nuovi dati** dello stesso tipo e per formulare **previsioni affidabili** ha permesso di fare passi in avanti radicali nella ricerca e nelle applicazioni



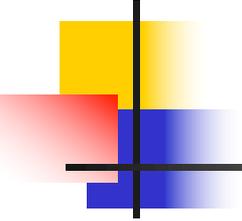
# Machine learning

---

**Machine learning** può essere definito come l'insieme di metodi, tecniche e strumenti che consentono ad un sistema di migliorare (in termini di copertura, prestazioni o qualità) la sua capacità di eseguire certi compiti imparando dall'esperienza

Dal punto di vista operativo, ogni sistema di apprendimento automatico, prima di diventare operativo, deve attraversare una fase di **training** seguita da una di **valutazione**

**Diversi algoritmi** di machine learning proposti in letteratura



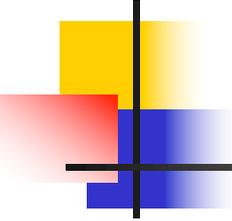
# Learning supervisionato - 1

---

**Diverse tipologie** di learning: supervisionato, non supervisionato e reinforcement learning

Per apprendere una funzione in grado di mappare ogni input nel corrispondente output, il sistema è inizialmente “alimentato” con un certo numero di coppie (input,output) di esempio (dati di training etichettati -- **training set**)

A partire da tali coppie, viene derivato un modello che consenta di predire, per ogni input, un valore di output il più vicino possibile alle etichette date



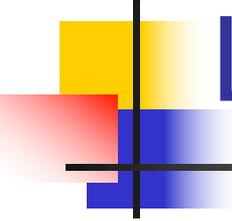
# Learning supervisionato - 2

---

Per determinare la capacità di generalizzazione del modello, occorre una fase di validazione su un opportuno **test set** (insieme di coppie di input-output disgiunte da quelle appartenenti al training set)

Gli input possono essere diversi da quelli incontrati durante la fase di training (ma dovrebbero provenire da una distribuzione di natura simile)

**Strumenti** comunemente utilizzati per il learning supervisionato sono gli alberi di decisione e la regressione lineare

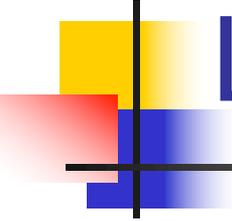


# Learning non supervisionato - 1

---

L'obiettivo non è definire una funzione di input-output su un insieme noto di istanze, ma determinare **ogni possibile correlazione** esistente fra gli elementi di un certo dataset

Sfruttando un numero sufficientemente ampio di esempi, il sistema impara a **classificare ogni nuovo input** (in base alla similarità con input precedenti)



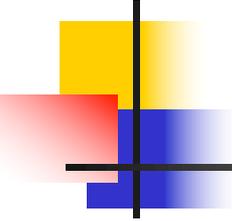
# Learning non supervisionato - 2

---

A differenza del learning supervisionato, non si distingue tra dati di training e dati di test

Rispetto al learning supervisionato, il learning non supervisionato è molto più **generale e flessibile**, ma anche molto più complesso.

Per il learning non supervisionato sono comunemente utilizzati metodi e algoritmi di clustering



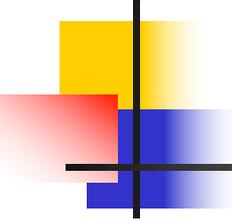
# Reinforcement learning

---

E' una forma intermedia di learning. I dati di training sono più ricchi di quelli di test, ma il sistema (learner) deve predire più informazione per questi ultimi

Si assume che il sistema sia provvisto di strumenti per interagire con l'ambiente (sensori, videocamere, GPS) in modo da meglio comprenderne le caratteristiche e raffinare le proprie capacità di apprendimento

L'ambiente è spesso modellato come un processo di decisione di Markov. Per scrivere algoritmi di reinforcement learning, vengono usate, ad esempio, tecniche di programmazione dinamica



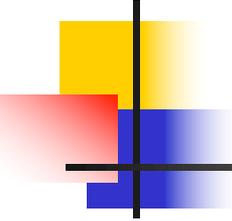
# Reti neurali e deep learning - 1

---

Lo strumento più efficace impiegato per il machine learning sono le **reti neurali**. La loro applicazione sistematica ha portato al cosiddetto deep learning

Le reti neurali programmabili consentono ad un sistema di prendere delle decisioni molto sofisticate e accurate senza alcun intervento da parte di un essere umano

Soluzione alternativa: **algoritmi genetici**



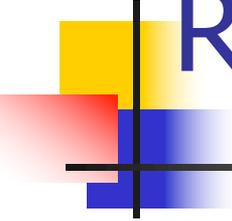
# Reti neurali e deep learning - 2

---

Le reti neurali sono state concepite come modelli artificiali del comportamento della **rete dei neuroni** della persona umana

Esse definiscono delle **gerarchie di concetti** che consentono ad un sistema di imparare concetti complessi a partire da concetti più semplici

