2° CORSO
FONDAMENTI DI
INTELLIGENZA ARTIFICIALE

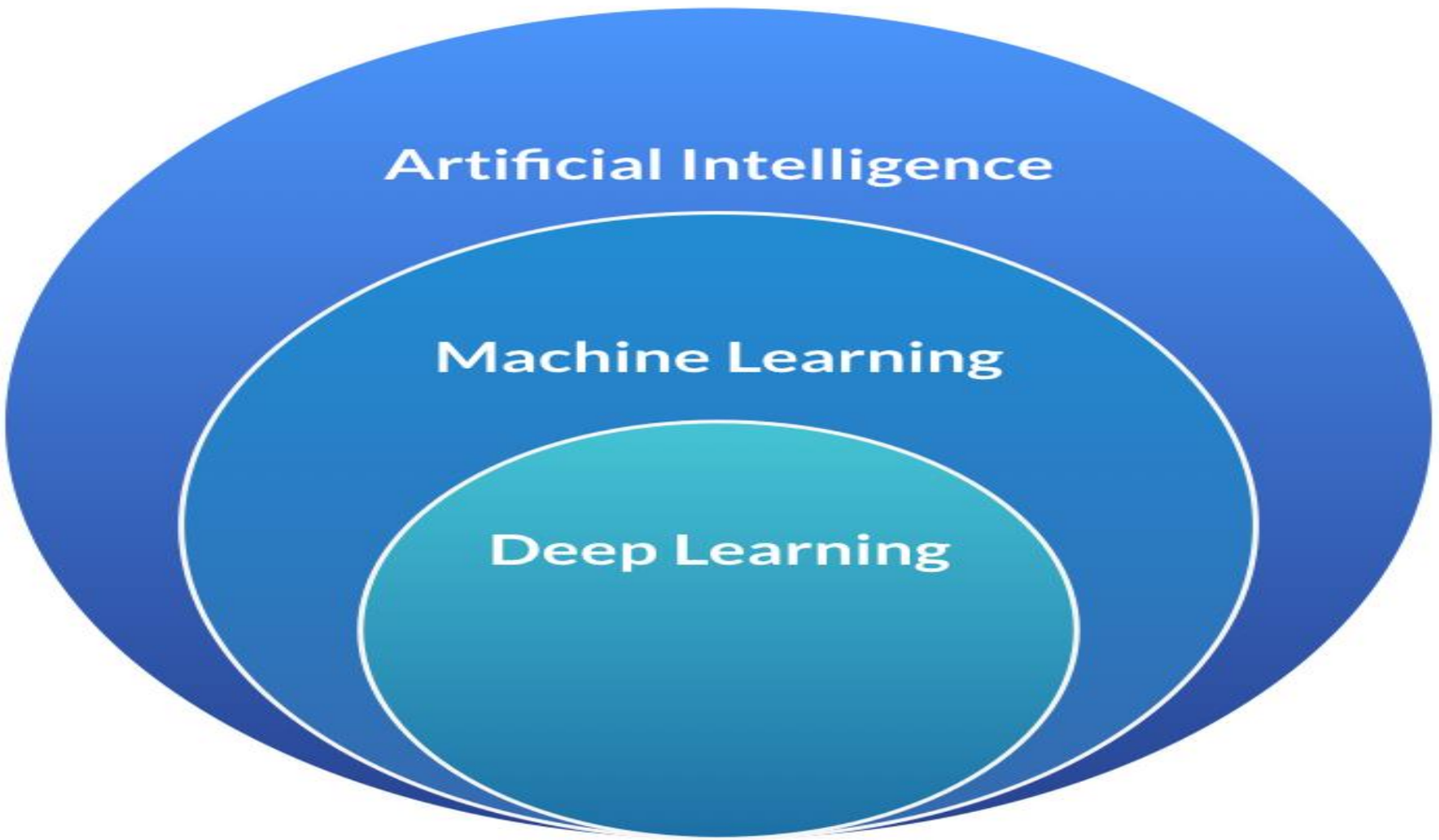
Col. ing. s.SM Emanuele SCANO

22 24 novembre 2022

AGENDA

1. INTELLIGENZA ARTIFICIALE: PRINCIPI
2. RETI NEURALI ARTIFICIALI
3. IL COMBUSTIBILE DELL' IA: BIG DATA, CLOUD COMPUTING, INTERNET OF THINGS
4. MACHINE LEARNING
5. DEEP LEARNING

INTELLIGENZA ARTIFICIALE: PRINCIPI



DEFINIZIONI

Intelligenza: Complesso di facoltà psichiche e mentali che consentono all'uomo di pensare, comprendere o spiegare i fatti o le azioni, elaborare modelli astratti della realtà, intendere e farsi intendere dagli altri, giudicare, e lo rendono insieme capace di adattarsi a situazioni nuove e di modificare la situazione stessa quando questa presenta ostacoli all'adattamento.

DEFINIZIONI

Intelligenza artificiale: Riproduzione parziale dell'attività intellettuale propria dell'uomo (con particolare riguardo ai processi di apprendimento, di riconoscimento, di scelta) realizzata o attraverso l'elaborazione di modelli ideali, o, concretamente, con la messa a punto di macchine che utilizzano per lo più a tale fine elaboratori elettronici.

DEFINIZIONI

Pensare, comprendere, elaborare → Ragionare

RAGIONAMENTO

Deduttivo - Aristotele (384 aC – 322 aC)

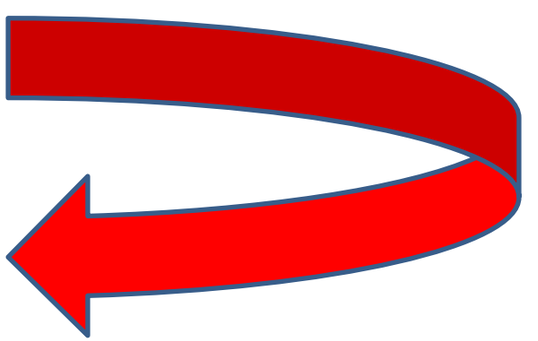
Nel ragionamento deduttivo (o sillogismo) la verità delle premesse (*caso generale*) garantisce la verità della conclusione (*caso particolare*).

REGOLA ($C \rightarrow R$): Tutti gli uomini sono mortali

CASO (C_1): Socrate è un uomo

quindi

RISULTATO (R_1): Socrate è mortale



Il ragionamento deduttivo è il fondamento di gran parte delle dimostrazioni e teoremi della matematica, … ma non ci permette di scoprire o prevedere fatti nuovi e quindi di ampliare le nostre conoscenze, compiendo un 'salto' dal noto all'ignoto.

RAGIONAMENTO

Induttivo - Francis Bacon (1561-1626) filosofo, e per quello sperimentale e scientifico, da Leonardo da Vinci (1452-1519) e Galileo Galilei (1564-1642) ... fino a Isaac Newton (1642-1727).

RAGIONAMENTO INDUTTIVO

Nel ragionamento induttivo, diversamente da quello deduttivo, le premesse (*caso/i particolare/i*) forniscono un'evidenza più o meno forte a sostegno della conclusione (*generalizzazione*), ma non ne garantiscono necessariamente la verità.

I ragionamenti induttivi comportano quindi un rischio da cui sono esenti quelli deduttivi: possono portare da premesse vere a conclusioni false. Il ragionamento induttivo è quindi un ragionamento probabilistico, le cui conclusioni dipendono dal grado di probabilità delle informazioni contenute nelle premesse.

CASO (C_1): Socrate era un uomo

RISULTATO (R_1): Socrate morì

quindi

REGOLA ($C \rightarrow R$): Tutti gli uomini sono mortali

RAGIONAMENTO

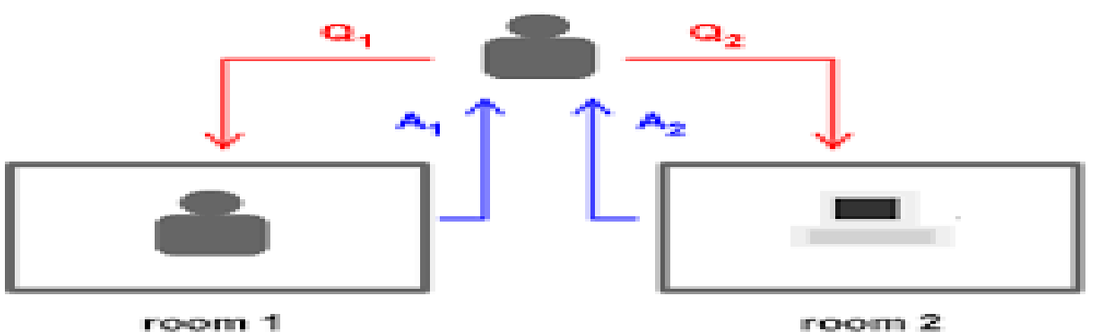
La forma più comune di ragionamento induttivo è la generalizzazione, con cui otteniamo informazioni su un gruppo di cose, persone, eventi, oggetti e così via, esaminando una porzione – o campione – di quel gruppo.

NASCITA DELL'INTELLIGENZA ARTIFICIALE

Bletchley Park

Alan Turing (Matematico inglese)

Test Turing (1950)



NASCITA DELL'INTELLIGENZA ARTIFICIALE

1956: nascono due linee di ricerca nel settore IA

- **Impostazione simbolica**
- **Impostazione connessionista**

NASCITA DELL'INTELLIGENZA ARTIFICIALE

Marvin Minsky

Impostazione simbolica

Per riprodurre l'attività cognitiva del cervello si deve partire da modelli astratti espressi attraverso simboli: la mente quindi come una macchina il cui comportamento può essere studiato e replicato da una macchina.

Prime applicazioni: programmi in grado di riprodurre attività intelligenti quali dimostrazioni matematiche

NASCITA DELL'INTELLIGENZA ARTIFICIALE

Impostazione simbolica – Logic Theorist

Nel ragionamento deduttivo (o sillogismo) la verità delle premesse (*caso generale*) garantisce la verità della conclusione (*caso particolare*).

REGOLA ($C \rightarrow R$): Tutti gli uomini sono mortali

CASO (C_1): Socrate è un uomo

quindi

RISULTATO (R_1): Socrate è mortale

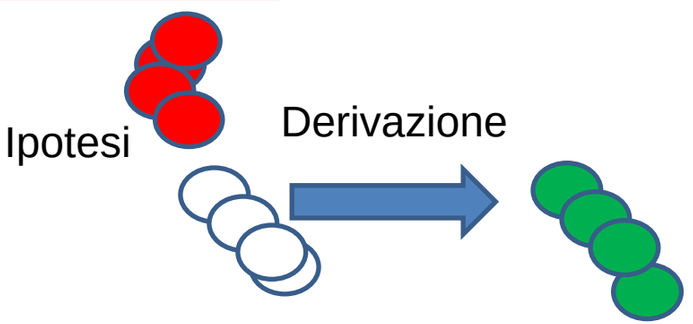
A: Socrate
B: uomo
A -> B (=Socrate è un uomo)
C: mortale
B -> C (=gli uomini sono mortali)

REGOLA DI DERIVAZIONE
Se A -> B e B -> C, quindi A -> C ossia Socrate è mortale

NASCITA DELL'INTELLIGENZA ARTIFICIALE

Impostazione simbolica – Logic Theorist

DIFFICOLTA'



- Elevato numero di regole di derivazione

- Richiesta elevata capacità calcolo

- Capacità del cervello di parallelizzare: computer non allo stesso livello

NASCITA DELL'INTELLIGENZA ARTIFICIALE

Frank Rosenblatt

Impostazione connessionista

Considera l'intelligenza come una proprietà funzionale del cervello biologico e, per simularla, cerca di riprodurre la struttura del cervello, ispirandosi al funzionamento del nostro sistema nervoso.

Prime applicazioni: sviluppo delle Reti Neurali Artificiali (RNA), i primi programmi capaci di imparare in modo autonomo.

NASCITA DELL'INTELLIGENZA ARTIFICIALE

Impostazione connessionista – Reti Neurali Artificiali (RNA)

DIFFICOLTA'

Limiti tecnologici legati alla realizzazione del neurone artificiale e delle RNA.

NASCITA DELL'INTELLIGENZA ARTIFICIALE

«INVERNO» DELL'IA
1970-1985

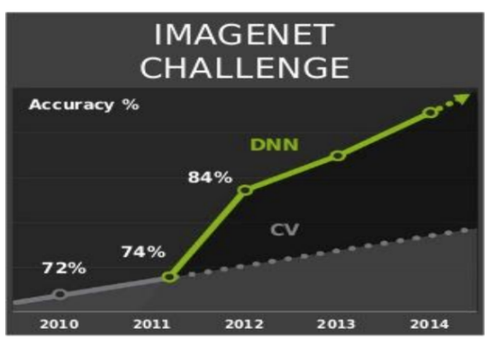
- Eccessiva complessità e potenza di calcolo necessaria;
- Limiti tecnologici legati alla realizzazione del neurone artificiale e delle RNA.

SVILUPPI DELL'INTELLIGENZA ARTIFICIALE

- 1993-2011 – Tempi moderni
  - Hardware sempre più potente.
  - Bayesian Networks, Intelligent Agents.
  - Classificatori robusti (SVM), Multi-classificatori (Random Forest, Adaboost)
  - Hidden Markov Models (HMM).
  - Maturità tecniche di feature extraction (hand-crafted) in diversi domini, (es. SIFT, Dictionaries & Bag of Words).
  - Deep Blue, Watson, Darpa Grand Challenge (guida automatica).
  - Successi in numerose discipline: visione, sistemi biometrici, riconoscimento del parlato, robotica, guida automatica, diagnosi mediche, data mining, motori di ricerca, videogames.

SVILUPPI DELL'INTELLIGENZA ARTIFICIALE

- 2011-oggi – Deep learning
  - CNN (Convolutional Neural Network) introdotte da Yan LeCun nel 1989, ma risultati inferiori ad altre tecniche: mancavano due ingredienti fondamentali, big data & potenza calcolo, grazie ai quali è possibile addestrare reti con molti livelli (deep) e milioni di parametri.
  - Nel 2012 rivoluzione in Computer Vision: una CNN denominata AlexNet vince (con ampio margine) ImageNet challenge: object classification and detection su milioni di immagini e 1000 classi.

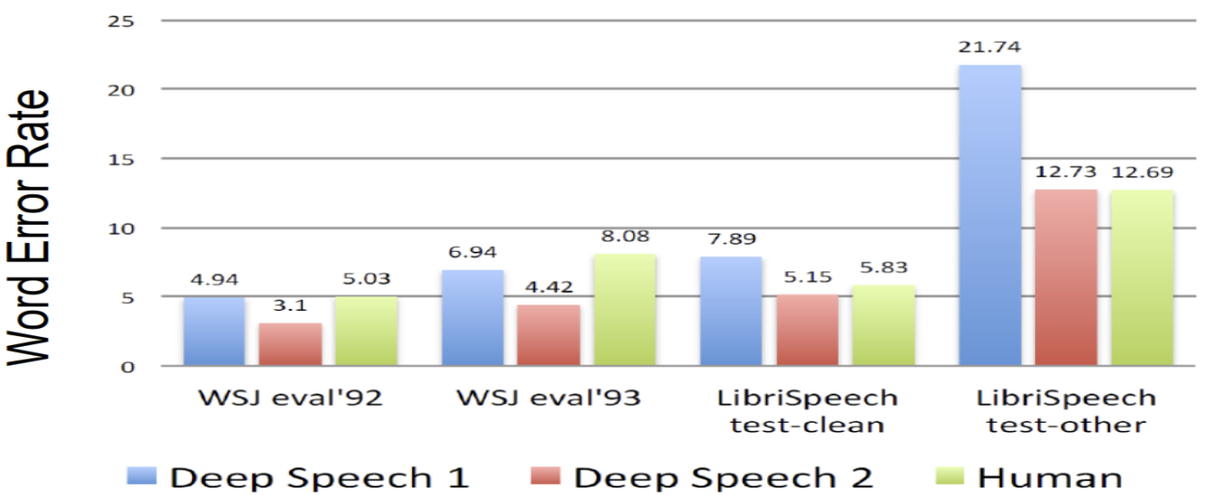


- Google acquisisce la tecnologia, ingaggia gli autori (Goffrey Hinton + Alex Krizhevsky: Univ. Toronto) e in sei mesi la incorpora nei propri prodotti (es. Google – Immagini, Street view).

SVILUPPI DELL'INTELLIGENZA ARTIFICIALE

- 2016 - Speech Recognition (es: Siri, Google Now…) in lingua inglese ha oramai raggiunto e superato prestazioni umane (ref. Baidu - Deep Speech 2).

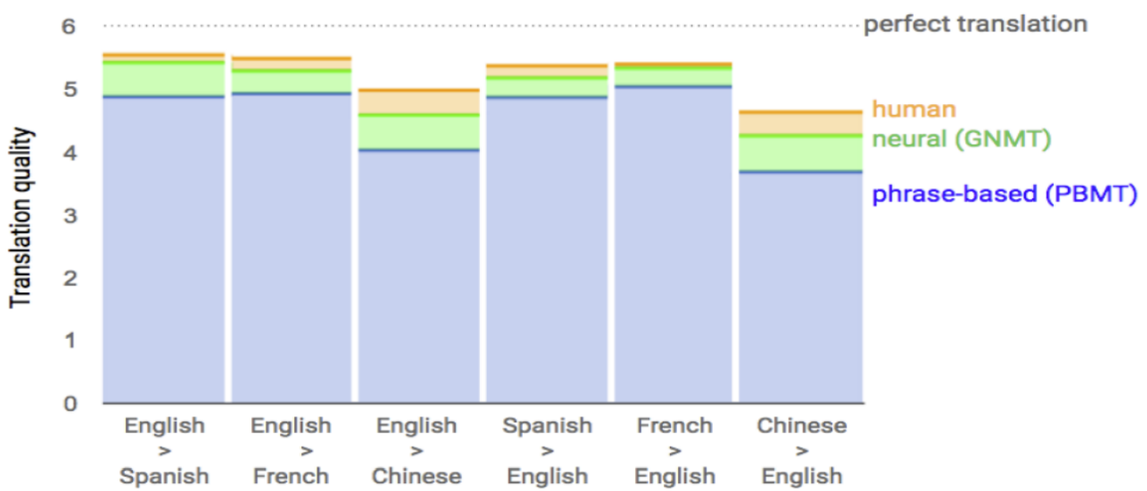
  > 10,000 ore di parlato (milioni di utenti) per il training



SVILUPPI DELL'INTELLIGENZA ARTIFICIALE

- 2016 – Language Translation per alcune lingue eguaglia prestazioni umane (ref. Google - Neural Machine Translation System).

  - 36 milioni di coppie di frasi per il training

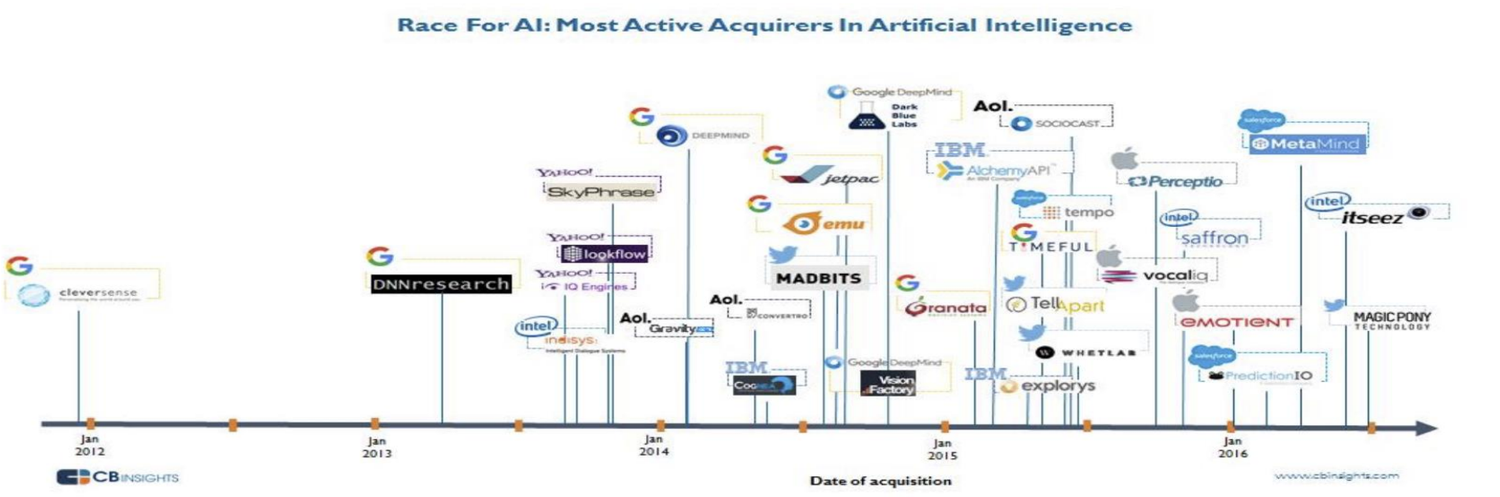


SVILUPPI DELL'INTELLIGENZA ARTIFICIALE

- A partire dal 2011, tecniche di deep learning raggiungono e superano lo stato dell'arte in molteplici applicazioni:
  - Object detection and localization (es. Yolo)
  - Face Recognition, Pedestrian Detection, Traffic Sign Detection
  - Speech Recognition, Language Translation
  - Natural Language Processing
  - Medical Image analysis (es. CheXnet)
  - Autonomous Car (es. PilotNet) and Drones (es. TrailNet)
  - Recommendation systems
  - Arts (es. Deep Dream, Style Transfer)

SVILUPPI DELL'INTELLIGENZA ARTIFICIALE

- I big dell'ICT (Microsoft, Apple, Facebook, Google, Amazon, Baidu, IBM, Nec, Samsung, Yahoo, …) investono molto nel settore reclutando talenti e acquisendo start-up. Negli USA la migrazione da Academia ad aziende (*grab of talents*) è per alcuni piuttosto preoccupante:
  - G. Hinton, A. Krizhevsky (Toronoto) → Google
  - Y. LeCun, M. Ranzato (New York) → Facebook
  - A. Ng, A. Coates (Stanford) → (ex) Baidu
  - A. Karpathy (Stanford, OpenAI) → Tesla



SVILUPPI DELL'INTELLIGENZA ARTIFICIALE

Ricercatori e specialisti del settore sono molto richiesti e ben pagati (New York Times, 2018):

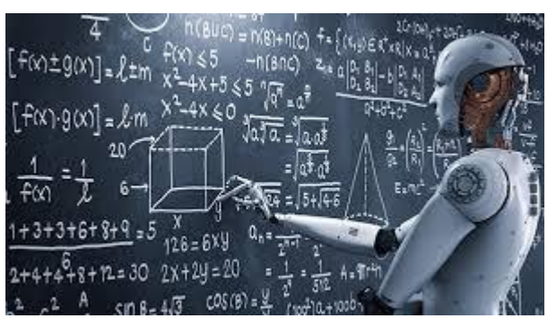
- A.I. specialists with little or no industry experience can make between $300,000 and $500,000 a year in salary and stock. Top names can receive compensation packages that extend into the millions.
- At DeepMind, a London A.I. lab now owned by Google, costs for 400 employees totaled $138 million in 2016, according to the company's annual financial filings in Britain. That translates to $345,000 per employee, including researchers and other staff.
- OpenAI paid its top researcher, Ilya Sutskever, more than $1.9 million in 2016. It paid another leading researcher, Ian Goodfellow, more than $800,000 — even though he was not hired until March of that year. Both were recruited from Google.

SVILUPPI DELL'INTELLIGENZA ARTIFICIALE
TRE FILONI

Ragionamento e acquisizione della conoscenza
«simulare ragionamento e processo decisionale, con dati incompleti e ambigui»

Comunicazione e interazione con ambiente, esseri viventi compresi
«macchina vede e riconosce gli oggetti, comprende il linguaggio umano e comunica con esseri umani: robot, NLP»

Apprendimento
«principalmente basato sulle RNA e sulla disponibilità di dati»



PERICOLI DELL'INTELLIGENZA ARTIFICIALE

Bletchley Park

Alan Turing (Matematico inglese)

Irvin John Good (Matematico inglese)

1965

«Definiamo super intelligente una macchina che può superare le capacità intellettuali di un essere umano»

«questa macchina sarà in grado di progettare macchine migliori di se stessa»

«la prima macchina super intelligente sarà l'ultima che l'uomo realizzerà»

PERICOLI DELL'INTELLIGENZA ARTIFICIALE

Nick Bostrom (Filosofo Università Oxford)

2014

«TESI ORTOGONALITA': una Intelligenza Artificiale, intesa come capace di fare previsioni, pianificazioni e in generale un ragionamento mezzi/fin avrà scopi compatibili con quelli umani»

«TESI CONVERGENZA STRUMENTALE: ci si aspetta che una Intelligenza, seppur artificiale, condivida con quella umana alcuni valori quali: autoconservazione, l'automiglioramento e l'interesse ad acquisire risorse ossia valori per raggiungere scopi, qualsiasi essi siano.

«**CONCLUSIONE**: si potrebbe verificare che la IA concentri i suoi poteri per raccogliere tutte le risorse disponibili per garantire la propria conservazione e costante proprio miglioramento…»

PERICOLI DELL'INTELLIGENZA ARTIFICIALE

2014 Nick Bostrom (Filosofo Università Oxford)

1965 Irvin John Good (Matematico inglese)

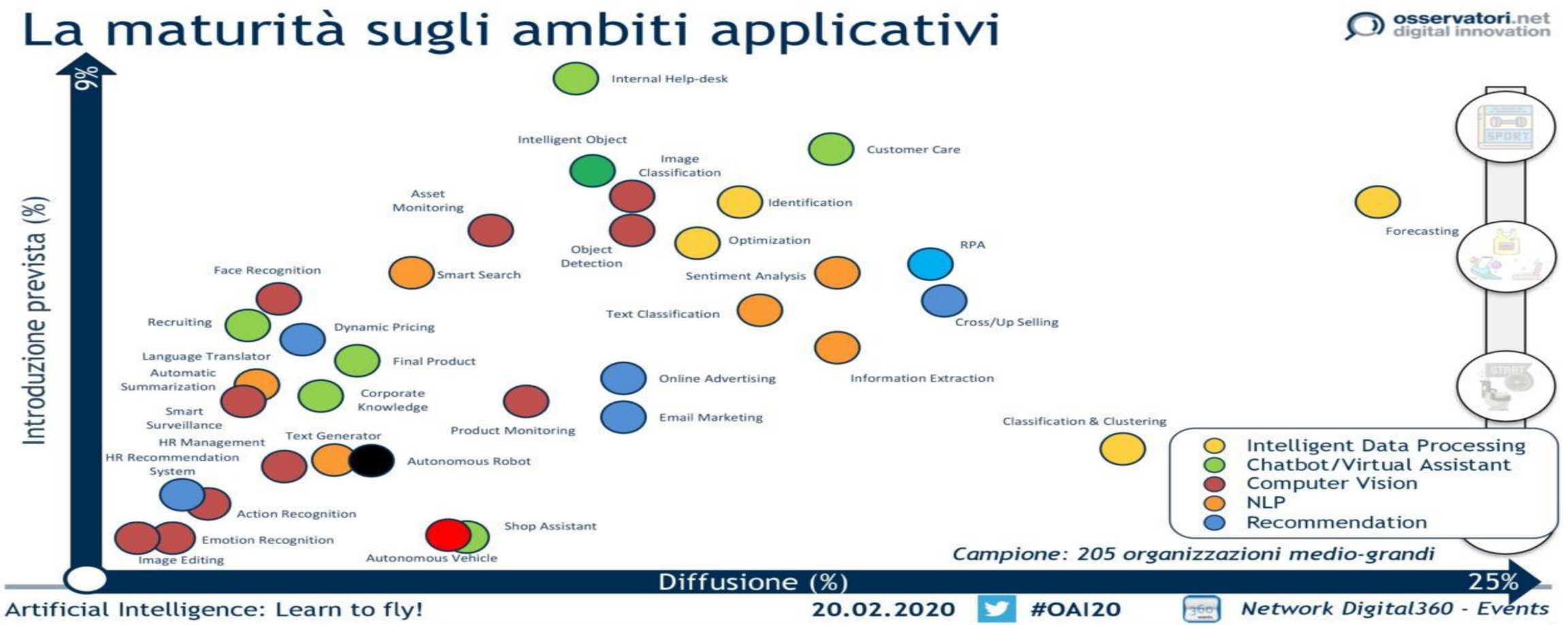
ALTRI PERICOLI?

