



Fondamenti di Intelligenza Artificiale

A. a. 2025/2026

Davide Fazio 07.04.2026

Informazioni di servizio

Orari, Ricevimenti, Testi, E-learning, Modalità d'esame,

Orari

Martedì 15:30 - 17:30 (aula 20)

Mercoledì 15:30 - 17:30 (aula 9)

Giovedì 15:30 - 17:30 (aula 9)

Ricevimenti

Per ricevimenti concordare un appuntamento con il docente scrivendo all'indirizzo

dfazio2@unite.it

Informazioni di servizio

Orari, Ricevimenti, Testi, E-learning, Modalità d'esame,

- Testi di riferimento:

S. Russell, P. Norvig, **Intelligenza Artificiale: un approccio moderno**, Pearson Prentice Hill, vol.1, III ed. (o successive), 2010. (Solo introduzione)

M. Mitchell, **L'intelligenza artificiale: una guida per esseri umani pensanti**, Einaudi.

- Testo di approfondimento

M. Davis, **Il calcolatore universale**, Adelphi, 2012, ISBN: 9788845927416

Esame

- **Esame orale** tradizionale oppure (se i numeri lo consentono) **presentazione** su un tema **non** trattato nel corso ma **affrontabile con gli strumenti concettuali maturati.**

Sinossi del corso

Moduli didattici

- **Modulo 1: Che cos'è l'intelligenza artificiale?**
 - Che cosa è l'Intelligenza artificiale?
 - Brevi cenni storici sull'origine dell'IA;
 - Il concetto di agente intelligente;
 - Ambienti operativi
 - Strutture degli agenti e loro classificazione



Modulo 1

Che cos'è l'IA?

Dal pensare umano all'agire razionale

- 4 possibili definizioni di IA.
 1. L' **IA** è la scienza che si occupa della simulazione del **pensiero** umano (approccio delle **scienze cognitive**).
 2. L' **IA** è la scienza che si occupa della simulazione del **comportamento** umano (approccio di Turing)
 3. L' **IA** è la scienza che si occupa della simulazione del **pensiero** razionale
 4. L' **IA** è la scienza che si occupa della simulazione del **comportamento** razionale.

Che cos'è l'IA?

- L' IA come scienza che si occupa della **simulazione del pensiero umano** si basa sull'idea che questa debba basarsi su e mettere in evidenza i meccanismi di funzionamento dell'intelligenza umana con i suoi pregi e i suoi difetti.
- L' IA come scienza che si occupa della simulazione del comportamento umano è ben rappresentata dal noto **TEST DI TURING**.

Requisiti:

- Capacità di interpretazione del linguaggio naturale;
- Capacità di memorizzazione e rappresentazione della conoscenza;
- Capacità di ragionamento;
- Capacità di apprendimento.

Che cos'è l'IA?

- L' **IA** come scienza che si occupa della **simulazione del pensiero razionale** (approccio logico-matematico) si basa sull'idea per cui si debba provvedere ad un account del pensiero razionale rigoroso attraverso la creazione di sistemi formali complessi. Simulare il pensiero significa dunque **costruire sistemi formali**.
- L' **IA** come scienza che si occupa della **simulazione del comportamento razionale** si basa sull'idea per cui si debba provvedere a simulare (e prototipare) agenti artificiali che svolgano, pensino, decidano la “cosa giusta”.

Che cos'è l'IA?

- Oggi per IA intendiamo la branca dell'Informatica che si occupa della prototipazione di agenti razionali. Il vantaggio fornito dalla nozione di razionalità risiede nel fatto questa sia **facilmente trattabile da un punto di vista matematico**.

Le origini dell'IA

Cenni storici

- Reti neurali (W. McCulloch, W. Pitts, 1943). Costruire un modello matematico del funzionamento del cervello come rete di neuroni che si attivano sulla base di stimoli di una data entità.
- Conferenza di A. Turing per la London Mathematical Society (1947).
- A. Turing, *Computing machinery and intelligence* (1950). -> Test di Turing e idea del “Programma bambino”.
- Workshop al Dartmouth College (1956) - Programma di sviluppo dell'IA.
 - A. Newell e H. Simon (*Logic theorist*) -> le facoltà umane sono di principio simulabili e implementabili in una macchina (*è solo questione di tempo...*) -> General Problem Solver (progettato per imitare il problem solving umano)

Le origini dell'IA

Cenni storici

- John Mc Carthy inventa l'Advice Taker (1958). Idea fondamentale: separare la “base di conoscenza” dalla logica per manipolarla -> Macchine non più vincolate da determinati contenuti.
- **Micromondi:** progettare sistemi capaci di simulare facoltà umane specifiche. Esempio: ANALOGY di T. Evans (1968) e STUDENT di D. Bobrow (1967).

Le origini dell'IA

Cenni storici

- Difficoltà dell'IA:
 1. I primi sistemi di intelligenza artificiale si basavano essenzialmente su manipolazioni sintattiche, non erano in grado di avere a che fare con la semantica.
 2. Problemi di complessità computazionale. I primi sistemi procedevano per tentativi (**metodi deboli**) -> *esplosione combinatoria* -> limiti nella simulazione di fenomeni di apprendimento.

Sistemi basati sulla conoscenza

DENDRAL e MYCIN

- DENDRAL (1969): rappresentare la struttura di una molecola data la sua formula elementare e lo spettro di massa. (basato su una base di conoscenza fissata)

La versione: generare possibili strutture e confrontarle con lo spettro.

La versione: **Includere nel programma** informazioni circa possibili strutture che possono tipicamente evincersi dallo spettro di massa.

- MYCIN (1975): sistemi capaci di effettuare diagnosi su alcune patologie. Il sistema **NON è equipaggiato con una base di conoscenza determinata** ma può essere arricchito con i dati degli esperti e con i dati estrapolati dai casi concreti. -> Il sistema deve essere in grado di gestire l'incertezza.

Agenti razionali

- Sistemi esperti *normativi* (Pearl, 1982) -> i sistemi **NON** devono essere prototipati in modo da ricalcare le facoltà umane -> devono “ragionare” secondo strategie ottimali come definite dalla teoria delle decisioni.
- **IN GENERALE:** l'IA si allontana dal programma di riprodurre l'intelligenza nella sua interezza limitandosi a considerare la riproduzione di singole funzioni.
- **TUTTAVIA:** l'ambito dell'Intelligenza Generale Artificiale intende sopperire a tale mancanza -> algoritmi capaci di implementare *qualsiasi* comportamento razionale.

Cos'è un agente?

Definizione

- Un **agente** è un'entità dotata di **sensori** e di **attuatori** (interagisce con un ambiente)
 - un agente deve essere in grado di percepire.

Percezioni: input percettivi in un dato istante

Sequenza percettiva: la sequenza delle percezioni dell'agente dalla sua nascita (o eventualmente in un intervallo temporale precedente prefissato) ad un dato istante.

- un agente deve essere in grado di agire in base a percezioni. Agisce sulla base delle sue percezioni in base ad un **programma agente** (algoritmo) che implementa una **funzione agente** (oggetto matematico).

Che cos'è un agente razionale?

- Un agente si dice razionale quando “fa la cosa giusta”.
- Fare la cosa giusta significa massimizzare una **misura di prestazione** che quantifica la “bontà” del risultato che l’agente ha ottenuto.
- L’intelligenza/razionalità dell’agente dipenderà dunque da 4 fattori:
 - a) ciò che l’agente percepisce,
 - b) le ulteriori informazioni che ha a disposizione,
 - c) ciò che l’agente **può** fare,
 - d) la misura di prestazione.

Definizione: un agente si dice **razionale** quando, **per ogni data sequenza percettiva, sceglie un’azione che massimizzi la misura di prestazione** sulla base delle proprie percezioni e delle proprie conoscenze pregresse.

Che cosa è un agente razionale?

- **IMPORTANTE:** un agente razionale **non deve essere perfetto**. Deve **massimizzare il risultato atteso** sulla base delle risorse di cui dispone.

TUTTAVIA, deve eventualmente essere in grado di procurarsi le informazioni per svolgere al meglio il proprio compito (*information gathering*).

- Per specificare un agente razionale dobbiamo tener conto di:
 - a) La misura di prestazione (*Performance*);
 - b) L'ambiente in cui opera (*Environment*);
 - c) Gli attuatori (*Actuators*)
 - d) I sensori (*Sensors*)
- a) - d) costituiscono l'**ambiente operativo** dell'agente.

Ambienti operativi

Classificazione

- **Completamente osservabili/Parzialmente osservabili:** Un ambiente si dice completamente osservabile quando l'agente può acquisire in esso tutte le informazioni necessarie per operare al meglio.
- **Agente singolo/Multiagente:** l'ambiente comprende solo l'agente o anche altre entità riconoscibili come agenti. Come si riconosce un altro agente? -> **Entità che si comporta come se dovesse massimizzare una misura di prestazione.** Gli ambienti possono essere **cooperativi** o **competitivi.**
- **Deterministico/Stocastico:** un ambiente è deterministico quando gli effetti delle azioni dell'agente sono **univocamente determinati** dallo stato attuale dell'ambiente e della specifica azione compiuta. Altrimenti sarà:
 - **Non deterministico:** il risultato è costituito da un insieme di possibilità, oppure
 - **Stocastico:** ad ogni esito è associato un valore di probabilità.

Ambienti operativi

Classificazione

- **Episodico/Sequenziale:** L'agente può percepire l'ambiente attraverso pezzi di informazione "atomici" oppure attraverso **sequenze di percezioni** più o meno lunghe.
- **Statico/Dinamico:** l'ambiente può essere stabile oppure può evolversi. L'evoluzione del sistema può essere **discreta** oppure **continua**.
- **Noto/Ignoto:** se l'agente conosce l'esito delle proprie azioni, l'ambiente si dirà **noto**, altrimenti si dirà **ignoto**.

Struttura degli agenti

Programmi agente e loro classificazione

- Un agente è costituito da due componenti:
 - a) Programma agente;
 - b) Architettura (il supporto fisico).
- Gli agenti possono essere classificati sulla base di come **percepiscono** e **reagiscono** agli stimoli esterni.

Struttura degli agenti

- **Agenti reattivi semplici:** un agente si dice reattivo semplice quando agisce soltanto in base alla **percezione corrente**. Le sue azioni sono basate su **regole di condizione/azione (if - then)**-> l'azione è completamente determinata dalla percezione attuale

Problema: se il sistema non è completamente osservabile l'agente può incorrere in loops infiniti.

- **Agenti reattivi basati su modello:** agente che agisce sulla base della storia delle proprie percezioni (preferibile in caso di osservabilità parziale). Si crea una "immagine" del mondo, detta **modello**. L'azione è determinata da due tipi di conoscenza:

a) Informazioni sull'ambiente pregresse

b) Conoscenza degli esiti delle proprie azioni

Struttura degli agenti

- **Agenti basati su obiettivi:** agenti che agiscono non in base ad uno stato di cose percepito più o meno attualmente ma attraverso obiettivi specifici sulla base dei quali pianificano le azioni più convenienti.
- **Agenti basati sull'utilità:** agenti che internalizzano una misura di prestazione attraverso **la misura di utilità**. L'agente agisce per **massimizzare l'utilità attesa** degli esiti delle proprie azioni (dipende dalla utilità del singolo esito e dalla sua probabilità).
- **Agenti basati sull'apprendimento:** un agente basato sull'apprendimento modifica/apprende strategie ottimali d'azione per massimizzare la funzione di prestazione. La loro struttura si basa su quattro elementi:
 - a) **Elemento esecutivo:** si occupa della scelta e dell'implementazione delle azioni dell'agente;
 - b) **Elemento critico:** valuta l'operato dell'elemento esecutivo sulla base di uno **standard di prestazione** prefissato;
 - c) **Elemento di apprendimento:** appronta modifiche al fine di massimizzare la prestazione alla luce della valutazione dell'elemento critico.
 - d) **Generatore di problemi:** elemento la cui funzione è suggerire strategie che portino ad esperienze nuove e significative in base alle quali l'elemento di apprendimento può determinare come modificare l'elemento esecutivo.

Come un agente può “rappresentare” uno stato di cose?

- **Rappresentazione atomica:** l'agente rappresenta lo stato di cose come una totalità compatta (acceso/spento, sporco/pulito, etc...)
- **Rappresentazione fattorizzata:** l'agente rappresenta lo stato di cose come un insieme di parametri con determinati valori.
- **Rappresentazione strutturata:** l'agente si rappresenta la realtà come un insieme di oggetti.



Modulo 2

Intelligenza Artificiale Simbolica e Subsimbolica

Manipolare informazioni richiede “logica”(!)

Ma come entra in gioco la logica e, soprattutto, cos'è?

- La logica è la scienza che studia relazioni tra (insiemi di) enunciati.
- La logica si preoccupa di rispondere a domande quali, ad esempio, “Cosa significa affermare che una conclusione segue da determinate premesse?”, “In che modo dovremmo ragionare su tipi particolari di oggetti?”, “Quali sono le regole valide che ci permettono di dimostrare un teorema?”.
- Bisogna distinguere immediatamente due nozioni di logica.
 - a) Logica deduttiva:** riguarda la deduzione di conclusioni a partire da premesse -> la logica deduttiva estrae informazioni da “bacini” di informazioni.
 - b) Logica induttiva:** ha una funzione **ampliativa**, ovvero ci permette di acquisire **nuove** informazioni.

Logica deduttiva e logica induttiva

Logica deduttiva

A) Se piove, il pavimento è bagnato

B) Piove

Conclusione: Il pavimento è bagnato

Logica induttiva

In un campione arbitrario di 200.000 corvi, ogni esemplare è nero.

Conclusione: Tutti i corvi sono neri

Differenze?

Logica e Monotonia

- La differenza essenziale tra la logica deduttiva e la logica induttiva è che la prima è “**stabile**” rispetto all’aggiunta di nuove premesse (**monotona**).
- La logica induttiva è meno stabile: l’aggiunta di nuove premesse può falsificare la conclusione (**non monotonia**).
- Sebbene l’IA si avvalga sia di logiche monotone che non monotone, durante questo corso ci occuperemo prevalentemente di logiche monotone, in particolare richiameremo quella che oggi è meglio nota come logica classica (**CL**).

Logica Deduttiva: su cosa ragionare?

- Sulla base di come un agente si rappresenta uno stato di cose, (in modo atomico, fattorizzato, strutturato...) sarà opportuno utilizzare un linguaggio piuttosto che un altro. Detto altrimenti, l'uso di un linguaggio è sempre funzione di come vogliamo rappresentare e ragionare su uno stato di cose e di quanto vogliamo essere dettagliati.

Esempio: Se intendiamo rappresentare enunciati quali “La spia è accesa”, “Marco è al mare”; “I mercati sono in ribasso”, e non siamo interessati alla struttura interna di questi enunciati ma solo alle loro condizioni di verità, potremo limitarci a ragionare con la logica **proposizionale**.

Se, invece, vogliamo rappresentare stati di cose in cui rendere esplicite **tutte** le informazioni di cui disponiamo, ad esempio, se volessimo appresentare l'enunciato “Esiste almeno un numero primo che è pari” è necessario considerare un linguaggio **predicativo**.

Cos'è una conseguenza (logica) deduttiva?

- Una relazione di conseguenza **deduttiva** \Vdash in un linguaggio L (o semplicemente **conseguenza logica su L**) è una collezione di inferenze in L con la seguente proprietà:
- (Riflessività) Se X è un insieme di enunciati e $A \in X$, allora $X \vdash A \in \Vdash$
- (Transitività) Se $X \vdash A \in \Vdash$, per ogni $A \in Y$, e $Y \vdash B \in \Vdash$, allora $X \vdash B \in \Vdash$
- (**Monotonia**) Se $X \vdash A \in \Vdash$ e $X \subseteq Y$, allora $Y \vdash A \in \Vdash$.

Precisazioni

- Una conseguenza deduttiva \sqsubset su un linguaggio L specifica un possibile modo di “ragionare” all’interno di L (**cosa posso concludere a partire da cosa**).
- O anche: una conseguenza deduttiva specifica che informazioni posso estrarre da certi insiemi di informazioni.
- Quanti “possibili modi di ragionare” (a.k.a. **conseguenze deduttive**) in un dato linguaggio ci sono? **Infiniti (tanti quanti sono i numeri reali...)**.
- **[Ricorda] Dedurre significa sempre estrarre informazioni. La deduzione estrae informazioni, non ne aggiunge di nuove.**

Induzione

- I ragionamenti che comunemente utilizziamo per ottenere informazioni non estraibili da ciò che creiamo hanno un'altra natura.
- **Ragionamenti induttivi**
- **[Esempio]** “Dal momento che l'85% del **campione della popolazione** scelto presenta la proprietà A, possiamo concludere, **con una certa probabilità**, che **la maggior parte della popolazione** ha la proprietà A.
- Caratteristiche:
 - I ragionamenti induttivi sono (e devono!) essere **probabilistici**.
 - I ragionamenti induttivi **non sono mai conclusivi** (non vale la monotonia)

Abduzione

- Inference to the best explanation. Mentre nel ragionamento deduttivo estraiamo informazioni da un set di premesse, nel ragionamento abduttivo cerchiamo una serie di premesse che supportino la conclusione.
- **[Esempio]** “Ho tre sacchi di fagioli. Il sacco A contiene fagioli bianchi, il sacco B fagioli neri, il sacco C fagioli rossi. questi fagioli sono rossi, Quindi, questi fagioli sono stati presi dal sacco C.
- In un ragionamento abduttivo passo dalla conclusione alle premesse. Uso un ragionamento per motivare qualcosa che considero vera (ha a che fare con la nozione di **spiegazione**).