Tali gerarchie vengono strutturate in grafi che possono presentare una molteplicità di strati



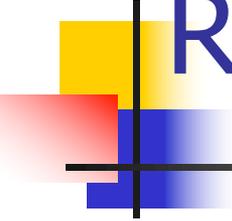
# Reti neurali: una lunga storia - 1

---

Primo modello matematico dei neuroni umani: **McCulloch** e **Pitts** (anni 40), in seguito raffinato mediante la cosiddetta regola di Hebb

**Modello del neurone:** un elemento con  $N$  input binari e 1 output binario. I 2 valori rappresentano i 2 possibili stati del neurone (at rest o firing). Un peso è associato a ciascun input per modellare le diverse intensità dei contatti sinaptici

Il neurone computa la somma pesata degli input (**attivazione**), mimando il comportamento dei neuroni umani. Se l'attivazione è maggiore o uguale ad un certo valore di soglia, **il neurone “si accende”**

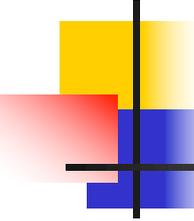


# Reti neurali: una lunga storia - 2

---

Tale modello fa una serie di **assunzioni** semplificatrici (input e output binari, somma lineare, ritardo fisso, sincronia, nessuna relazione tra intensità dello stimolo e “frequenza delle accensioni”)

L'**obiettivo** è costruire un sistema artificiale in grado di gestire enormi quantità di informazioni in modo effettivo ed efficiente per eseguire compiti tipici dell'uomo, come il riconoscimento del parlato e la visione, non di riprodurre fedelmente struttura e comportamento dei neuroni



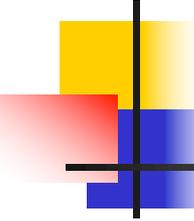
# Reti neuronali e apprendimento

---

Come si può costruire **un sistema che apprende** a partire dal modello di McCulloch e Pitts?

Mettendo in **rete** un enorme numero di neuroni che operano in parallelo: anche se la loro struttura di base è la stessa, comportamenti diversi si possono ottenere associando pesi diversi ai diversi neuroni

In linea di principio, per ogni problema, esiste un opportuno insieme di neuroni, con pesi appropriati, il cui comportamento combinato risolve il problema



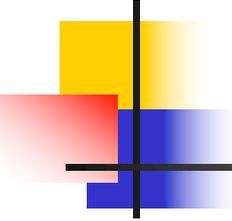
# Il problema dell'apprendimento

---

Il problema è che **non esiste** un **metodo generale** per determinare tali pesi

L'**apprendimento** può essere visto come il tentativo di determinare i pesi giusti, ossia un modo di **programmare le reti neurali artificiali**

Esso è realizzato fornendo un insieme di esempi di soluzione (training set) a partire dai quali il processo di learning (supervisionato) genera dei pesi appropriati per i neuroni coinvolti (pesi che consentano alla rete di calcolare le soluzioni giuste / di minimizzare un'opportuna funzione di errore)



# Le principali tappe - 1

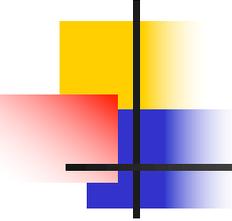
---

Il **Perceptron** (Rosenblatt 1958-1962): un neurone accoppiato con una semplice regola di apprendimento che esegue un progressivo **raffinamento dei pesi**

Introduzione di funzioni d'errore continue e differenziabili (la tecnica della **discesa del gradiente**)

Sostituzione del singolo Perceptron con una rete di Perceptron: **feed-forward neural network**.

L'informazione fluisce in una sola direzione, ma vengono introdotti uno strato di input, uno strato di output e uno o più strati interni nascosti (hidden layer)



# Le principali tappe - 2

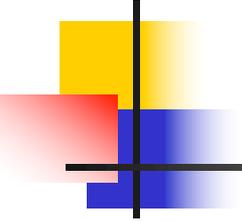
---

L'applicazione della discesa del gradiente alle feed-forward neural network ha portato alla proposta del fondamentale **algoritmo di back propagation**, ancora in uso

**Hopfield network** (anni '80). L'informazione fluisce in entrambe le direzioni: **symmetric recursive neural network**.

**Generative adversarial network (GAN) e transformer**

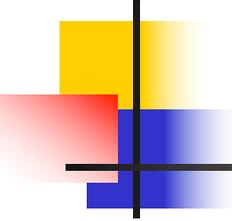
Domini applicativi: elaborazione del linguaggio naturale, sistemi di raccomandazione, bioinformatica, videogame



# Alcune questioni tecniche

---

- Mancanza di dati
- Distribuzione dei dati
- Overfitting
- Interpretabilità (in alcuni casi più importante della performance predittiva)

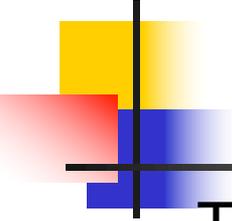


# Alcune questioni filosofiche - 1

---

Le ragioni del machine learning: dal momento che non possiamo fornire ad un sistema un insieme di conoscenze completo e imm modificabile, a partire dal quale ogni altra conoscenza di interesse possa essere derivata mediante opportune procedure, in un unico passo iniziale, il sistema deve essere in grado di **estendere** in modo automatico, ed eventualmente **rivedere**, la propria conoscenza (per mezzo di esempi, esperimenti ed esperienze)

Vari **approcci**: tecniche simboliche di learning, reti neurali, algoritmi evolutivi.