PERICOLI DELL'INTELLIGENZA ARTIFICIALE

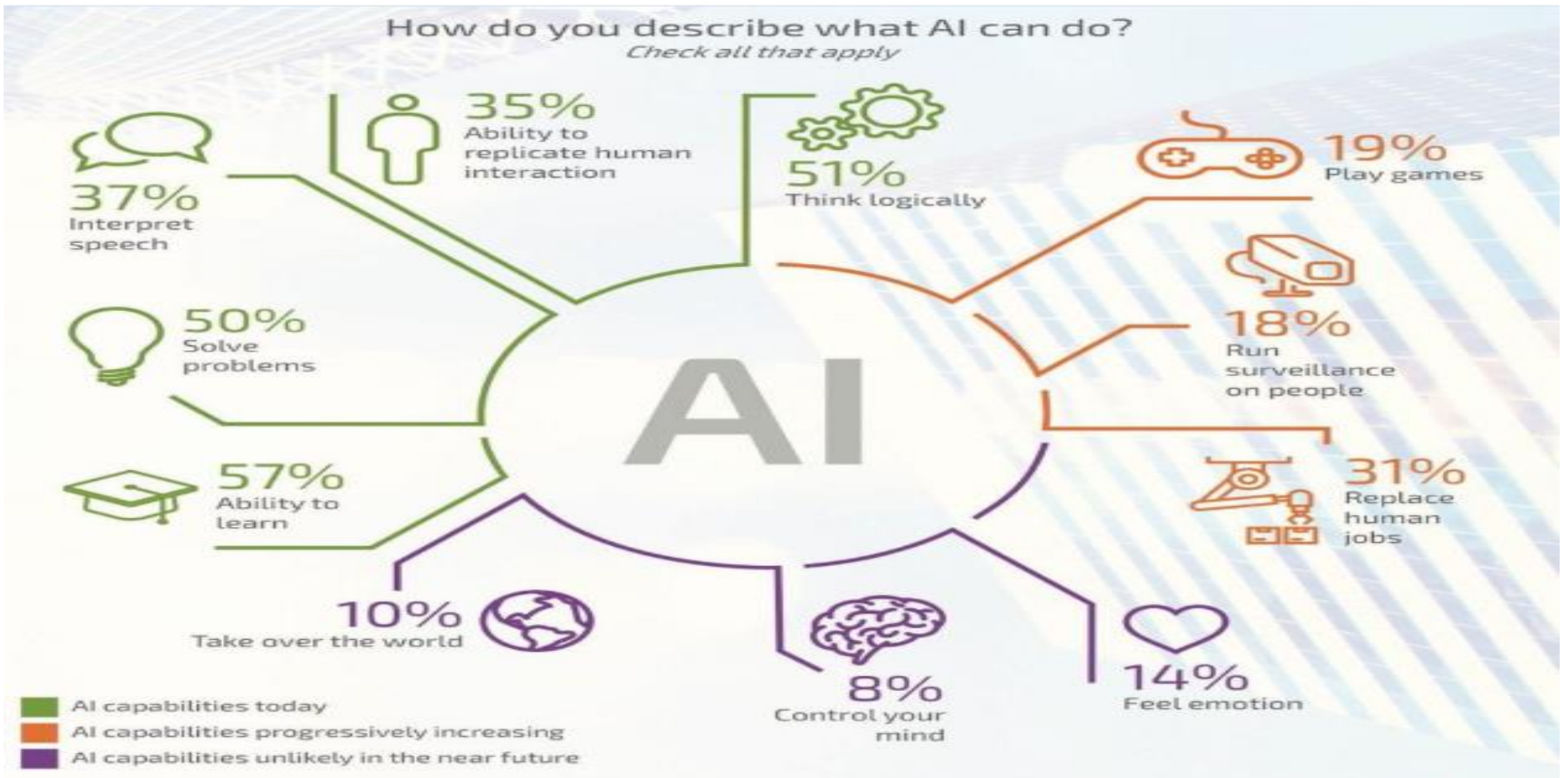
2014

DOCUMENTO con cui sostengono IA ma avvertono che i sistemi «devono fare quello che noi chiediamo che facciano»

INTELLIGENZA ARTIFICIALE

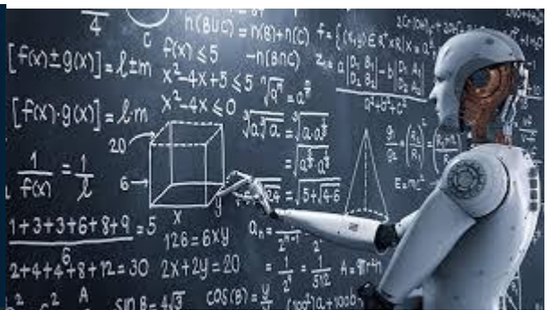


INTELLIGENZA ARTIFICIALE



«ARCHITETTURA» DELL'INTELLIGENZA ARTIFICIALE

Ragionamento e acquisizione della conoscenza



Apprendimento

Comunicazione e interazione con ambiente, esseri viventi compresi

SINTETIZZANDO CONVERGENZA SU RNA

RIFERIMENTI NAZIONALI

Programma Strategico Intelligenza Artificiale 2022-2024

Governo Italiano

a cura del Ministero dell'Università e della Ricerca, del Ministero dello Sviluppo Economico e del Ministro per l'Innovazione tecnologica e la Transizione Digitale

Per la redazione del Programma strategico per l'Intelligenza Artificiale si ringrazia il gruppo di lavoro sulla Strategia Nazionale per l'Intelligenza Artificiale, composto da:

Barbara Caputo, Isabella Castiglioni, Marco Conti, Rita Cucchiara, Juan Carlos de Martin, Fosca Giannotti, Giuseppe Magnifico, Michela Milano, Giovanni Miragliotta.

RIFERIMENTI NAZIONALI

A tal fine, **per il triennio 2022-2024**, questo Programma Strategico contiene:

- **6 obiettivi:** che indicano le *ambizioni* della strategia italiana,
- **11 settori prioritari:** che indicano *dove* l'Italia intende concentrare gli investimenti,
- **3 aree di intervento:** che indicano *come* il Paese si propone di raggiungere gli obiettivi dichiarati.

Queste tre aree di intervento si delineano in:

- **Rafforzare le competenze e attrarre talenti** per sviluppare un ecosistema dell'intelligenza artificiale in Italia.
- Aumentare i **finanziamenti per la ricerca avanzata nell'IA**
- **Incentivare l'adozione dell'IA e delle sue applicazioni**, sia nella pubblica amministrazione (PA) che nei settore produttivi in generale.

RIFERIMENTI NAZIONALI

Riepilogo delle principali politiche previste

Talenti e Competenze

A.1 Rafforzare il programma Nazionale di Dottorato
Aumentare il numero di dottorati di ricerca

A.2 Attrarre e trattenere i ricercatori
Attrarre giovani ricercatori beneficiari di borse di ricerca internazionali di alto profilo come l'ERC

A.3 Rafforzare le competenze di IA nella Pubblica Amministrazione
Attivare tre cicli di nuovi corsi di dottorato specificamente progettati per le esigenze generali della PA

A.4 Promuovere corsi e carriere in materie STEM
Integrare attività, metodologie e contenuti finalizzati allo sviluppo delle materie STEM nei curricula di tutti i cicli scolastici

A.5 Espandere l'IA negli ITS ("Istituti Tecnici Superiori")
Espandere i corsi di programmazione e includere corsi e stage di IA applicata in tutti i curricula ITS

Ricerca

B.1 Rafforzare l'ecosistema italiano della ricerca sull'IA
Creare un'architettura di ricerca su base hub & spoke con competenze territoriali

B.2 Lanciare la piattaforma italiana di dati e software per la ricerca sull'IA
Creare una connessione strutturale di piattaforme nuove ed esistenti, dati e infrastrutture informatiche dedicate all'IA, con librerie open-source

B.3 Creare cattedre italiane di ricerca sull'IA
Allocare fondi specifici per un unico Principal Investigator (PI), già iscritto ad università o centri di ricerca nazionali, per favorire la collaborazione con industrie ed enti pubblici

B.4 Creare iniziative IA-PRIN per ricerca fondamentale
Promuovere bandi dedicati alla ricerca fondamentale sull'IA e sull'IA affidabile

B.5 Promuovere campioni nazionali IA multidisciplinari
Lanciare sfide su temi specifici con concorrenti valutati sulla base di risultati misurabili

B.6 Lanciare bandi di ricerca-innovazione IA per collaborazioni pubblico-private
Promuovere progetti su settori prioritari ma con proposte di libera iniziativa volte a trasferire competenze dalla ricerca alle industrie

C.1 Finanziare ricerca e applicazioni dell'IA creativa
Finanziare progetti che integrano la ricerca accademica nel campo di frontiera dell'IA creativa assieme alle sue applicazioni industriali

C.2 Promuovere progetti bilaterali per incentivare il rientro in Italia di professionisti
Lanciare bandi per progetti incentrati su temi specifici definiti da priorità italiane cofinanziati da un altro Paese con rientro in Italia di almeno un ricercatore

Applicazioni

Per le aziende

D.1 Fare dell'IA un pilastro a supporto della Transizione 4.0 delle imprese
Introdurre crediti d'imposta o voucher per l'assunzione di profili STEM; aggiornamento dell'elenco delle spese software e hardware ammissibili agli incentivi transizione 4.0

D.2 Sostenere la crescita di spin-off innovativi e start-up
Promuovere la collaborazione all'interno degli ecosistemi delle start-up: offrire appalti pubblici alle start-up per l'acquisto di beni e servizi

D.3 Promuovere il go-to-market delle tecnologie IA
Promuovere Sperimentazione Italia, uno strumento che consente sperimentazioni attraverso un'esenzione temporanea dalla normativa vigente

D.4 Supportare le imprese nella certificazione dei prodotti IA
Definire un sistema di governance nazionale a supporto della certificazione dei prodotti di IA che si affacciano sul mercato in ambiti con profilo di rischio elevato

D.5 Promuovere campagne di informazione sull'IA per le imprese
Organizzare azioni di comunicazione e sensibilizzazione sull'IA. Le campagne includeranno la diffusione del Programma strategico nazionale per l'IA agli imprenditori

Per la PA

E.1 Creare interoperabilità e dati aperti per favorire la creazione di modelli di IA
Creare interoperabilità tra le banche dati della PA e mantenere aggiornate le linee guida per Open Data riutilizzabili per modelli di IA con dataset estesi e annotati

E.2 Rafforzare le soluzioni IA nella PA e nell'ecosistema GovTech in Italia
Introdurre bandi periodici per identificare e supportare le start-up con potenziali soluzioni basate sull'intelligenza artificiale per efficientare la PA e migliorarne i servizi

E.3 Creare un dataset comune di lingua italiana per lo sviluppo dell'IA
Creare una risorsa linguistica aperta e condivisa-una raccolta strutturata di dati digitali da documenti in italiano, disponibili a tutti gratuitamente

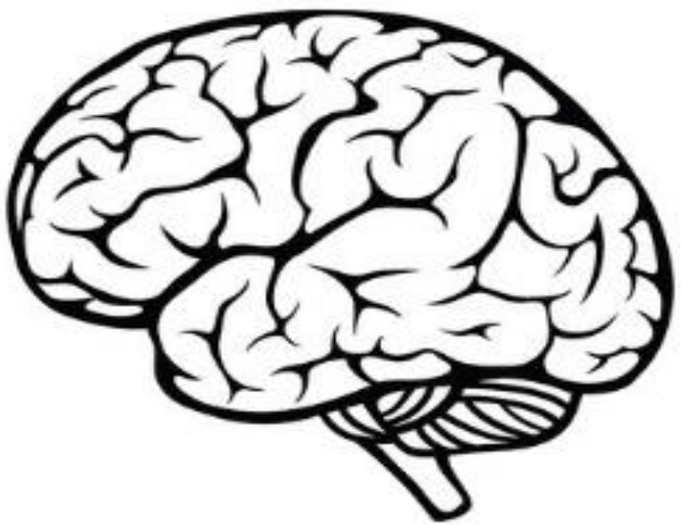
E.4 Creare banche dati e analisi basate su IA/NLP per feedback/ miglioramento dei servizi
Creare dataset annotati e anonimizzati interazioni cittadini-PA per supportare lo sviluppo/integrazione dei fornitori di IA nello sviluppo di servizi PA innovativi

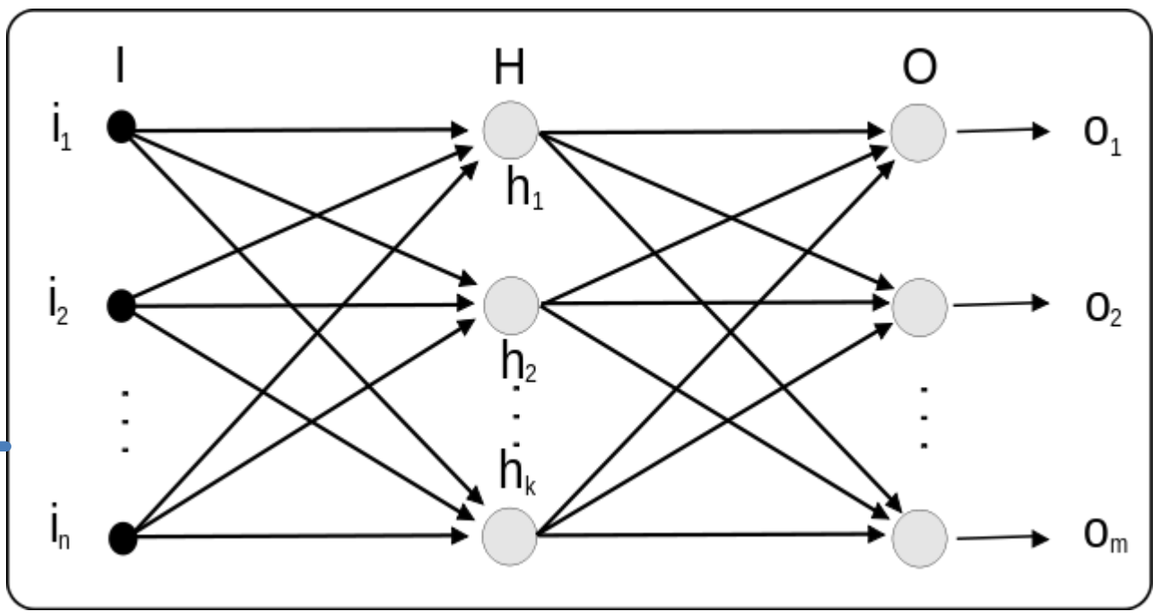
E.5 Creare banca dati IA/Computer Vision per il miglioramento dei servizi nella PA
Creare un dataset annotato di grandi dimensioni con immagini satellitari di paesaggi urbani e rurali, incluse immagini catastali digitalizzate

E.6 Introdurre tecnologie per condivisione e risoluzione casi trasversali a varie autorità
Introdurre tecnologie basate sull'IA per automatizzare lo smistamento e la preparazione delle richieste per l'elaborazione

RETI NEURALI ARTIFICIALI

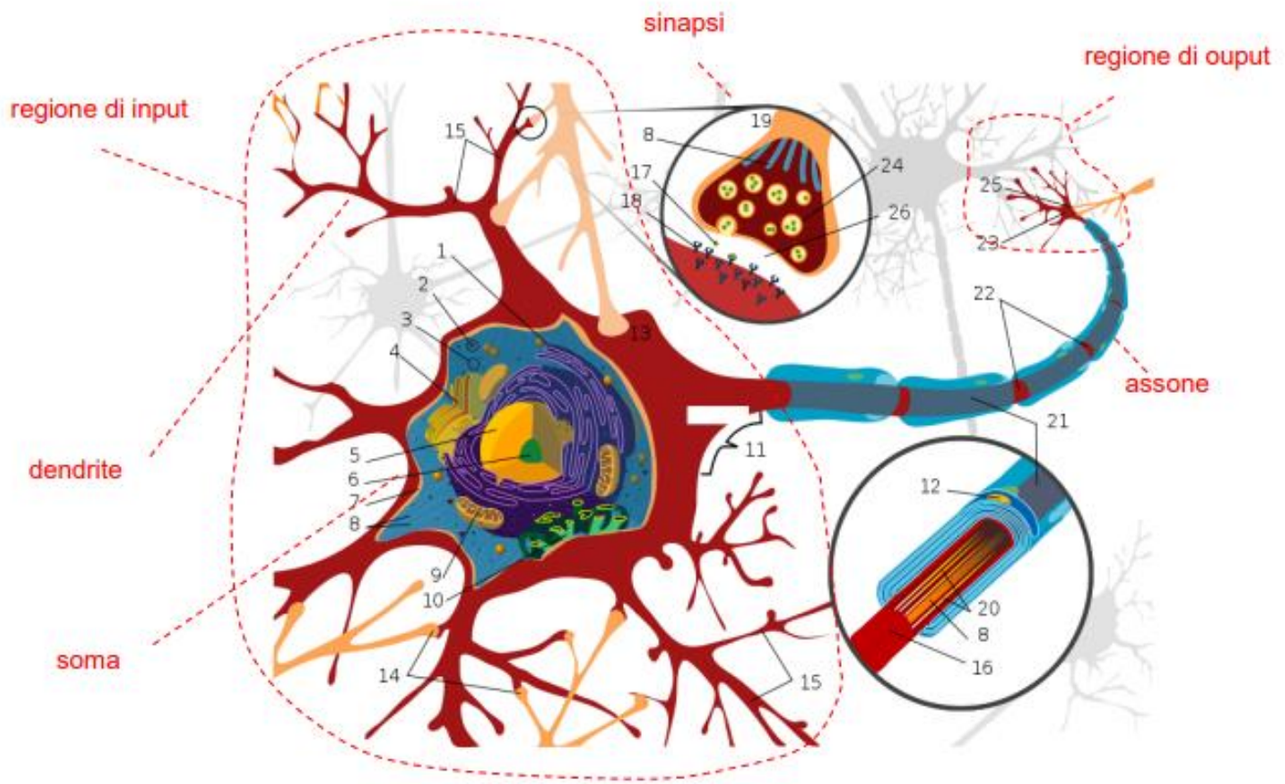
RETI NEURALI ARTIFICIALI





CON COSA RAGIONA L'UOMO?

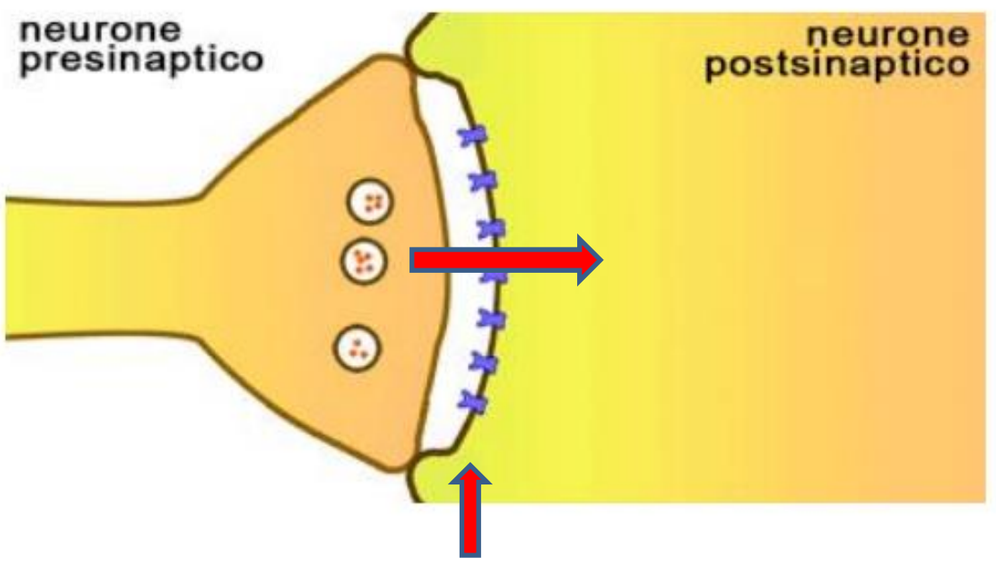
NEURONI BIOLOGICI



- Le connessioni sinaptiche o (sinapsi) agiscono come porte di collegamento per il passaggio dell'informazione tra neuroni.
- I dendriti sono fibre minori che si ramificano a partire dal corpo cellulare del neurone (detto soma). Attraverso le sinapsi i dendriti raccolgono input da neuroni afferenti e li propagano verso il soma.
- L'assone è la fibra principale che parte dal soma e si allontana da esso per portare ad altri neuroni (anche distanti) l'output.
- I neuroni sono le più importanti cellule del sistema nervoso.

NEURONI BIOLOGICI

- Il passaggio delle informazioni attraverso le sinapsi avviene con processi elettro-chimici.



- L'ingresso di ioni attraverso le sinapsi dei dendriti determina la formazione di una differenza di potenziale tra il corpo del neurone e l'esterno. Quando questo potenziale supera una certa soglia si produce uno spike (impulso): il neurone propaga un breve segnale elettrico detto potenziale d'azione lungo il proprio assone: questo potenziale determina il rilascio di ioni dalle sinapsi dell'assone.

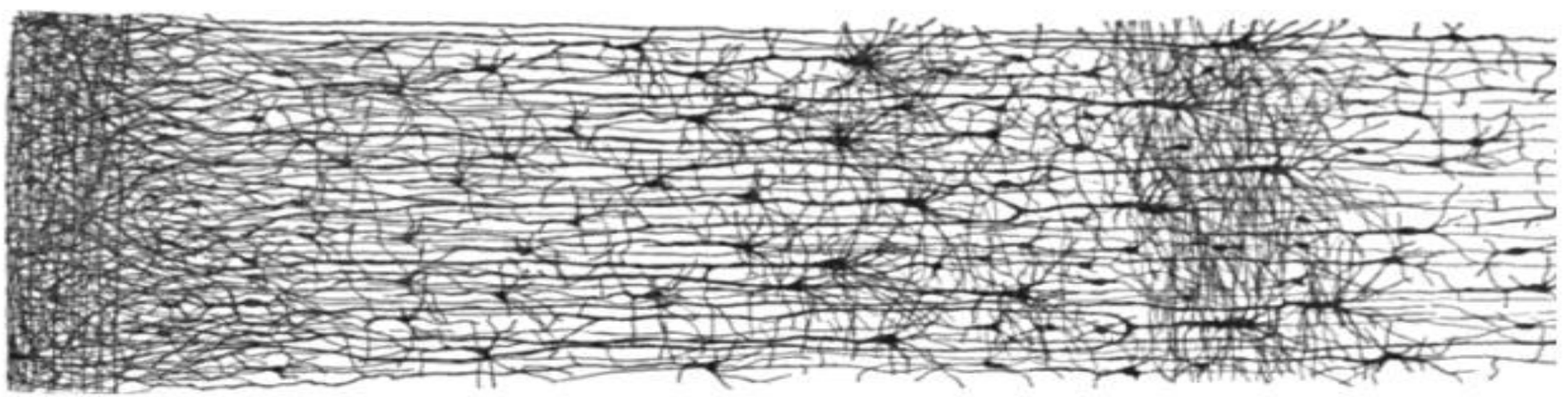
- Il reweighting delle sinapsi (ovvero la modifica della loro efficacia di trasmissione) è direttamente collegato a processi di apprendimento e memoria in accordo con la regola di Hebb.

  Hebbian rule: se due neuroni, tra loro connessi da una o più sinapsi, sono ripetutamente attivati simultaneamente allora le sinapsi che li connettono sono rinforzate.

RETI NEURALI BIOLOGICHE

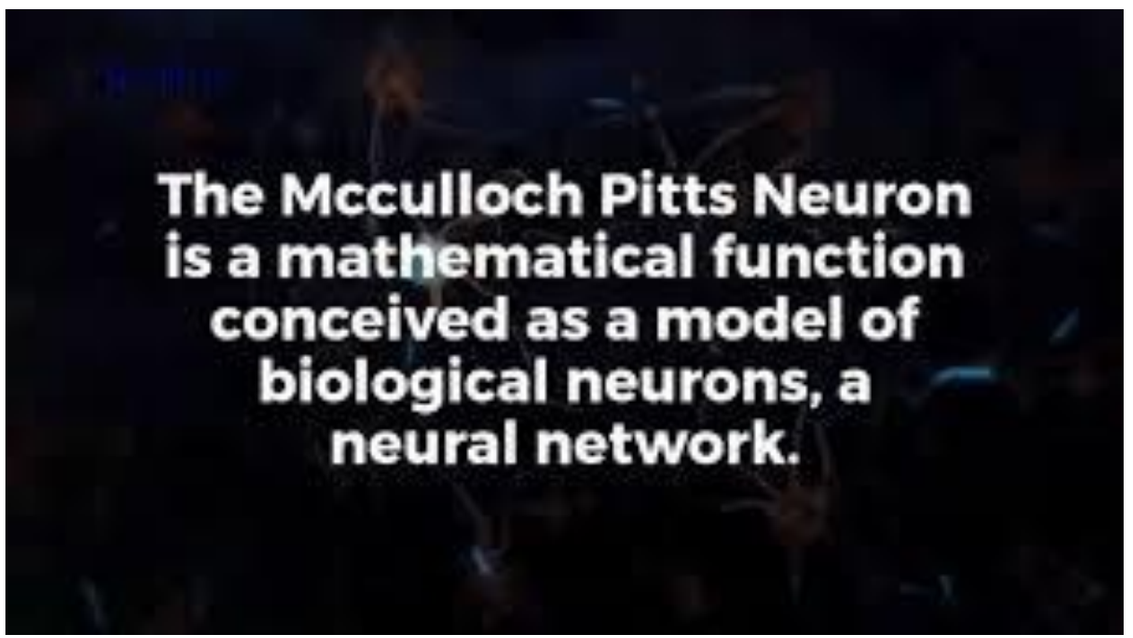
- Il cervello umano contiene circa 100 miliardi di neuroni ciascuno dei quali connesso con circa altri 1000 neuroni (10^{14} sinapsi).

- La corteccia celebrale (sede delle funzioni nobili del cervello umano) è uno strato laminare continuo di 2-4 mm, una sorta di lenzuolo che avvolge il nostro cervello formando numerose circonvoluzioni per acquisire maggiore superficie. Sebbene i neuroni siano disposti in modo «abbastanza» ordinato in livelli consecutivi, l'intreccio di dendriti e assoni ricorda una foresta fitta e impenetrabile.

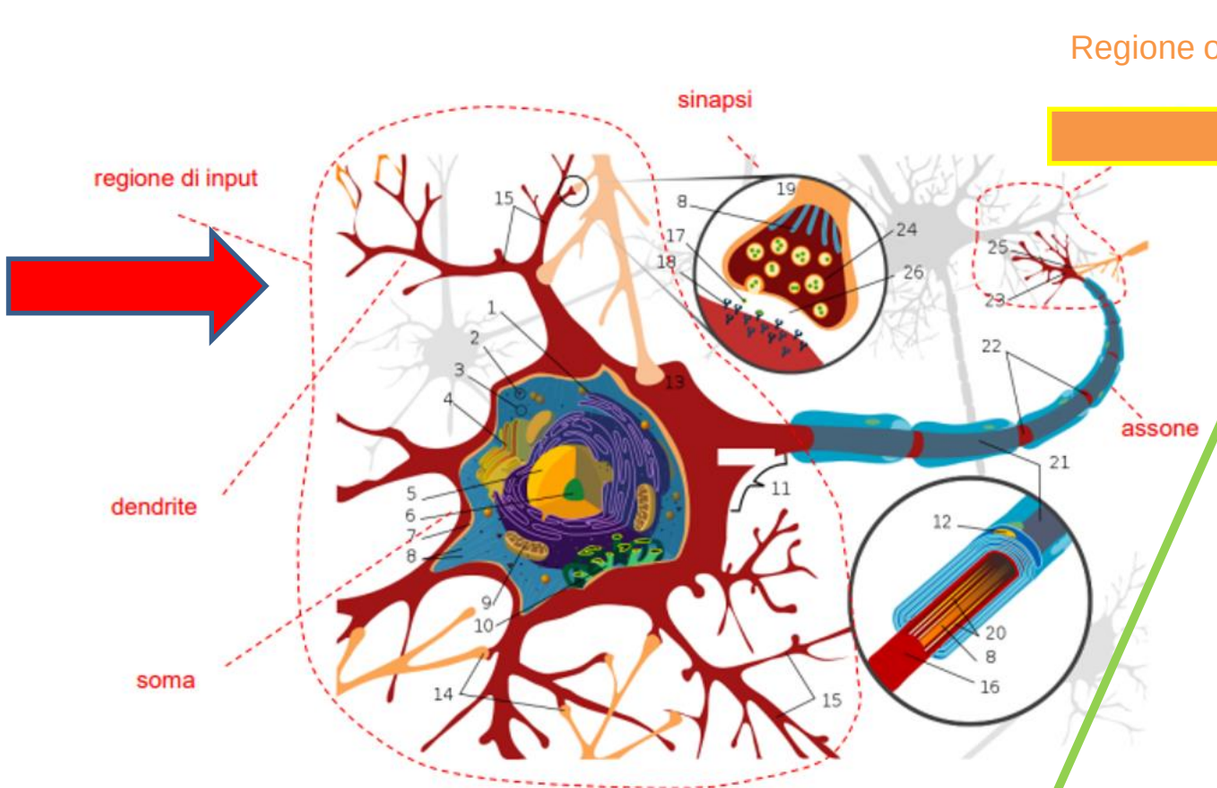


NEURONE ARTIFICIALE

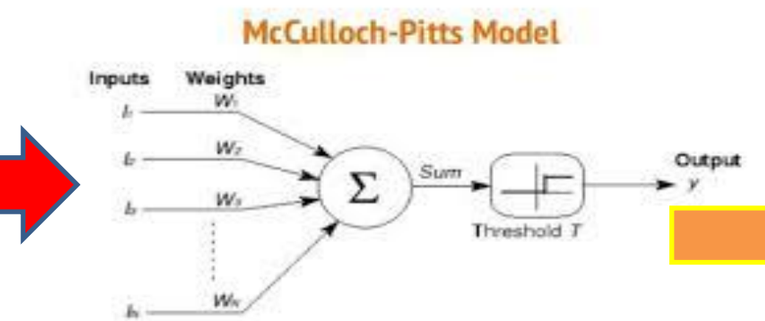
- Primo modello del 1943 di McCulloch and Pitts. Con input e output binari era in grado di eseguire computazioni logiche.

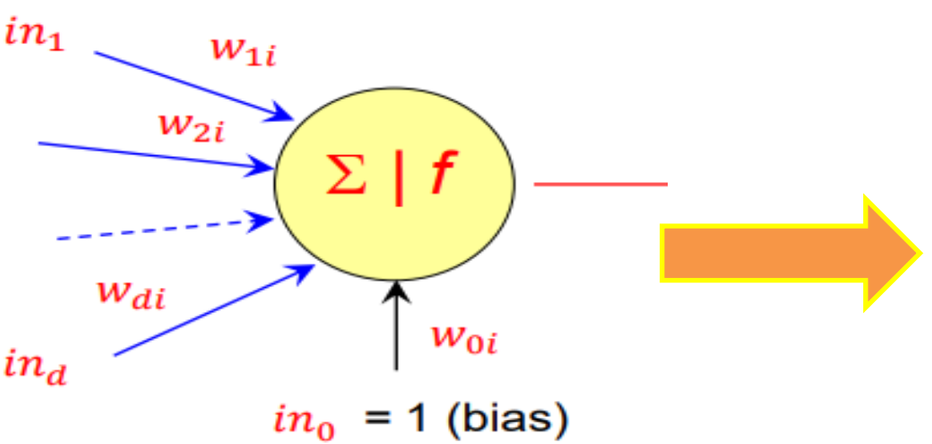


NEURONE BIOLOGICO vs NEURONE ARTIFICIALE

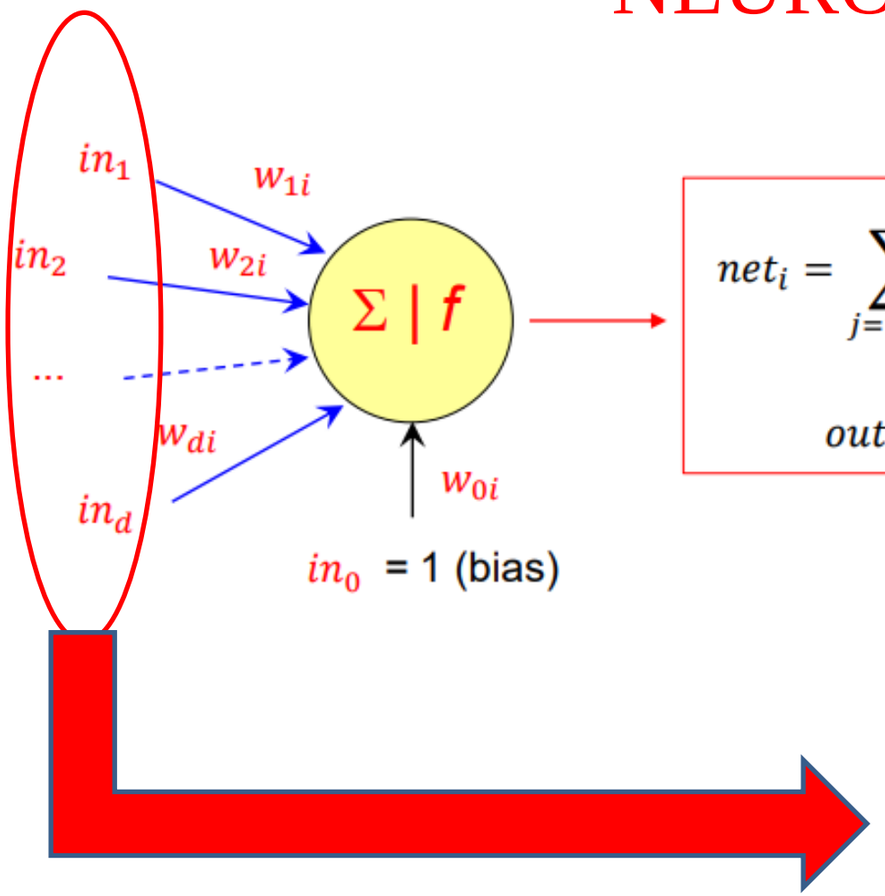


Regione output





NEURONE ARTIFICIALE

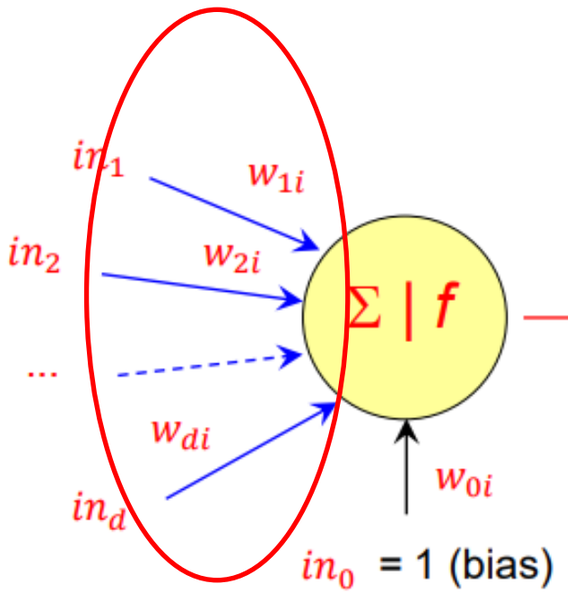


$$net_i = \sum_{j=1\ldots d} w_{ji} \cdot in_j + w_{0i}$$

$$out_i = f(net_i)$$

- in_1, in_2, … in_d sono i d ingressi che il neurone i riceve da assoni di neuroni afferenti.