Tornando al Dartmouth College (1956)

- **Idea fondamentale:** l'Intelligenza consta essenzialmente della manipolazione di simboli. Comprendere le leggi che permettono di manipolare i simboli è sufficiente per modellare i meccanismi di funzionamento dell'intelligenza **autentica** degli esseri umani.
- L'intelligenza una "parola valigia" (M. Minsky), ovvero è un contenitore con una miriade di significati determinati dal contesto e che non è semplice separare.
- Cos'è l'intelligenza umana? Può essere unidimensionale o multidimensionale, binaria (qualcosa è o non è intelligente) oppure distribuirsi su un continuo (graduata).
- Già al Dartmouth College, i ricercatori si schierarono a favore di vari metodi. Alcuni sostenevano l'idea di ragionare su sistemi formali mediante un approccio logico-matematico. Altri sostenevano la necessità di fare appello alla biologia e alla psicologia per comprendere le dinamiche del pensiero umano -> aspetti **normativi** e **descrittivi** dell'IA.

IA simbolica e subsimbolica

- Oggi il paradigma dominante è dato dal **deep learning** e dalle **deep neural networks**, un ventaglio di metodi basati sul **machine learning**.
- Tuttavia, i metodi logico matematici continuano ad essere utilizzati in ambiti in cui il machine learning non ha ancora ottenuto i risultati sperati.
- In generale, dobbiamo distinguere tra IA **simbolica** e **subsimbolica**.
- Secondo l'IA simbolica, la conoscenza può essere formalizzata attraverso simboli (frasi e parole).
- Esempio: **General Problem Solver**

IA simbolica e subsimbolica

- Il **GPS** si basa essenzialmente su regole di manipolazione di simboli. Descrivo l'input e il problema da risolvere (ovvero l'output desiderato) e modifico l'input attraverso regole che codificano le "mosse" che fare e i vincoli cui sono soggetto per avvicinarmi sempre di più all'output. **Esempio:**

Input:

Riva destra: [3 missionari, 3 cannibali, 1 barca];

Riva sinistra: [vuota]

Stato desiderato:

Riva destra: [vuota];

Riva sinistra: [3 missionari, 3 cannibali, 1 barca]

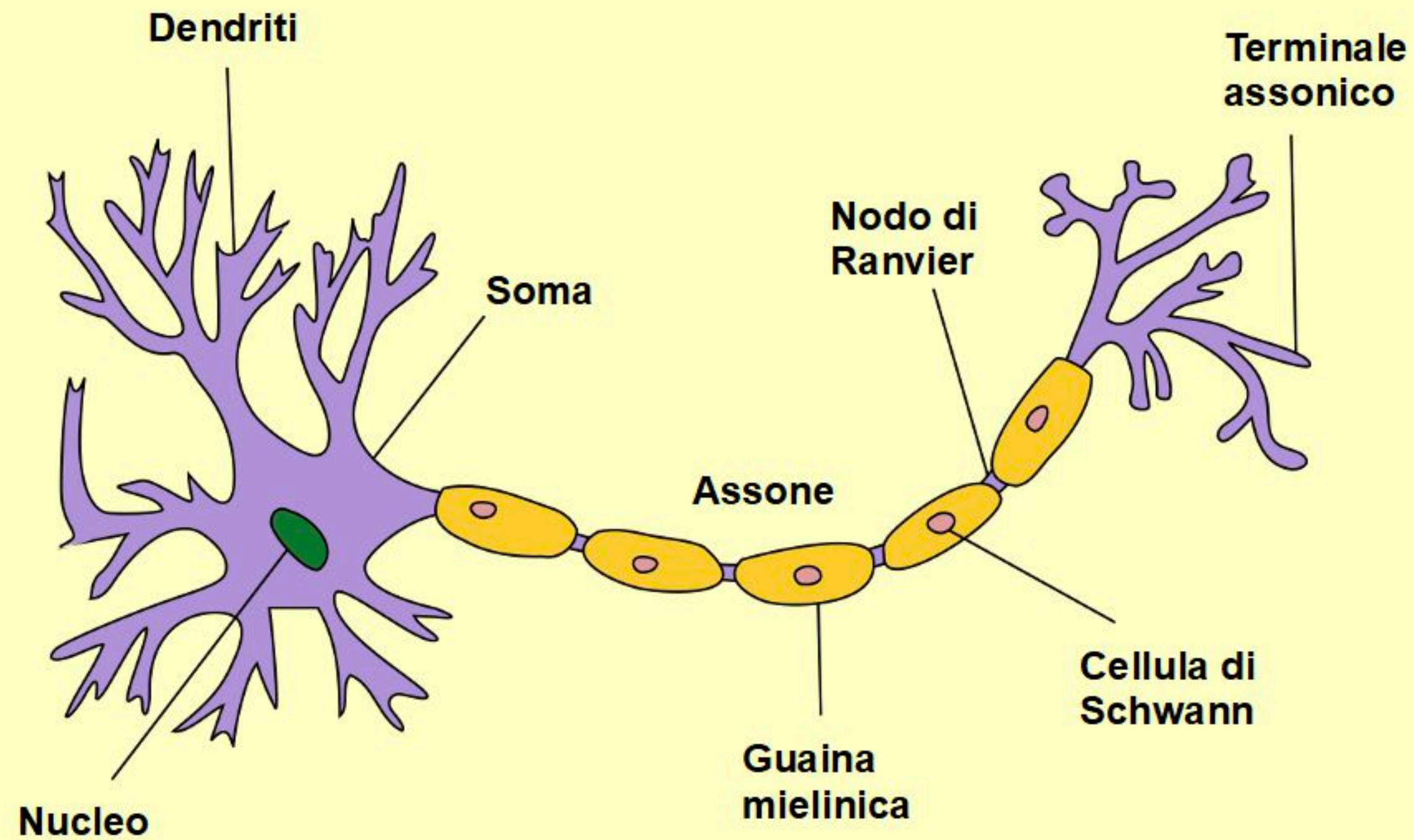
Regola di manipolazione: SPOSTA[#missionari,#cannibali, da #riva a #riva]

IA simbolica e subsimbolica

- **Fondamentale:** Il significato dei simboli è completamente determinato dalle regole per manipolarli.
- L'IA subsimbolica si basa sui risultati delle neuroscienze. -> I processi mentali non sono del tutto consapevoli. I processi mentali si basano solo superficialmente su manipolazioni di concetti esprimibili attraverso unità sintattiche ma si devono a processi che non sono codificabili attraverso regole sintattiche.
- I primo esempio di programma subsimbolico è il **percettrone** di Frank Rosenblatt. Il percettrone il progenitore delle deep neural networks.

Neuroni

Schema della struttura di un Neurone



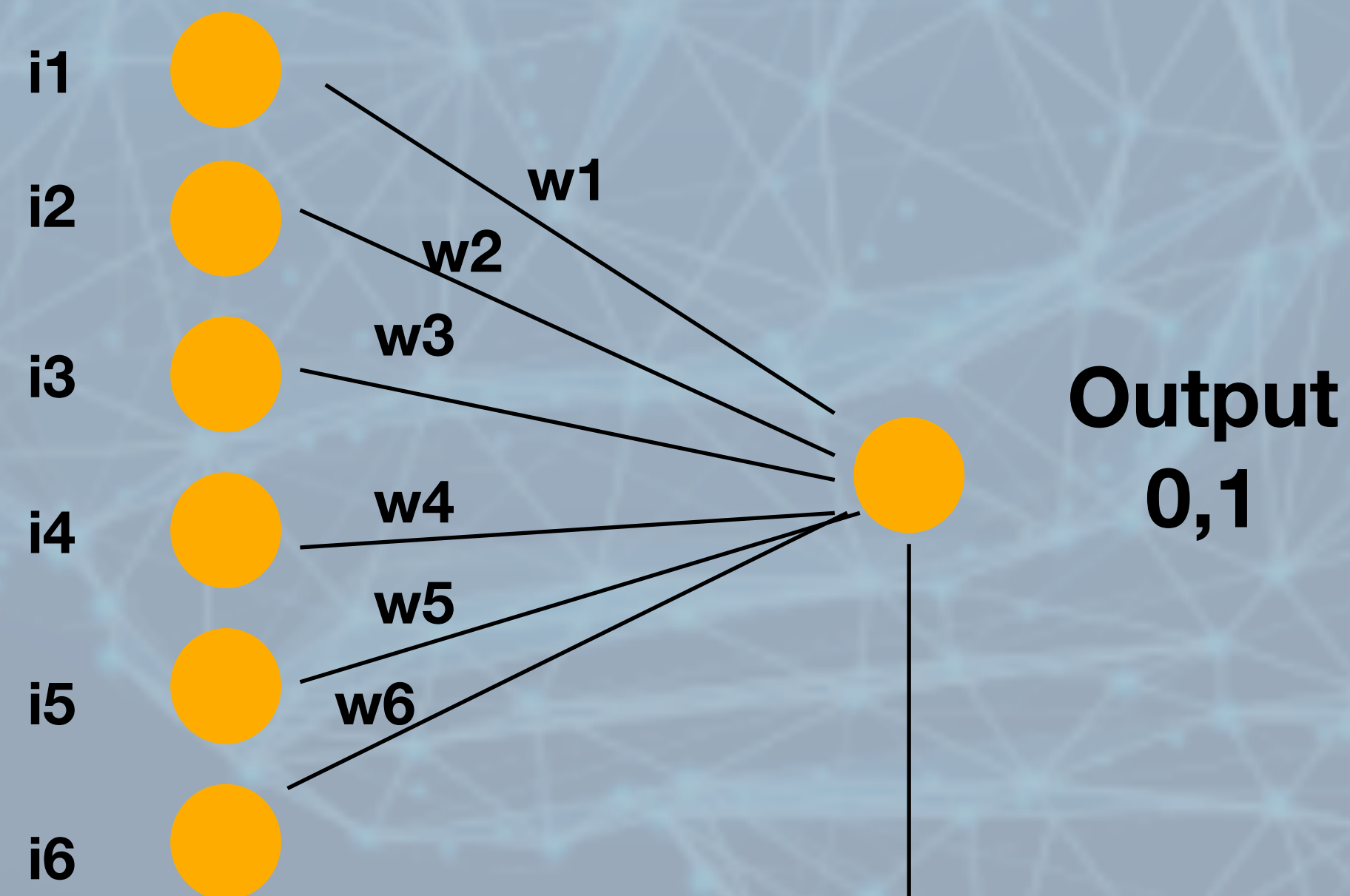
Neuroni



- Un neurone riceve in input stimoli elettrici o chimici da altri neuroni attraverso i **dendriti**, se la somma degli stimoli supera una certa **soglia**, il neurone scarica un impulso tramite l'**assone**.
- Il valore dell'impulso in entrata proveniente dagli altri neuroni tende a variare sulla base della forza delle **connessioni sinaptiche** (le connessioni ad alcuni neuroni saranno più "pesanti" rispetto ad altre).
- La forza delle connessioni sinaptiche è un fattore fondamentale dell'apprendimento

I percettrone

Input numerici



$$h = \sum_{1 \leq j \leq 6} (i_j \cdot w_j)$$

Se $h \geq n$ allora l'output è 1,
altrimenti 0

BIAS (o soglia)
 n
(viene stabilito
dal
programmatore)

Applicazioni dei perceptron

Riconoscimento di immagini

- Rosenblatt aveva ipotizzato la possibilità che i perceptron potessero svolgere compiti di riconoscimento di cifre scritte a mano.
- Possiamo concepire l'area occupata da una cifra come una griglia di pixels ognuno con un determinato valore di intensità. Ogni cella della griglia può essere visto come un nodo di input.
- Rosenblatt dimostra che data una specificazione adeguata di pesi e soglia, il neurone può riconoscere cifre scritte a mano.
- Cosa determina i pesi e la soglia? Il perceptrone dovrebbe impararli **autonomamente**. Come? **Attraverso il condizionamento.**

Applicazioni dei perceptron

Riconoscimento di immagini

- Un perceptrone deve essere addestrato in base ad esempi. Deve essere premiato quando scarica in modo corretto e punito quando sbaglia (**supervised learning**).
- Il perceptrone va allenato attraverso un **training set**, ovvero un insieme di input. In base alle risposte che fornisce, il perceptrone riceve un feedback di supervisione che gli comunica lo scostamento dell'output fornito da quello reale. Il feedback viene utilizzato per aggiornare i pesi e la soglia. -> **N.B. per essere addestrato sono necessari un gran numero di esempi positivi e negativi**
- Vi è un secondo insieme di input detto **insieme di valutazione (test set)** sulla base del quale si valutano i “progressi” del perceptrone, ovvero in che misura ha imparato a svolgere il compito assegnatogli.

Algoritmo di apprendimento del perceptrone

- Allo stato iniziale, il perceptrone ha una distribuzione **random** di pesi compresi tra -1 e 1, ed ha una data soglia (la soglia è introdotta con un nodo di input con valore costante di 1 avente come peso *-soglia*)
- Gli viene somministrato il primo esempio de training set. Il perceptrone fa la somma pesata degli input e restituisce, dopo il confronto con la soglia, il valore 0 oppure 1. In altri termini, il perceptrone scarica se la somma pesata degli input è maggiore o uguale a 0 (**questo significa che la somma pesata degli input supera la soglia**).
- L'output viene confrontato con il **valore reale (fornito dall'uomo)**
- Se il perceptrone risponde in modo corretto, non vi è alcuna variazione.
- Se il perceptrone sbaglia, i pesi e la soglia vengono modificati.

Algoritmo di apprendimento del perceptrone

- Come si aggiornano i pesi?

$$w_j := w_j \leftarrow w_j + \eta(t - y)x_j$$

PESO originario

Tasso di
apprendimento

output effettivo

output fornito dal perceptrone

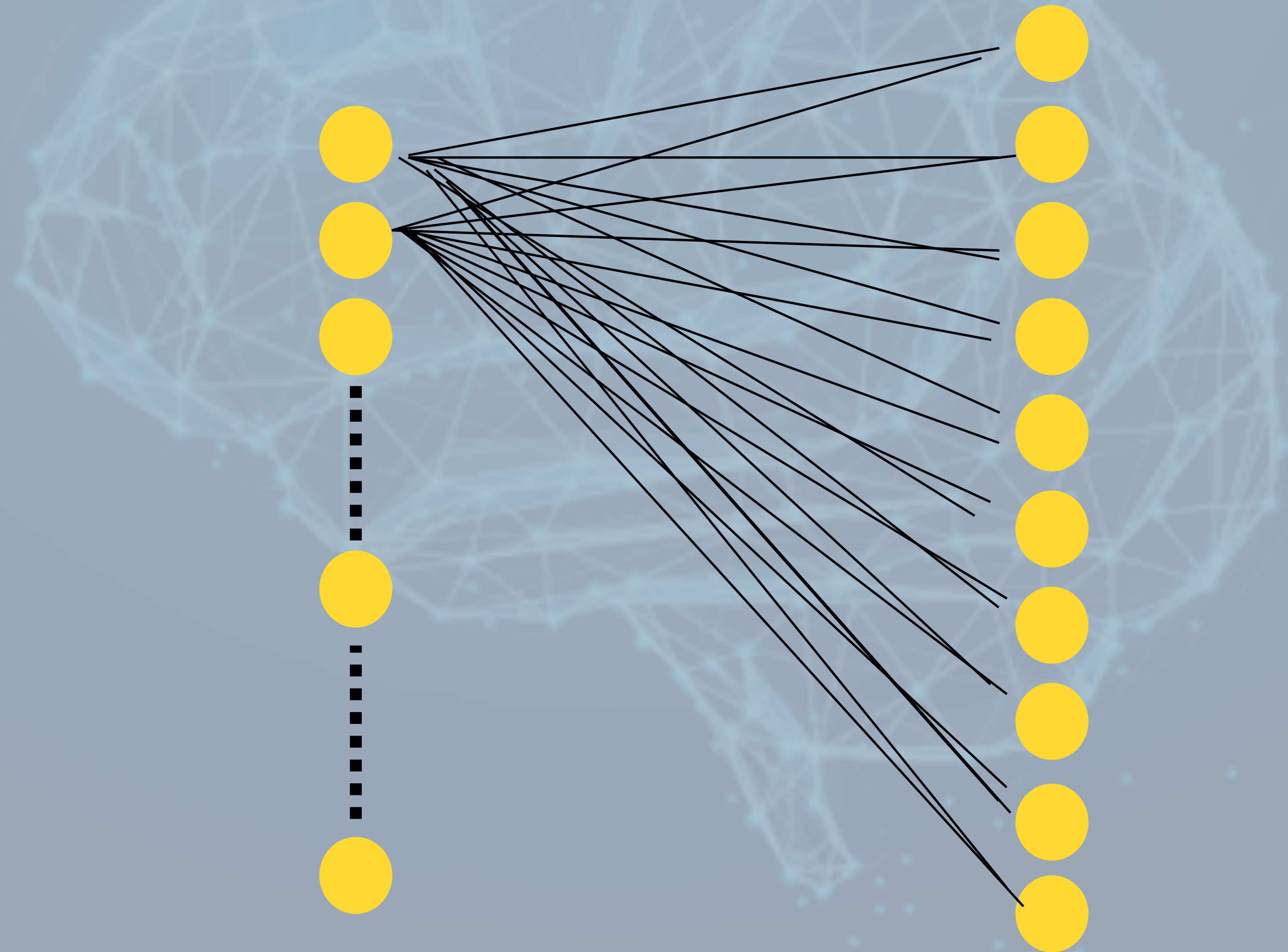
input associato a peso w_j

La soglia si aggiorna come un qualsiasi altro peso.

Ancora sui percettroni

- **N.B.** La variazione dei pesi non deve essere eccessiva altrimenti potrebbero occorrere distorsioni.
- Possiamo concludere che il percettrone “ha imparato” quando i suoi pesi e la sua soglia si assestano su valori per cui gli output sull’insieme di addestramento sono corretti.
- Un percettrone come quello descritto può riconoscere e.g. una sola cifra. Tuttavia si può immaginare un percettrone con dieci output ognuno dei quali corrispondente ad un certo numero compreso tra 0 e 9.