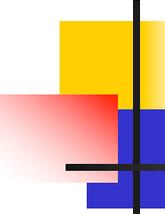
# Alcune questioni filosofiche - 2

---

Tutte le soluzioni sono in qualche misura problematiche a causa di problemi legati alla definizione del **bias induttivo** e alla scelta del **training set**.

**Bias induttivo:** insieme di ipotesi formulate dal learner per prevedere gli output di input futuri (dati non appartenenti al **training set**). Esso è inevitabile e, in una certa misura, necessario per rendere l'algoritmo di apprendimento efficace e di successo

Il **problema dell'induzione:** possono i sistemi indurre regole generali, da usare per future predizioni, sulla base delle regolarità rilevate fino ad un certo punto?



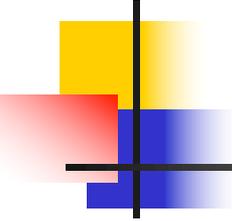
# Un nuovo paradigma scientifico?

---

Una posizione emergente.

Machine learning come un nuovo **paradigma scientifico** guidato dai dati: disponendo di dati sufficienti con adeguata garanzia di significatività statistica, è possibile estrarne automaticamente leggi e modelli di valenza generale, mediante algoritmi di apprendimento e/o analisi statistiche

Una scienza costruita sui dati storici disponibili è una scienza che guarda solo al passato, e quindi intrinsecamente **conservatrice**, in quanto essa è “strutturalmente” costretta a escludere le novità



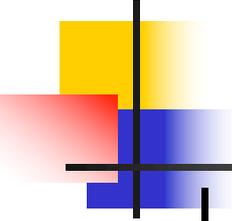
# Passato e futuro

---

E' una scienza che costruisce modelli delle persone e dei sistemi artificiali analizzando il loro comportamento passato e li utilizza per prevedere quello futuro

**Problema:** il comportamento futuro non è necessitato da quello passato. La novità è una caratteristica distintiva del comportamento umano e della storia umana e, a meno che il futuro non sia completamente determinato dal passato, è anche una caratteristica distintiva della natura

Per quanto riguarda le macchine, si può ipotizzare una qualche forma di ripetizione rigida per i sistemi più semplici (e chiusi), ma non è una caratteristica dei sistemi aperti complessi



# Il problema dei dati

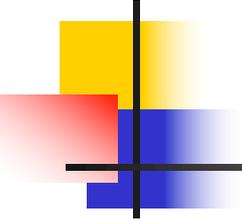
---

I dati non esistono di per sé. I dati dipendono dall'**osservatore**, uomo o macchina, che li raccoglie.

Dipendono anche dai **dispositivi di misurazione**, poiché si possono raccogliere solo dati che possono essere effettivamente misurati/acquisiti.

Per stabilire quali sono i dati rilevanti da raccogliere (se è possibile raccogliarli), sono necessari un obiettivo e un piano per raggiungerlo. Entrambi possono essere formulati solo con riferimento ad un esplicito o modello implicito del mondo (una **teoria**).

Non esistono "dati neutri" e i dati "esistenti" sono sia di fatto che in linea di principio incompleti

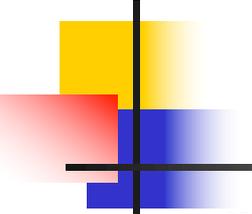


# Alcune considerazioni finali

---

- Creatività vs. **non controllabilità** delle macchine
- **Autonomia** (limitata) delle macchine
- **Intenzionalità** (derivata) delle macchine

IA basata sui modelli e IA guidata dai dati: l'IA guidata dai dati contribuisce all'efficacia dei sistemi; l'IA basata sui modelli aiuta a interpretare quanto ottenuto dall'IA guidata dai dati (la frontiera dell'**explainable AI**)



# Intelligenza e intenzionalità

---

Nella riflessione filosofica (Brentano, Husserl, Carnap), **l'intenzionalità** viene riconosciuta quale elemento distintivo della coscienza (in generale, di ogni fenomeno psichico)

**Intenzionalità:** riferimento interno di un atto o di uno stato mentale a un determinato oggetto, ossia connessione che l'atto o lo stato hanno, in virtù della loro identità, con un certo oggetto, indipendentemente dalla sussistenza di questo eventuale oggetto nella realtà esterna

**Esempio.** Dell'identità di uno stato emotivo di speranza fa parte ciò che è sperato, indipendentemente dal fatto che si realizzi oppure no