NEURONE ARTIFICIALE

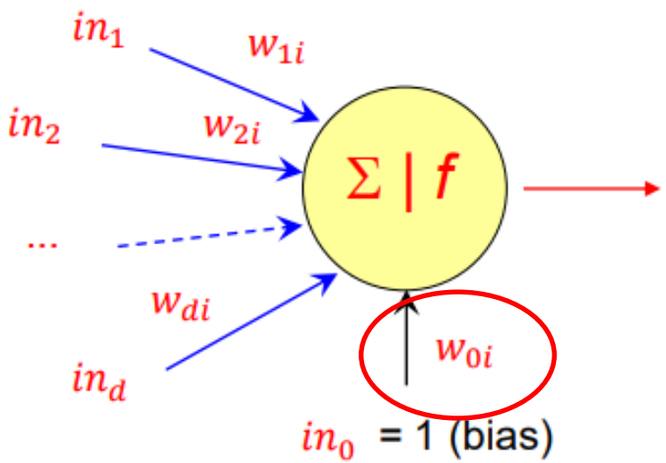


$$net_i = \sum_{j=1\ldots d} w_{ji} \cdot in_j + w_{0i}$$

$$out_i = f(net_i)$$

- w_{1i}, w_{2i}, … w_{di} sono i pesi (weight) che determinano l'efficacia delle connessioni sinaptiche dei dendriti (agiremo su questi valori durante l'apprendimento).

NEURONE ARTIFICIALE

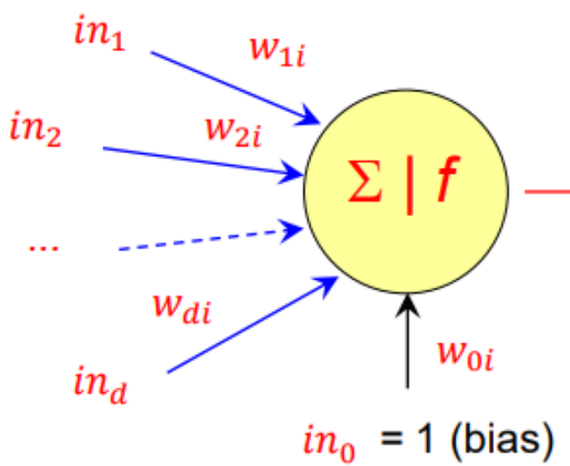


$$net_i = \sum_{j=1\ldots d} w_{ji} \cdot in_j + w_{0i}$$

$$out_i = f(net_i)$$

- w_{0i} (detto bias) è un ulteriore peso che si considera collegato a un input fittizio con valore sempre 1; questo peso è utile per «tarare» il punto di lavoro ottimale del neurone.

NEURONE ARTIFICIALE

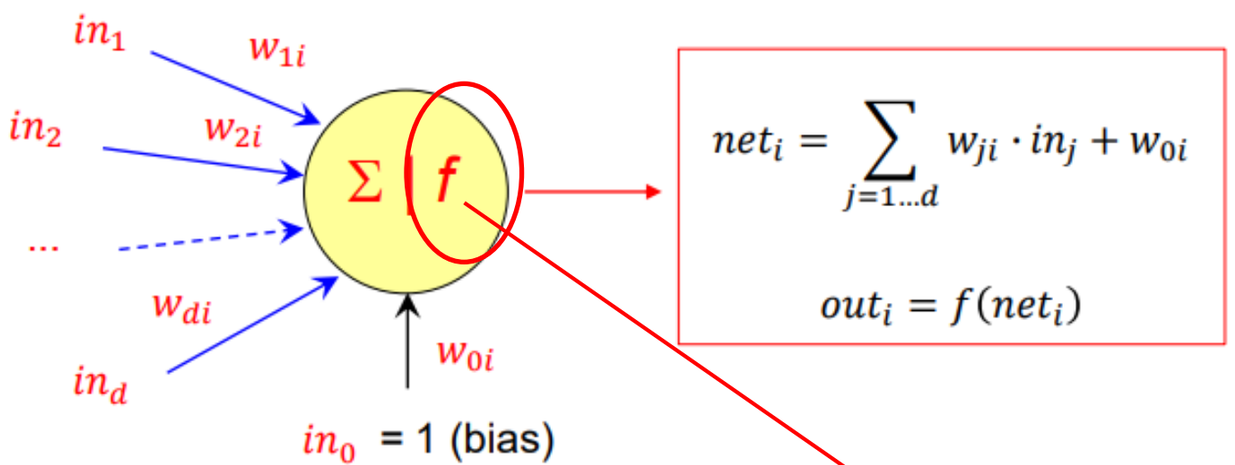


$$net_i = \sum_{j=1\ldots d} w_{ji} \cdot in_j + w_{0i}$$

$$out_i = f(net_i)$$

- net_i è il livello di eccitazione globale del neurone (potenziale interno);

NEURONE ARTIFICIALE

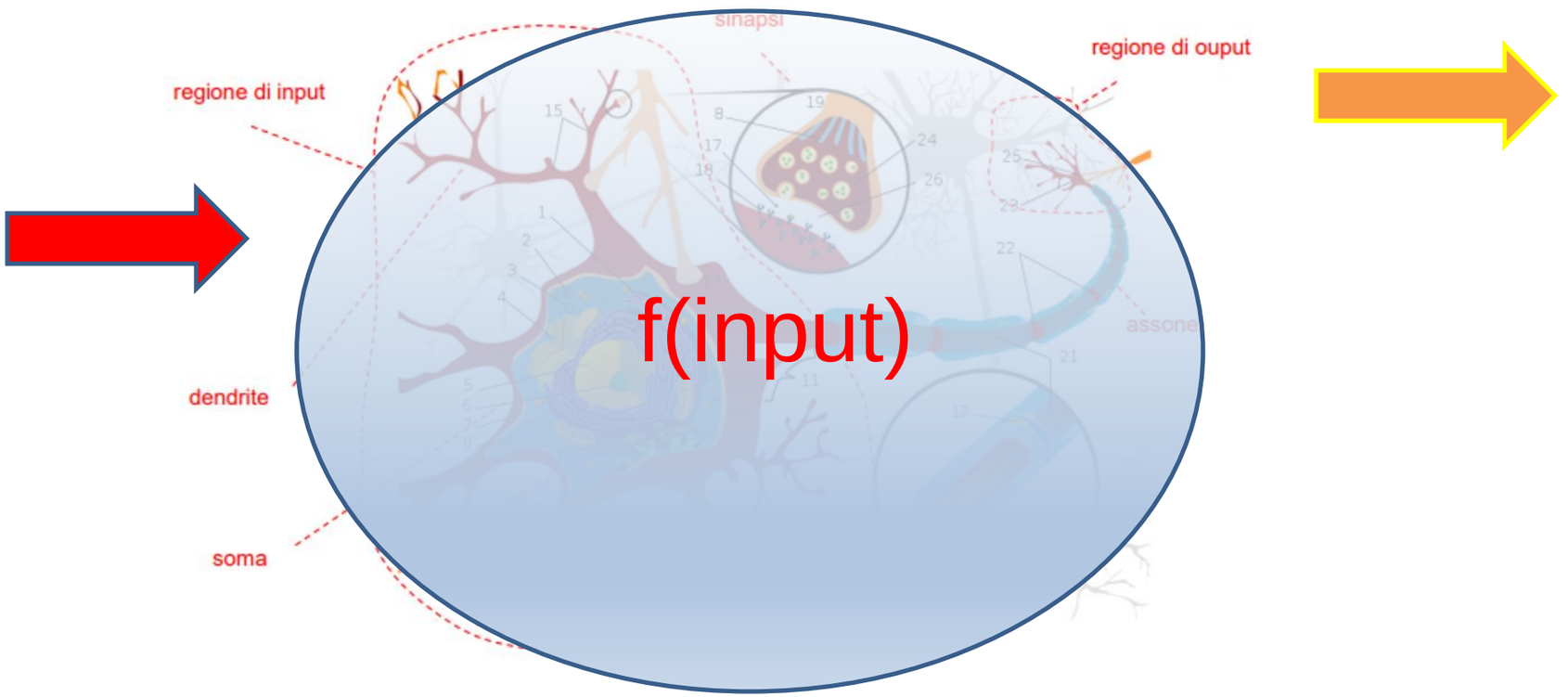


$$net_i = \sum_{j=1\ldots d} w_{ji} \cdot in_j + w_{0i}$$

$$out_i = f(net_i)$$

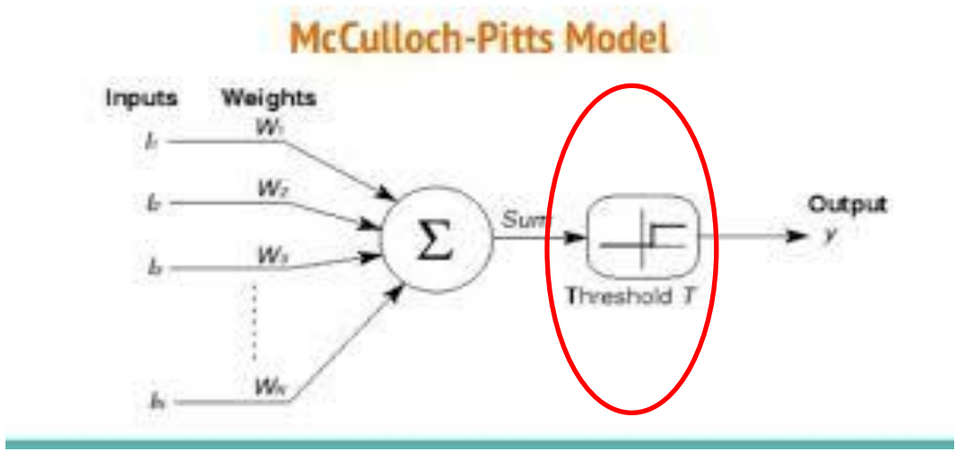
- $f(\cdot)$ è la funzione di attivazione che determina il comportamento del neurone (ovvero il suo output net_i) in funzione del suo livello di eccitazione net_i.

NEURONE BIOLOGICO: FUNZIONE ATTIVAZIONE



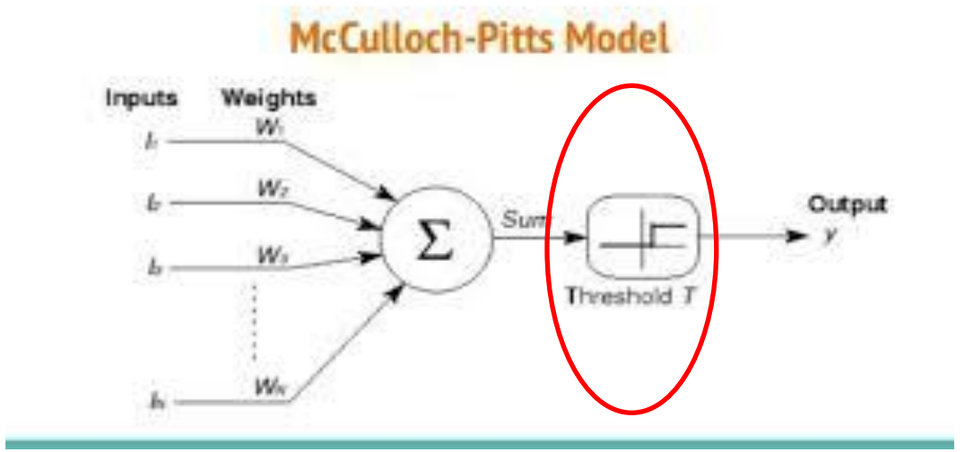
- Nei neuroni biologici $f(\cdot)$ è una funzione tutto-niente temporizzata: quando net_i supera una certa soglia, il neurone «spara» uno spike (impulso) per poi tornare a riposo.

NEURONE ARTIFICIALE: FUNZIONE ATTIVAZIONE



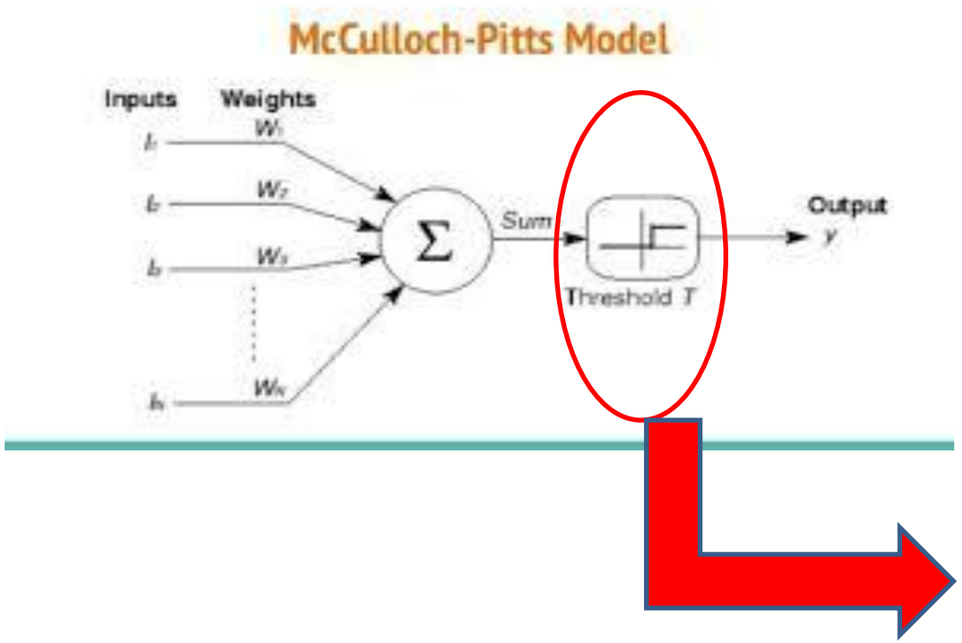
- Le reti neurali più comunemente utilizzate operano con livelli continui e $f(\cdot)$ è una funzione non-lineare ma continua e differenziabile (quasi ovunque).

NEURONE ARTIFICIALE: FUNZIONE ATTIVAZIONE



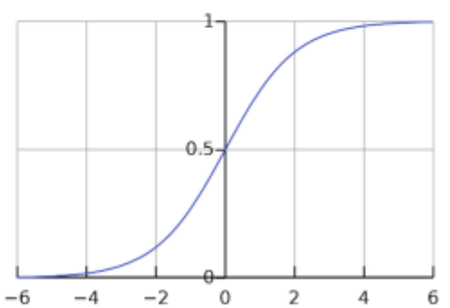
- Una delle funzioni di attivazione più comunemente utilizzata è la sigmoide nelle varianti:
  - standard logistic function (chiamata semplicemente sigmoid)
  - tangente iperbolica (tanh)

NEURONE ARTIFICIALE: FUNZIONE ATTIVAZIONE

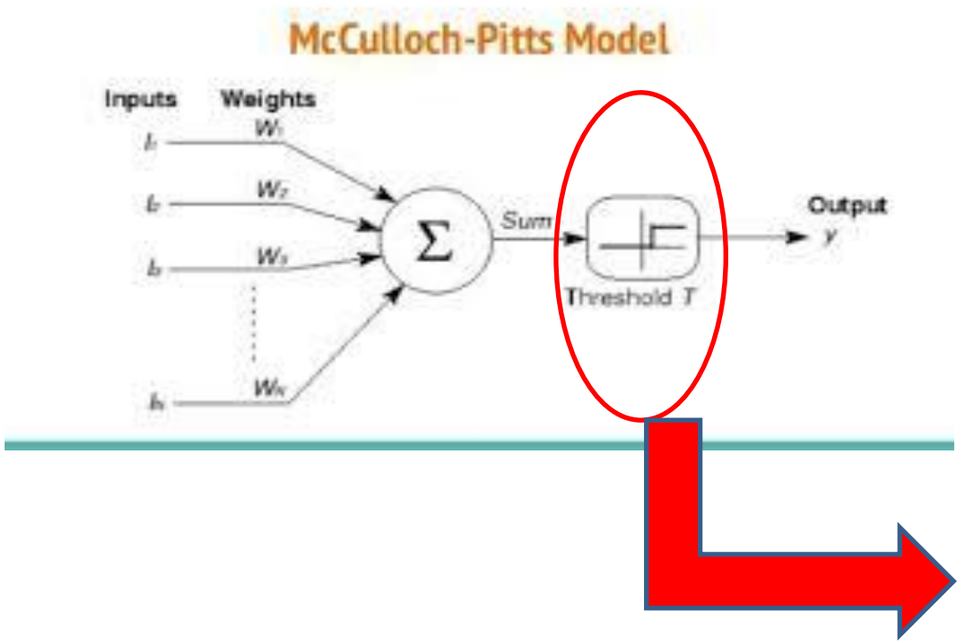


- Standard logistic function (Sigmoid), (valori in [0…1]):

$$f(net) = \sigma(net) = \frac{1}{1 + e^{-net}}$$



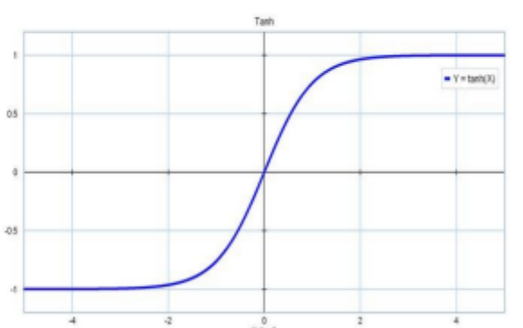
NEURONE ARTIFICIALE: FUNZIONE ATTIVAZIONE



- Tangente iperbolica (Tanh), (valori in [-1…1]):

  Può essere ottenuta dalla funzione precedente a seguito di trasformazione di scala (×2) e traslazione (-1).

$$f(net) = \tau(net) = 2\sigma(2 \cdot net) - 1$$



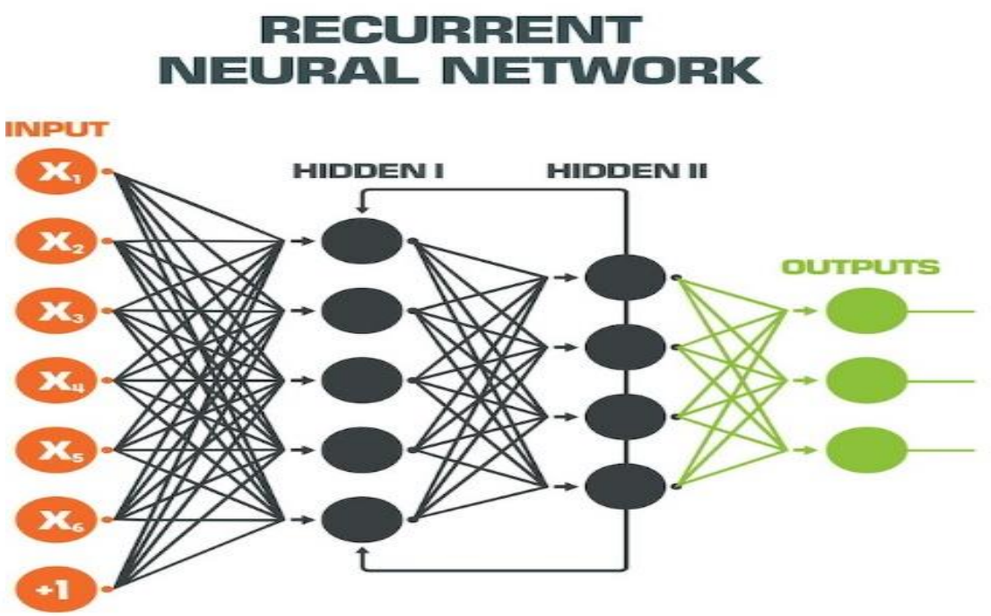
DEFINIZIONE DI RETI NEURALI ARTIFICIALI

Le reti neurali sono composte da gruppi di neuroni artificiali organizzati in livelli. Tipicamente sono presenti: un livello di input, un livello di output, e uno o più livelli intermedi o nascosti (hidden). Ogni livello contiene uno o più neuroni.

TIPOLOGIA DI RETI NEURALI ARTIFICIALI

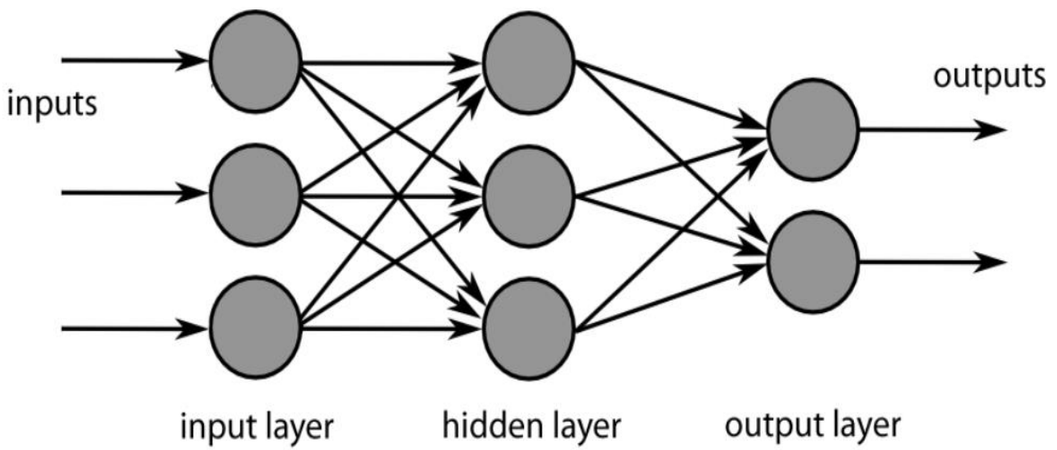
- RECURRENT

Recurrent: nelle reti ricorrenti sono previste connessioni di feedback (in genere verso neuroni dello stesso livello, ma anche all'indietro). Questo complica notevolmente il flusso delle informazioni e l'addestramento, richiedendo di considerare il comportamento in più istanti temporali (unfolding in time).



TIPOLOGIA DI RETI NEURALI ARTIFICIALI

FEED FORWARD: modello più diffuso e utilizzato

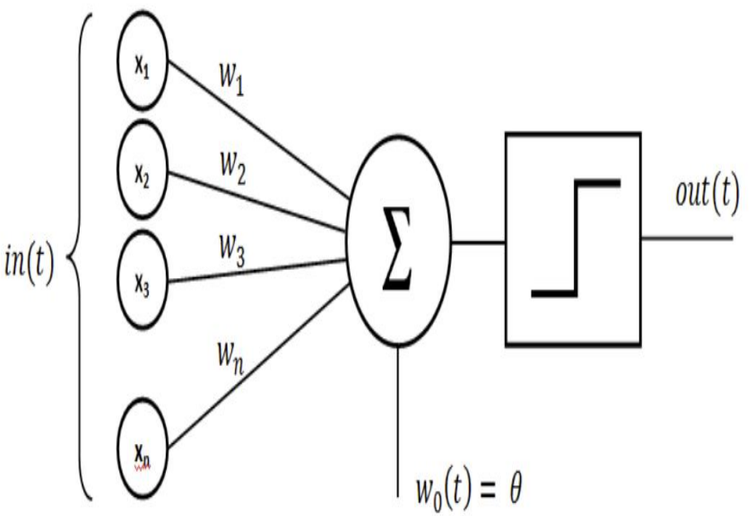


Feedforward: nelle reti feedforward («alimentazione in avanti») le connessioni collegano i neuroni di un livello con i neuroni di un livello successivo. Non sono consentite connessioni all'indietro o connessioni verso lo stesso livello. È di gran lunga il tipo di rete più utilizzata.

IL PERCETTRONE

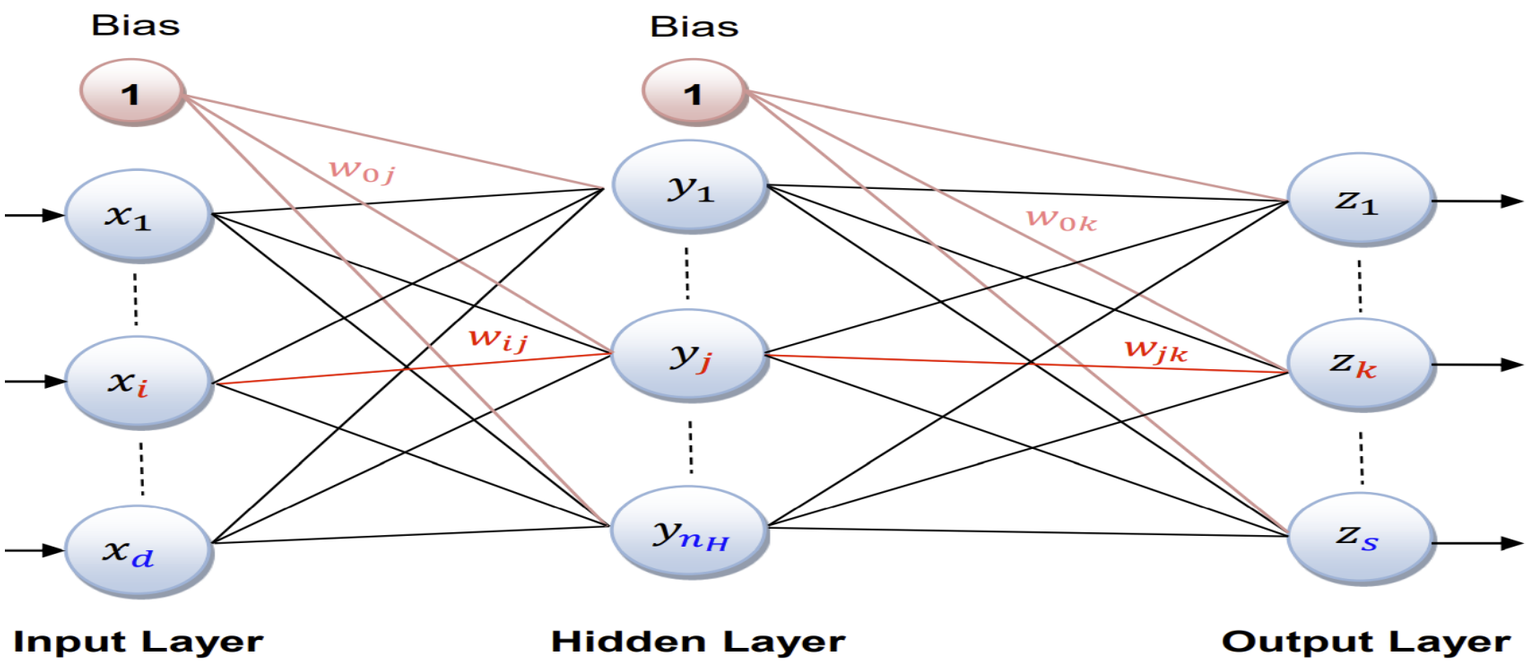
Il termine perceptron (percettrone) deriva dal modello di neurone proposto da Rosenblatt nel 1956.

Il percettrone utilizza una funzione di attivazione lineare a soglia (o scalino). Un singolo percettrone, o una rete d percettroni a due soli livelli (input e output), può essere addestrato con una semplice regola detta delta rule (ispirata alla regola di Hebb).



IL MULTI LAYER PERCEPTRON

Un Multilayer Perceptron (MLP) è una rete feedforward con almeno 3 livelli (almeno 1 hidden) e con funzioni di attivazione non lineari.

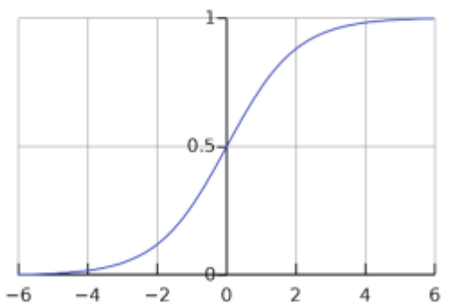


IL MULTI LAYER PERCEPTRON

Un Multilayer Perceptron (MLP) è una rete feedforward con almeno 3 livelli (almeno 1 hidden) e con funzioni di attivazione non lineari.

- Standard logistic function (Sigmoid), (valori in [0…1]):

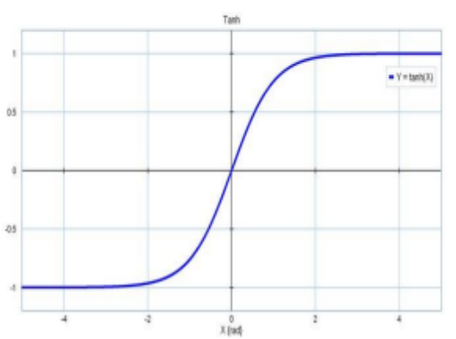
$$f(net) = \sigma(net) = \frac{1}{1 + e^{-net}}$$



- Tangente iperbolica (Tanh), (valori in [-1…1]):

Può essere ottenuta dalla funzione precedente a seguito di trasformazione di scala (×2) e traslazione (-1).