Outputs multipli



Perceptrons (Seymour Papert, Marvin Minsky, 1969)

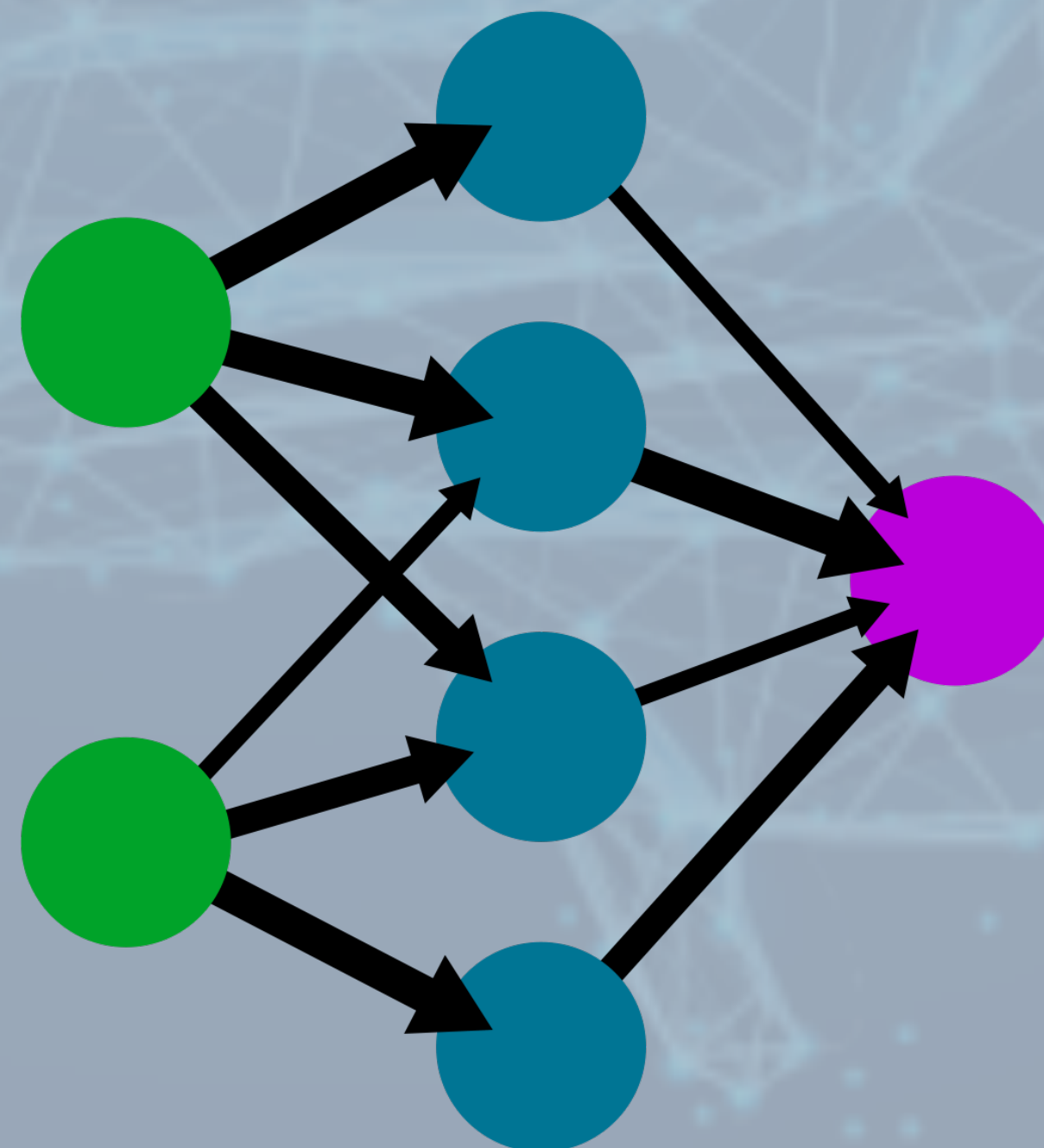
- *Perceptrons* dimostrò che i perceptroni avevano un potenziale limitato e che, aggiungendo uno strato intermedio si sarebbero potuti svolgere compiti molto più ampi. Tuttavia, non si era al corrente di estensioni dei risultati di Rosenblatt alle reti multistrato.
- L'indagine sull'IA subsimbolica fu abbandonato in favore di sistemi più "canonici", ovvero i sistemi esperti e le loro applicazioni.
- Il "fallimento" del programma di Rosenblatt produsse un nuovo inverno dell'IA a seguito della riduzione di fondi.

Reti Neurali Profonde

- Come abbiamo potuto notare, i percettroni constano di 2 soli strati. Tuttavia non è difficile immaginare reti composte da un numero superiore di strati

A simple neural network

input layer hidden layer output layer

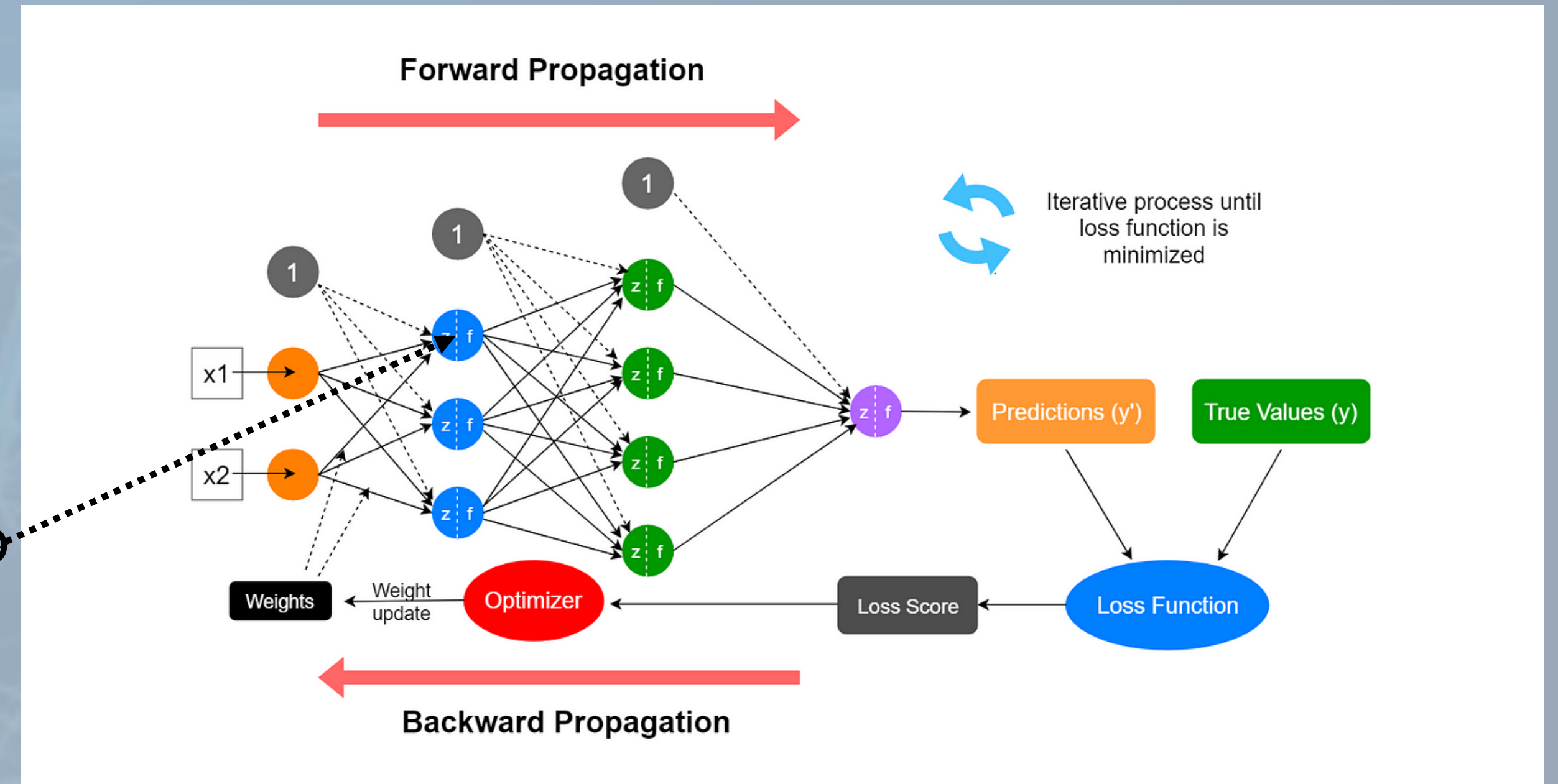


Reti neurali profonde

- Una rete neurale si dice **profonda** se ha almeno **quattro** strati (ovvero **almeno due strati nascosti**)
- Come per i percettroni le reti neurali profonde (o multistrato) sono composte di nodi e di connessioni pesate. Come nel caso dei percettroni, ogni nodo riceve valori pesati in input e li processa attraverso la loro somma pesata.
- **TUTTAVIA:** ora i nodi non producono più soltanto 0 oppure 1 ma producono un valore attraverso una **funzione di attivazione** applicata alla somma pesata degli input. La funzione di attivazione produce sempre un valore tra 0 e 1.
- I valori di attivazione divengono gli input per i neuroni cui il neurone è collegato.
- I valori di output misurano il “grado di confidenza” della rete circa un determinato risultato.

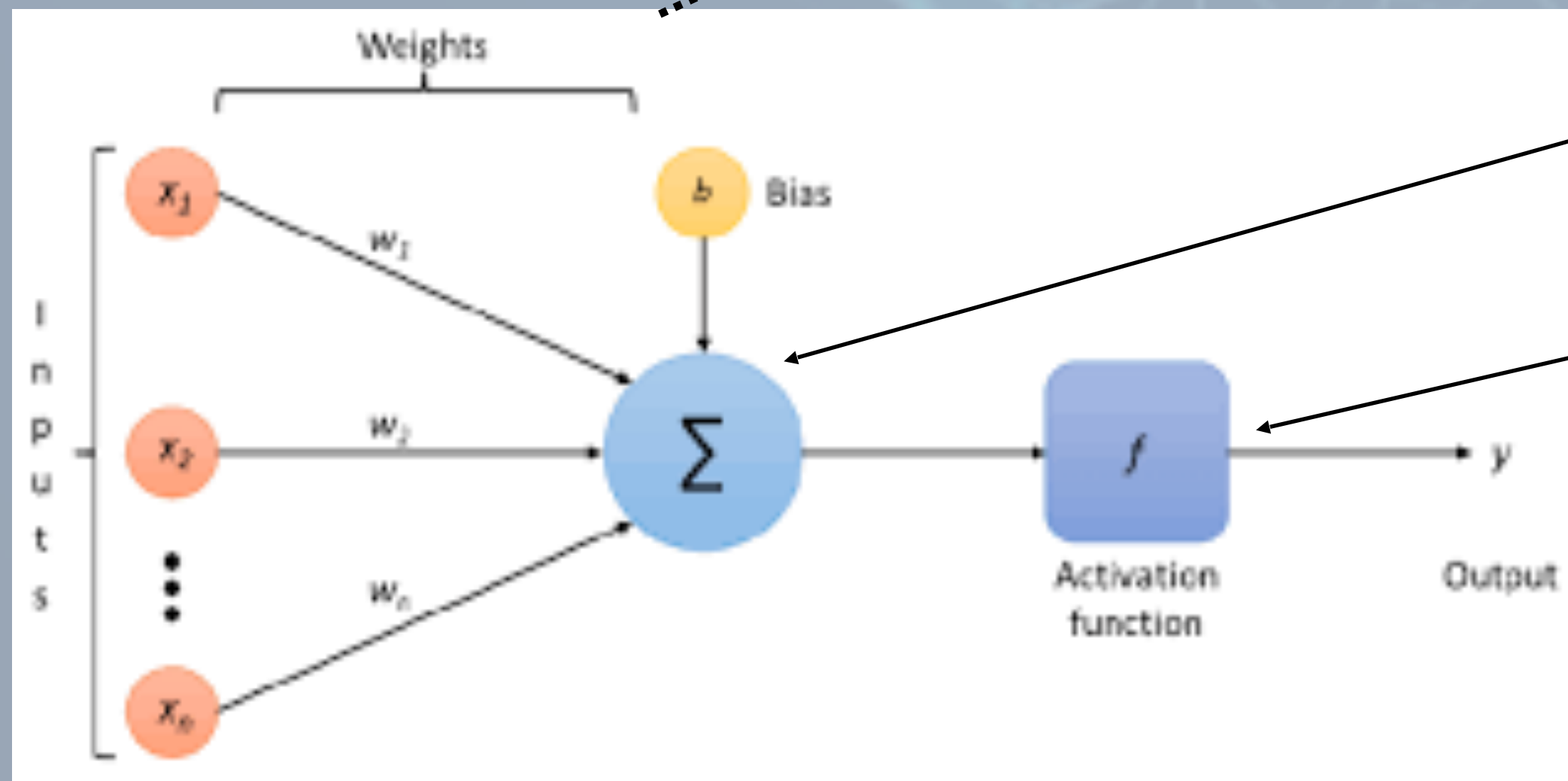
Come funziona una rete neurale?

- Una rete neurale ha l'obiettivo di calcolare una funzione. La funzione viene stabilita dopo cicli (**epoche**) di **allenamento**.



Valore di pre-attivazione

Valore di post-attivazione



Retropropagazione

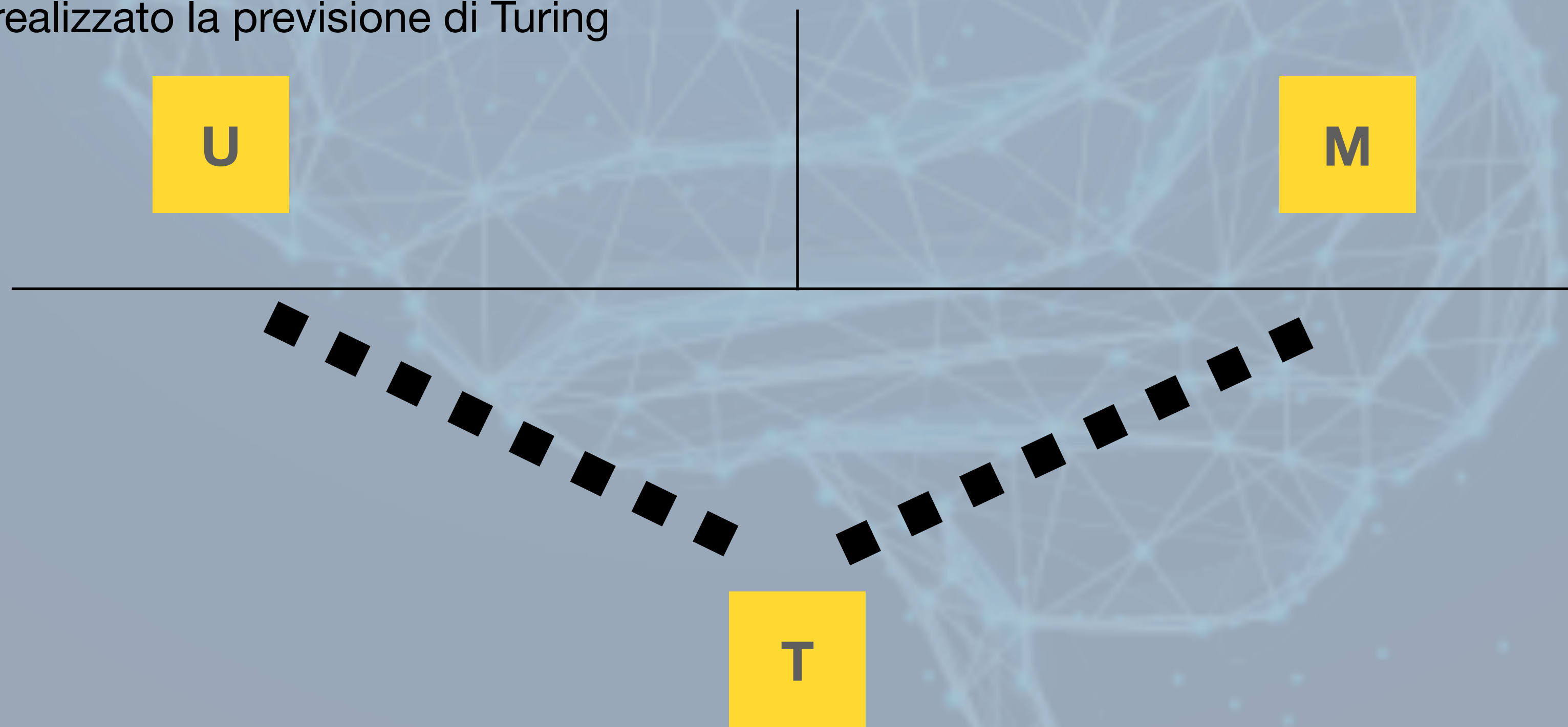
- Come per il perceptrone, una rete neurale profonda viene allenata attraverso meccanismi di *rewarding* e *punishment*.
- Considerato un dataset etichettato **D**, la rete prende i valori di input in **D** e li processa di strato in strato fino ai nodi di output. Una volta ottenuto il risultato, la rete lo confronta con il valore **reale** e, nel caso in cui questo è corretto, non fa nulla, altrimenti implementa una **funzione di perdita** che ha la funzione di misurare la “**gravità**” dell’errore.
- La responsabilità dell’errore è attribuita a ritroso attraverso l’**algoritmo di backpropagation** che ha la funzione di determinare di quanto modificare ciascun peso nella rete per ridurre l’errore.
- Una volta modificati i pesi, la procedura viene reiterata fino a quando non si ottiene un risultato accettabile (**ovvero un valore della funzione di perdita prossimo allo 0**) per ogni esempio nel dataset.

IA ristretta, generale, debole, forte

- L'IA **ristretta (o debole)** si riferisce ad un insieme di sistemi che svolgono **singoli** compiti che richiedono intelligenza;
- L'IA **forte** si riferisce a un insieme di sistemi che simulano l'intelligenza umana (Artificial General Intelligence).
- Ma come è possibile implementare un sistema di IA forte?
- Serve integrazione delle parti (i.e. delle varie facoltà cognitive)
- Serve **autocoscienza (da intendersi non solo come consapevolezza di sé, ma anche come capacità di provare emozioni, volizioni, avere aspirazioni, etc..)**
- È possibile creare una macchina autocosciente?