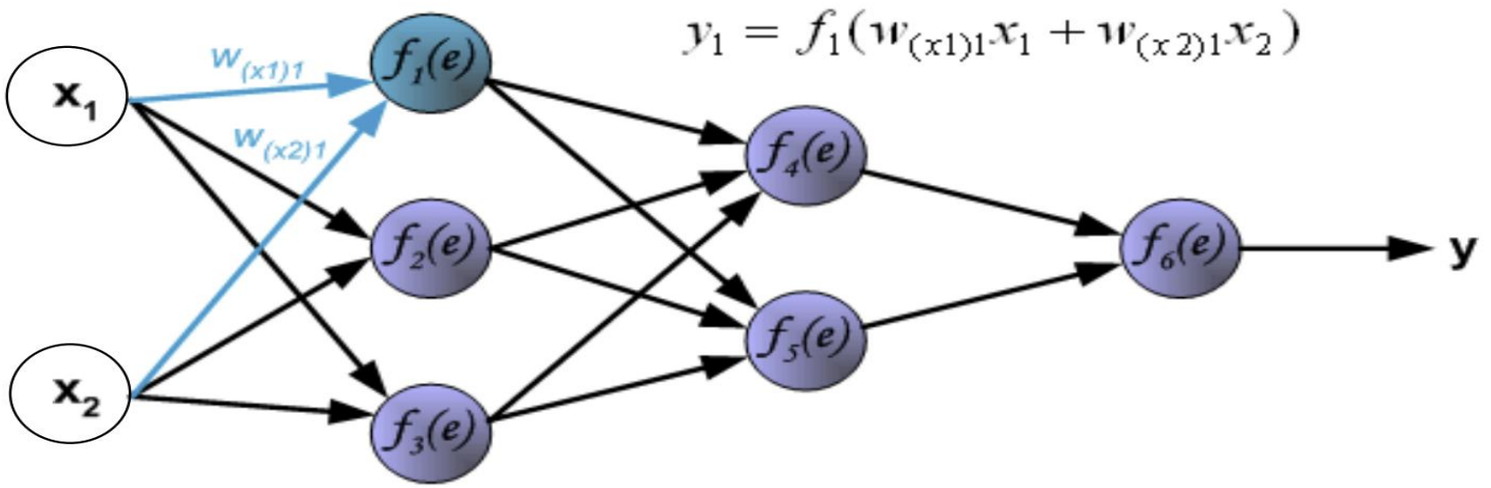
$$f(net) = \tau(net) = 2\sigma(2 \cdot net) - 1$$



IL MULTI LAYER PERCEPTRON

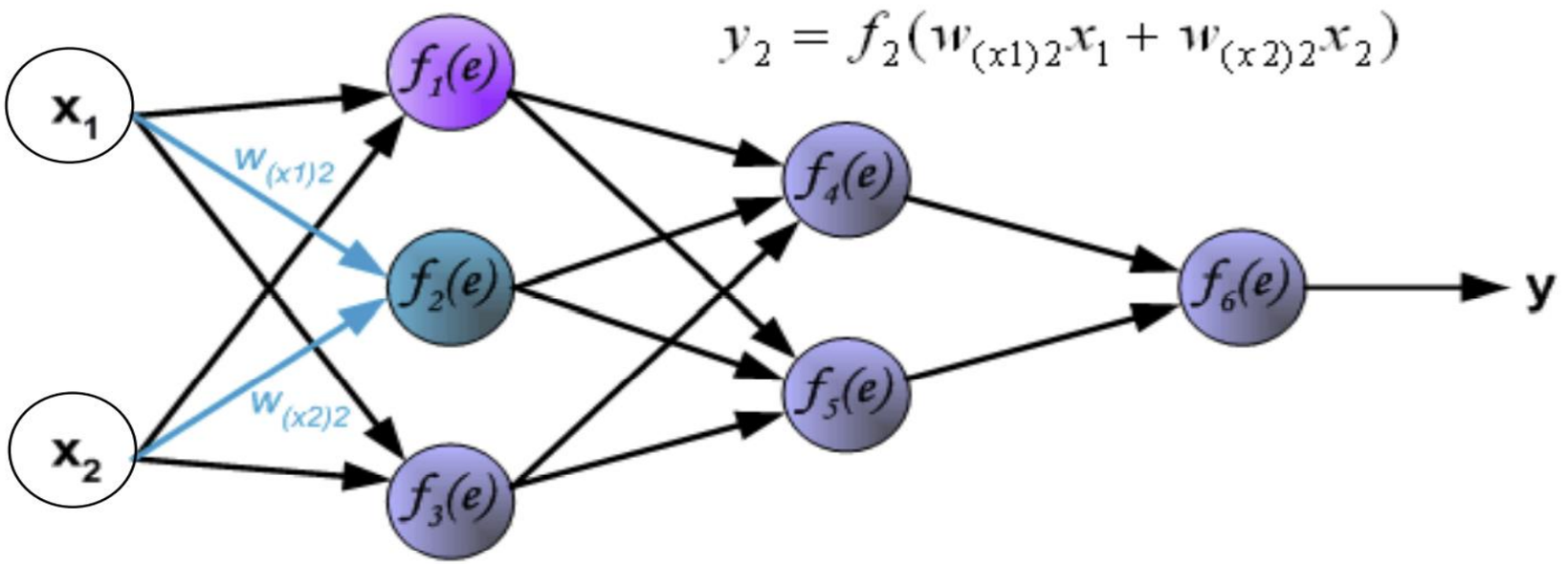
Con forward propagation (o inference) si intende la propagazione delle informazioni in avanti: dal livello di input a quello di output. Una volta addestrata, una rete neurale può semplicemente processare pattern attraverso forward propagation.

Forward propagation: esempio grafico



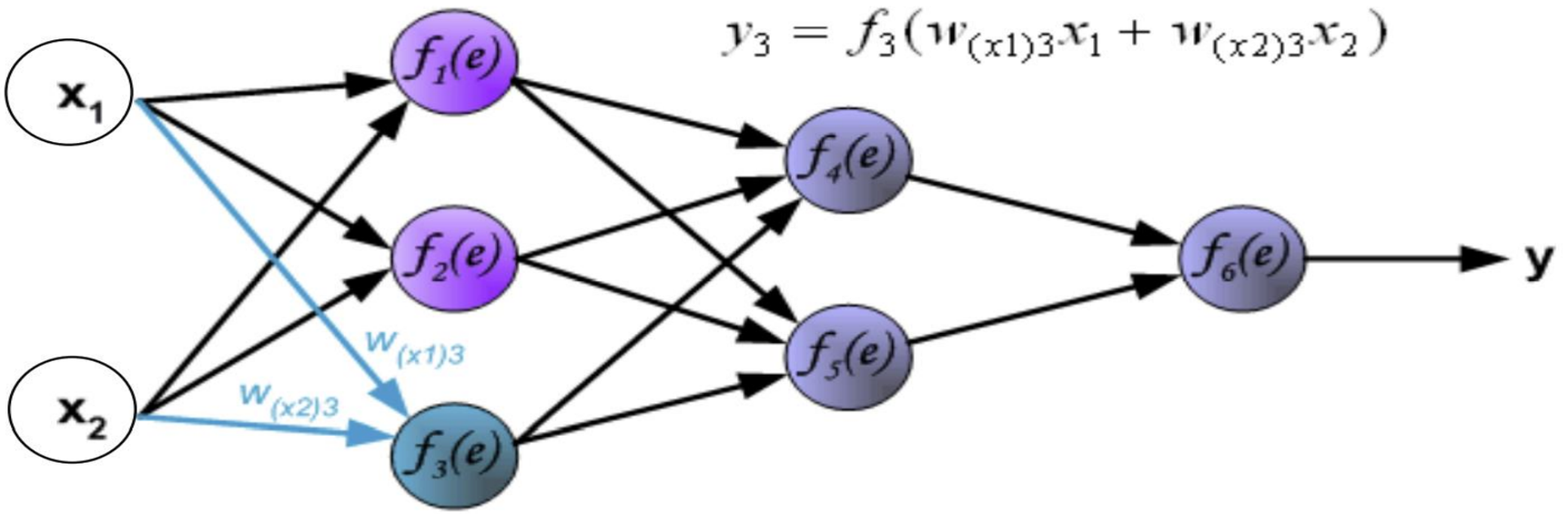
STEP 1

Forward propagation: esempio grafico



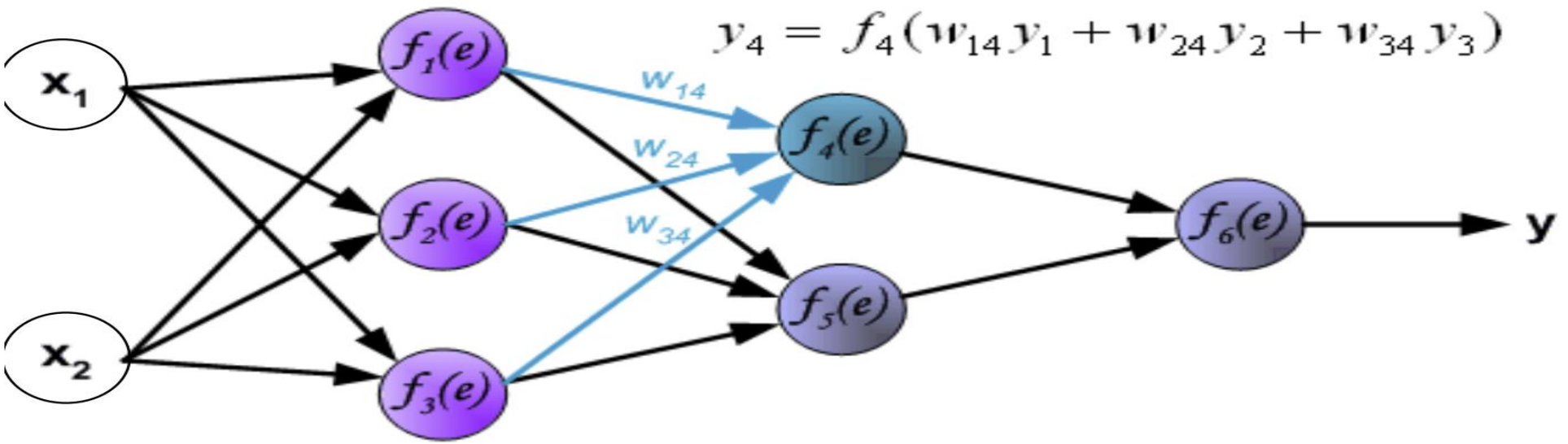
STEP 2

Forward propagation: esempio grafico



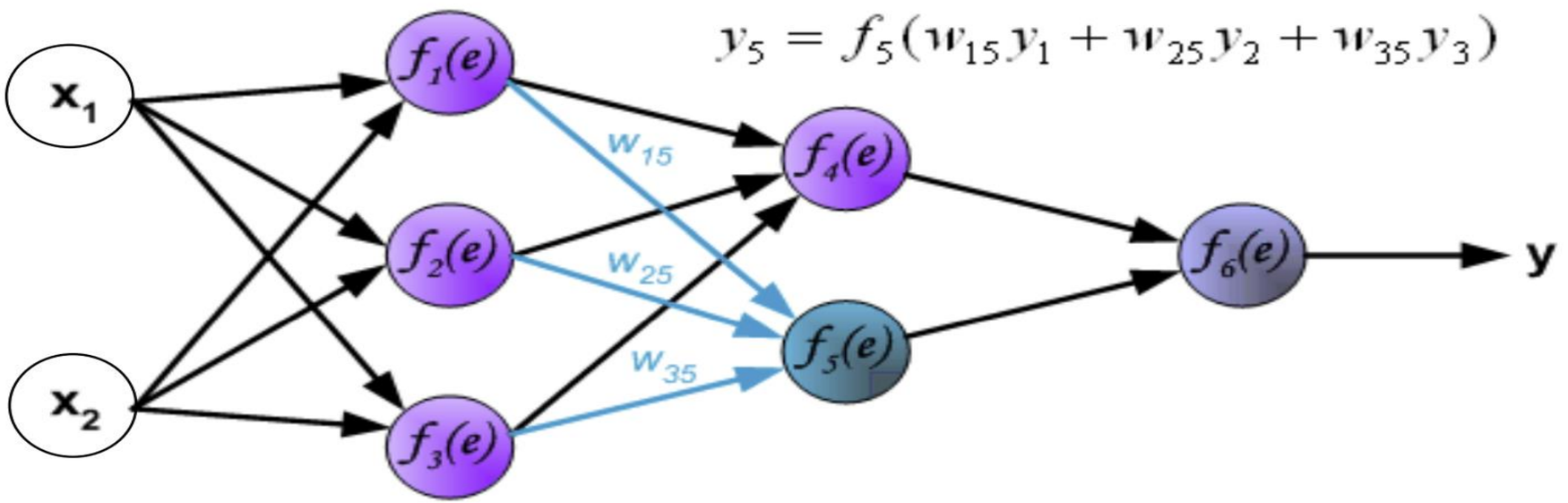
STEP 3

Forward propagation: esempio grafico



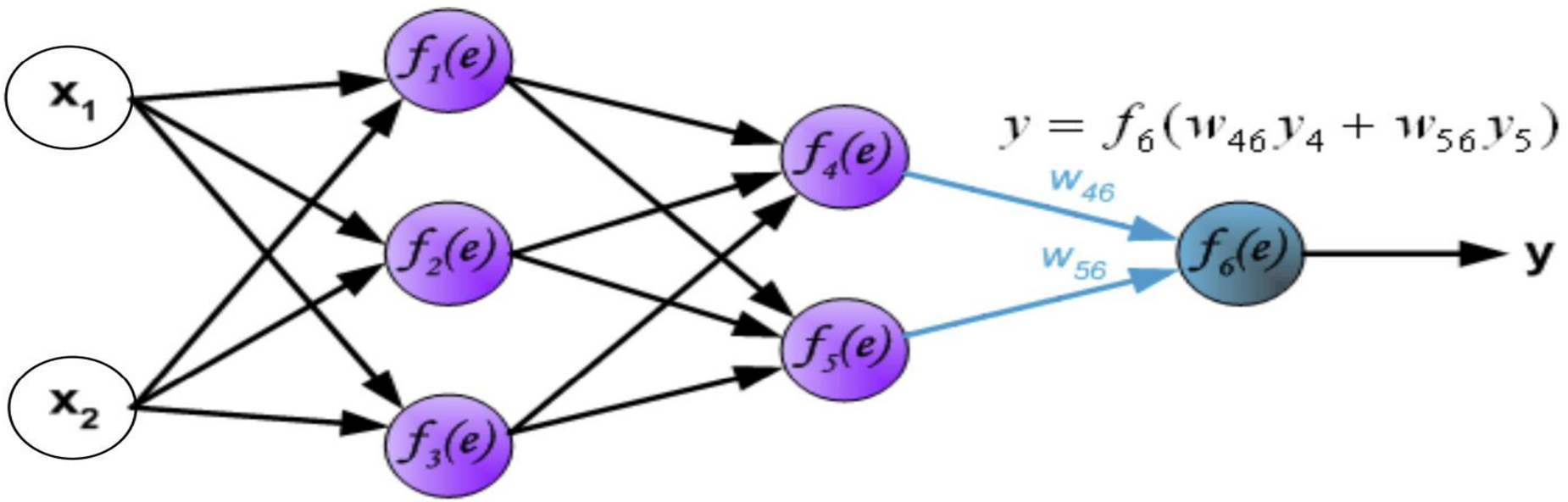
STEP 4

Forward propagation: esempio grafico



STEP 5

Forward propagation: esempio grafico

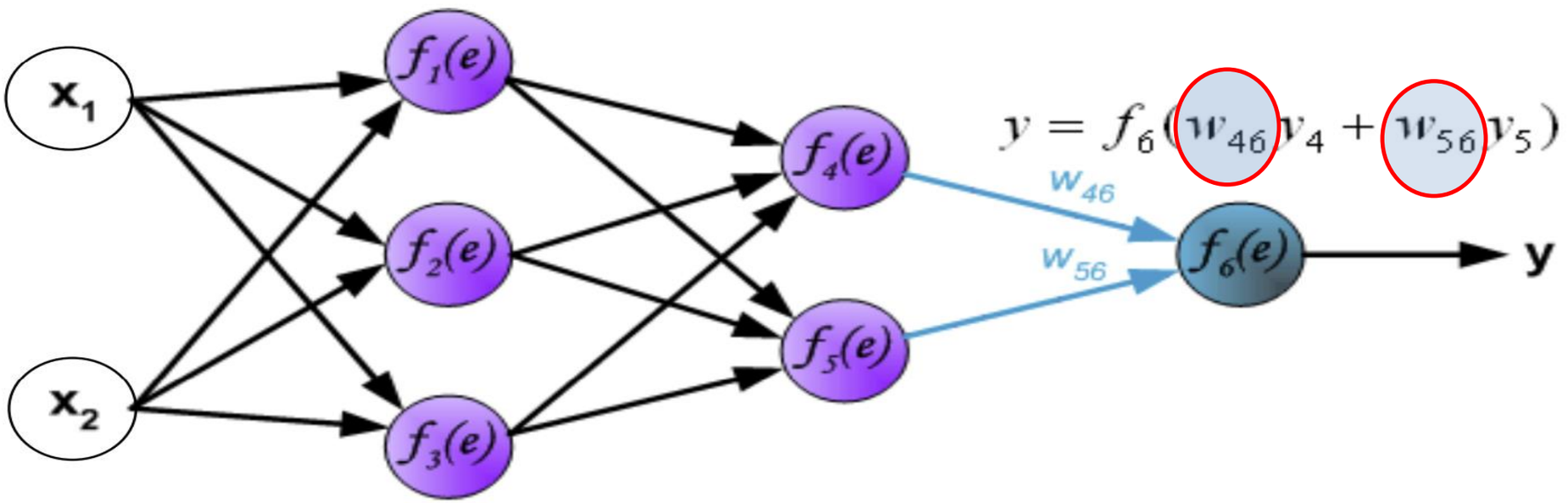


STEP 6

IL MULTI LAYER PERCEPTRON: ADDESTRAMENTO

Fissata la topologia (numero di livelli e neuroni), l'addestramento di una rete neurale consiste nel determinare il valore dei pesi **w** che determinano il mapping desiderato tra input e output.

IL MULTI LAYER PERCEPTRON: ADDESTRAMENTO



IL MULTI LAYER PERCEPTRON: MAPPING

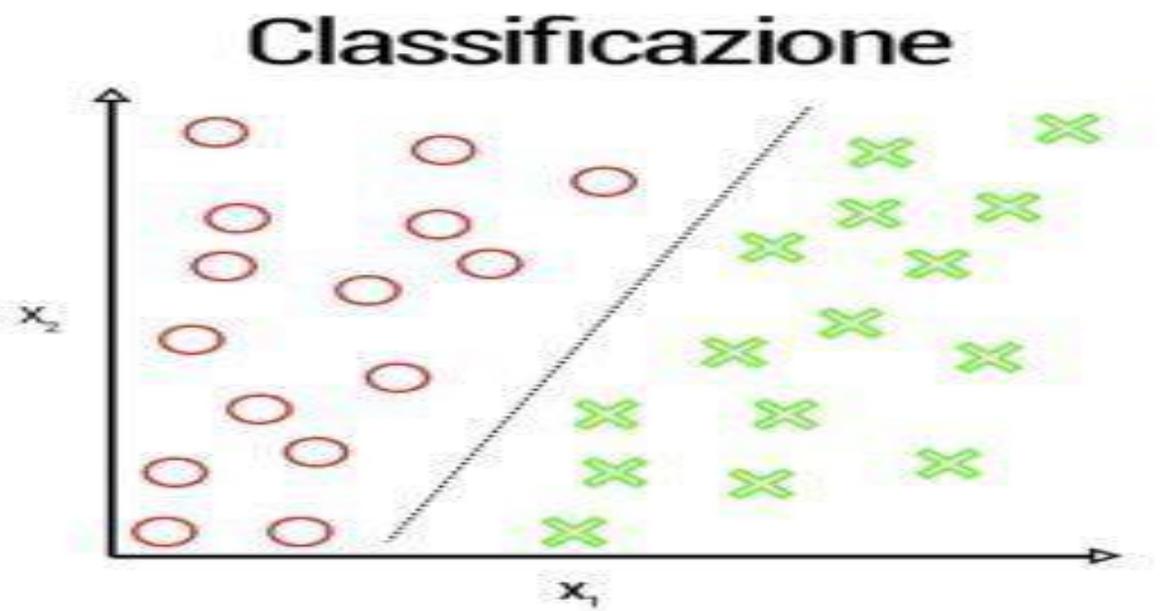
Che cosa intendiamo per mapping desiderato? Dipende dal problema che vogliamo risolvere:

- CLASSIFICAZIONE
- REGRESSIONE

IL MULTI LAYER PERCEPTRON: MAPPING

Che cosa intendiamo per mapping desiderato? Dipende dal problema che vogliamo risolvere:

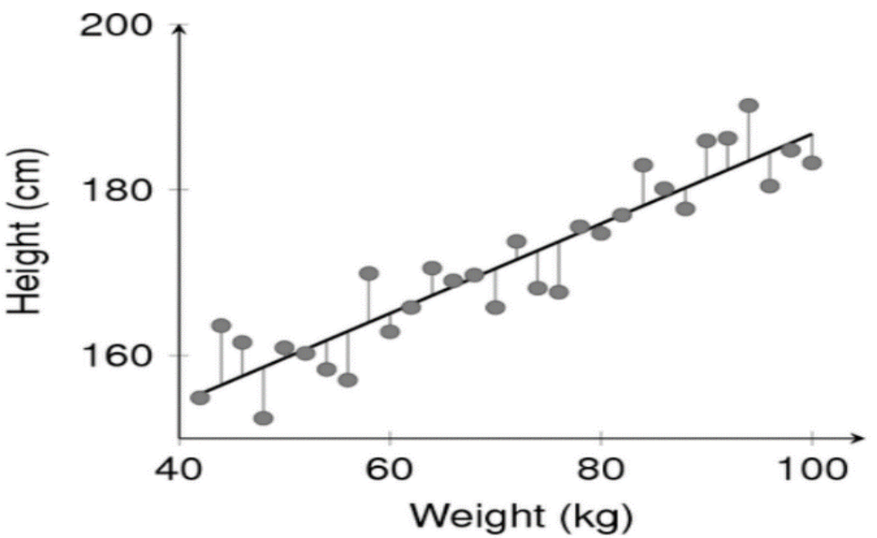
CLASSIFICAZIONE: assegnazione di un dato (pattern) ad un certo insieme (CLASSE)



IL MULTI LAYER PERCEPTRON: MAPPING

Che cosa intendiamo per mapping desiderato? Dipende dal problema che vogliamo risolvere:

REGRESSIONE: individuare valore di una variabile indipendente in funzione di una variabile in input



Es. *stima dell'altezza di una persona in base al peso*

IL MULTI LAYER PERCEPTRON: ADDESTRAMENTO

Sebbene i primi neuroni artificiali risalgano agli anni 40', fino a metà degli anni 80' non erano disponibili algoritmi di training efficaci.

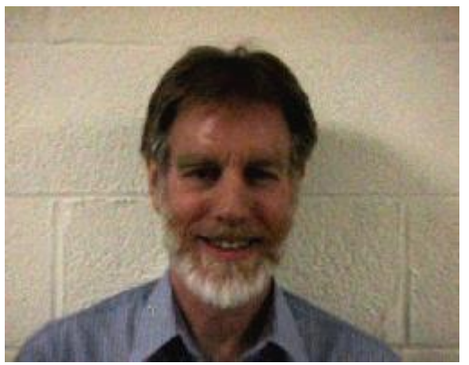
Nel 1986 Rumelhart, Hinton & Williams hanno introdotto l'algoritmo di Error Backpropagation suscitando grande attenzione nella comunità scientifica.