Il test di Turing

- Il test di Turing fornisce un criterio di attribuzione di intelligenza di natura “funzionalista” (o fenomenologico)-> un agente è intelligente se viene riconosciuto come tale da un tester nel “gioco dell’imitazione”.
- Previsione di Turing: entro 50 anni un tester umano non avrà più del 70% di possibilità di individuare l’umano dopo 5 minuti di domande.
- 2014: In una dimostrazione indetta dalla Royal Society , il ChatBot **Eugene Goostman** creato da programmatori russi e ucraini ha realizzato la previsione di Turing



L'ottimismo dei fautori dell'IA

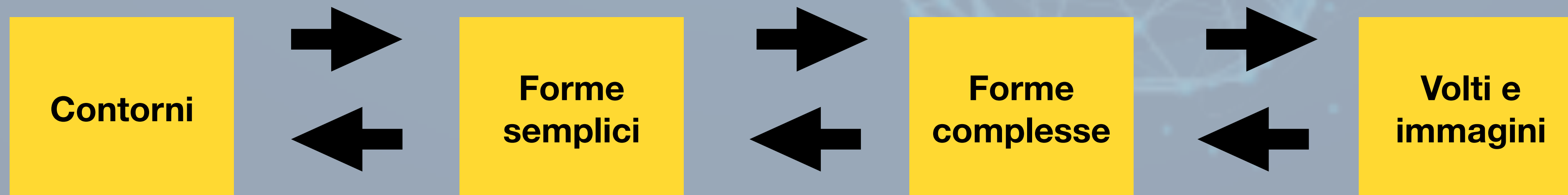
- Ray Kurzweil: verrà un tempo in cui saranno disponibili macchine di intelligenza superiore a quella umana. Sarà un evento con conseguenze singolari -> **Singolarità**
- La motivazione di Kurzweil riposa sulla constatazione della crescita esponenziale delle capacità dei moderni computers (legge di Moore—> il numero di componenti nel chip di un computer raddoppia ogni 2 anni —> **sviluppo tecnologico esponenziale**)
- Quando avremo macchine sufficientemente potenti, la creazione di una intelligenza “sovraumana” sarà solo un fatto di retroingegneria (bisognerà solo aspettare di aver compreso a sufficienza il funzionamento del cervello)
- Sarà sufficiente addestrare le macchine fornendo loro le competenze minime per apprendere da sole.
- Tuttavia, il fatto l'intelligenza artificiale non sia ancora **embodied** (in un corpo) preclude la possibilità di permettere alle macchine di fare alcune delle esperienze fortemente caratterizzanti per l'esperienza umana.

Computer Vision

- Come è possibile insegnare ad una macchina a vedere (guardare?) e riconoscere gli oggetti? (M. Minsky, S. Papert -> Summer Vision Project, 1966).
- Classificare un'immagine implica categorizzare i pixel che la costituiscono.
- Come posso riconoscere i pixel di un cane e distinguerli da quelli di un gatto?
- Come posso riconoscere un cane a prescindere dalla razza, dalla sua posizione, dalla qualità dell'immagine, dall'illuminazione, dal fatto una sua parte sia coperta o meno etc.?

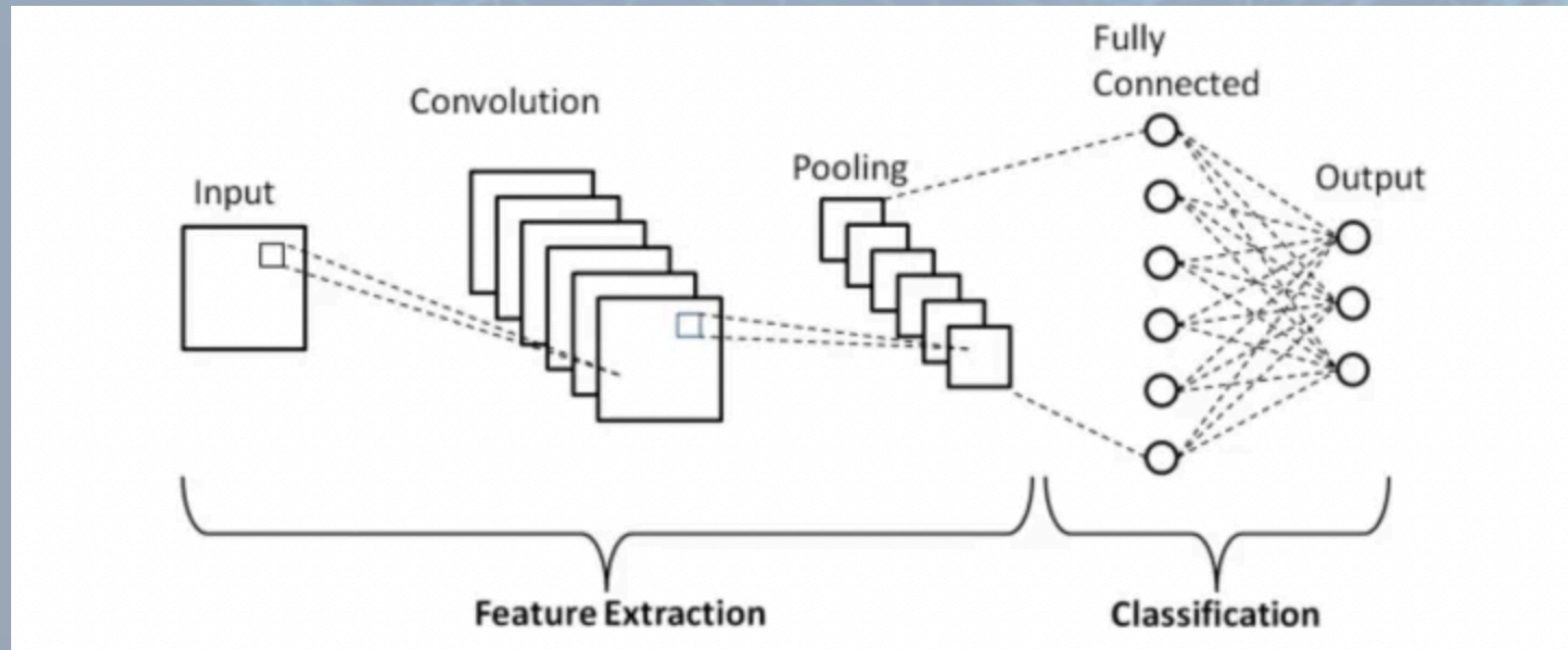
Computer Vision

- David Hubel, Torsten Weisel (1959) -> Studi sulla corteccia visiva
- La luce riflessa dagli oggetti e dalle superfici colpisce la retina, posta in fondo all'occhio. La retina può essere vista come una griglia di neuroni, ogni area della retina è deputata alla "scansione" di una porzione del campo visivo.
- Tramite il nervo ottico, le informazioni catturate dalla retina passano alla corteccia visiva.
- La corteccia visiva è suddivisa in "strati". Ogni strato è deputato a catturare proprietà delle immagini di livello sempre più alto.



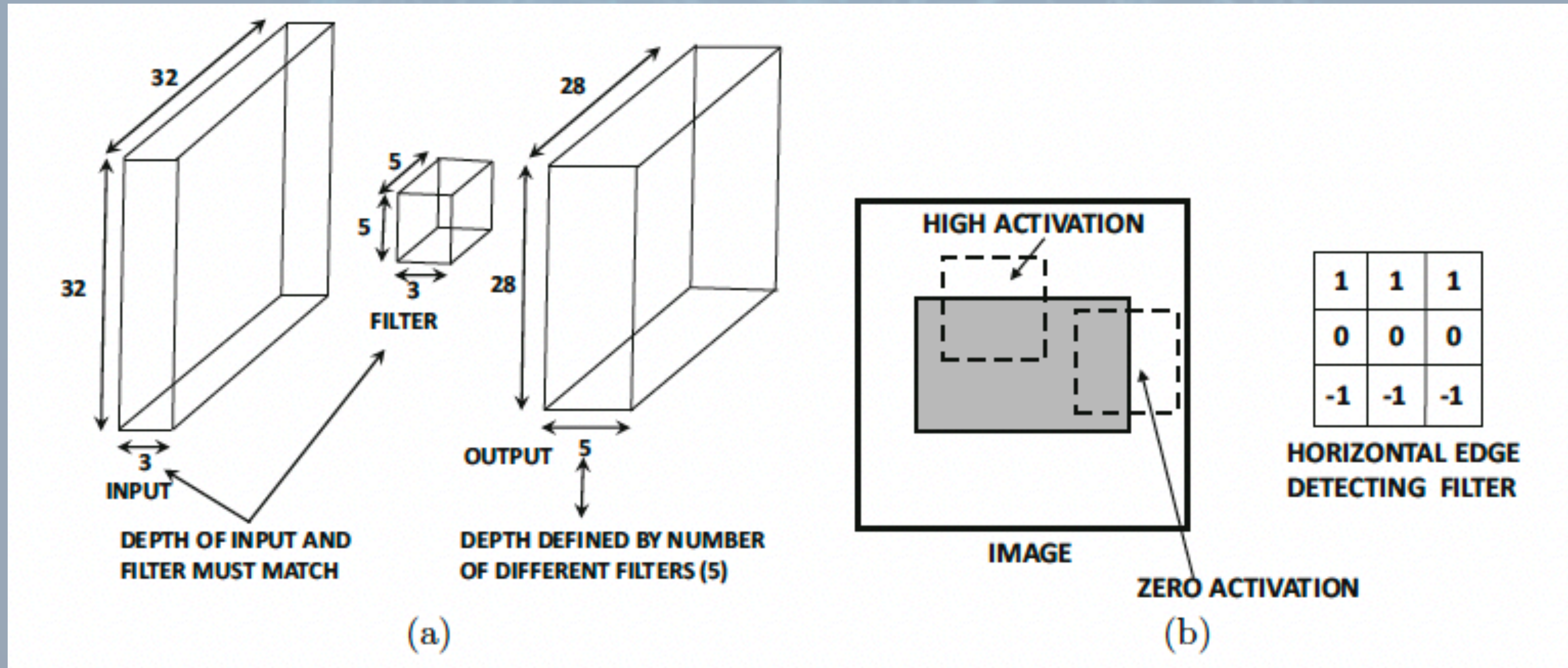
Dal Neocognitron alle CNN

- I lavori di Hubel e Weisel hanno ispirato il *Cognitron* e il *Neocognitron* di Kunihiro Fukushima.
- Il Neocognitron ha ispirato Yan LeCun che ha ideato le reti neurali convolutive.
- Come funziona una rete neurale convolutiva?



Convolutional Neural Networks

-



Le CNN

- Una rete neurale convoluzionale ha, in genere, una funzione di classificazione di un input rappresentabile come una matrice (e.g. un'immagine -> matrice dei valori dei pixel).
- Output: grado di confidenza nel fatto l'input appartenga ad una classe oppure ad un'altra.
- **Obiettivo:** Insegnare alla macchina a rilevare i tratti salienti degli oggetti che ne determinano l'appartenenza ad una classe oppure ad un'altra.
- Una CNN è divisa in **strati**. Ogni strato è diviso in **mappe di attivazione**. Ogni mappa di attivazione può essere vista come una griglia di neuroni con la funzione di “mappare” proprietà specifiche dell'input. Ogni neurone si attiva “osservando” una piccola “zona” dell'input detto **campo recettivo**.
- Come i neuroni di una mappa di attivazione calcolano il valore di attivazione?
- Ogni neurone è dotato di una griglia di pesi (detta **filtro**) della stessa dimensione del suo campo visivo.
- La distribuzione dei pesi determina cosa considerare più “rilevante” del campo visivo.
- Il valore di attivazione si ottiene moltiplicando ogni pixel per il peso corrispondente e poi sommando i risultati. Questa operazione si chiama **convoluzione**.

Le CNN

- **Importante:** Al termine delle operazioni di convoluzione, la mappa di attivazione restituisce un aspetto (tratto) dell'input filtrato in base ad un determinato criterio.
- Gli strati sono organizzati in modo **gerarchico**. Gli input dello strato 2 sono le mappe di attivazione dello strato 1.
- Ogni strato ha l'obiettivo di individuare aspetti sempre più complessi (e astratti) dell'immagine.
- Gli strati costituiti dalle mappe di attivazione sono detti **strati convolutivi**
- La parte terminale di una CNN è un modulo di classificazione che consta di una semplice rete neurale feed-forward che ha l'obiettivo di trasformare il risultato della processazione dell'input in un grado di confidenza nell'appartenenza dell'input ad una determinata classe.

Come allenare una CNN?

- Per verificare i progressi nell'ambito delle ricerche in AI, vengono organizzate periodiche “sfide” tra programmi per il riconoscimento di oggetti.
- 2005-2010: **Pascal Visual Object Classes**. 15.000 foto, 20 categorie. A tutti i concorrenti viene somministrato un training set di foto etichettate. L'obiettivo è, data una categoria, dire se l'immagine contiene un elemento della categoria. i programmi vengono testati su un insieme di valutazione.
- Problema: le foto sono **troppo** poche; le categorie sono troppo poche (quante possibili categorie di oggetti vi sono?)

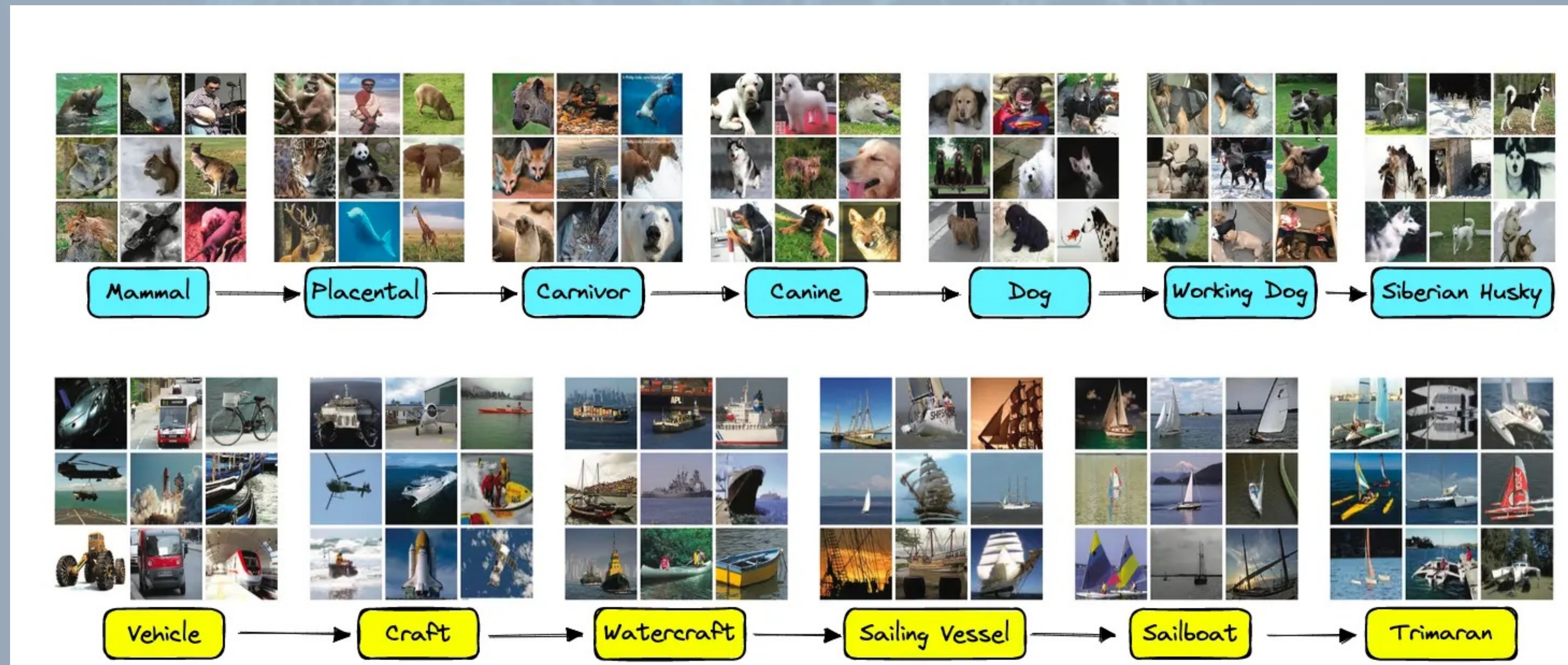
Come allenare una CNN?

ImageNet

- Fei Fei Li (Princeton): è necessario costruire un database contenente un insieme molto grande di foto (nell'ordine di milioni) etichettate (come fare??)
- **Il dataset viene STRUTTURATO** sulla base del modello WordNet, un sistema di classificazione dei sostantivi di ampio uso nell'ambito della linguistica.
- Per l'etichettatura vengono "arruolate" migliaia di persone attraverso Amazon Mechanical Turk. Le immagini vengono reperite in rete.

ImageNet

- Ad ogni sostantivo viene associato un insieme di immagini. I sostantivi sono ordinati “ontologicamente” dal particolare al generale:



ImageNet Large Scale Visual Recognition Challenge

- Competizione basata su ImageNet: 1.200.000 [c.ca](#) immagini; 1000 possibili categorie scelte da WordNet.
- Un programma dà una risposta esatta se e soltanto se attribuisce ad un'immagine **5** categorie una delle quali contiene effettivamente l'immagine (**misura di accuratezza top 5**)
- 2012: la prima CNN vince la ILSVRC (ImageNet Large Scale Visual Recognition Challenge) con una percentuale di risposte esatte pari all'85% (AlexNet).
- 2017: accuratezza **top 5** del 98%.
- **NB: l'accuratezza top 5 non ci dà una misura precisa di quanto la macchina sappia riconoscere gli oggetti.**

Alcune domande

- **Domande:**

- **Una CNN può davvero riconoscere un oggetto?**
- **Quando una CNN fornisce una risposta esatta, sta davvero riconoscendo l'oggetto nell'immagine (Esempio: se in tutte le figure in cui occorrono cani vi è un guinzaglio, siamo sicuri non stia riconoscendo il guinzaglio)?**
- **Nel 2017 la misura di accuratezza top 1 era all'82%, possiamo prendere come parametro di valutazione delle capacità di una macchina la misura top 5?**
- **Possiamo considerare le macchine più abili dell'uomo nel riconoscimento di oggetti?**

Alcune domande

- - Nel valutare le prestazioni di una macchina, dovremmo forse prendere in considerazione la natura degli errori che questa commette? Gli errori umani sono dovuti perlopiù ad ignoranza (per esempio, non riuscire a riconoscere una razza canina) ed alla presenza di più oggetti. Gli errori delle macchine sono più “grossolani”.
- Esiste una connessione tra categorizzazione di oggetti e immagini (si pensi ad immagini in cui si manifestano sentimenti) ed esperienza umana?