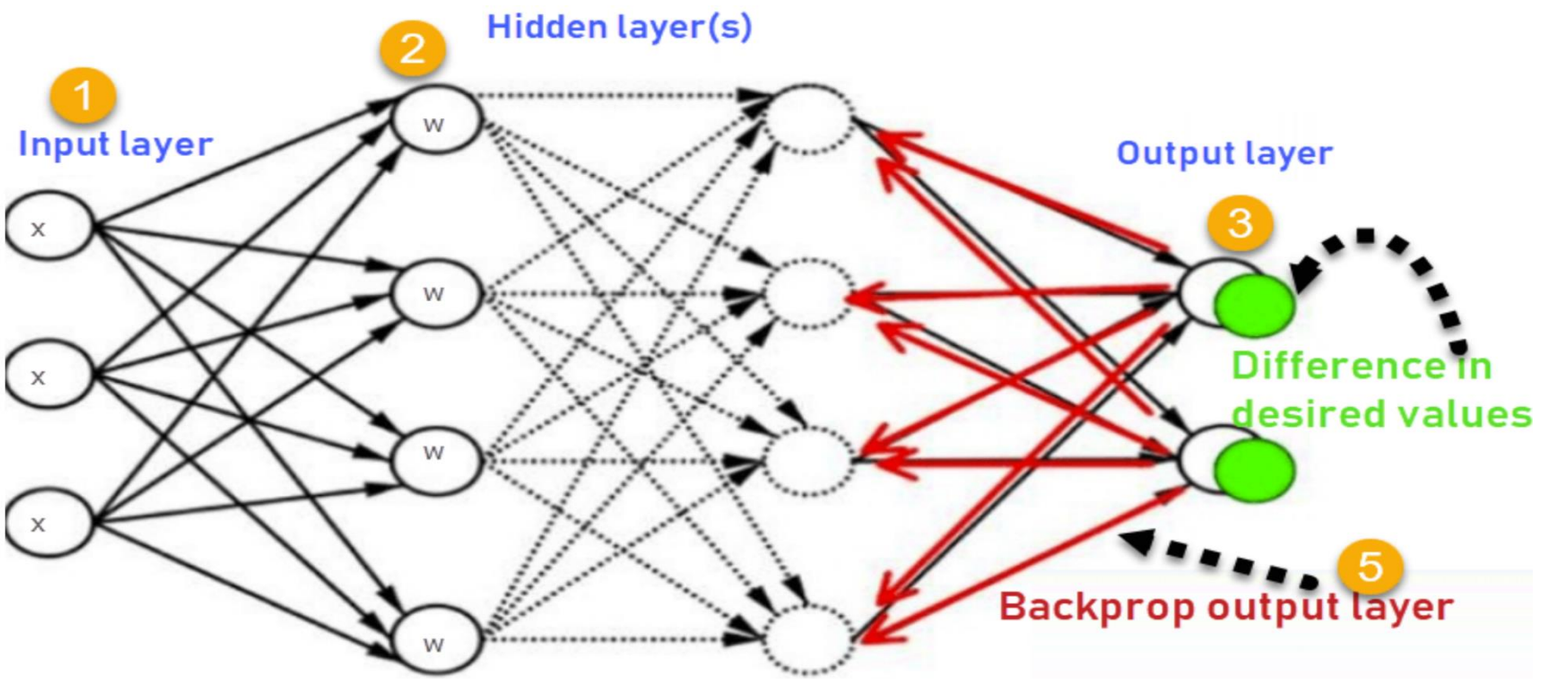
Hinton:informatico

Rumelhart:psicologo

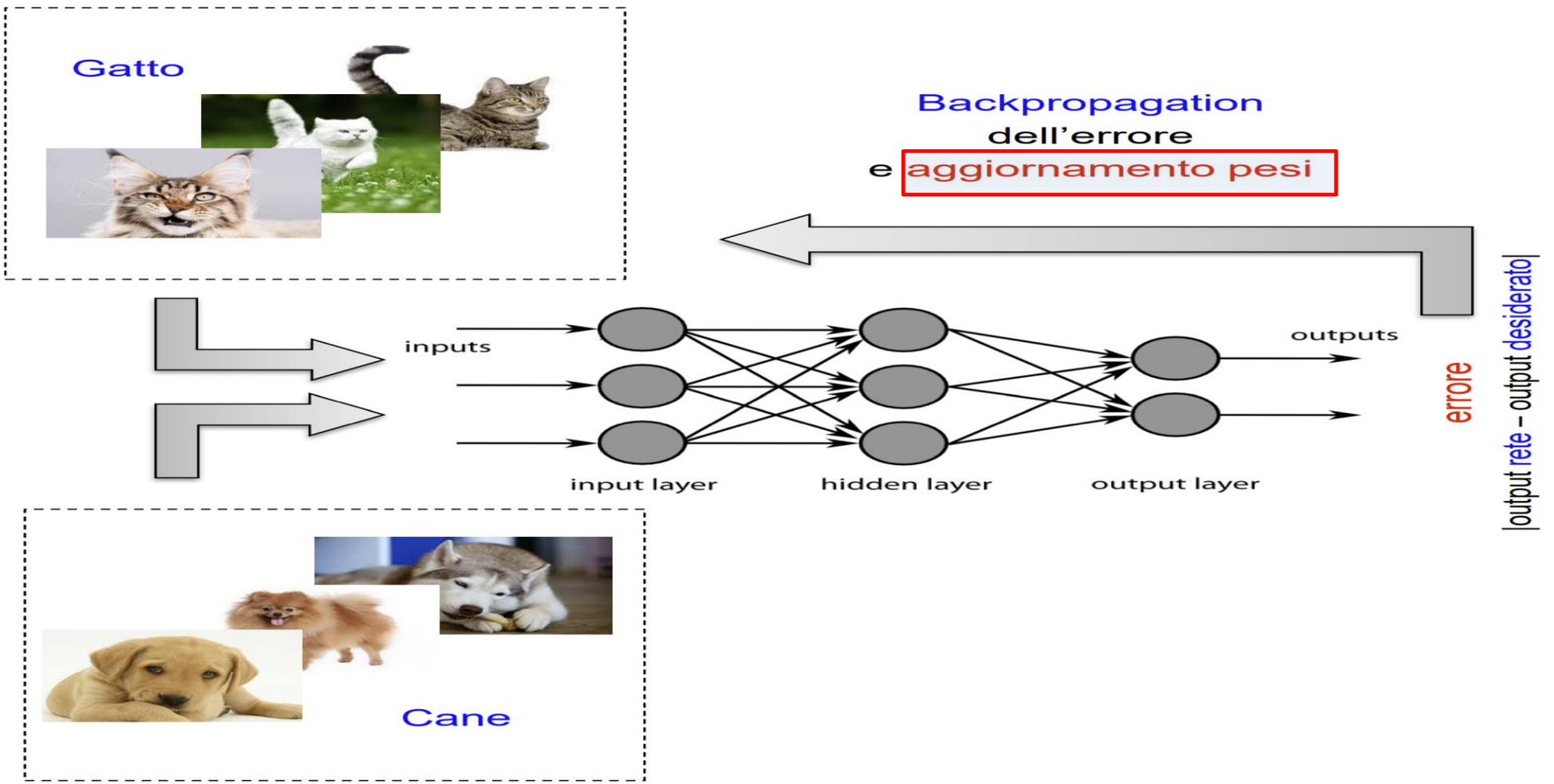
Williams:informatico

IL MULTI LAYER PERCEPTRON: BACK PROPAGATION





IL MULTI LAYER PERCEPTRON: BACK PROPAGATION

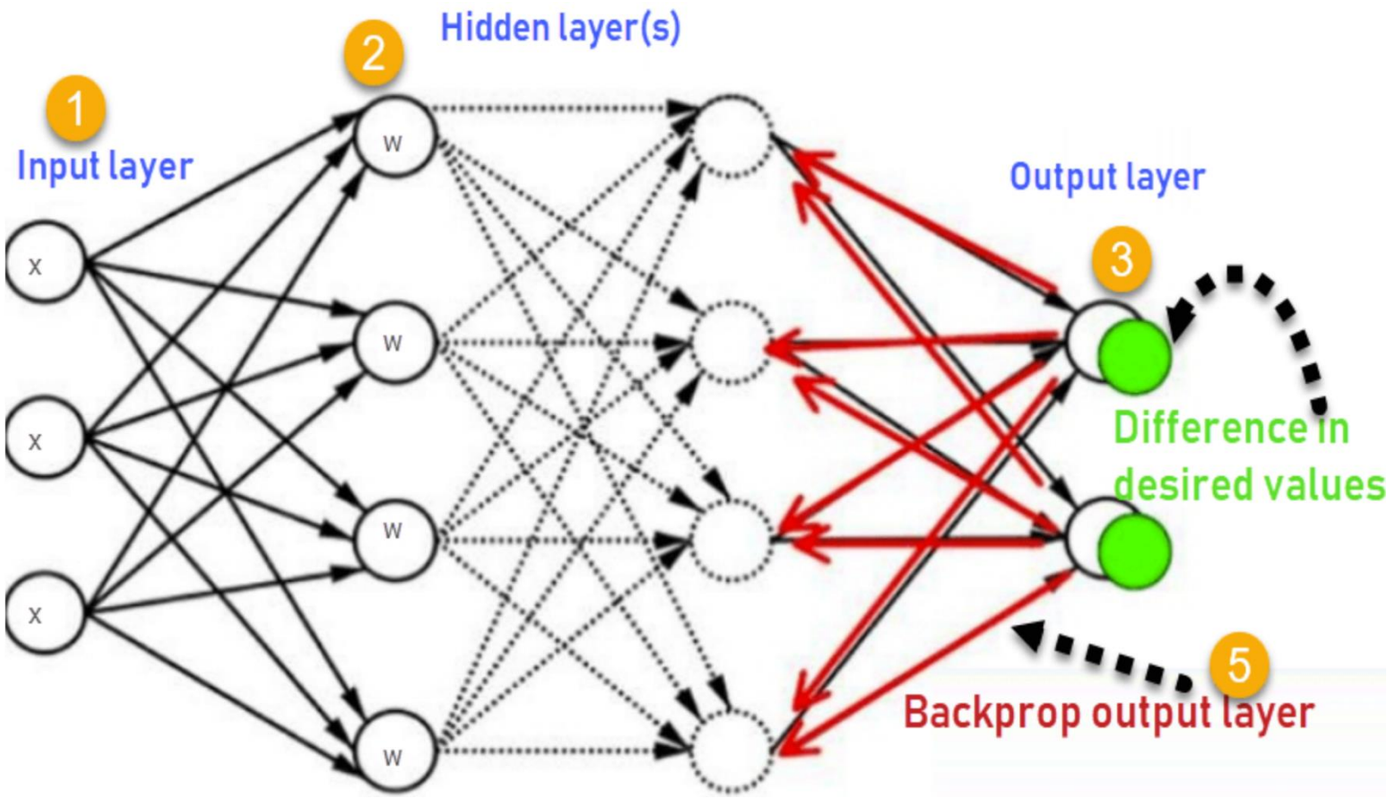


TRAINING DI UN CLASSIFICATORE
ESEMPIO

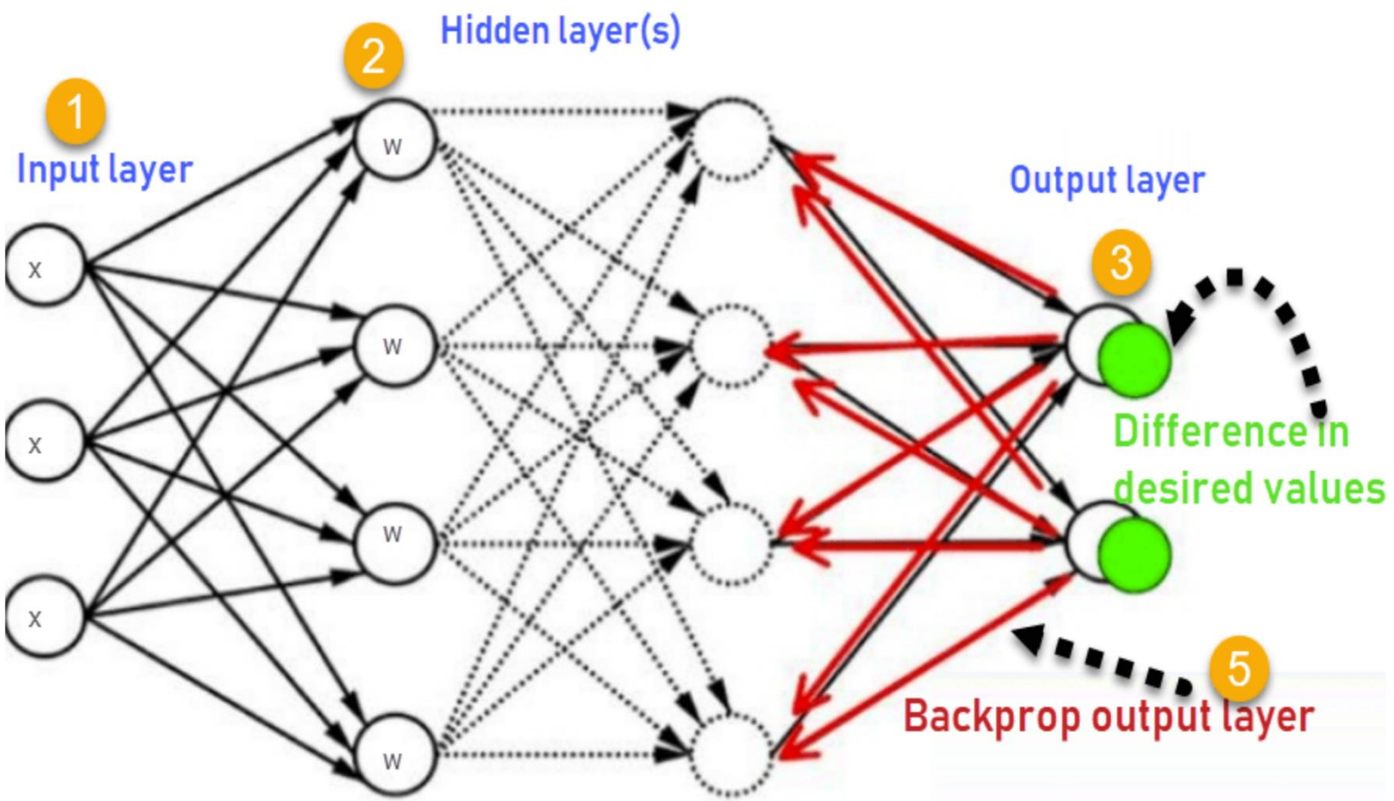
INPUT LAYER
X: dimensione del dato (pattern input)

Xc [0,0,1], [0,1,0],[1,0,0]

Xg [0,0,0], [1,1,1],[1,0,1]



TRAINING DI UN CLASSIFICATORE
ESEMPIO

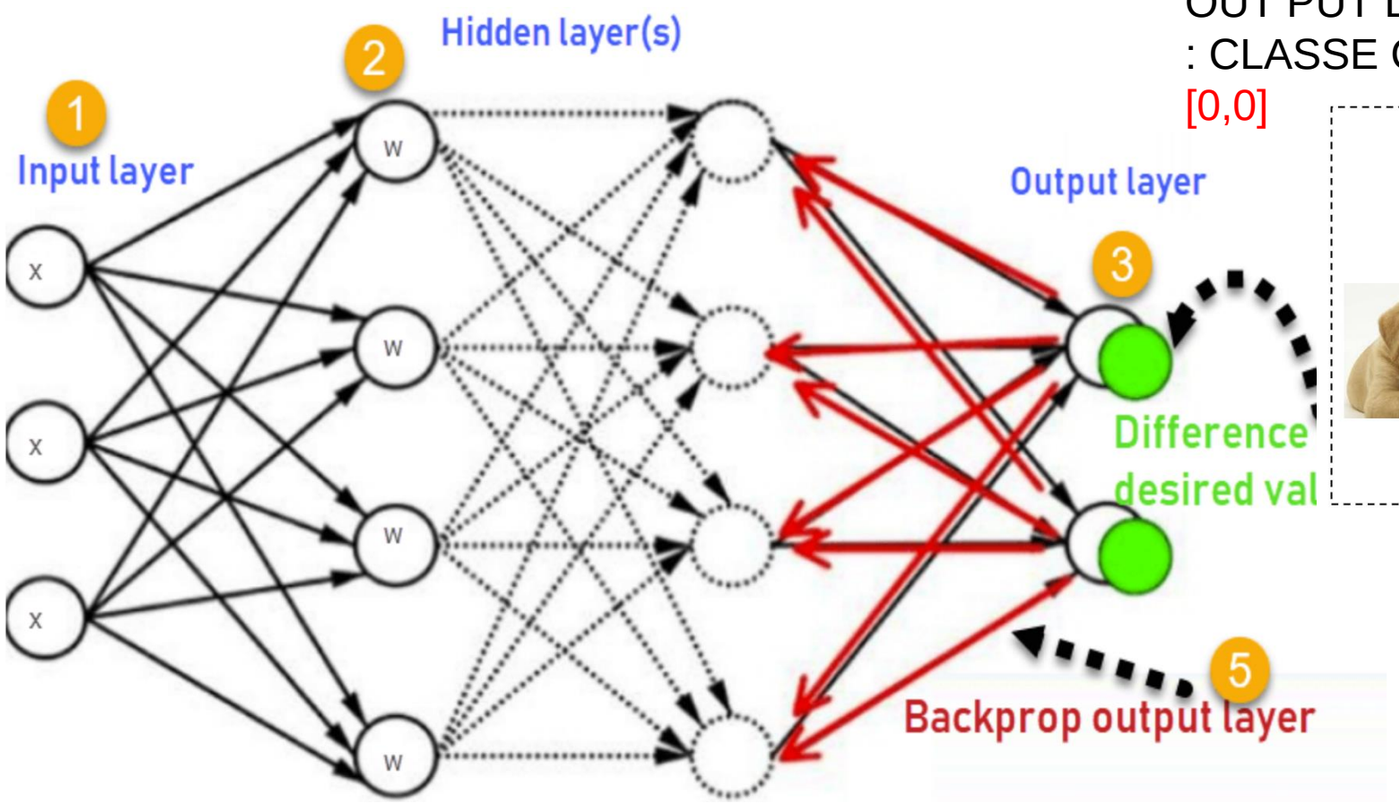


OUT PUT LAYER
: CLASSE CANE
[0,0]

OUTPUT LAYER
: CLASSE GATTO
[0,1]

TRAINING DI UN CLASSIFICATORE
ESEMPIO

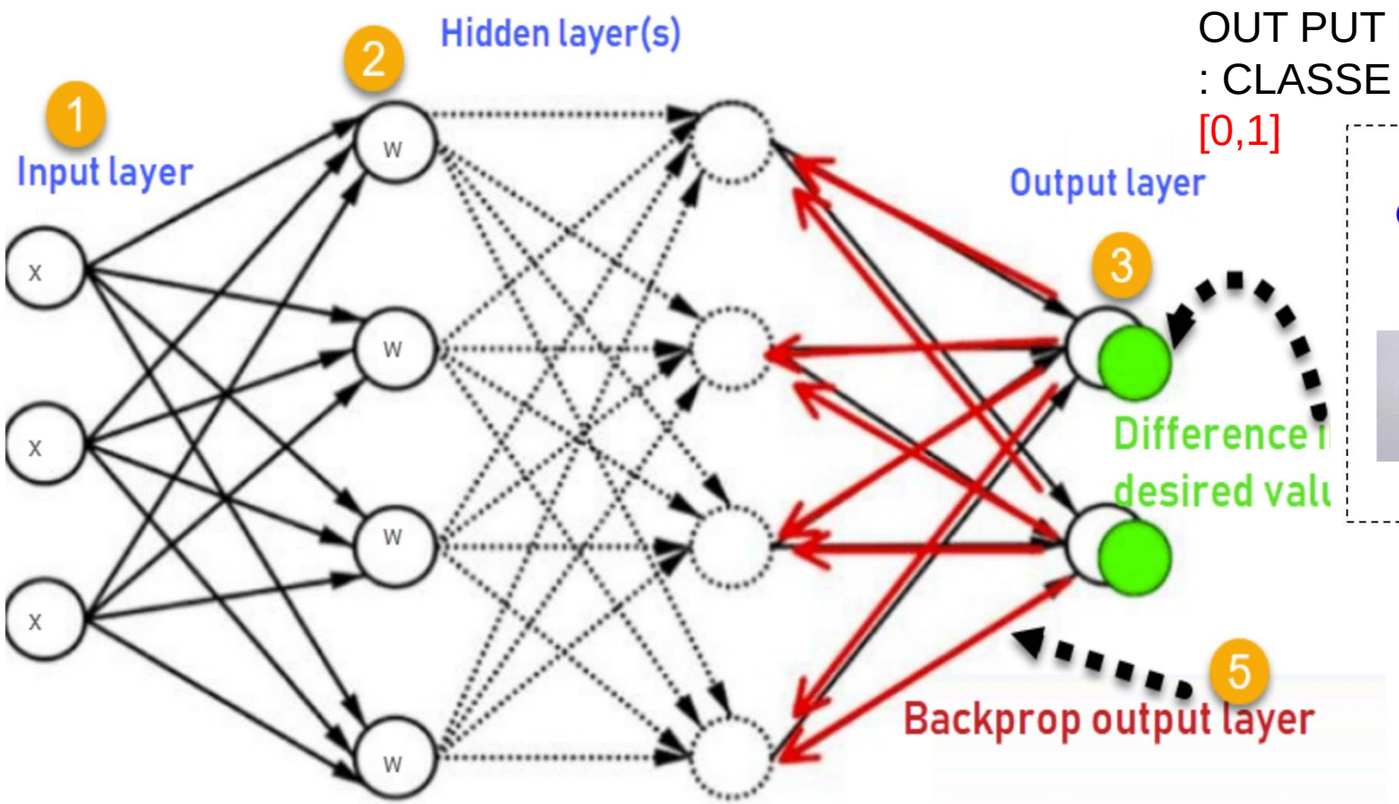
Xc [0,0,1]



OUT PUT LAYER
: CLASSE CANE
[0,0]

TRAINING DI UN CLASSIFICATORE
ESEMPIO

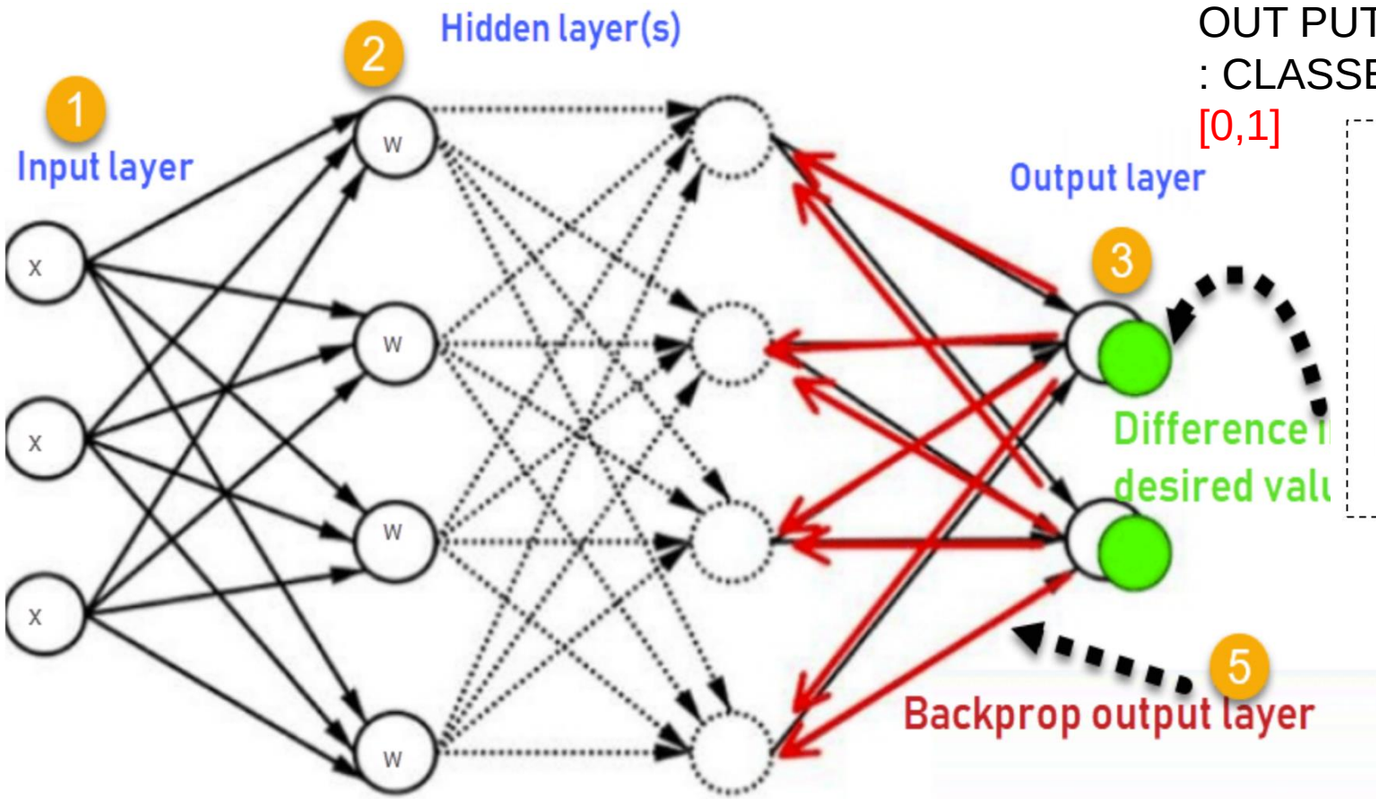
Xg [0,0,0]



OUT PUT LAYER
: CLASSE GATTO
[0,1]

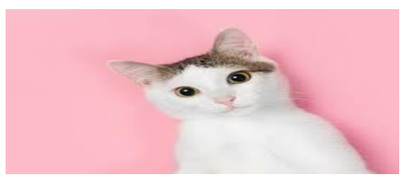
TRAINING DI UN CLASSIFICATORE
ESEMPIO

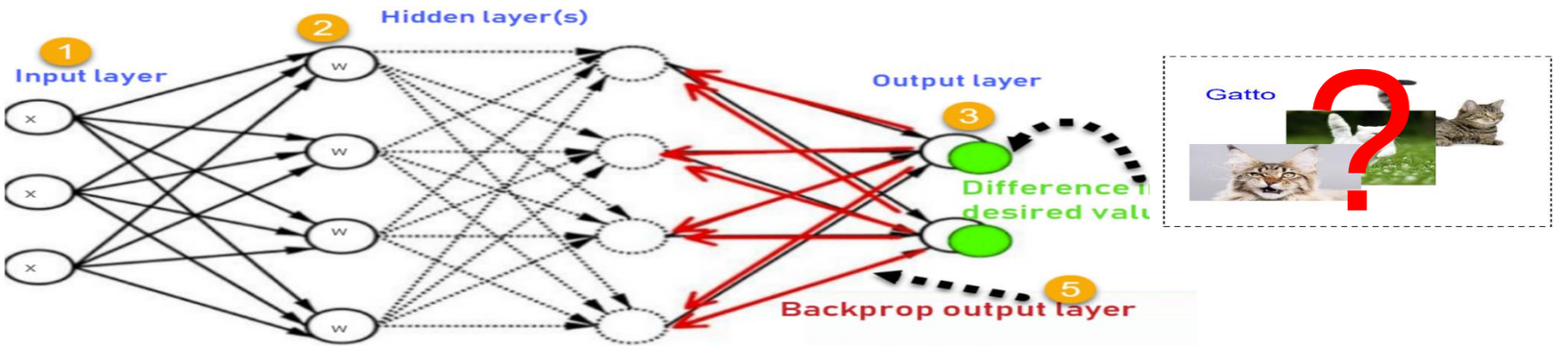
Xg [0,1,1]



OUT PUT LAYER
: CLASSE GATTO
[0,1]

CONCLUSIONI





Tanto maggiore è la quantità di dati con cui posso addestrare la mia rete, tanto migliorano le prestazioni della stessa

LA DISPONIBILITA' DI GRANDI QUANTITA' DI DATI E' FONDAMENTALE

RETI NEURALI
PER GIOCARE UN PO….

RIFERIMENTI

HTTP://PLAYGROUND.TENSORFLOW.ORG

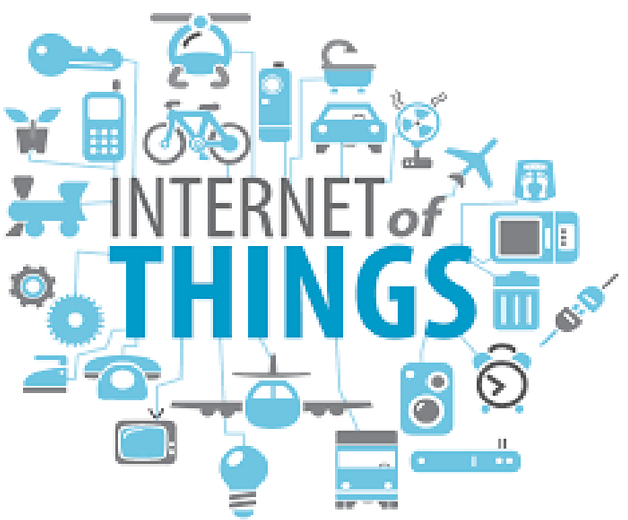
IL COMBUSTIBILE DELL'IA

BIG DATA, CLOUD COMPUTING, INTERNET OF THINGS

QUALI I FATTORI CARDINE DELLA RIVOLUZIONE DELL'AI?

- Big Data

- Internet of things



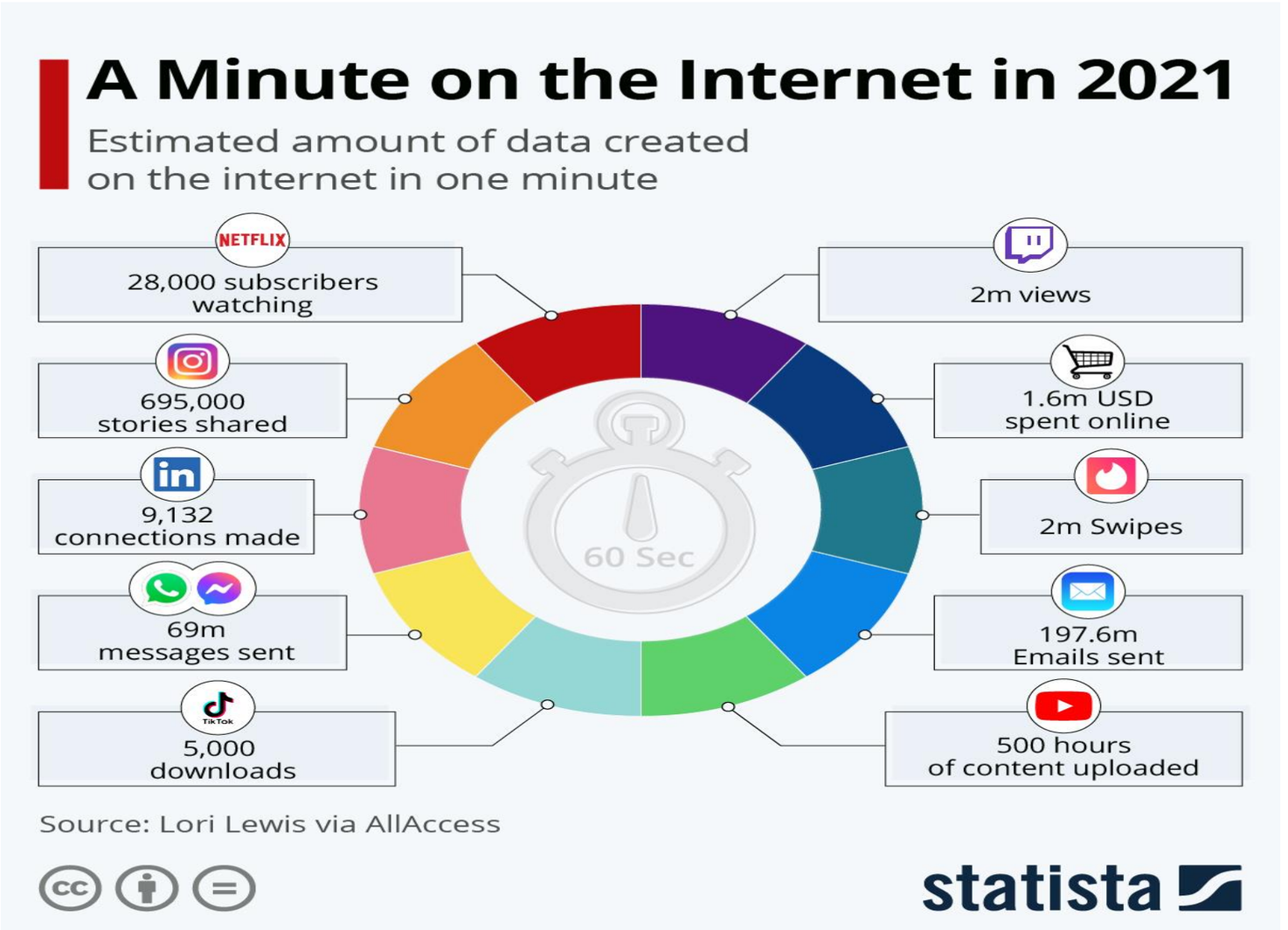
- Cloud Computing

BIG DATA

DEFINIZIONE DI BIG DATA

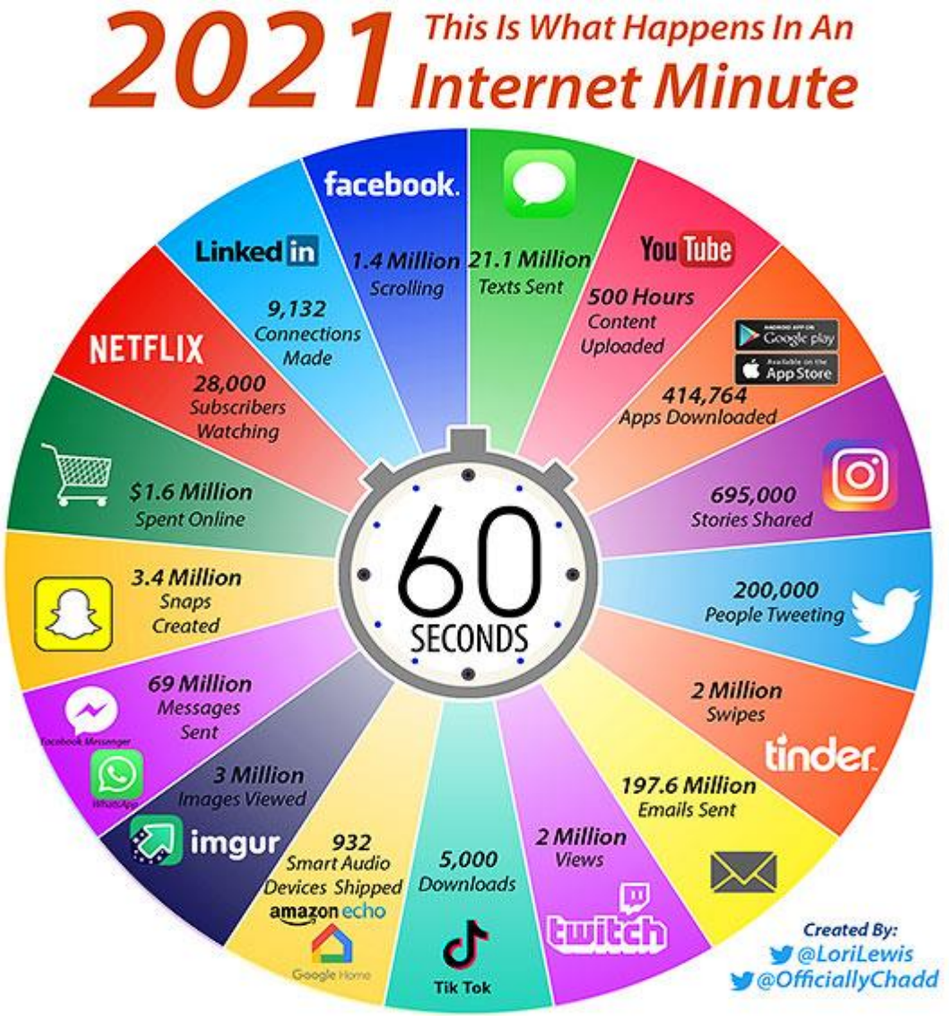
I **Big data** sono insiemi di dati dal volume talmente elevato da non poter essere gestiti dagli strumenti convenzionali, bensì da tecnologie e metodi innovativi in grado di raccoglierli, elaborarli e analizzarli, in modo da poterli sfruttare per fare previsioni su trend di comportamento, per esempio, e così prendere delle decisioni più efficienti.