Le CNN imparano autonomamente?

- Una CNN apprende come classificare un input come appartenente ad un dato insieme di categorie dopo un lungo addestramento (richiede *big data*).
- Un bambino impara a classificare oggetti come appartenenti ad un insieme molto più grande di categorie dopo averne visti pochi esemplari.
- Possiamo dire che le reti neurali convolutive apprendono da sole? **Non proprio.**
 - Ogni CNN va addestrata sulla base di un training set dato dal programmatore (*supervised learning*).
 - Ogni CNN va impostata rispetto ai propri **iperparametri** (numero degli strati, grandezza del campo recettivo, *tasso di apprendimento*).
- Il settaggio degli iperparametri è detto **tuning** degli iperparametri.

Problemi di apprendimento

- L'apprendimento supervisionato, per essere efficace, richiede una lunga serie di esempi. Ma è possibile catturare **tutti** gli esempi necessari per fare in modo che una CNN impari a classificare tutte le possibili immagini che le si presentano?
- *Long tail problem*: gli eventi “rari” possono manifestarsi ma di essi vi sono pochissimi esempi. Come posso insegnare ad una CNN a riconoscerli?



- Cosa dire degli esseri umani?

Overfitting



- **Fatto:** è possibile che una rete neurale si comporti benissimo rispetto al proprio insieme di addestramento ma di fatto non riesca a generalizzare quanto ha imparato a casi che non hanno esattamente le caratteristiche di quelli presenti nell'insieme di addestramento.
- Questo fenomeno si chiama **overfitting**.
- **IN GENERALE:** Una rete neurale come quelle che abbiamo visto apprendono “correlazioni statistiche”. Può capitare che si adattino “troppo” ai dati del training set.

Bias

- È possibile che un sistema di IA sia portatore di *bias* ovvero di pregiudizi. Per esempio, alcuni sistemi tenderanno a classificare come “essere umano in carriera” un uomo bianco e non una donna o un afroamericano.
- Chi è il colpevole?
- Un sistema di IA ad apprendimento **supervisionato** tenderà ad attribuire etichette sulla base del training set con cui è stato allenato.
- Il bias è **contenuto nel training set e non** nella macchina!
- Spesso i pregiudizi di cui soffre un sistema di AI sono gli stessi pregiudizi di cui soffriamo noi umani.

Explanability

- La vera novità dei sistemi basati di IA basati sul machine learning è che riescono a risolvere problemi che potenzialmente noi (il programmatore!) non sapremmo risolvere.
- Come possiamo essere sicuri che la procedura che una rete neurale apprende sia corretta?
- Per esserne sicuri dovremmo essere in grado di tracciare i passi attraverso i quali la rete neurale, una volta ben allenata, giunge ad una conclusione.
- **Tuttavia**, i passi fatti da una rete neurale (possibilmente) constano di miliardi di operazioni aritmetiche. **Come posso riconoscere tra queste operazioni aritmetiche quelle che corrispondono ad un “passo di ragionamento”?**
- **Explainable AI:** Progettare sistemi capaci di “estrarre” i “passi di ragionamento” (ovvero le procedure di decisione) da un sistema di AI.

Adversarial learning

- Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, Rob Fergus, *Intriguing properties of neural networks*, 2013
- **Le reti neurali possono essere ingannate attraverso una lieve perturbazione dell'input (e.g. alterando in modo impercettibile i pixel di un'immagine).**
- I risultati di queste perturbazioni sono chiamati **esempi avversari**.
- L'adversarial learning è lo studio di come produrre “attacchi” ad una rete neurale e difendere la rete da attacchi.
- Alla luce dell'esistenza di esempi avversari, possiamo concludere che una rete neurale apprenda come un essere umano? O ancora, cosa impara una rete neurale?
- La comprensione umana è differente dalle azioni di una deep neural network (feed-forward o CNN). Gli umani hanno conoscenza di contesto, sono in grado di catturare la forma degli oggetti, etc..

Reinforcement learning

- Supponiamo di voler progettare un agente che svolga al meglio una determinata funzione in una gamma ampia di situazioni (e.g. avvicinarsi ad una porta ed aprirla). Cosa facciamo?
- Usare una CNN non sarebbe ottimale (quanto ampio dovrebbe essere il training set?)
- Si può pensare di allenare un agente non a “far bene i compiti” ma a massimizzare i guadagni derivanti dalle proprie azioni.
- Se l’agente raggiunge l’obiettivo che vogliamo raggiunga, sarà premiato. Altrimenti, sarà ignorato (**condizionamento operante -> reinforcement learning**)

L'idea

- Per progettare un agente capace di svolgere un dato compito utilizzando strategie flessibili (cioè, capaci di tener conto di eventuali imprevisti e far bene il proprio lavoro) si può pensare di costruire un sistema che proceda per tentativi e, quando fa la cosa giusta, riceve un premio, detto **rinforzo** (e.g. un numero reale positivo).
- L'agente impara a compiere quelle azioni che lo portano ad avere un premio.
- Per progettare tale agente, dovremo incorporare in esso una **rappresentazione della conoscenza**, ovvero fare in modo che l'agente incorpori una visione dello stato di cose in cui deve operare -> **stato**
- In ogni stato, l'agente può compiere un certo numero di azioni (e.g. *andare avanti, indietro, portare il braccio in avanti, girare il polso, etc...*)

L'idea

- L'apprendimento avviene attraverso una serie (anche molto grande!) di **episodi** di apprendimento ognuno dei quali consta di una serie di **iterazioni** (esecuzioni di azioni).
- Ad ogni iterazione, l'agente, **considerato lo stato** sceglie un'azione.
- Inizialmente l'agente sceglie azioni in modo **casuale**. Dopo numerose iterazioni, se l'agente in uno stato **x** esegue l'azione **a** e **raggiunge l'obiettivo**, riceverà una ricompensa. In questo caso imparerà ad associare allo stato **x** l'azione **a** come “qualcosa che è una buona idea fare”.
- **IMPORTANTE:** l'agente sa solo che, nello stato **x** fare **a** è **vantaggioso**. Tuttavia **non sa** se fare qualcos'altro può portare ad un vantaggio ancora maggiore o meno. Questo si impara in episodi successivi. (-> **l'apprendimento deve essere graduale**)

L'idea

- Un concetto importante è il **valore associato ad una azione A in uno stato X** -> **stima della probabilità che eseguendo l'azione A nello stato X e, eseguendo azioni di valore elevato negli stati successivi si arrivi ad ottenere la ricompensa.**
- Un esempio tipico di apprendimento di rinforzo è il **Q-learning**.
- Nel Q-learning, l'agente è equipaggiato con una "tabella" (**tabella Q**) che mette in corrispondenza ogni stato con i valori associati alle possibili azioni in quello stato.
- Inizialmente, i valori sono tutti pari a zero. -> **l'agente sceglie in modo casuale**
- Quando riceve una ricompensa dopo aver eseguito un'azione A in uno stato X, l'agente inserirà un valore positivo nella "colonna corrispondente ad A in corrispondenza di X.

L'idea

- **IN GENERALE:** L'attribuzione dei valori avviene **a ritroso**. L'agente agisce in uno stato X modo casuale quando non ha azioni di valore elevato corrispondenti ad X . Se, eseguendo casualmente A , l'agente si ritrova in uno stato con un'azione di valore elevato, l'agente attribuisce ad A in X un valore superiore a quello che aveva prima.
- **Gli agenti operano sempre in un ambiente (vi ricordate l'ambiente operativo?).** Come può un agente gestire l'osservabilità parziale dell'ambiente, il suo essere eventualmente non-deterministico, etc.?
- **NON è sempre detto che un'azione ad alto valore sia sempre la migliore.** Esplorando nuove soluzioni si potrebbe arrivare ad una ricompensa maggiore.
- **È sempre necessario un equilibrio tra esplorazione e sfruttamento.**

I problemi del Q-learning

- Quanti possibili stati possono esservi nel **mondo reale**? Quanto dovrebbe essere grande la tabella Q?
- Si può rimpiazzare la tabella Q con una rete neurale -> **vantaggio**: permette di effettuare stime dei valori delle azioni che si possono compiere in uno stato **X** quando questo è completamente nuovo. (**Input -> Stato, Output-> valori**).
Obiettivo: classificare stati come più o meno “promettenti”.
- **PROBLEMA**: le risorse in termini di tempo necessarie al Q-learning sono **enormi** -> **si ricorre ad ambienti simulati**.
- **IN GENERALE**: costruire agenti che sappiano operare nel mondo reale richiede di tener conto di un numero molto grande di **gradi di libertà (possibili variabili di cui tener conto)**

Macchine in grado di giocare

- Uno dei maggiori problemi dell'apprendimento per rinforzo è la gestione dell'imprevisto e degli innumerevoli gradi di libertà della realtà.
- Per superare il problema ed ottenere risultati si potrebbe provare ad inserire un'agente in ambienti più o meno osservabile con un numero di gradi di libertà ridotto: **i videogiochi.**
- **DeepMind Technologies** -> Addestrare algoritmi di IA a giocare ai videogiochi Atari, e.g. Breakout (Arkanoid), Pong, etc...
- **Obiettivo** di DeepMind Technologies: **combinare apprendimento per rinforzo e deep neural networks.**

Deep Q-learning

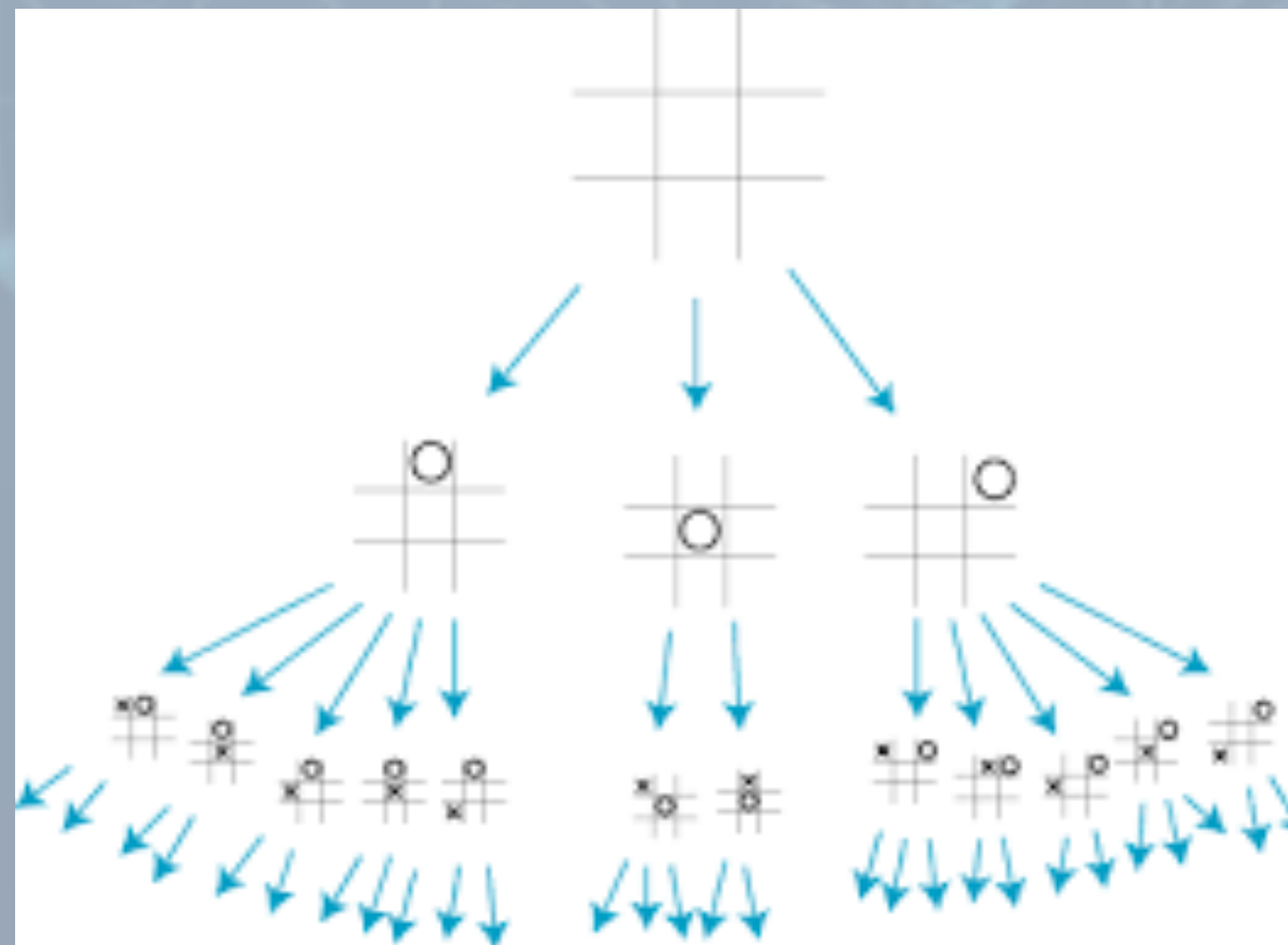
- **Idea di fondo:** rimpiazzare la tabella Q con una rete neurale (convolutiva) profonda.
- ESEMPIO tratto da M. Mitchell: **Una Deep Q-Network per stimare il valore delle azioni nel gioco Breakout.**
- **Input:** Il **frame (snapshot)** corrente più i tre frame anteriori. **Output:** attribuzione di valori alle azioni (andare a sinistra, a destra, stare fermi).
- Cosa apprende la rete? **I PESI**
- L'algoritmo inserisce lo stato nell'input della rete, ottiene un output e sceglie un'azione con un valore alto (**non sempre**).
- Se a seguito dell'azione la palla colpisce un mattoncino, allora la macchina riceve una ricompensa ed aggiorna la CNN per retropropagazione.
- Come si aggiornano i pesi? In fin dei conti non vi sono etichette!

Deep Q-learning

- Per aggiornare i pesi si procede come segue. Si suppone che le stime che vengono effettuate verso la fine di un episodio siano più accurate di quelle che vengono effettuate all'inizio.
- Per aggiornare i pesi si minimizza la differenza tra la stima nello stato corrente e la stima nello stato immediatamente precedente.
- **Continua a valere la strategia di apprendimento a ritroso.**
- **Summary:** Il sistema in uno stato X inserisce X nella CNN. Ottiene dei valori per ogni azione. Sceglie l'azione. Eseguce l'azione e giunge in un nuovo stato Y . Inserisce Y nella CNN e ottiene una nuova stima di valori.
- Lo “scostamento” tra i valori dati da X ed Y è l'errore da minimizzare aggiornando i pesi.

Giochi competitivi

- Arthur Samuel (1949): primo programma in grado di giocare a dama (Samuel è l'inventore della parola *machine learning*).
- Il programma sfrutta un metodo di ricerca in un albero di gioco.
- Un **albero di gioco per la dama** è un albero avente alla radice la situazione iniziale della scacchiera. Dalla radice si dipanano tutti i possibili scenari.



Giochi competitivi

- Si può pensare di utilizzare un albero di gioco per scegliere la mossa migliore da fare prevedendo gli scenari possibili che potrebbero dipanarsi facendo quella mossa.
- Posso esplorare tutto l'albero di gioco? **NO**
- Data la configurazione corrente della scacchiera, il programma di Samuel generava un piccolo albero di gioco (e.g. di 5 stage). In seguito, attribuiva un valore alle configurazioni finali in base a certi criteri attraverso una **funzione di valutazione**. Il valore rifletteva la probabilità con cui quella configurazione finale avrebbe portato alla vittoria.
- I risultati della valutazione venivano usati per attribuire valori alle configurazioni precedenti fino ad arrivare a quelle immediatamente successive alla configurazione iniziale.

Scacchi e Go

- 1997: Deep Blue batte Garry Kasparov in una partita a scacchi.
- DeepBlue è equipaggiato con uno programma simile a quello di Samuel ma non utilizza il machine learning. Usa l'algoritmo minimax.
- 2016: AlphaGo batte Fen Hui e Lee Sedol nel gioco del Go.
- AlphaGo: Deep Q-Learning + ricerca ad albero di Montecarlo.
- Metodo di Montecarlo: utilizzare l'aleatorietà per risolvere problemi di matematica complessi.
- Ricerca ad albero Montecarlo: generare "casualmente" un piccolo numero di conseguenze dalle possibili mosse da una data configurazione in modo aleatorio.
- Il sistema immagina "a caso" una possibile mossa del suo rivale e costruisce gli scenari fino alla fine della partita

La ricerca ad albero Montecarlo

- **Obiettivo:** attribuire un punteggio (un valore di “preferibilità”) a tutte le mosse che possono essere fatte a partire da una data configurazione.
- Data una configurazione, il programma sceglie delle mosse attuabili a partire da quella configurazione in modo casuale.
- Per ogni mossa scelta, il programma ipotizza aleatoriamente una serie di scenari scegliendo in modo casuale la mossa successiva dell’avversario (**roll-out**).
- In base agli esiti delle simulazioni, il programma attribuisce valori alle mosse iniziali (calcola con quanto probabilità le mosse selezionate portano ad una vittoria o ad una sconfitta).
- Il programma conserva statistiche sulle mosse intermedie da riutilizzare in momenti successivi della partita.

Reti neurali convolutive e ricerca ad albero Montecarlo

- Si può utilizzare una CNN per attribuire dei valori alle mosse a partire da una data configurazione.
- Il sistema sceglie le mosse di cui fare il roll-out prendendo in considerazione i valori della CNN.
- Gli esiti della ricerca ad albero vengono poi utilizzati per aggiornare i pesi della CNN.
- La CNN simula “l’intuito” del programma.

Giochi e AGI

- I sistemi come AlphaGo **NON** sono multitasking. Sono ancora sistemi di IA debole. -> Problema del trasferimento della conoscenza.
- Anche nell'apprendimento per rinforzo è necessario **settare gli iperparametri**.
- I sistemi basati sul deep Q-learning sono ancora soggetti ad **esempi avversari**.
- **Problemi con i gradi di libertà**, i.e. a differenza dei giochi, operare in ambienti reali implica dover tenere in considerazione molti gradi di libertà. I sistemi come AlphaGo non sono in grado di avere una panoramica completa di tutti i gradi di libertà fondamentali per eseguire un'azione in un ambiente reale.



NLP: Natural Language Processing

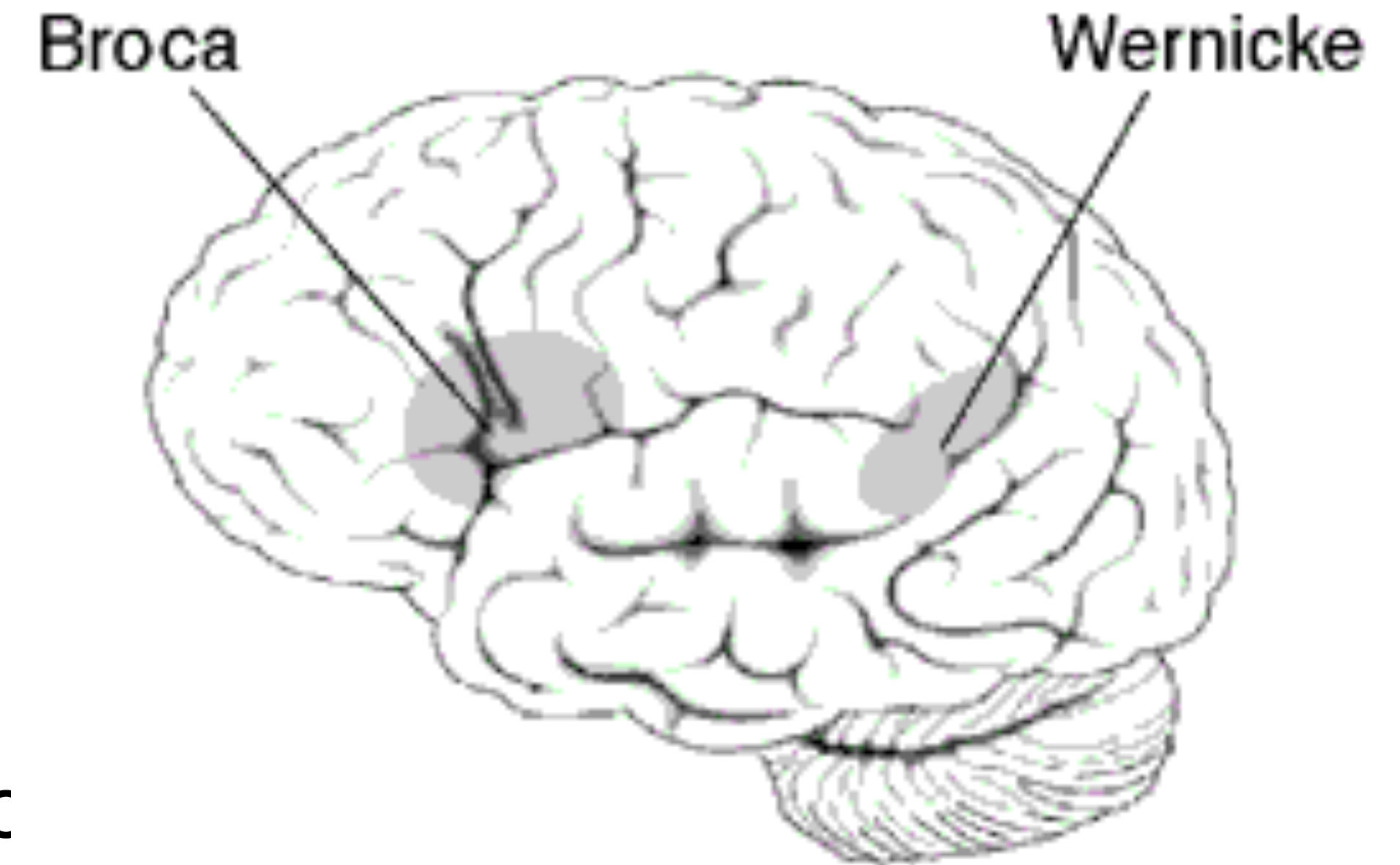
La comunicazione come impiego di codici

Il modello del Codice

- Comunicare significa codificare contenuti mentali attraverso un opportuno codice (linguaggio) e trasmetterlo ad un ricevente.
- Aristotele (*De Interpretatione*): Il **linguaggio** è costituito da **simboli** delle **affezioni dell'anima**.
 - Le affezioni dell'anima sono **causate da stati di cose esterni**.
 - Gli stati di cose esterni causano le medesime affezioni in **tutti** gli esseri umani.
 - Una volta adottato un codice comune, gli esseri umani possono comunicare senza fraintendimenti.
- J. Locke (*Essay concerning Human Understanding*):
 - gli esseri umani si formano idee elementari attraverso l'esperienza (questo bianco, questa sedia, etc..);
 - L'intelletto umano crea concetti generali per **astrazione** (il bianco, la sedia, etc..)
 - I concetti astratti possono essere comunicati attraverso il linguaggio

Il modello del codice e il suo potere esplicativo

- Evidenze neuroscientifiche
 - Area di Broca (1861): area del cervello deputata alla produzione del linguaggio.
 - Area di Wernicke (1874): area del cervello deputata alla comprensione del linguaggio.
- La produzione e processazione del linguaggio si focalizza in queste due aree. Potremmo dunque concludere che il nostro cervello si comporta come un encoder ed un decoder.
 - Le nostre rappresentazioni sono codificate dall'area di Broca che le trasforma in linguaggio
 - Il linguaggio è decodificato dall'area di Wernicke che lo trasforma in rappresentazioni mentali.



Il modello del codice

- La teoria è stata sviluppata in due contributi fondamentali pubblicati tra il 1948 e il 1949:
 - C. Shannon, “La teoria matematica delle comunicazioni”, 1948
 - W. Weaver , “Recenti contributi alla teoria matematica delle comunicazioni”, 1949
- Obiettivo: fornire una descrizione dei processi attraverso cui una sorgente (un essere umano, una macchina) può trasferire **un messaggio** (una frase, un melodia, un’immagine) ad un ricevente attraverso un medium.
- Rappresenta la fondazione matematica delle moderne tecnologie di trasmissione delle informazioni (dal telegrafo al modem)
- In ambito filosofico-comunicativo è stato recepito da alcuni come un modello della comunicazione tra gli esseri umani.

La comunicazione e le sue componenti

- Recenti contributi alla teoria matematica delle comunicazioni
- Definizione:
 - (1) La comunicazione è il processo attraverso cui un pensiero può influenzarne un altro.
 - (2) La comunicazione è il processo attraverso cui un “meccanismo” entra in contatto con un altro meccanismo.
- Quali problemi affrontare per chiarire il concetto di “comunicazione”?
 - [**Problema Tecnico**] Con quanta esattezza un messaggio può essere trasferito dalla sorgente al destinatario in modo tale che il messaggio originario non subisca “distorsioni”?
 - [**Problema Semantico**] Con quanta esattezza il messaggio veicola il significato inteso dalla sorgente al destinatario?
 - [**Problema dell’efficacia/pragmatico**] Con quanta esattezza il messaggio sortisce sul destinatario l’effetto inteso dalla sorgente?

La comunicazione e le sue componenti

- **[Problema tecnico]**
 - Un messaggio può consistere in una serie di simboli (linguaggio scritto), o di una funzione continua (lingua parlata), o di una funzione continua in due coordinate spaziali (immagini).
 - Per essere trasmesso, il messaggio dovrà essere codificato o come un segnale **discreto** (simboli nettamente divisi tra loro); oppure come una funzione continua (e.g. un'”onda”) oppure come una funzione continua in due coordinate spaziali (che caratterizzi e.g. il colore di pixel in uno schermo televisivo).

La comunicazione e le sue componenti

- **[Problema semantico]**
 - Riguarda il rapporto tra il significato attribuito al messaggio dal ricevente e di quello inteso dalla fonte.
 - Porta in campo il problema della **comprensione**: come può la sergente appurare con certezza di essere stata compresa dal ricevente? Il problema può essere ridotto attraverso spiegazioni in un linguaggio condiviso.
- **[Problema pragmatico]**

L'obiettivo della comunicazione è **sempre** suscitare un effetto nel ricevente (un comportamento o una rappresentazione mentale).

Come funziona un sistema di comunicazione?

Anatomia

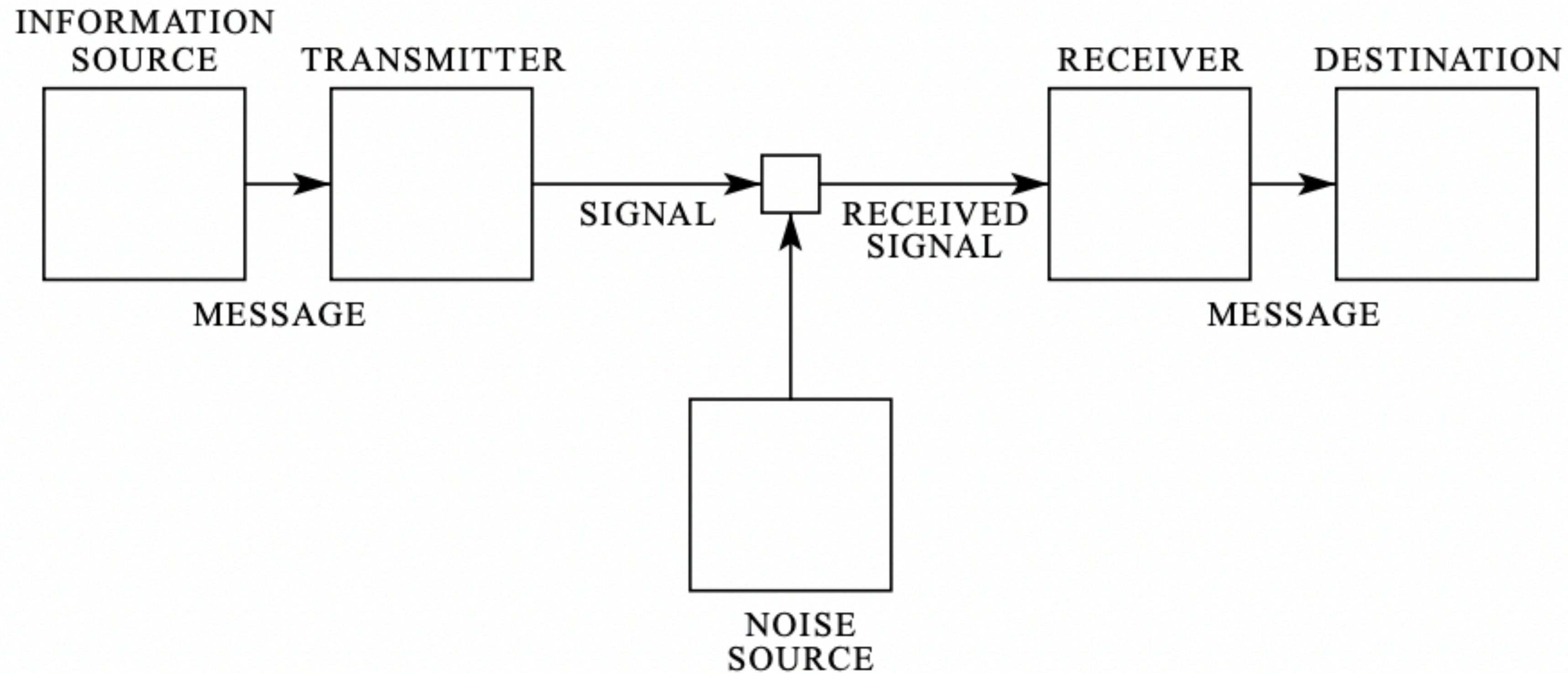


Fig. 1 — Schematic diagram of a general communication system.

Glossario

- [**Sorgente di informazione**] Sceglie un messaggio **tra una gamma di messaggi possibili** e lo invia al trasmettitore.
- [**Trasmettitore**] Riceve il messaggio dalla sorgente, lo codifica in un **segnale** e lo invia al ricevente tramite un canale.
- [**Canale**] il segnale passa attraverso un canale lungo il quale possono avvenire distorsioni del messaggio a causa di **disturbi (noise)**.
- [**Ricevitore**] Il ricevitore riceve il segnale e lo decodifica in un messaggio che arriva alla destinazione.
- [**Destinazione**] Riceve il messaggio ed eventualmente lo elabora.

Quali problemi bisognerebbe porsi nello studio di questo modello?

- È possibile misurare la quantità di informazione portata da un messaggio? È possibile misurare il grado di “informatività” di una sorgente? Ma soprattutto, **che cos'è l'informazione?**
- Come si misura la capacità di un canale?
- Esistono strategie ottimali di codifica di un messaggio (in grado di rendere il messaggio facilmente recuperabile)?
- In cosa consiste il disturbo e qual è il suo effetto sul segnale?

Informazione

- Il concetto di **informazione** indagato dalla teoria non va identificato con quello di **significato**. Per esempio, due enunciati con significati diversi possono essere identici dal punto di vista dell'informazione. **In un certo senso, la teoria di Shannon si preoccupa di specificare il concetto di "informatività".**
- **Esempio:** Consideriamo i seguenti messaggi. Quale dei due è una news?
 - a) La terra gira intorno al sole
 - b) Il presidente degli Stati Uniti è stato arrestato.
- **L'informazione è la misura della libertà di scelta di un messaggio tra un insieme di messaggi possibili. (Hartley)**
- La misura della libertà di scelta di un messaggio è data dalla **probabilità di scegliere il messaggio** all'interno di un insieme di messaggi.
- **Esempio:** Se una fonte emette un messaggio **A**, quanto è maggiore il numero di messaggi che avrebbe potuto emettere in alternativa ad **A** tanto più **A** è informativo.

Informazione

- Quanto più il messaggio emesso dalla fonte era probabile, **tanto meno è sorprendente**. Viceversa, quanto meno il messaggio è sorprendente, tanto più il fatto venisse emesso era probabile.
- Come si quantifica l'”essere sorprendente” di un messaggio? -> L'essere sorprendente di un messaggio è l'inverso della sua probabilità. Se x_i è il messaggio, e la probabilità venisse emesso dalla fonte era p_i , l'essere sorprendente di x_i è dato da

$$S(x_i) = \frac{1}{p_i}$$

Informazione

- **Fatto 1.** Ogni alfabeto può essere rappresentato solo con due caratteri: **0** e **1** (**alfabeto binario**)
Esempio. La “a” può essere rappresentata con 01100001, la “A” con “01000001” ->
La rappresentazione binaria dei simboli è alla base del funzionamento dei computer.
- Se ho un alfabeto che consta di 0 e 1, una **stringa** è una sequenza di caratteri giustapposti (esempio: 0000, 010001010101, 000000000000)
- La lunghezza di una stringa è il numero di caratteri giustapposti (esempio: la lunghezza della stringa 0000 è 4, la lunghezza della stringa 01 è 2, la lunghezza di 01111111 è 9)

Informazione

- **Fatto 2.** Se ho un alfabeto composto da 0 e 1, e k è un numero arbitrario, il numero di stringhe distinte di lunghezza k è 2^k
- **Esempio:** Se $k = 1$, avrò in totale 2 stringhe cioè 0 e 1. Se $k = 2$, avrò in totale $2^2 = 4$ stringhe, cioè 00, 01, 10, 11.
- Se ho un insieme di k elementi tutti equiprobabili, la probabilità di pescarne uno è $\frac{1}{k}$
- **Se ho un alfabeto composto da soli 0 e 1, la probabilità di pescare una qualsiasi delle stringhe di lunghezza k composte da 0 e 1 è $\frac{1}{2^k}$**

Informazione

- **Definizione:** la quantità di informazione (autoinformazione) contenuta in un messaggio x_i è k (scritto $\mathcal{I}(x_i) = k$) se e soltanto se la probabilità di avere x_i (cioè p_i) è uguale alla probabilità di avere un elemento qualsiasi dell'insieme di tutte le stringhe (composte di soli 0 e 1) di lunghezza k (supponendo siano tutte equiprobabili).

- $\mathcal{I}(x_i) = k$ se e solo se $p_i = \frac{1}{2^k}$, ovvero se $2^k = \frac{1}{p_i}$. Dal momento che k è il numero a cui elevare 2 per avere $\frac{1}{p_i}$ abbiamo che

$$\mathcal{I}(x_i) = k = \log_2 \frac{1}{p_i}$$

- L'unità di misura dell'informazione è il **bit**. Se $\mathcal{I}(x_i) = k$, diciamo che la quantità di informazione di x_i è di k bit

Informazione

- In generale, quando una sorgente compone un messaggio, ha a disposizione un insieme di **simboli elementari** (ad esempio, le lettere) attraverso i quali si compongono stringhe (ad esempio, parole) che vanno a comporre insiemi di stringhe (ad esempio, frasi) che costituiscono il messaggio.
- Ogni simbolo, ogni parola, ogni frase hanno una certa probabilità di essere scelte. **Osservazione:** la probabilità con cui scegliamo una lettera per comporre una parola dipende strettamente dalla probabilità con cui abbiamo scelto la lettera precedente. Lo stesso dicasi per le parole in una frase o per le frasi in un testo.
- un processo che produce un insieme di simboli ognuno con una certa probabilità è detto **processo stocastico**. Quando la probabilità di scelta di un simbolo dipende dalle scelte precedenti parliamo di **processo stocastico markoviano**

Produzione di informazioni

- La produzione di un messaggio può essere vista come un processo markoviano.
- [**Domanda**] Quanta informazione viene prodotta da un processo di produzione di un messaggio?
- La quantità di informazione prodotta da un processo di produzione di un messaggio si misura in termini di **entropia**. Questa tiene conto delle probabilità di arrivare a certi stati della produzione di un messaggio e delle probabilità che a questi stati ne seguano altri di un certo tipo.
- Perché l'entropia? In **termodinamica** ha la funzione di misurare il grado di "casualità" nell'evoluzione di un sistema fisico, quantifica quanto il sistema tenda ad evolversi in modo sempre meno organizzato.
- Perché è ragionevole pensare all'entropia all'interno di una teoria della comunicazione?