BIG DATA



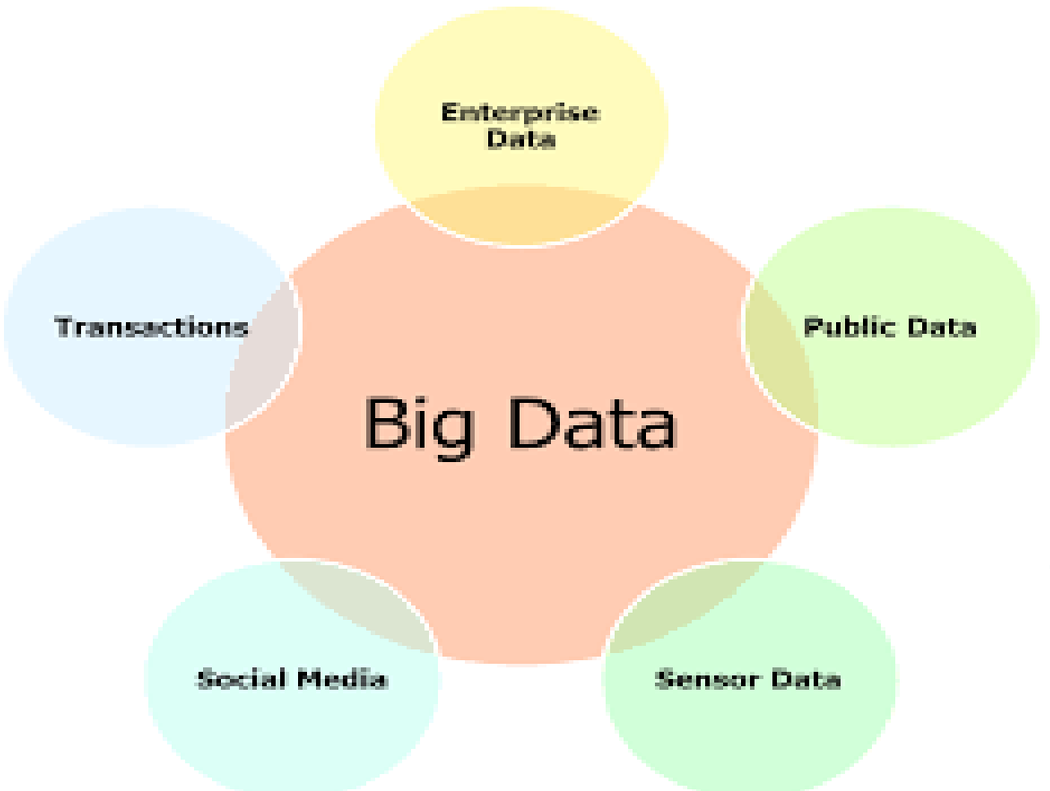
La proliferazione di dati disponibili grazie al boom di Internet ha creato un archivio sterminato di documenti, video e immagini che permettono di ***addestrare le reti neurali*** *e i loro derivati a un livello senza precedenti.*

BIG DATA



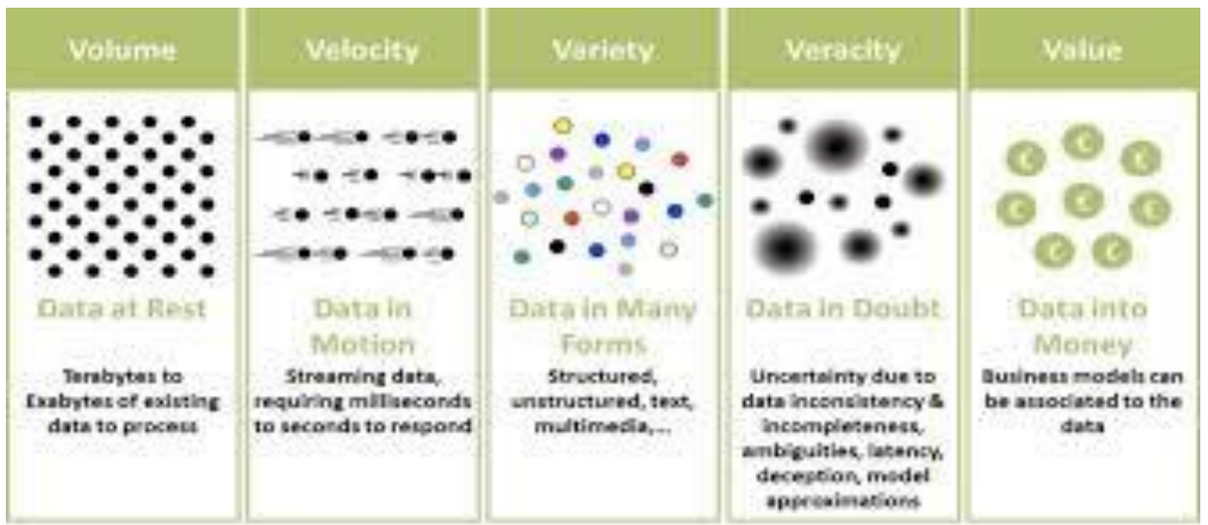
I dati sono prodotti da innumerevoli fonti e servizi disponibili su Internet.

BIG DATA: DEFINIZIONI



big data come un insieme di dati il cui volume è talmente grande «*da superare la capacità dei convenzionali strumenti di gestione di dati di raccoglierli, immagazzinarli, gestirli e analizzarli*» (McKinsey, 2011).

BIG DATA

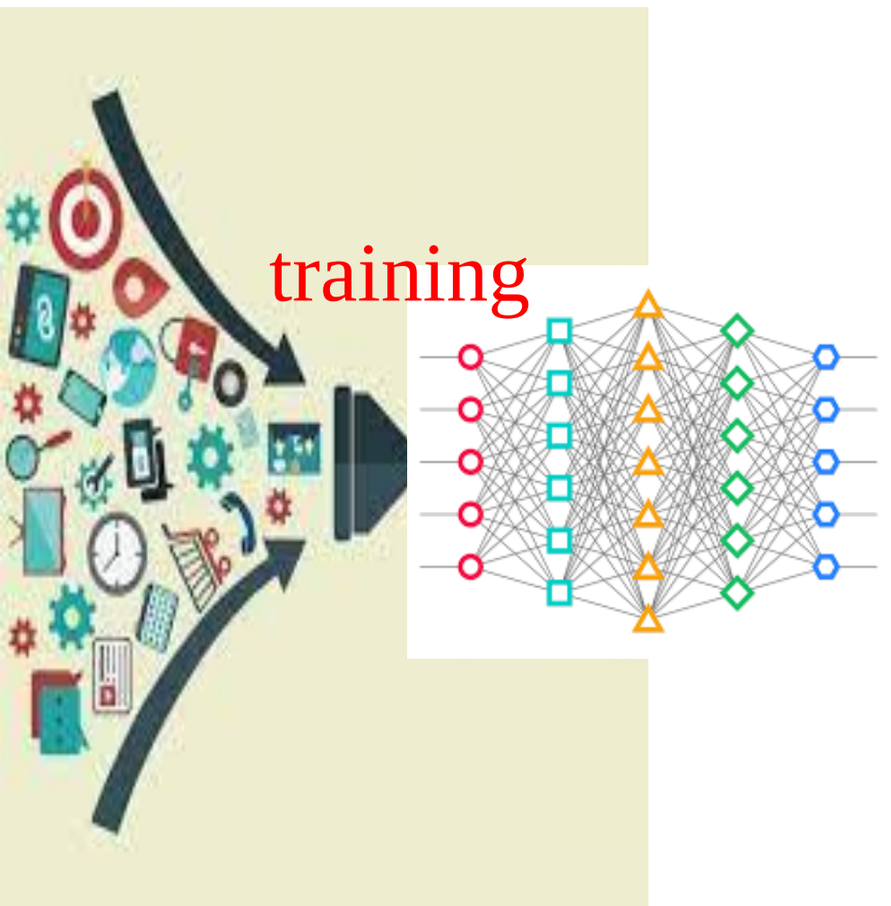


COSA SONO I BIG DATA?

L'espressione "big data" può essere impiegata sia in riferimento alla grande velocità con cui vengono attualmente generati i dati che alla capacità sempre crescente di immagazzinarli, elaborarli ed analizzarli, come si legge all'interno di un articolo di IBM. Anche noti come "**megadati**", i big data sono stati definiti da Gartner, nel **2001**, come «***risorse informative*** *a* ***elevato volume***, ***velocità*** *e* ***varietà*** *che richiedono* ***forme di elaborazione*** *delle informazioni economiche e innovative per potenziare la* ***comprensione***, *la presa di* ***decisioni*** *e l'****automazione*** *dei processi*».

BIG DATA

“La prima e più importante ragione dietro il legame tra AI e Big Data è che AI necessita grandi quantità di dati da caricare per costruire la sua intelligenza, tramite l’addestramento e al tempo stesso Big Data li rende disponibili ottenenedo algoritmi di analisi molto performanti

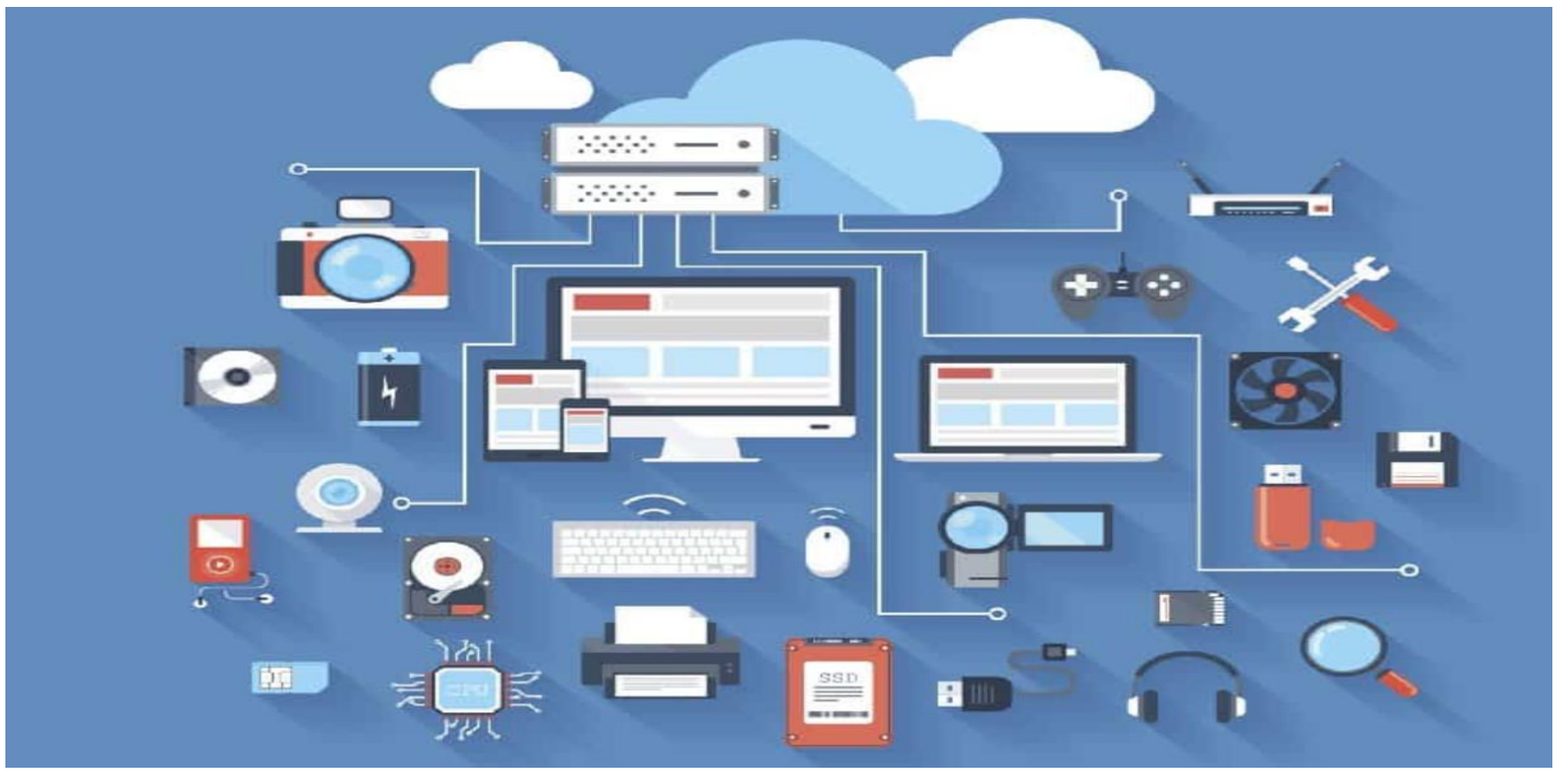


CLOUD COMPUTING

CLOUD COMPUTING

Il calcolo nella «nuvola informatica» o «cloud computing»: questa tecnologia ha incrementato la capacità di calcolo e si è rivelata imprescindibile ***per addestrare le reti neurali****. Se tuttavia, un giorno, si arriverà alla computazione quantistica il cloud computing potrà essere ampiamente superato.*

INTERNET OF THINGS



L'esplosione di dispositivi e sensori connessi tra loro mediante Internet non solo costituisce di per se una fonte di informazione, e ***di addestramento per le reti neurali****, ma ha aperto un canale di comunicazione in tempo reale con l'utente, innescando il bisogno di applicazioni di Intelligenza Artificiale che possano interagire con lui.*

A IOT

Edifici ed esercizi commerciali

Un primo esempio di applicazione dell'AIoT può essere legato ad edifici con uffici intelligenti, dotati di una rete di sensori ambientali in grado di rilevare il personale presente e regolare, di conseguenza, i parametri ambientali (per esempio, temperatura ed illuminazione) per migliorare l'efficienza energetica dei singoli uffici e dell'intero edificio commerciale. Un altro aspetto che può beneficiare dell'AIoT può riguardare il controllo degli accessi all'edificio intelligente attraverso una tecnologia di riconoscimento facciale dei visitatori e del personale dipendente. La combinazione di telecamere connesse, le cui immagini in uscita sono analizzate per mezzo di modelli di AI in grado di confrontare le immagini scattate in tempo reale con un database per determinare chi dovrebbe essere autorizzato ad accedere nell'edificio, rappresenta un esempio perfetto di applicazione del concetto dell'AIoT.

A IOT

Gestione di flotte di veicoli autonomi

L'AIoT può essere utilizzato anche per la gestione di flotte di veicoli, per il monitoraggio dei veicoli in tema di manutenzioni previste e, per quanto possibile, anche con un approccio predittivo, per ridurre i costi del carburante necessario per i singoli veicoli, ed anche per identificare comportamenti anomali o non sicuri dei conducenti. L'utilizzo di dispositivi IoT quali GPS ed altri sensori, unitamente ad un sistema AI in grado di elaborare i dati in arrivo dalle varie componenti del sistema, favorisce l'AIoT anche in questi contesti. I . .

Un altro utilizzo del concetto di AIoT è legato ai veicoli autonomi, come ad esempio i sistemi di guida automatica presenti a bordo di autovetture, e definiti da diverse aziende (es. **Amazon**, **Apple**, **Audi**, **Baidu**, **Ford**, **Hyundai**, **Microsoft**, **Nvidia**, **Samsung**, **Tesla**, **Toyota**, **Uber**, **Waymo**), che utilizzano radar, sonar, GPS e fotocamere per raccogliere dati sulle condizioni di guida e, in accoppiata con un sistema di AI, prendono decisioni sulle eventuali azioni da mettere in campo in situazioni di pericolo.

IBM WATSON

Jeopardy, molto famoso negli USA, utilizza una modalità di gioco «a rovescio» rispetto ai quiz televisivi classici: invece di rispondere a domande, è necessario fornire domande alle risposte fornite.

Quiz: *Napoleone Bonaparte*

Possibile risposta: *Chi morì in esilio a Sant'Elena?*

Durante il quiz, Watson aveva accesso a 200 milioni di pagine di contenuti (4 terabytes) tutte caricate in RAM: enciclopedie (tra cui Wikipedia), dizionari, thesauri, tassonomie, ontologie (es. Wordnet) e articoli giornale.

Watson non era connesso a Internet (sarebbe stato in ogni caso troppo lento lanciare ricerche online).

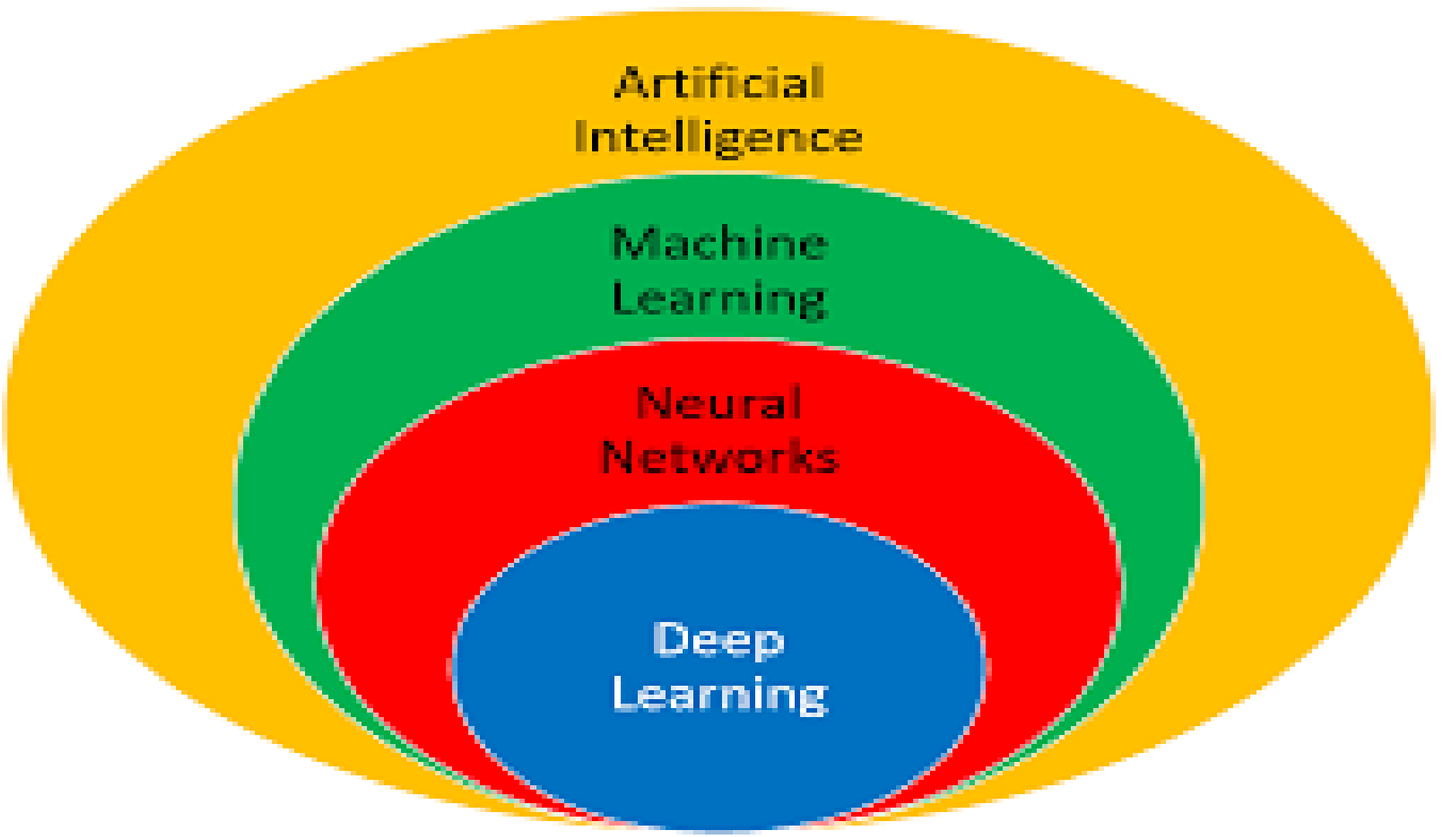
Secondo IBM Watson utilizza tecniche basate su analisi linguaggio, metodi ragionamento e apprendimento automatico: se a questa versatilità aggiungiamo i tre precedenti fattori quali scenari possiamo immaginare….??

ALPHA GO

Google DeepMind vince a Go

- 2016 – AlphaGo (Goggle) batte il campione Lee Sedol (9 dan).
  - Go è un antico gioco cinese, con regole semplici ma molte più mosse possibili rispetto agli scacchi, cosa che richiede più intuizione e lo rende più difficilmente suscettibile ad approcci forza bruta.
  - Mentre Deep Blue usa strategie di ricerca in profondità ed euristici, AlphaGo è basato principalmente su tecniche di machine learning.
  - Inizialmente sono addestrate in modo supervisionato due deep neural network, cercando di imitare le mosse di professionisti a partire da partite memorizzate e rese disponibili dai Go Server su Internet (30 milioni di mosse).
  - Poi il sistema gioca milioni di partite contro sé stesso utilizzando reinforcement learning per migliorare la strategia.
  - Nella partita finale utilizza 1202 CPU e 176 GPU.

MACHINE LEARNING



DEFINIZIONE

Machine Learning (ML)

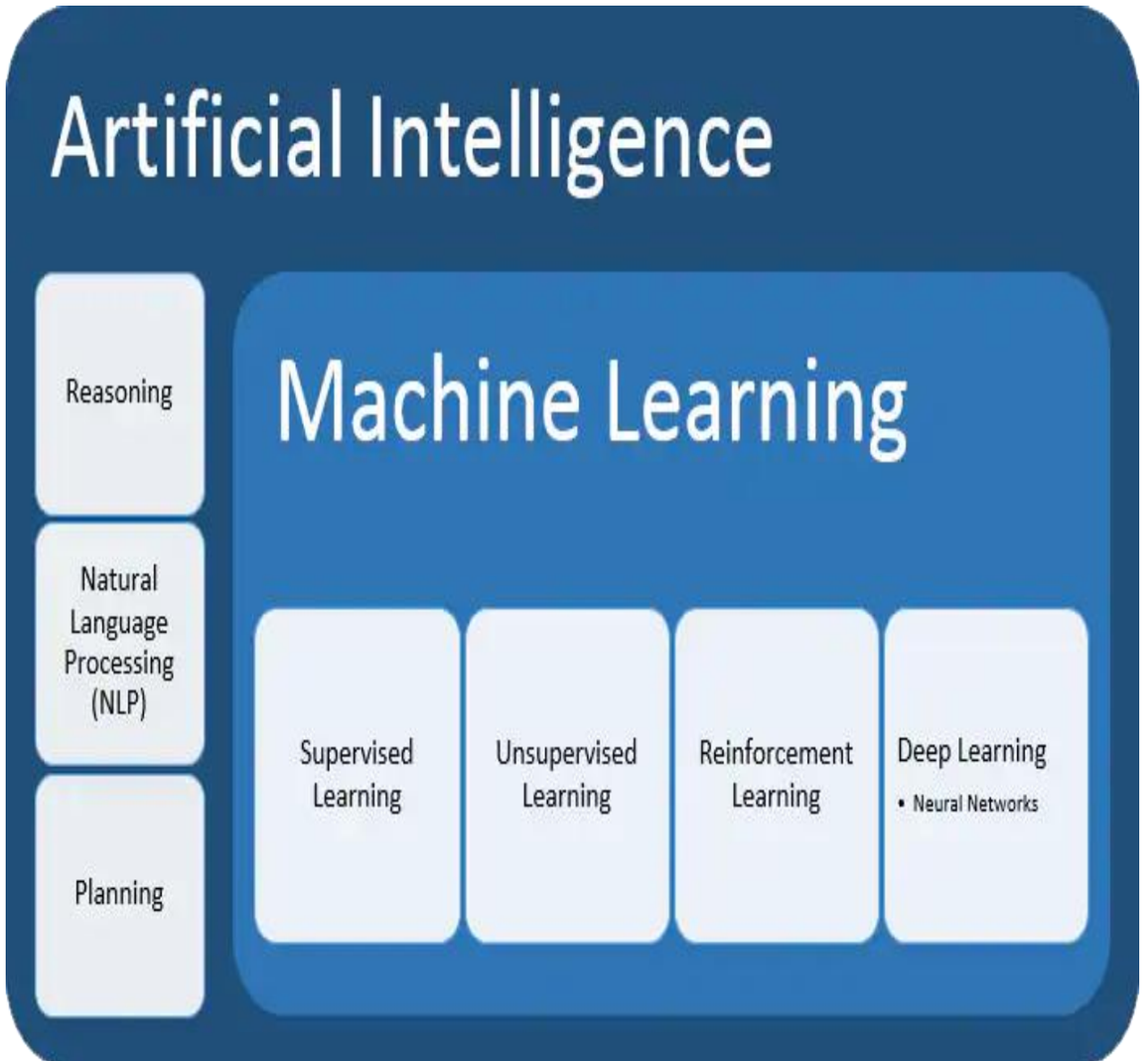
- Un sistema di Machine Learning (apprendimento automatico) durante la fase di training apprende a partire da esempi (in modo più o meno supervisionato). Successivamente è in grado di generalizzare e gestire nuovi dati nello stesso dominio applicativo.

Il Machine Learning (ML) è un sottoinsieme dell'intelligenza artificiale (AI) che si occupa di creare sistemi che apprendono—o migliorano le performance—in base ai dati che utilizzano.

Machine Learning

Un sistema di Machine Learning impara dagli esempi a migliorare le proprie prestazioni per la gestione di nuovi dati provenienti dalla stessa sorgente.

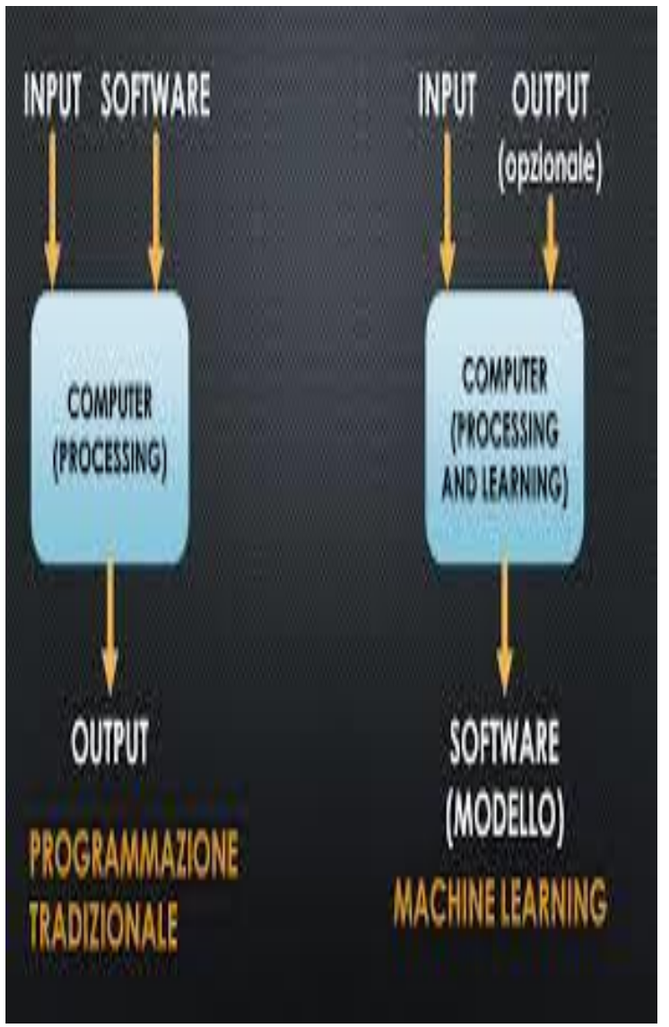
DEFINIZIONE



Il Machine Learning (ML) è un sottoinsieme dell'intelligenza artificiale (AI) che si occupa di creare sistemi che apprendono—o migliorano le performance—in base ai dati che utilizzano.

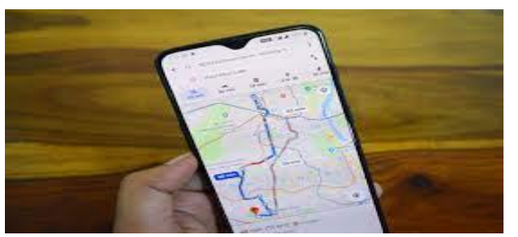
Perché Machine Learning ?

- Machine Learning è oggi ritenuto uno dei approcci più importanti dell'intelligenza artificiale.
  - L'apprendimento è una componente chiave del ragionamento
  - Apprendere → migliorare, evolvere
  - Consente di gestisce la complessità di applicazioni reali, talvolta troppo complesse per poter essere modellate efficacemente.



ESEMPI APPLICAZIONI DEL ML NELLA VITA QUOTIDIANA

STIMA DEL PERCORSO MIGLIORE



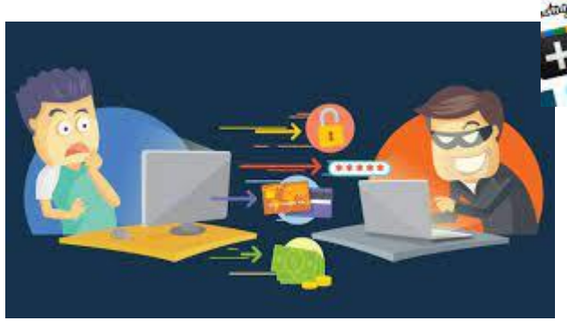
EMAIL INTEL

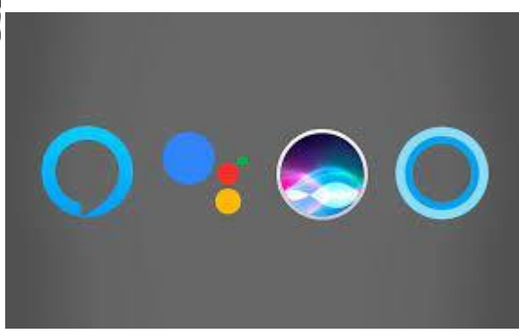
BANK INTEL

ASSISTENZA MEDICA

SOCIAL NETWORK

SMART ASSISTANT





ESEMPI APPLICAZIONI DEL ML NELLA VITA QUOTIDIANA

STIMA DEL PERCORSO MIGLIORE E OTTIMIZZAZIONE

GOOGLE MAPS: usando i dati di posizione provenienti dai cellulari Google Maps analizza le condizioni di traffico in tempo reale e utilizzando il questa enorme mole di dati per alimentare gli algoritmi di ML, ottimizza i tempi di percorrenza indicando la strada più veloce.

UBER: utilizza gli algoritmi di ML per determinare il costo della corsa, minimizzare i tempi di attesa, ottimizzare i percorsi in funzione del numero di passeggeri su tratte simili.

ESEMPI APPLICAZIONI DEL ML NELLA VITA QUOTIDIANA

EMAIL INTEL

SPAM FILTER: filtraggi basati su «rules-based» non sono utili quando per esempio si ricevono messaggi contenenti testo tipo “online consultancy”, “ online pharmacy”, o da “unknown address”. Gmail utilizza tale tipologia di filtri attivamente

EMAIL CLASSIFICATION: Gmail utilizza la classificazione via ML

ESEMPI APPLICAZIONI DEL ML NELLA VITA QUOTIDIANA

BANK INTEL

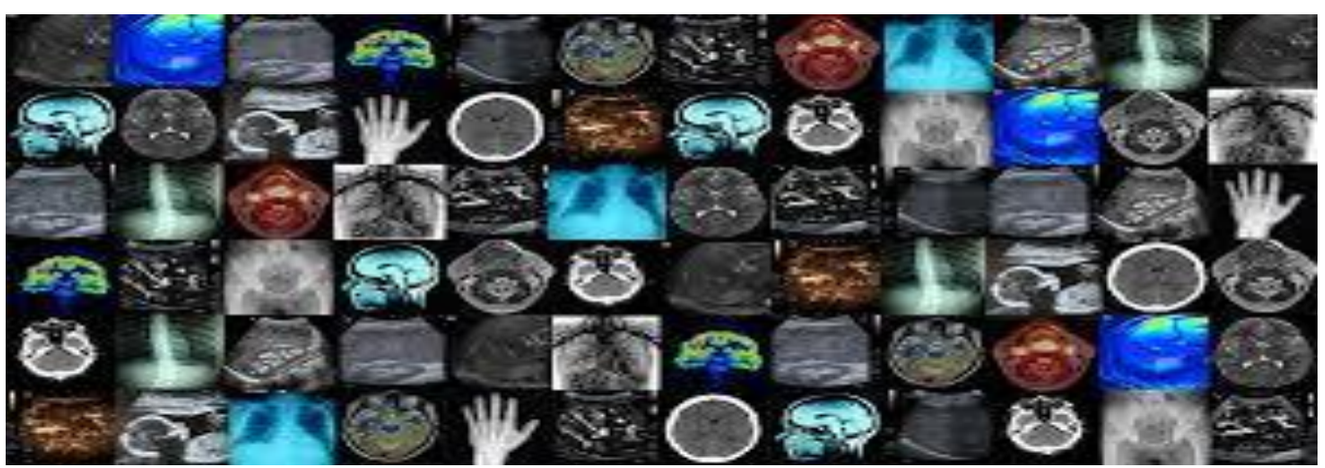
FRAUD DETECTION: l'elevato numero di transazioni giornaliere non permetteva un controllo puntuale delle stesse. Grazie al ML e alle capacità di apprendimento i sistemi bancari sono in grado di individuare comportamenti fraudolenti.