- **Esempio intuitivo:** Supponiamo la sorgente produca messaggi aggiungendo parole la cui probabilità è indipendente da quella delle altre. Consideriamo il messaggio A= “La mela è rossa” e il messaggio B = “La”. Quale dei due è più informativo?
- A consta di 4 parole;
B consta di 1 parola.
- Supponiamo la lingua italiana consti di 100 parole. Qual è la probabilità che il processo di produzione del messaggio mi dia proprio B?

$$P_B = \frac{1}{100} = 0,01 .$$

- Qual è la probabilità che il processo mi dia proprio A? In generale, avrò 100^4 possibili frasi di 4 parole. Quindi la probabilità di avere proprio A è, in assenza di altre informazioni:

$$P_A = \frac{1}{100^4} = 1 \times 10^{-8}$$

- **Maggiore informatività coincide con maggiore casualità. Quanto più un processo è stabile, organizzato, “scontato”, tanto meno è informativo.**

Produzione di informazioni

- La produzione di un messaggio non sempre veicola esattamente la quantità di informazione che il mittente intende trasmettere. A volte ne produce di più. Questa informazione aggiuntiva è detta **ridondanza**.
- **Esempio:** A = “Arriva treno”, B=“Arriva **un** treno”. B è più corretta da un punto di vista grammaticale ma A è **sufficiente** per convogliare l’informazione desiderata. (**NB: stiamo esemplificando un concetto, non identifichiamo informazione e significato!**)
- La ridondanza misura quanto della produzione del messaggio non dipende dalle “**esigenze informative**” della fonte.

Precisazione doverosa

- Per come lo abbiamo definito, il concetto di entropia/quantità di informazione di un processo **non si applica a singoli messaggi** ma a una **procedura** di produzione di uno messaggio arbitrario tra **tutti** i messaggi che quella fonte può produrre.
- **L'entropia di un processo tiene conto della “libertà” con cui quel processo produce un messaggio.** Quanto più il processo è libero, tanto più produrrà informazione (tanto meno un processo è libero, tanto meno questo produrrà informazione).

Codifica e trasmissione di un messaggio

- Nel modello di un sistema di comunicazione visto in precedenza si vede che un messaggio “viaggia” da un mittente ad un ricevente attraverso un canale.
 - Come misuro la capacità di un canale di trasmettere un messaggio? Se ciò che conta del messaggio è l'**informazione** allora dovrò fare una valutazione in termini di informazione.
- [**Caso semplice**] Supponiamo che la sorgente possa comunicare messaggi tali che ognuno di essi “porta con sé” s bit di informazione. Se tutti i messaggi hanno la stessa durata e il canale mi permetterà di inviare n messaggi al secondo, avrò che la capacità del canale è di $n \times s$ bit al secondo.

Il trasmettitore

- Il trasmettitore e il ricevitore hanno una funzione essenziale nella trasmissione e ricezione di un messaggio: il primo trasforma il messaggio in un codice trasmissibile attraverso il canale, il secondo decodifica il codice e lo trasforma in un messaggio.
- **Come deve essere fatto un trasmettitore ottimale?** Un trasmettitore ottimale trasforma il messaggio in un segnale che preserva l'entropia del messaggio il più possibile (nei limiti della capacità e dell'adeguatezza del canale).

Il disturbo (noise)

- Il disturbo di un segnale può essere dovuto a vari motivi. Si pensi al fruscio durante una comunicazione telefonica, oppure al rumore durante una conversazione, oppure ad un errore occorso nella trasformazione dell'impulso di una tastiera nell'immagine di una parola su un monitor.
- In generale, **il disturbo trasforma il messaggio rendendolo ancora più informativo** (sic!) cioè casuale.
- Tuttavia, **l'informazione aggiunta dal rumore al messaggio codificato è tutt'altro che desiderabile (è una "falsa" informazione).**

Il disturbo (noise)

- Il grado di “disturbo” di un segnale si misura in termini di **equivocazione**. L'**equivocazione** è una stima di quanto sia incerto che il messaggio ricevuto dal ricevente corrisponda a quello effettivamente inviato dalla fonte (più alta è l'equivocazione, maggiore è il rumore).

Perché il modello del codice dovrebbe essere utile a spiegare la comunicazione umana?

- Nel suo saggio, W. Weaver sostiene che la soluzione al problema
A) Con quanta esattezza possono essere trasmessi i simboli di un messaggio da una fonte ad un ricevente?
permette di risolvere il problema
B) Con quanta esattezza un messaggio riesce a veicolare il significato che la fonte intende esprimere?
C) Con quanta esattezza il messaggio riesce a sortire l'effetto sul ricevente inteso dalla fonte?

Perché il modello del codice dovrebbe essere utile a spiegare la comunicazione umana?

- Le motivazioni di Weaver:
 - a) Il modello di comunicazione offerto dalla teoria di Shannon è molto generale e non fa assunzioni sulla natura dei simboli e dei messaggi coinvolti. Quindi: permette di tener conto di tutti i tipi di messaggi coinvolti nella comunicazione umana.**
 - b) La teoria della comunicazione sviluppata potrebbe essere utile alla progettazione di sistemi intelligenti (se il sistema prototipato è intelligente, evidentemente la teoria coglie aspetti dell'intelligenza).**

Perché il modello del codice dovrebbe essere utile a spiegare la comunicazione umana?

- **c) Per tener conto degli aspetti semantici, basta arricchire il modello con ulteriori componenti** come sistemi di codifica/decodifica semantica (avremo, ad esempio, anche un “rumore semantico”).
- d) Una teoria della comunicazione umana non può prescindere dal concetto di informazione.**
- e) L'utilizzo dei processi di Markov è utile a spiegare la componente contestuale della comunicazione umana** (l'effetto del contesto sulla comunicazione).

È convincente?

Cos'è l'NLP?

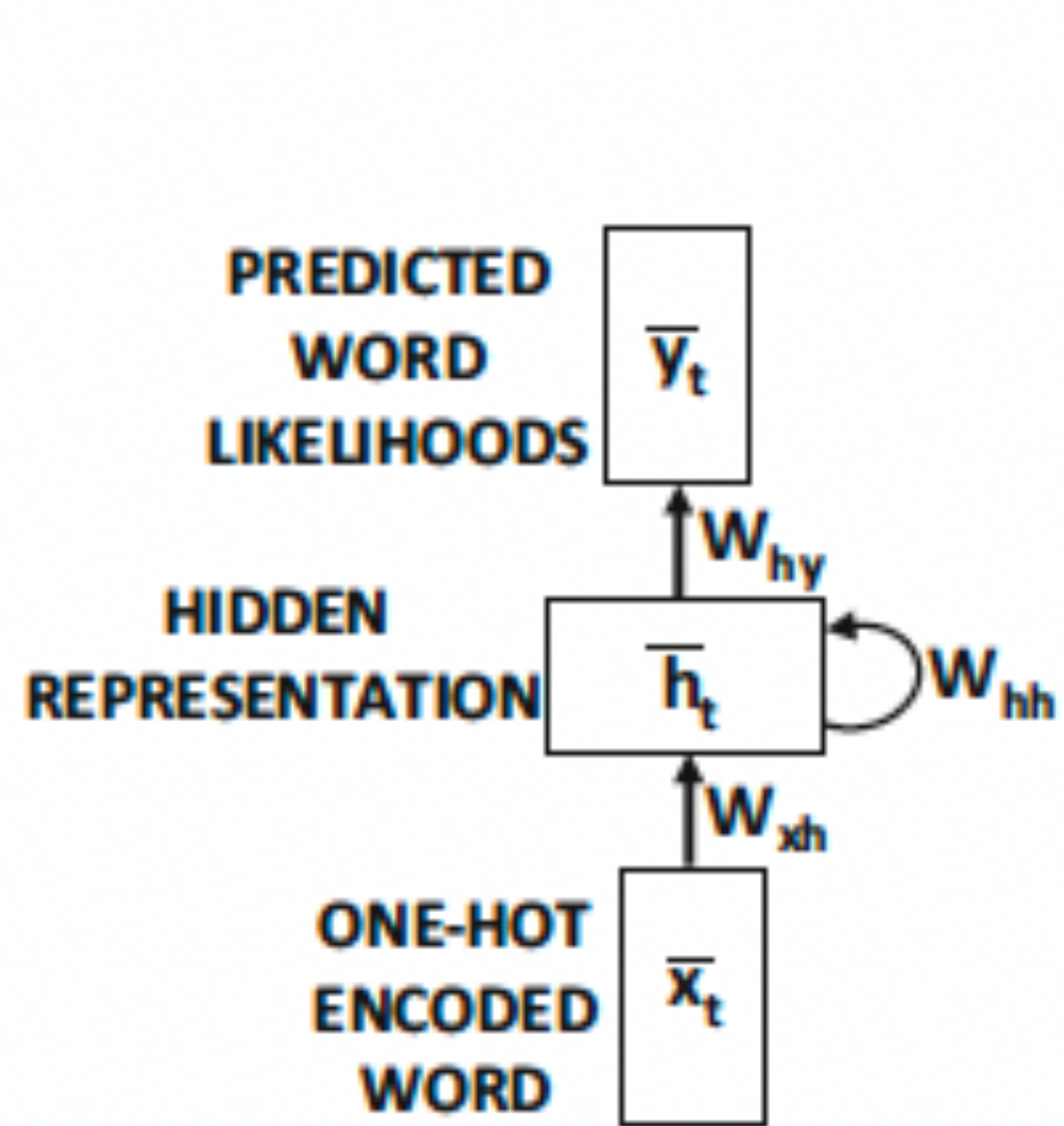
- Natural Language Processing è un campo dell'IA finalizzato alla progettazione di sistemi di elaborazione del linguaggio naturale, cioè umano.
- i sistemi progettati includono strumenti per la traduzione automatica, chatbots, descrizione automatica di immagini, miglioramento dei linguaggi di programmazione, etc..
- **Problema (filosofico) di fondo:** È possibile tradurre un testo, rispondere ad una domanda, descrivere un'immagine **senza** comprendere?
- I primi sistemi di processazione del linguaggio naturale si basano sull'IA simbolica -> regole sintattiche che codificano regole grammaticali e linguistiche.
- I metodi sintattici sono stati presto rimpiazzati da metodi statistici -> machine learning.

Applicazioni

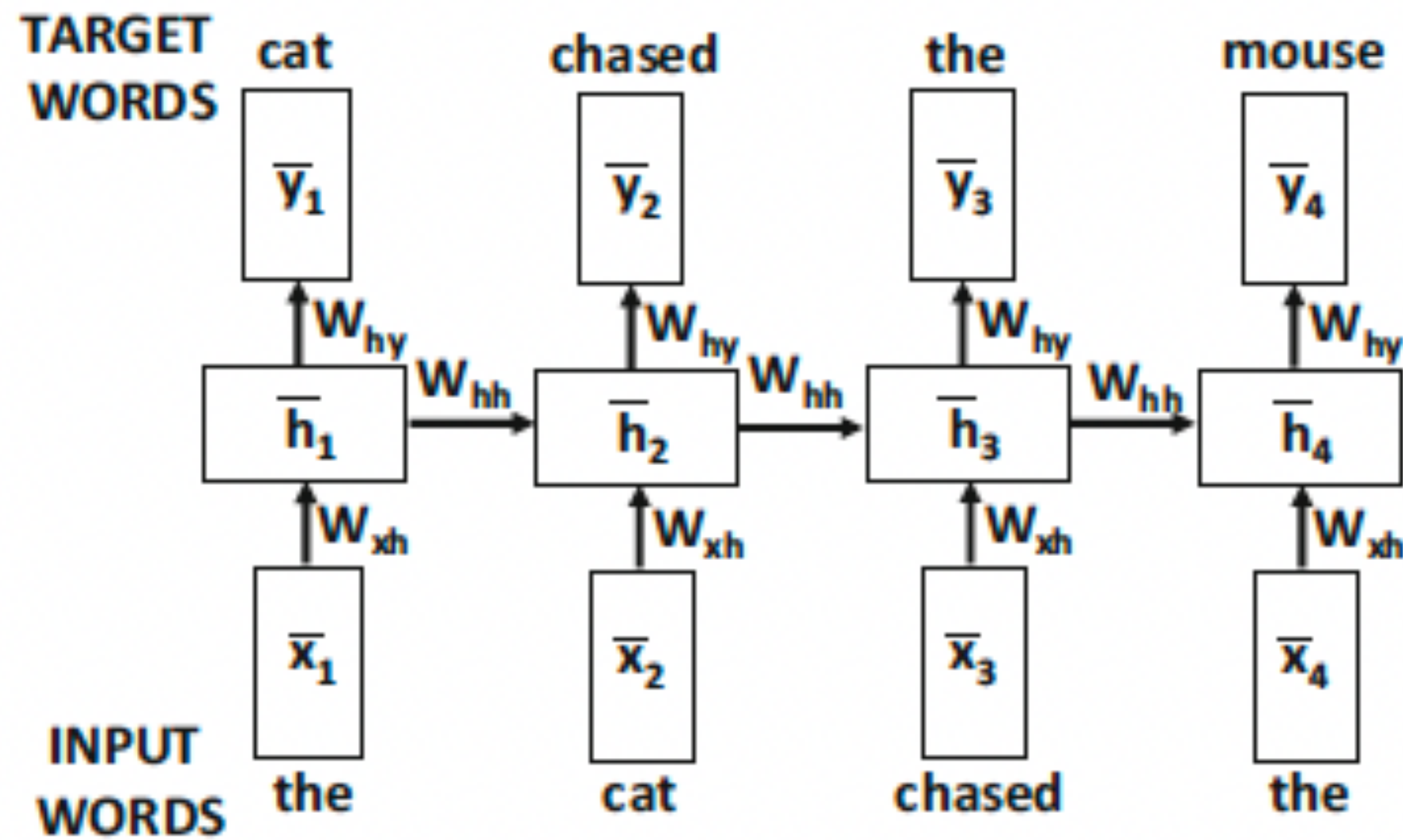
- Utilizzo delle DNN nel **riconoscimento vocale** (trascrizione automatica da parlato) -> i primi successi si sono avuti nel 2012 grazie ai team di Google.
- Problemi: presenza di rumore che non permette di cogliere correttamente i fonemi -> incapacità di cogliere (o completare) le parole in base al contesto; incapacità di cogliere parole insolite.
- **Sentiment mining**: dato un testo, classificare il suo tono (positivo, negativo, aggressivo, compiacente, etc...)
- È possibile affidarsi al riconoscimento di singole parole (es. brutto, terribile, orrendo, etc...)? **No.**
- È necessario prendere in considerazione più parole e le loro interdipendenze -> processare dati sequenziali

Reti Neurali Ricorrenti (RNN)

-



(a) RNN



(b) Time-layered representation of (a)

differenze con le reti neurali tradizionali

- I nodi negli strati nascosti sono interconnessi in modo da permettere, insieme all'input corrente, l'uso di valori di attivazione negli strati nascosti ottenuti nel passo temporale precedente per calcolare quelli nel passo temporale corrente.
- Ad ogni passo temporale, le attivazioni degli strati nascosti costituiscono la codifica della porzione della frase considerata fino a quel momento.
- Per segnalare al sistema che la frase è terminata è necessario apporre il simbolo END.
- L'output più importante è l'output che si ottiene alla fine dei assi temporali.
- Gli input sono **numeri**.

Come codificare le parole?

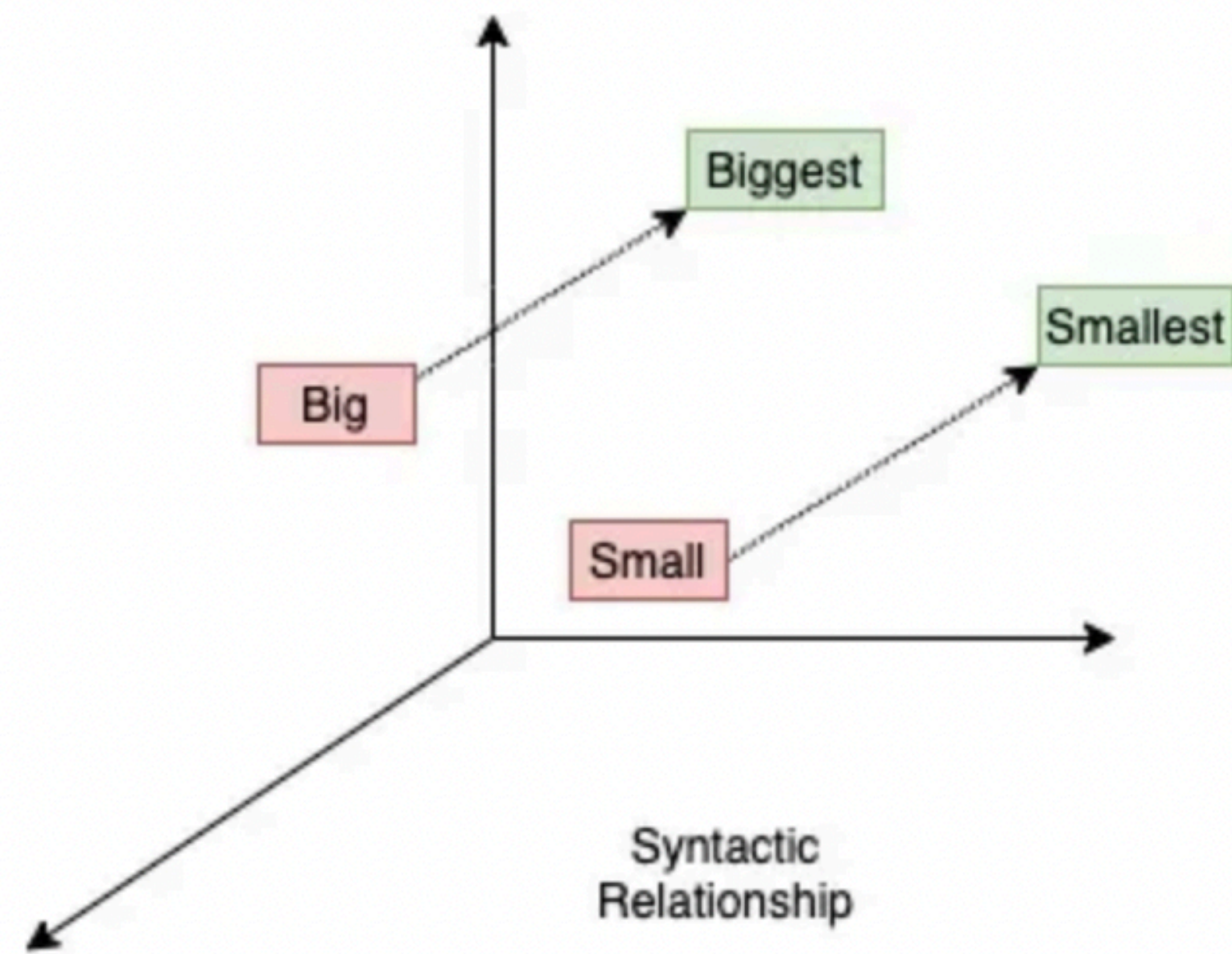
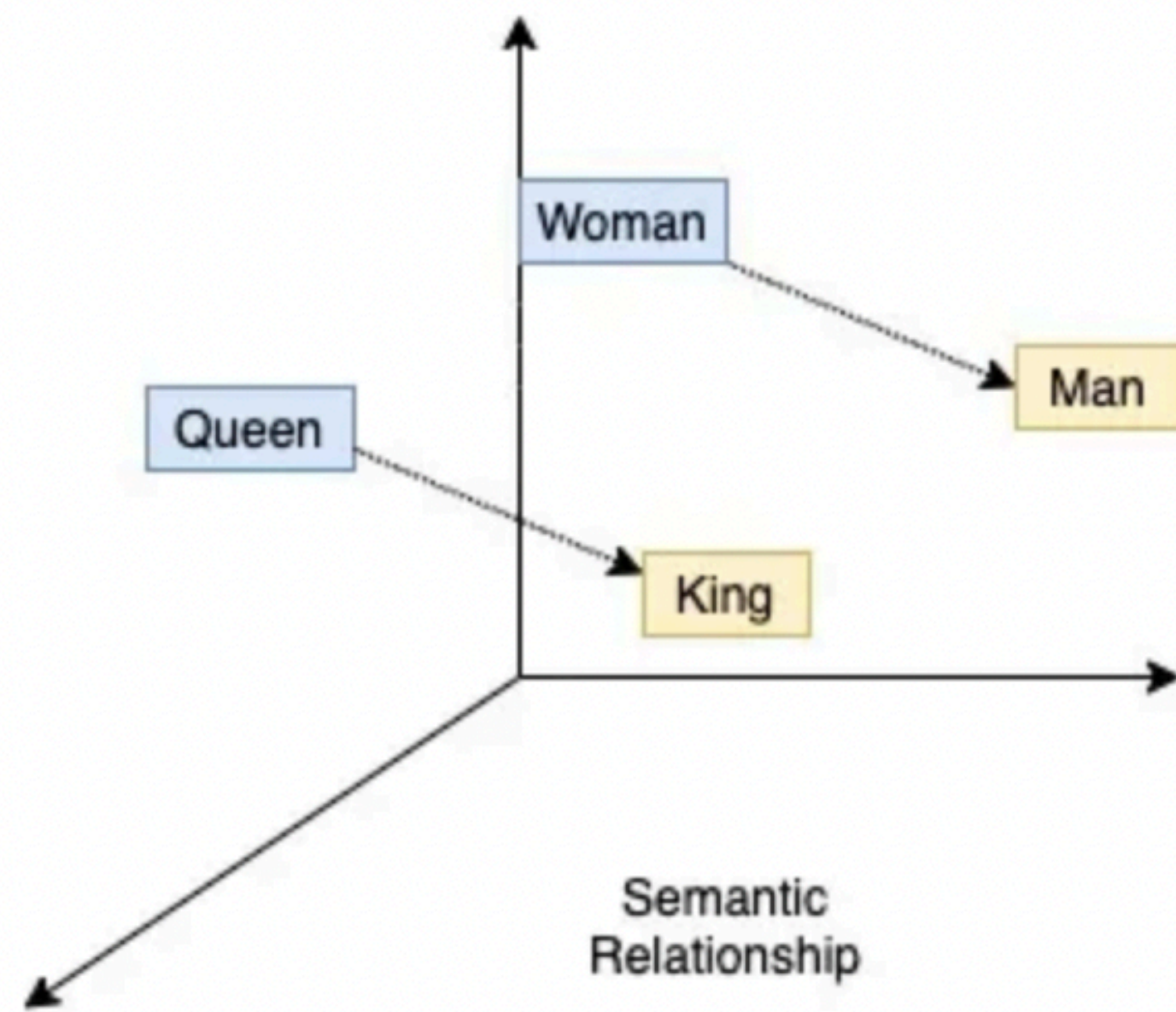
- Devo scegliere un vocabolario, ovvero un insieme di n (con n abbastanza grande) di parole.
- Il modo più semplice di codifica di un vocabolario è la codifica **one-hot** ovvero l'attribuzione di un numero diverso ad ogni elemento del vocabolario
- Esempio: Ho 50000 parole. Ad ogni parola assegno un numero diverso compreso tra 0 e 50000.
- One hot -> Tutti i bit sono posti a zero tranne uno.
- Problema: questa codifica non coglie il significato e le relazioni tra le parole.

Semantica distributiva e Word2vec

- Il significato delle parole è dato dagli insiemi di parole con cui tendono a presentarsi. In altri termini, il significato di una parola è dato dai contesti linguistici in cui la parola è pertinente.
- Due parole sono tanto più simili dal punto di vista semantico quanto più i contesti in cui occorrono sono simili.
- **Semantica distributiva**
- Il complesso dei contesti linguistici **si chiama spazio semantico**
- Lo spazio semantico può essere visualizzato come un vero e proprio spazio geometrico. Gli spazi semantici hanno più dimensioni che dipendono dai contesti linguistici e dai criteri in base ai quali le parole sono vicine tra loro.
- La vicinanza tra le parole coincide con la distanza geometrica

Semantica distributiva e Word2vec

- Il miglior modo di rappresentare matematicamente uno spazio geometrico a molte dimensioni è dato dagli **spazi vettoriali**.
- Gli elementi di uno spazio vettoriale si chiamano **vettori** e sono sequenze ordinate di numeri (di solito reali). Come nello spazio bidimensionale, i vettori specificano la posizione di un punto.
- Le coordinate rappresentano la “posizione” di un punto rispetto ad una determinata dimensione.
- Il Word2Vec permette la rappresentazione di parole attraverso vettori che siano in grado di codificare il significato delle parole attraverso le relazioni tra di essi



Come funziona il Word2Vec?

- Il Word2Vec utilizza una rete neurale tradizionale per imparare i vettori delle parole appartenenti ad un vocabolario fissato.
- I dati di addestramento sono coppie di parole tali che la prima occorre vicino all'altra in un corpus di documenti.
- Vengono scartate parole estremamente frequenti come *il, la, di, etc..*
- Come è fatta la rete?
- Strato di input: un nodo per ciascuna parola del vocabolario.
- Strato di output: un nodo per ciascuna parola del vocabolario
- Nel training viene somministrata una parola assegnando valore 1 al nodo corrispondente in input.
- Ogni nodo di output produce la probabilità che la parola corrispondente al nodo si trovi vicina a quella inserita in input.
- Quali saranno le coordinate del vettore associato ad una parola? I pesi delle connessioni che collegano il nodo corrispondente allo strato nascosto così come ottenuti alla fine dell'addestramento.

Come funziona il Word2Vec?

- **Problema: Il Word2vec** è efficace ma favorisce la propagazione di Bias.
- Tuttavia: il Word2Vec rappresenta una tecnologia promettente anche per risolvere problemi di ragionamento. Ha il vantaggio di sfruttare il potenziale degli ambienti matematici della geometria per trattare il significato.

Traduzione automatica

- Come tradurre un testo automaticamente?
- Primi sistemi basati sull'IA simbolica: equipaggiare il sistema con sistemi di regole in grado di trasformare le strutture sintattiche in una lingua **sorgente** nelle strutture sintattiche corrispondenti nella **lingua di arrivo**. Equipaggiare il sistema con regole capaci di mettere in corrispondenza il lessico.
- **Traduzione automatica statistica**: Allenare sistemi di machine learning con coppie di frasi (o meglio, **sintagmi**), costituite da una frase nella lingua sorgente e dalla frase corrispondente nella lingua di arrivo. **Obiettivo**: Ottenere valori di probabilità circa la corrispondenza di un sintagma nella lingua sorgente con il sintagma corretto nella lingua di arrivo
- **Traduzione automatica neurale (Google translate)**: la traduzione avviene per codifica e decodifica. Il codificatore e il decodificatore sono due reti neurali ricorrenti. la prima prende in input le parole codificate con Word2Vec. La seconda i valori di output del codificatore. Ad ogni passo temporale del decodificatore, la rete prende in input il valore precedente.

Problemi della traduzione automatica

- I sistemi di questo tipo tendono a gestire frasi molto lunghe con difficoltà: perdono il filo del discorso molto velocemente.
- **Soluzione: Long-Short-Term Memory.** Equipaggiare i nodi negli strati profondi delle reti neurali ricorrenti con pesi speciali in grado di permettere la propagazione delle informazioni “acquisite” in precedenza.
- Come si valuta l’adeguatezza di una traduzione automatica?
 - **Bilingual evaluation understudy:** confronto automatico tra la traduzione effettuata da una macchina e la traduzione effettuata da un agente umano.
 - Controllo della traduzione automatica da parte di agenti umani -> Può basarsi sulla media: si attribuisce un punteggio alle traduzioni automatiche e a quelle umane. Si fanno le medie dei punteggi attribuiti e si confrontano.-> La media può essere forviante.

Il problema del significato

- Il vero problema dei primi sistemi di traduzione automatica è il superamento della barriera del significato. -> Watson (Ibm) -> Sistema in grado di fornire risposte in forma di domanda a domande contenenti giochi di parole.
- **test Comprensione della lettura:** Metodo SQuAD (Stanford Question Answering Dataset) -> estrarre una risposta su un testo dal testo. -> Primi sistemi basati su DNN.
- Test di Winograd per la comprensione del significato -> Rispondere a domande ambigue contenenti e.g. indicali .
- Il fatto che le macchine possano realmente comprendere è discutibile alla luce degli esempi avversari.

Comprendere le situazioni

- Gli esseri umani comprendono le situazioni in virtù di una
 - Fisica intuitiva;
 - Biologia Intuitiva;
 - Psicologia Intuitiva;
 - Capacità di prevedere futuri possibili anche in base al ragionamento plausibile
 - Capacità di maturare modelli mentali.
- In **generale**: quando comprendiamo un fenomeno, è possibile rappresentarcelo al livello mentale, siamo in grado di **modellarlo**
- **Lawrence Barsalou: Comprensione = simulazione mentale.**

Comprensione e metafore

- George Lakoff, Mar Johnson, **Metafora e vita quotidiana** (1980)
- La comprensione avviene per mezzo di metafore -> ricorriamo a metafore per “afferrare” concetti astratti -> attingiamo a metafore estratte dal mondo fisico (e.g. “cadere nella disperazione”), dal mondo economico (e.g. “un tempo ben speso”)
- Alcuni studi psicologici hanno sottolineato l’importanza delle sensazioni fisiche per l’attribuzione di senso e di proprietà astratte (il calore fisico è associato al calore sociale).

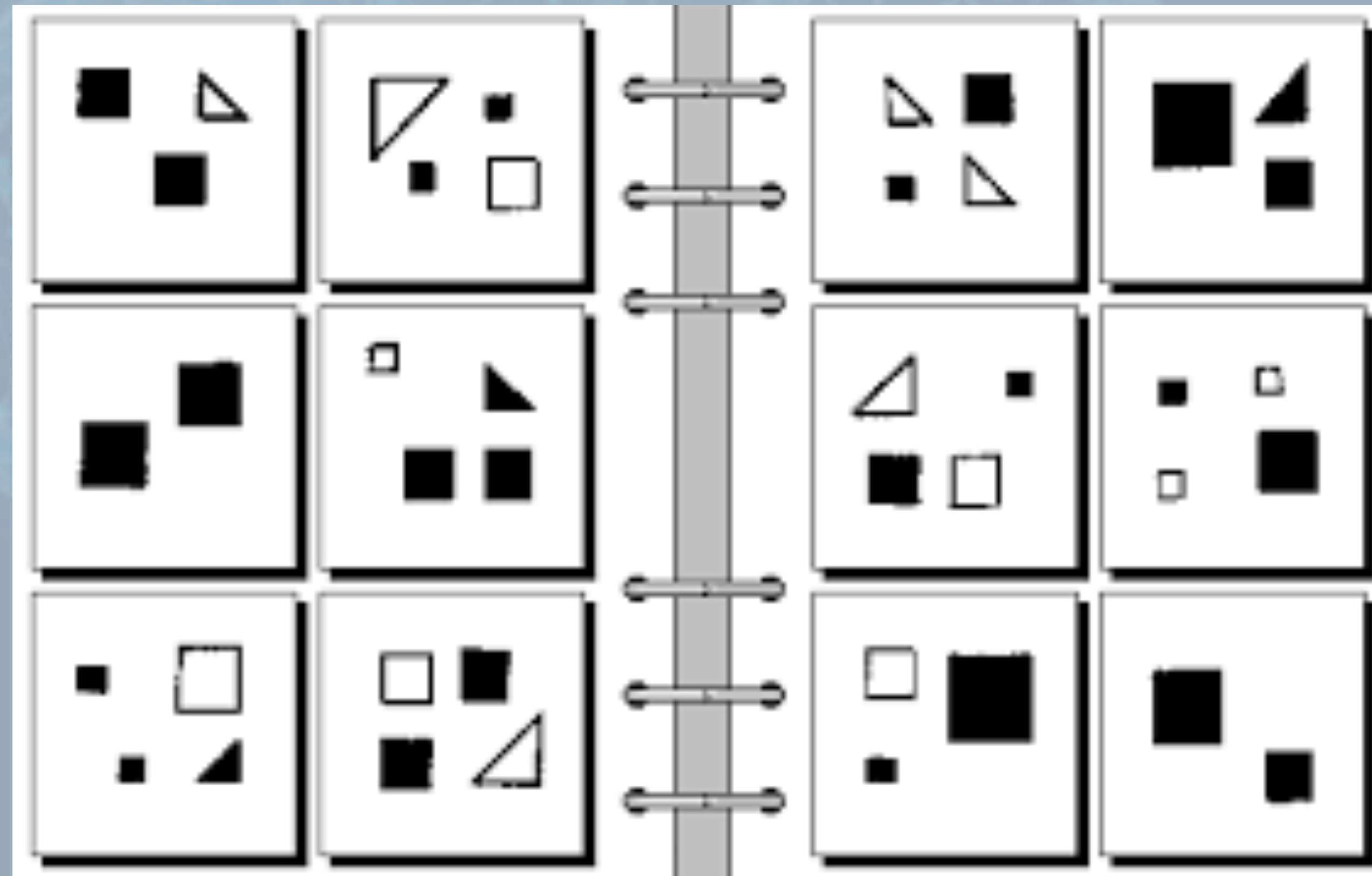
Astrazione e analogia



- Che cosa significa “astrarre”? -> Creazione di categorie generali a partire da un insieme limitato di casi concreti -> generalizzazione e creazione di modelli.
- L’astrazione è connessa alla creazione di analogie -> la possibilità di creare analogie dipende dalla capacità di cogliere proprietà astratte che non dipendono dall’oggetto specifico considerato.

Sistemi esperti

- Primi tentativi di dotare le macchine di “senso comune” -> Douglas Lenat (Cyc) -> codificare la conoscenza non scritta (e.g. consapevolezza di relazioni spaziali, mereologiche, etc...)
- **Problema:** Cyc tratta la conoscenza come esplicita. Solo quando la conoscenza è esplicita, i.e. l'agente ne è consapevole, allora è possibile manipolarla utilizzando regole di inferenza. **Tuttavia**, esiste una parte della conoscenza che è **implicita**.
- L'astrazione così come il riconoscimento di schemi spesso non poggia sul calcolo (Mikhail Bongard, *Pattern recognition*).
 - 12 riquadri in gruppi da 6. Il primo gruppo di riquadri esemplifica un concetto, l'altro esemplifica il suo opposto.
- **In generale:** riuscire in compiti di astrazione ed analogia richiede capacità di prevedere uno slittamento concettuale (fare in modo che determinati concetti e schemi non si applichino soltanto ad un numero finito di oggetti ma si applichino ad un numero potenzialmente infinito di oggetti date determinate precondizioni).



Un'osservazione finale



- Come abbiamo potuto notare durante il corso, buona parte delle nostre facoltà cognitive sono alimentate dall'esperienza e sono radicate profondamente nel fatto che questi siano performati all'interno di un **corpo**.
- **Secondo M. Mitchell, il problema dell'embodiment non può essere trascurato nella valutazione dei limiti della raggiungibilità dell'AGI.**