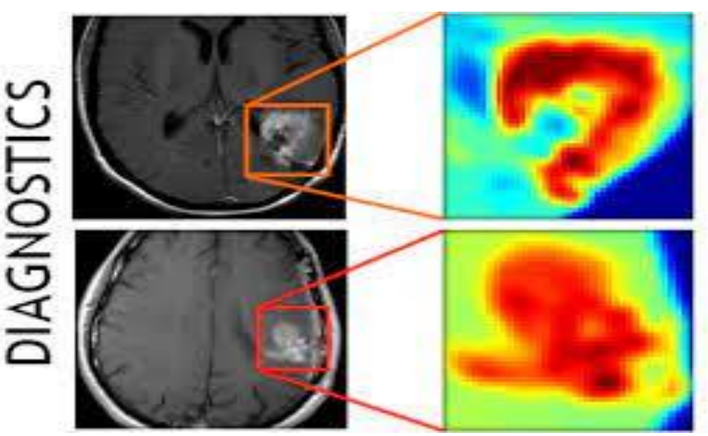
CREDIT DECISION: utilizzati quando la Banca deve rapidamente decidere, sulla base delle conoscenze pregresse se concedere un muto, attivare una carta di credito, etc.

ESEMPI APPLICAZIONI DEL ML NELLA VITA QUOTIDIANA

ASSISTENZA MEDICA

- Analisi di dati medici per individuare regolarità/irregolarità sugli stessi;
- Gestione di dati medici da esami parziali per effettuare diagnosi;
- Analizzare dati generate da strumenti medici;
- Monitorare classi di pazienti.





ESEMPI APPLICAZIONI DEL ML NELLA VITA QUOTIDIANA

SOCIAL NETWORK

Facebook usa ML per:
- aumentare le prestazioni dei sw di riconoscimento facciale;
- Personalizzare ifeed diretti agli utenti e ottimizzare i post per intrattenerlo
- Personalizzare gli inserti pubblicitari a seconda dei gusti espresso nel tempo dal cliente.

Pinterest usa ML per individuare nelle immagini i Pin e raccomandare Pin similari;

Snapchat usa ML per filtrare e modellare attività facciali per replicarle;

Instagram usa ML per individuare gli stati d'animo dietro ciascun emoji e raccomandarne altri

ESEMPI APPLICAZIONI DEL ML NELLA VITA QUOTIDIANA

SMART ASSISTANT

Questi “assistenti intelligenti” si basano su algoritmi di ML per:

- raccogliere informazioni;
- capire le preferenze dell’utente;
- migliorare le prestazioni sulla base delle interazioni con l’utente

COME FUNZIONA IL MACHINE LEARNING?

QUALCHE DEFINIZIONE

Dati e Pattern

- I dati sono un ingrediente fondamentale del machine learning, dove il comportamento degli algoritmi non è pre-programmato ma appreso dai dati stessi.

- Utilizzeremo spesso il termine Pattern per riferirci ai dati
  - Pattern può essere tradotto in italiano in vari modi: *forma*, *campione*, *esempio*, *modello*, ecc. [meglio non tradurlo].
  - S. Watanabe definisce un pattern come l'opposto del caos e come un entità vagamente definita cui può essere dato un nome.
  - Ad esempio un pattern può essere un volto, un carattere scritto a mano, un'impronta digitale, un segnale sonoro, un frammento di testo, l'andamento di un titolo di borsa.

COME FUNZIONA IL MACHINE LEARNING?
QUALCHE DEFINIZIONE

Dati e Pattern

- I dati sono un ingrediente fondamentale del machine learning, dove il comportamento degli algoritmi non è pre-programmato ma appreso dai dati stessi.

- Utilizzeremo spesso il termine Pattern per riferirci ai dati
  - Pattern può essere tradotto in italiano in vari modi: *forma*, *campione*, *esempio*, *modello*, ecc. [meglio non tradurlo].
  - S. Watanabe definisce un pattern come l'opposto del caos e come un entità vagamente definita cui può essere dato un nome.
  - Ad esempio un pattern può essere un volto, un carattere scritto a mano, un'impronta digitale, un segnale sonoro, un frammento di testo, l'andamento di un titolo di borsa.

PATTERN RECOGNITION

Disciplina che studia il riconoscimento dei pattern con tecniche di ML

COME FUNZIONA IL MACHINE LEARNING?
QUALCHE DEFINIZIONE

Tipi di Pattern

- Numerici: valori associati a caratteristiche misurabili o conteggi.
  - Tipicamente continui (ma anche discreti, es. interi), in ogni caso soggetti a ordinamento.
  - Rappresentabili naturalmente come vettori numerici nello spazio multidimensionale.
  - L'estrazione di caratteristiche da segnali (es., immagini, suoni) produce vettori numerici detti anche feature vectors.
  - Es. Persona: [*altezza*, *circonferenza toracica*, *circonferenza fianchi*, *lunghezza del piede*]

COME FUNZIONA IL MACHINE LEARNING??
QUALCHE DEFINIZIONE

Tipi di Pattern

- Categorici: valori associati a caratteristiche qualitative e alla presenza/assenza di una caratteristica (yes/no value).
  - Non «semanticamente» mappabili in valori numerici.
  - Es. Persona: [*sesso*, *maggiorenne*, *colore occhi*, *gruppo sanguigno*].
  - Talvolta soggetti a ordinamento (ordinali): es. temperatura ambiente: *alta*, *media* o *bassa*.
  - Normalmente gestiti da sistemi a regole e alberi di classificazione.
  - Molto utilizzati nell'ambito del data mining, spesso insieme a dati numerici (mixed).

COME FUNZIONA IL MACHINE LEARNING??

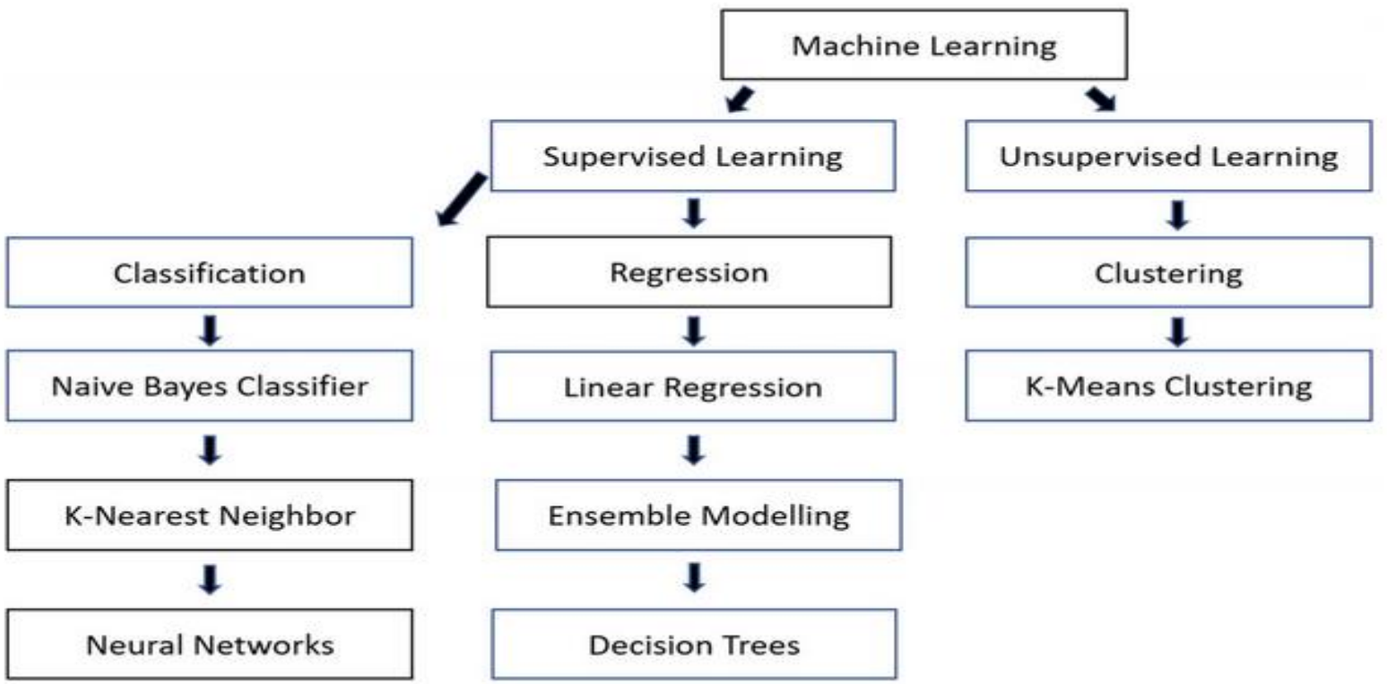
QUALCHE DEFINIZIONE

Sequenze e altri dati strutturati

Sequenze: pattern sequenziali con relazioni spaziali o temporali.

- Es. uno stream audio (sequenza di suoni) corrispondente alla pronuncia di una parola, una frase (sequenza di parole) in linguaggio naturale, un video (sequenza di frame).
- Spesso a lunghezza variabile
- La posizione nella sequenza e le relazioni con predecessori e successori sono importanti.
- Critico trattare sequenze come pattern numerici.
- Allineamento spaziale/temporale, e «memoria» per tener conto del passato.

IL FRAMEWORK MACHINE LEARNING



ALGORITMI DI MACHINE LEARNING

APPRENDIMENTO (TRAINING)

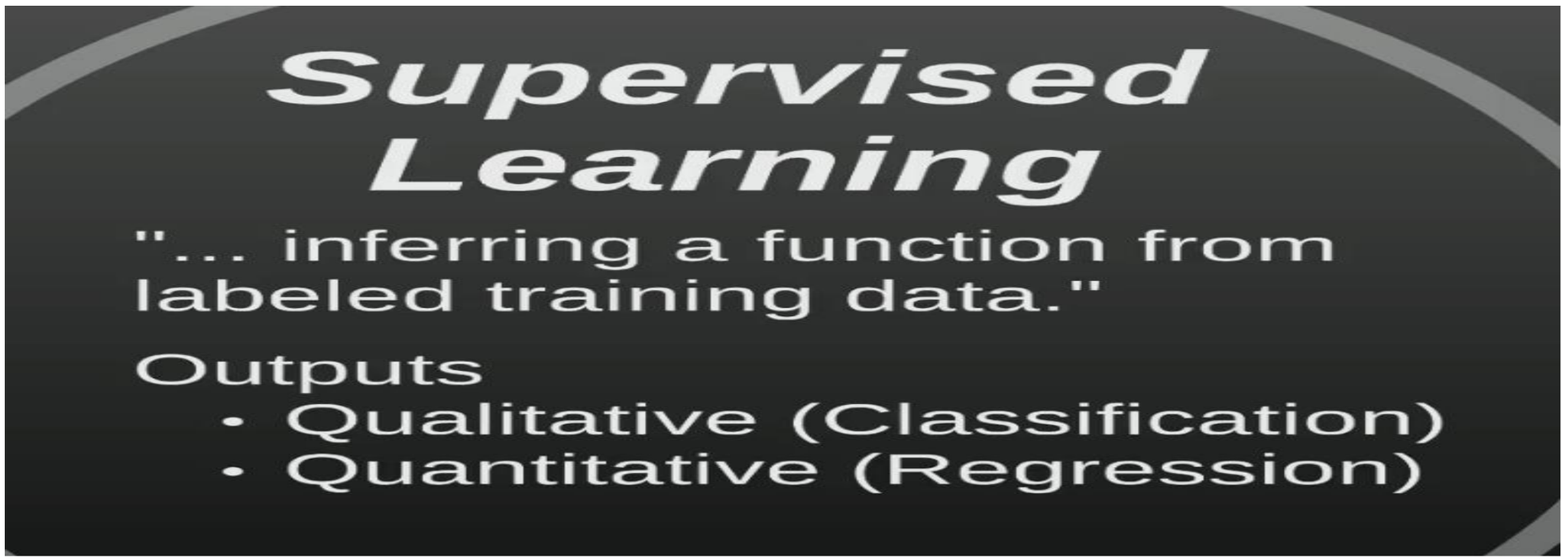
- Supervisionato (Supervised): sono note le classi dei patterr utilizzati per l'addestramento.
  - *il training set è etichettato*.
- Non Supervisionato (Unsupervised): non sono note le classi dei pattern utilizzati per l'addestramento.
  - *il training set non è etichettato*.

TRAINING SET: è l'insieme di pattern su cui addestrare il sistema neurale, trovando il valore ottimo per la funzione f

ALGORITMI DI MACHINE LEARNING

APPRENDIMENTO (TRAINING)

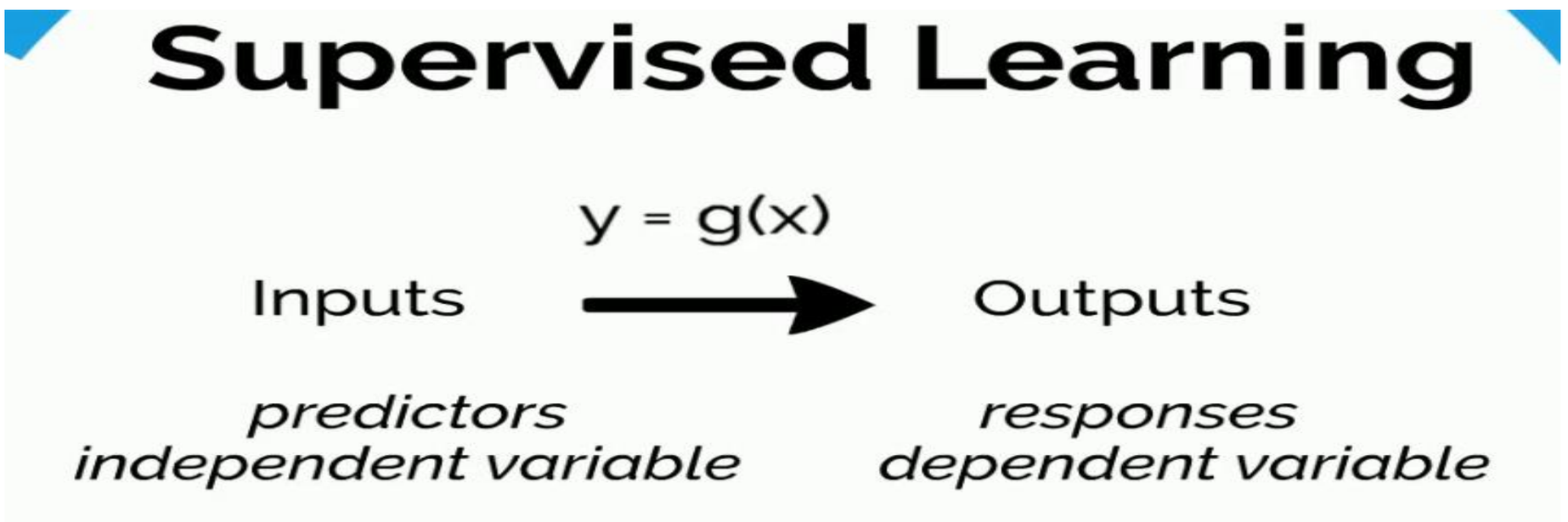
- Supervisionato (Supervised): sono note le classi dei patterr utilizzati per l'addestramento.
  - *il training set è etichettato*.



ALGORITMI DI MACHINE LEARNING

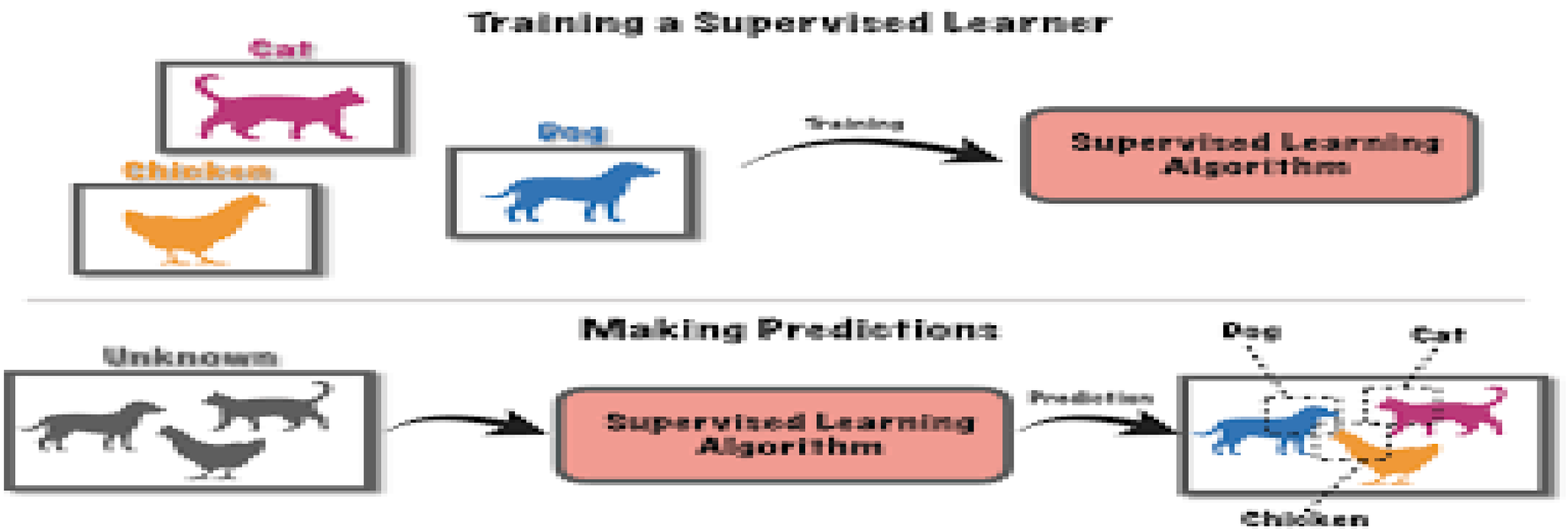
APPRENDIMENTO (TRAINING)

- Supervisionato (Supervised): sono note le classi dei patterr utilizzati per l'addestramento.
  - *il training set è etichettato*.



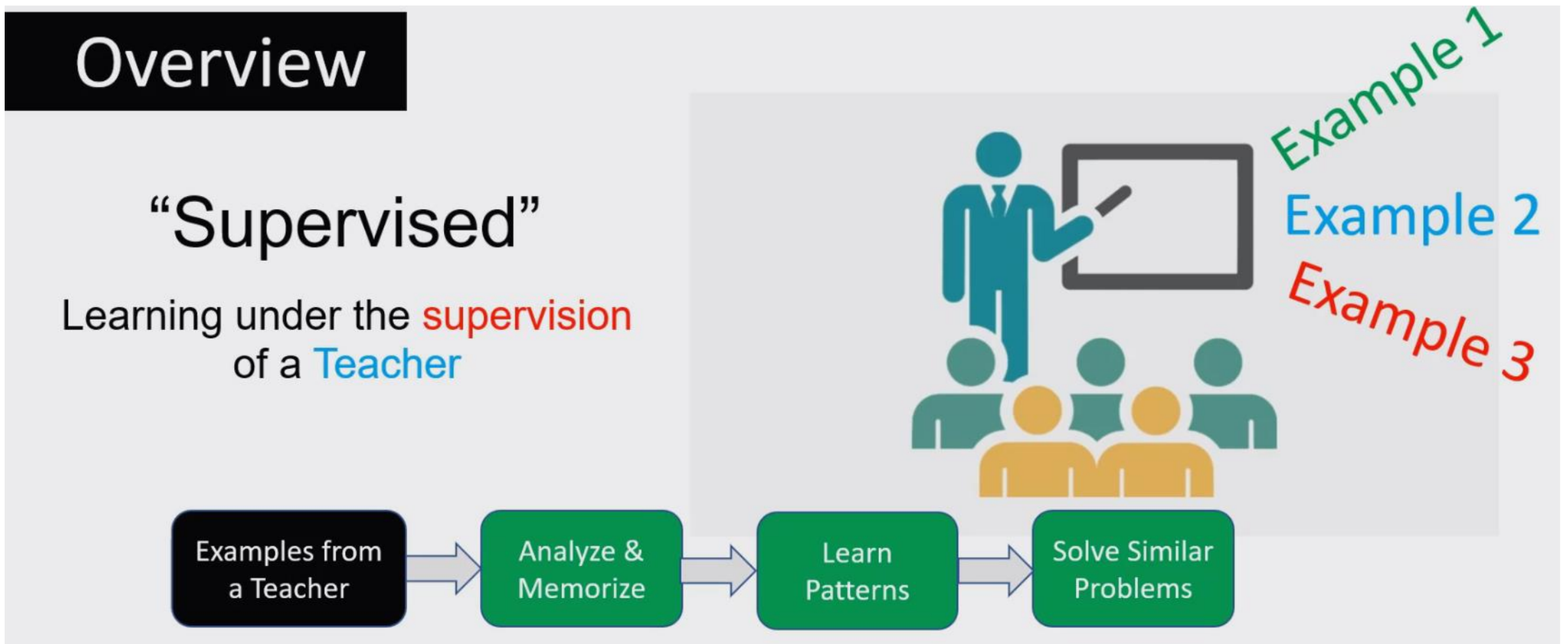
ALGORITMI DI MACHINE LEARNING

TEST SUPERVISIONATO

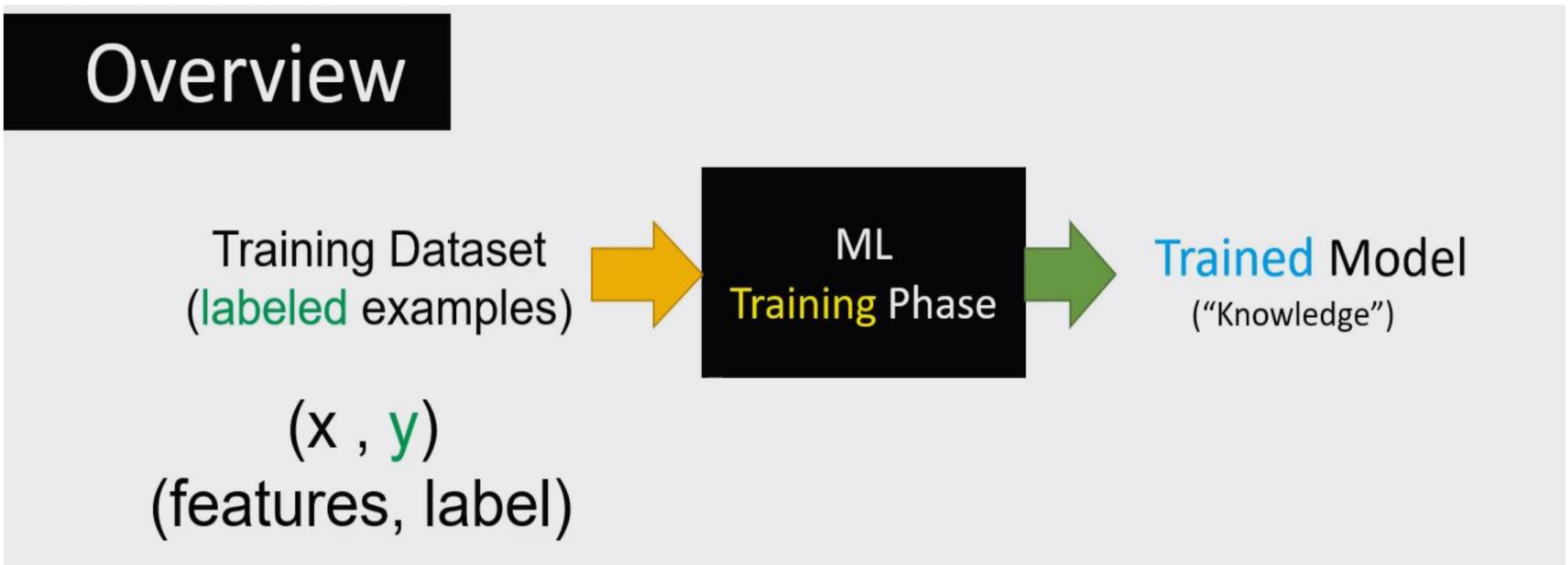


<u>TEST SET</u>: è l'insieme di pattern su cui valutare le prestazioni finali del sistema neurale.

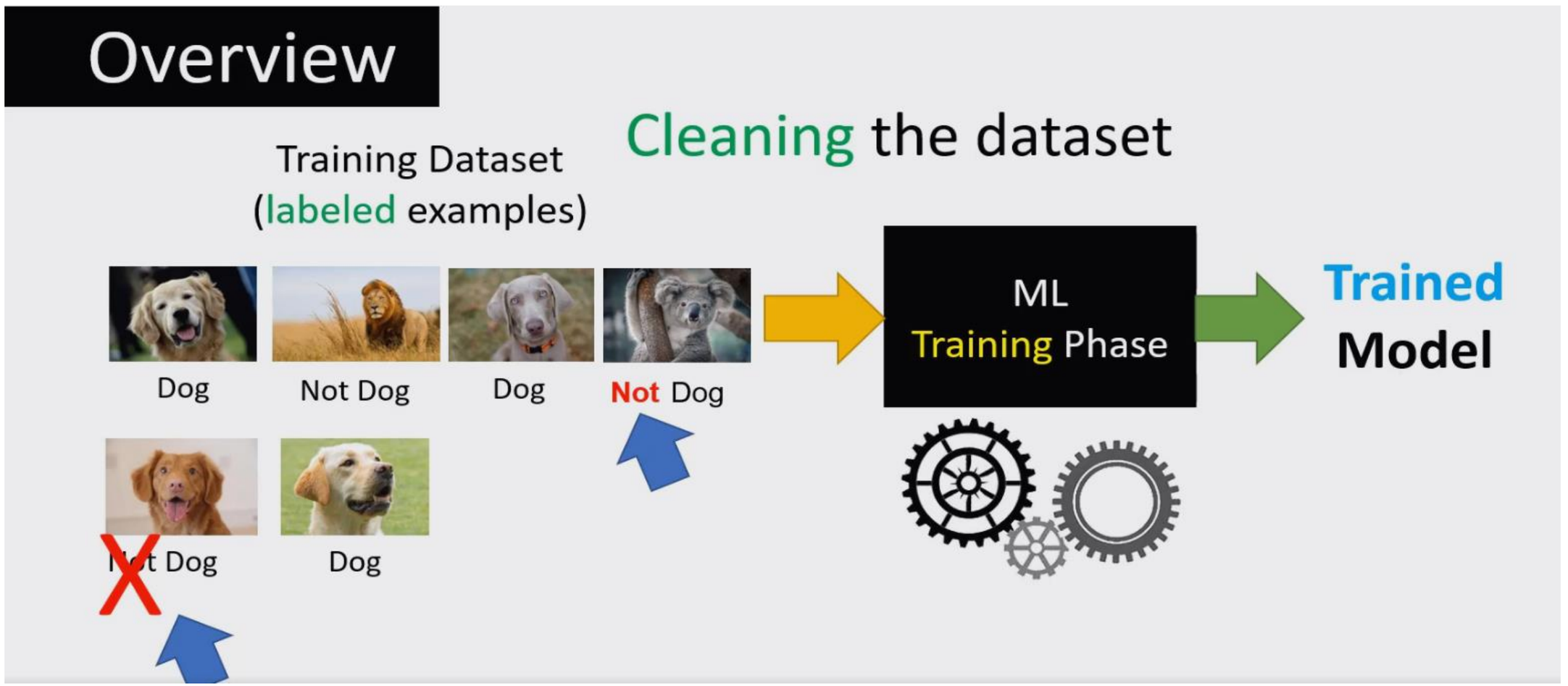
APPRENDIMENTO SUPERVISIONATO



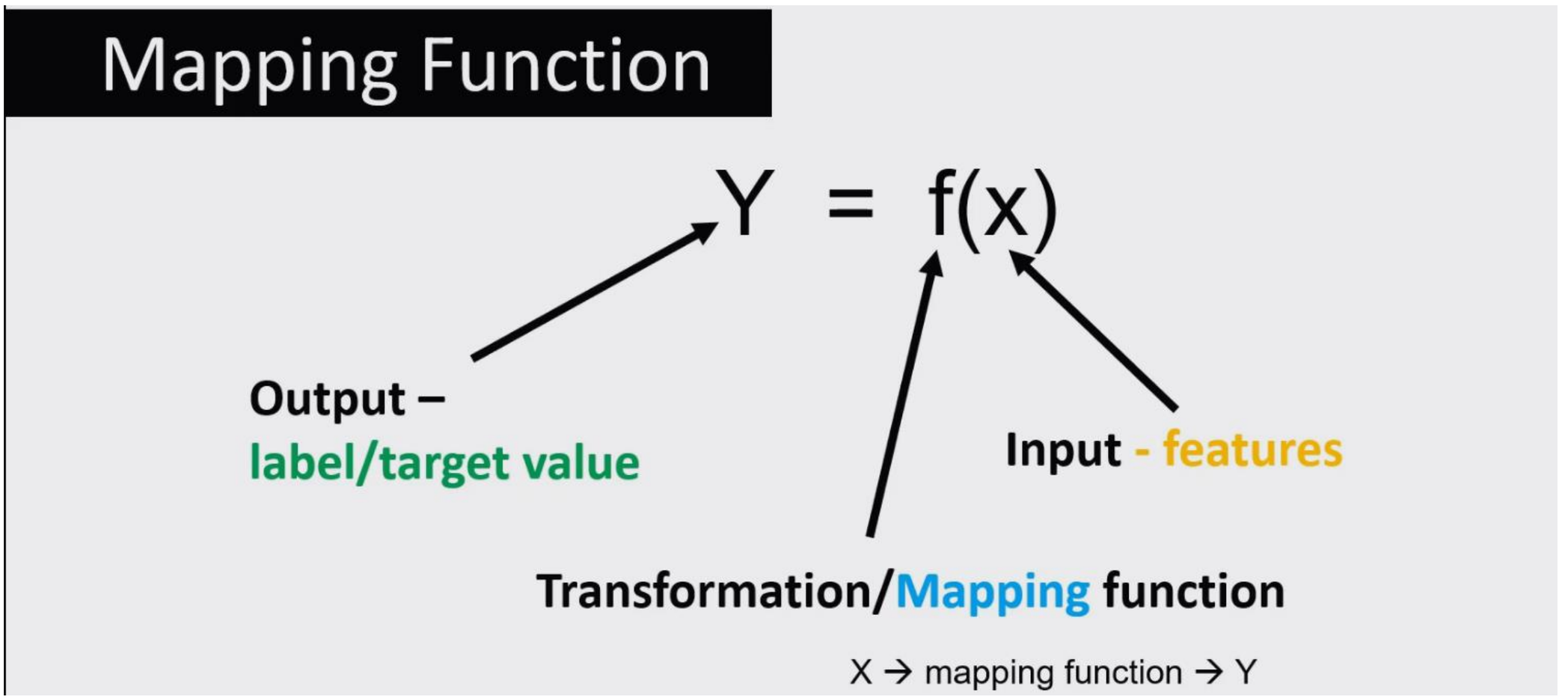
APPRENDIMENTO SUPERVISIONATO



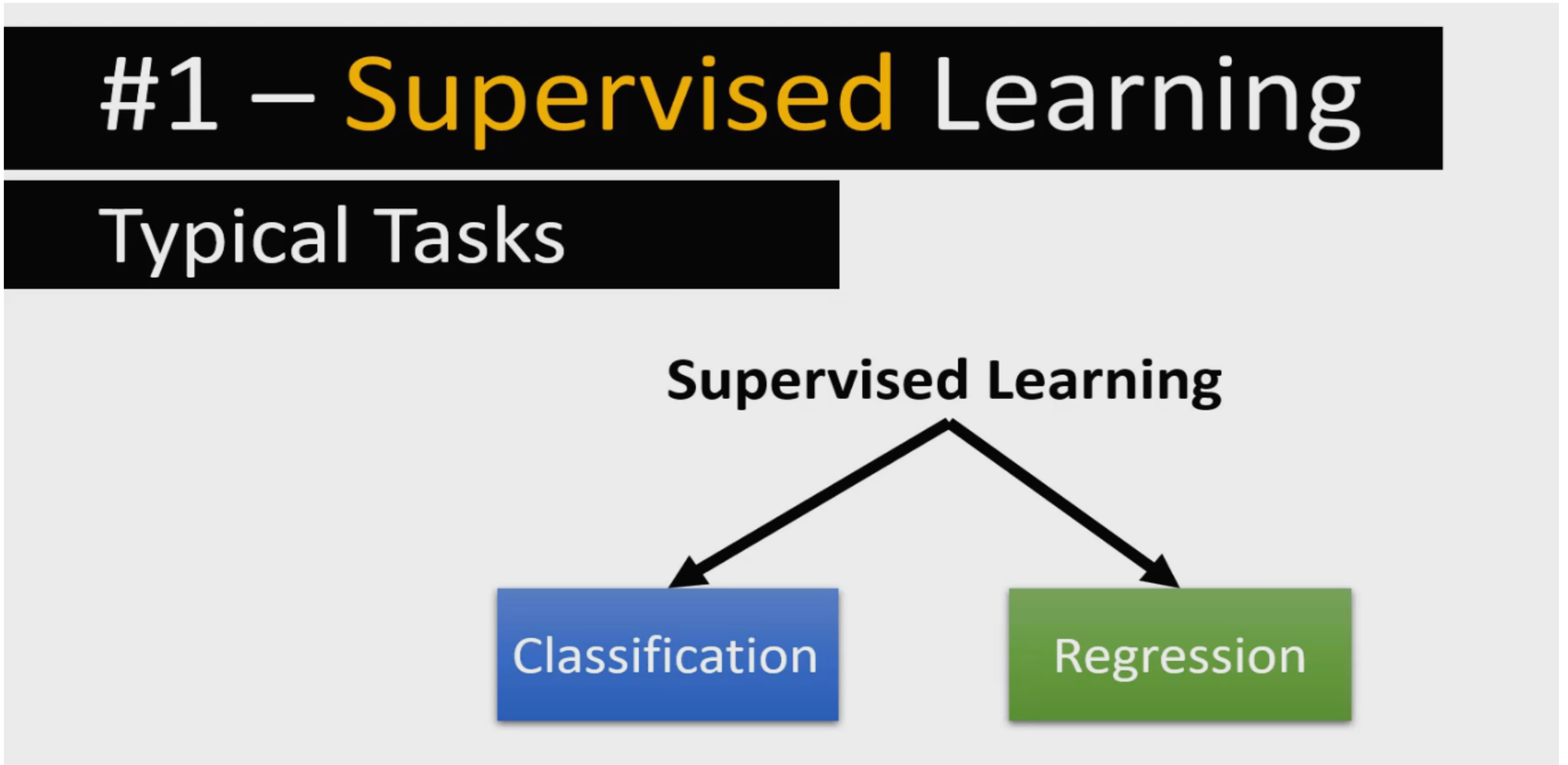
APPRENDIMENTO SUPERVISIONATO



APPRENDIMENTO SUPERVISIONATO



APPRENDIMENTO SUPERVISIONATO



COME FUNZIONA IL MACHINE LEARNING?

PROBLEMA DELLA CLASSIFICAZIONE

Classificazione

- Classificazione: assegna una classe a un pattern.
  - Necessario apprendere una funzione capace di eseguire il mapping dallo spazio dei pattern allo spazio delle classi
  - Si usa spesso anche il termine riconoscimento.
  - Nel caso di 2 sole classi si usa il termine binary classification, con più di due classi multi-class classification.

- Classe: insieme di pattern aventi proprietà comuni.
  - Es. i diversi modi in cui può essere scritto a mano libera il carattere A.
  - Il concetto di classe è semantico e dipende strettamente dall'applicazione:

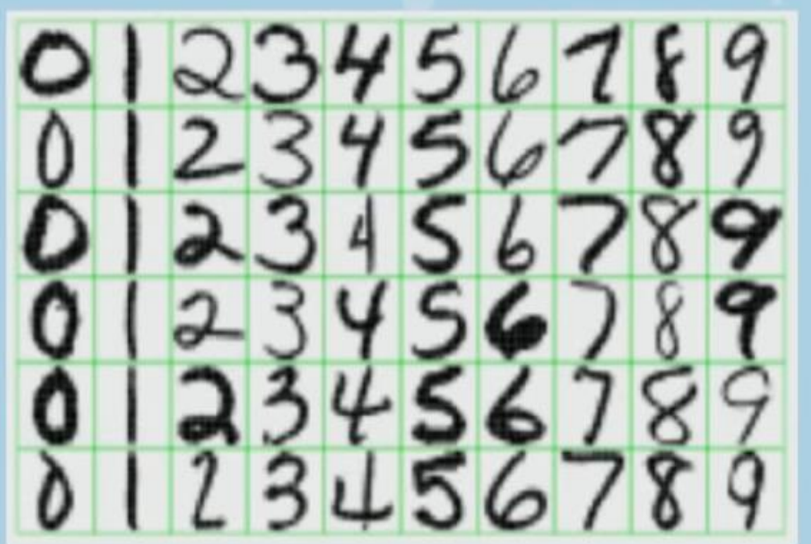
    - 21 classi per il riconoscimento di lettere dell'alfabeto

    - 2 classi per distinguere le lettere dell'alfabeto italiano da quello cirillico

COME FUNZIONA IL MACHINE LEARNING?

PROBLEMA DELLA CLASSIFICAZIONE

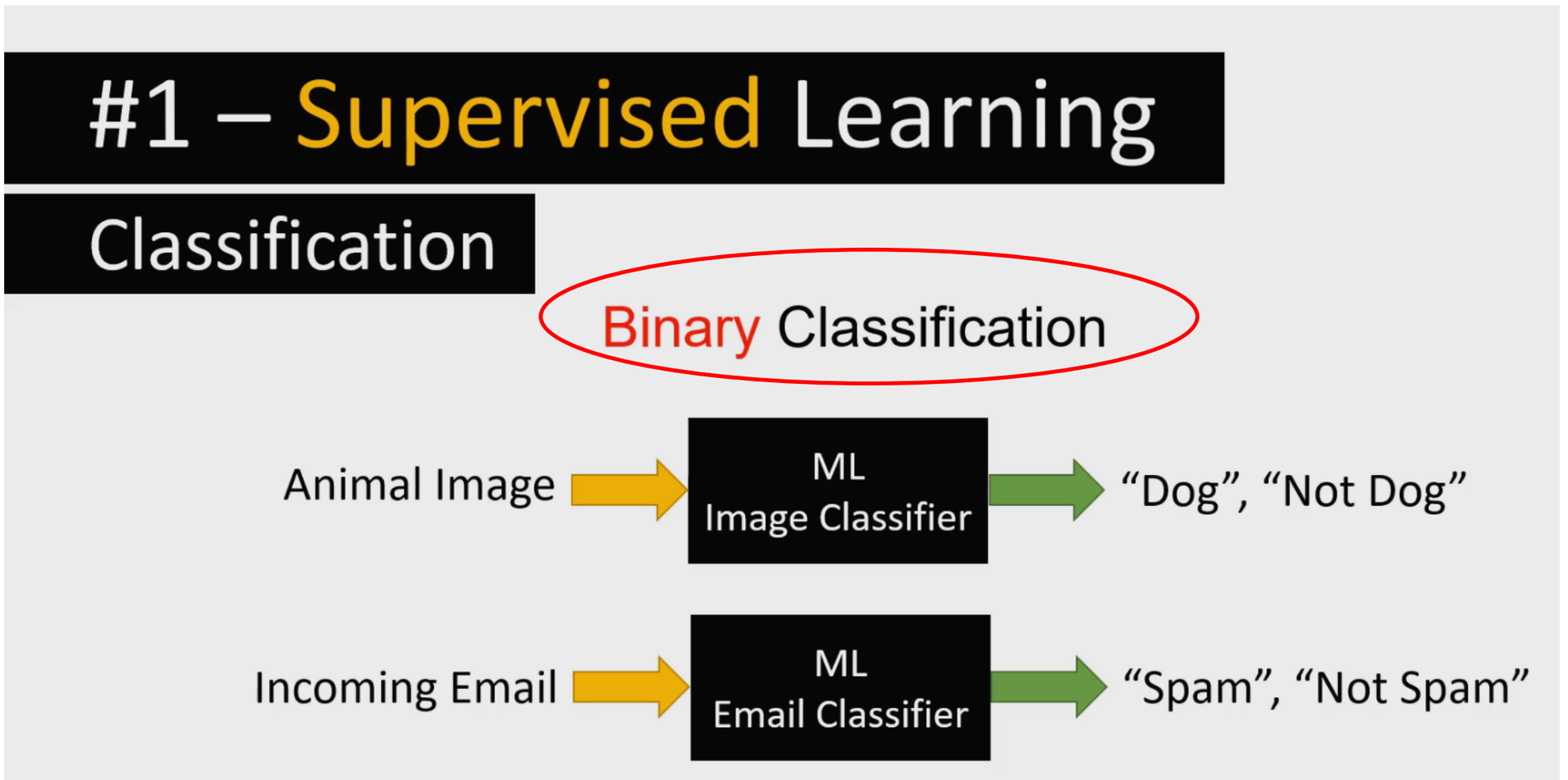
Handwritten ZIP codes on envelopes from US postal mail



- Output classes (0,1, ... , 9)
- Classification Problem

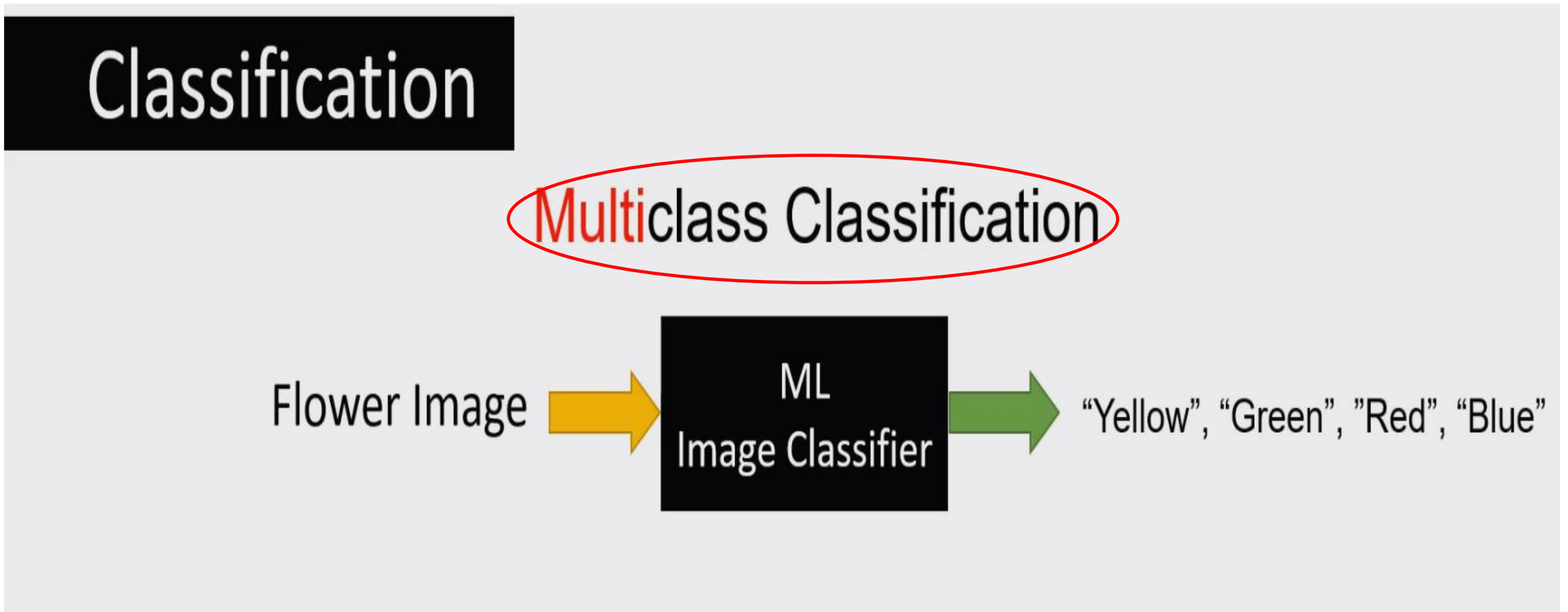
APPRENDIMENTO SUPERVISIONATO

CLASSIFICAZIONE



APPRENDIMENTO SUPERVISIONATO

CLASSIFICAZIONE

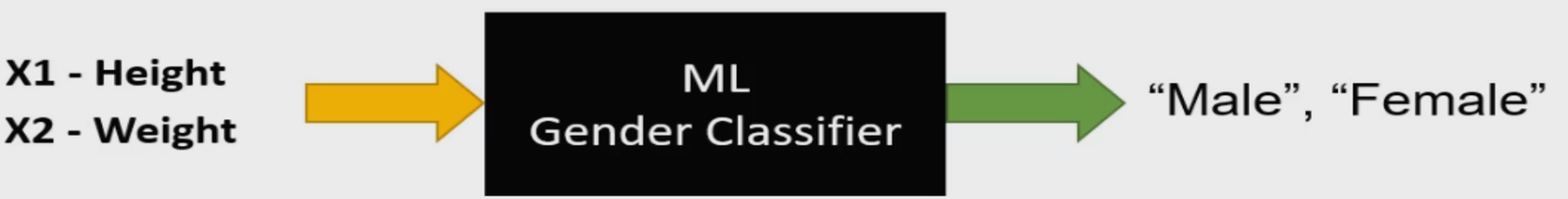


APPRENDIMENTO SUPERVISIONATO

CLASSIFICAZIONE

Classification

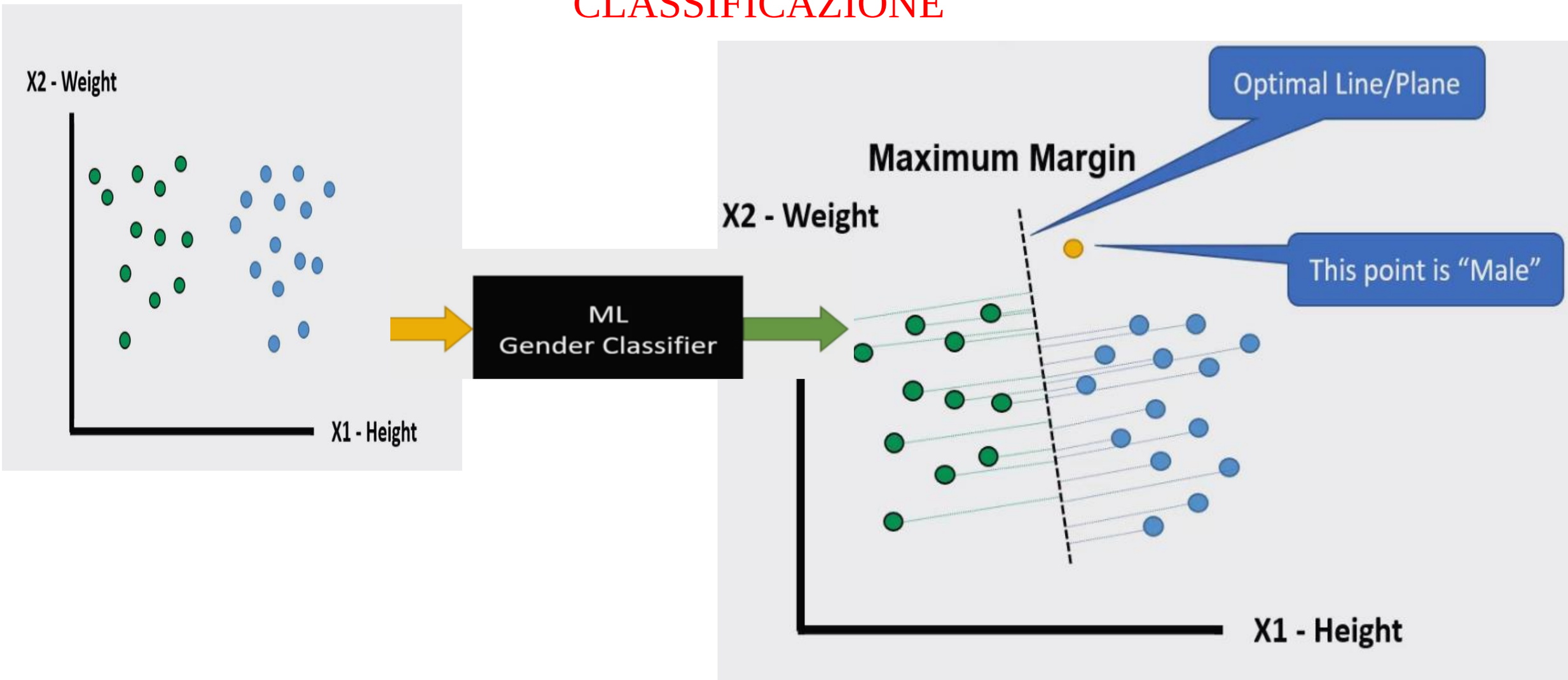
Support Vector Machines (SVMs)



X1 – H	X2 - W	Gender
160	56	F
167	74	M
183	85	M
.....		

APPRENDIMENTO SUPERVISIONATO

CLASSIFICAZIONE



APPRENDIMENTO SUPERVISIONATO

REGRESSIONE

Apprendimento per regressione

Nel machine learning la regressione lineare è una tecnica di classificazione degli esempi di un dataset (insieme di training) per consentire alla macchina di apprendere automaticamente un modello decisionale.

L'algoritmo assegna delle etichette (categorie) alle istanze utilizzando una funzione continua.

Ogni riga del dataset è un esempio composto da:

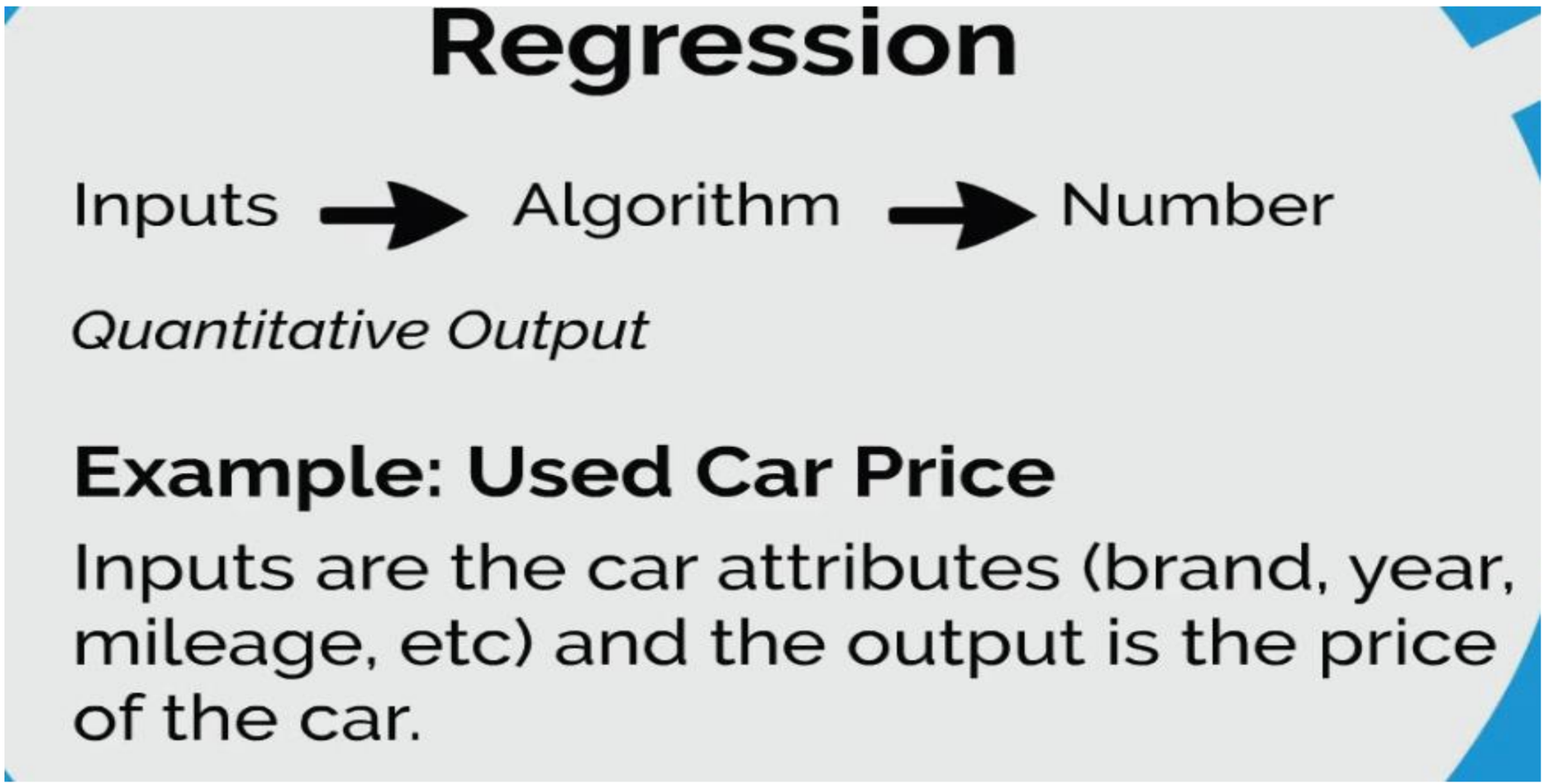
- **Gli attributi (X)**. Sono le variabili predittive che descrivono una categoria.
- **Il risultato (Y)**. E' la variabile target che indica alla macchina il risultato corretto se l'istanza appartenesse all'etichetta Z. E' il dato che istruisce la macchina a decidere in modo giusto.

L'algoritmo di machine learning deve trovare una relazione tra le variabili X e Y tramite la regressione.

$$Y=F(X)$$

APPRENDIMENTO SUPERVISIONATO

REGRESSIONE



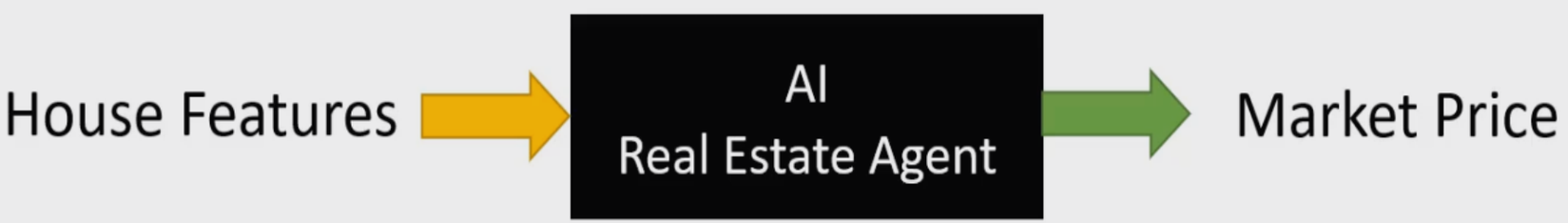
APPRENDIMENTO SUPERVISIONATO

REGRESSIONE

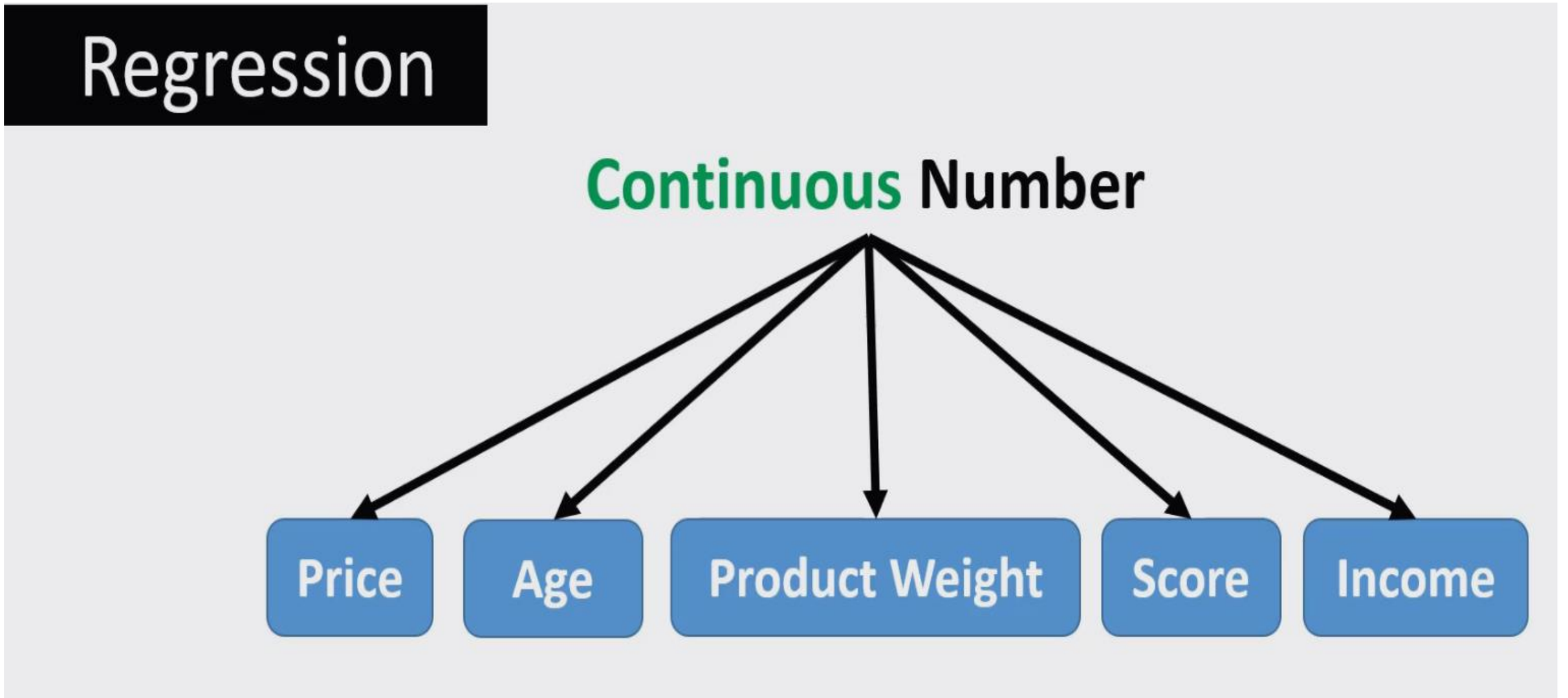
Regression

Statistical method to analyze and predict data

Predict a **continuous** number

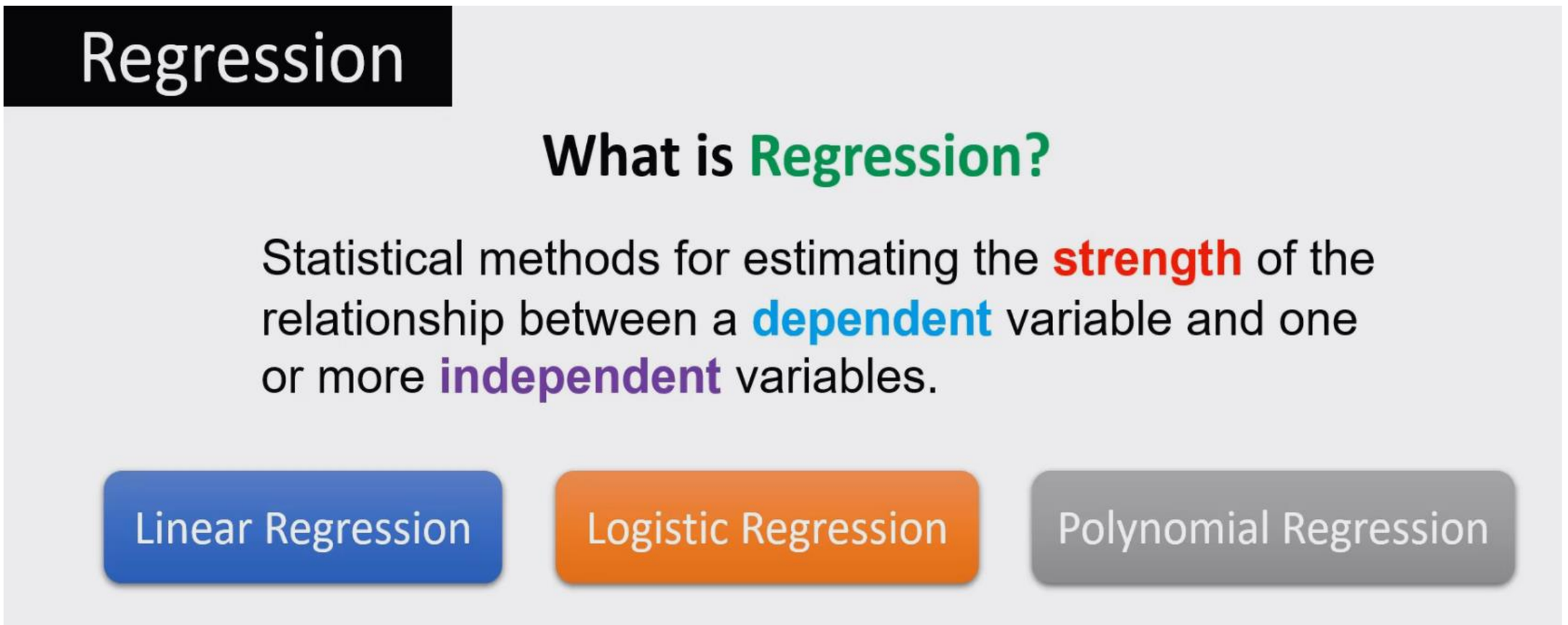


APPRENDIMENTO SUPERVISIONATO



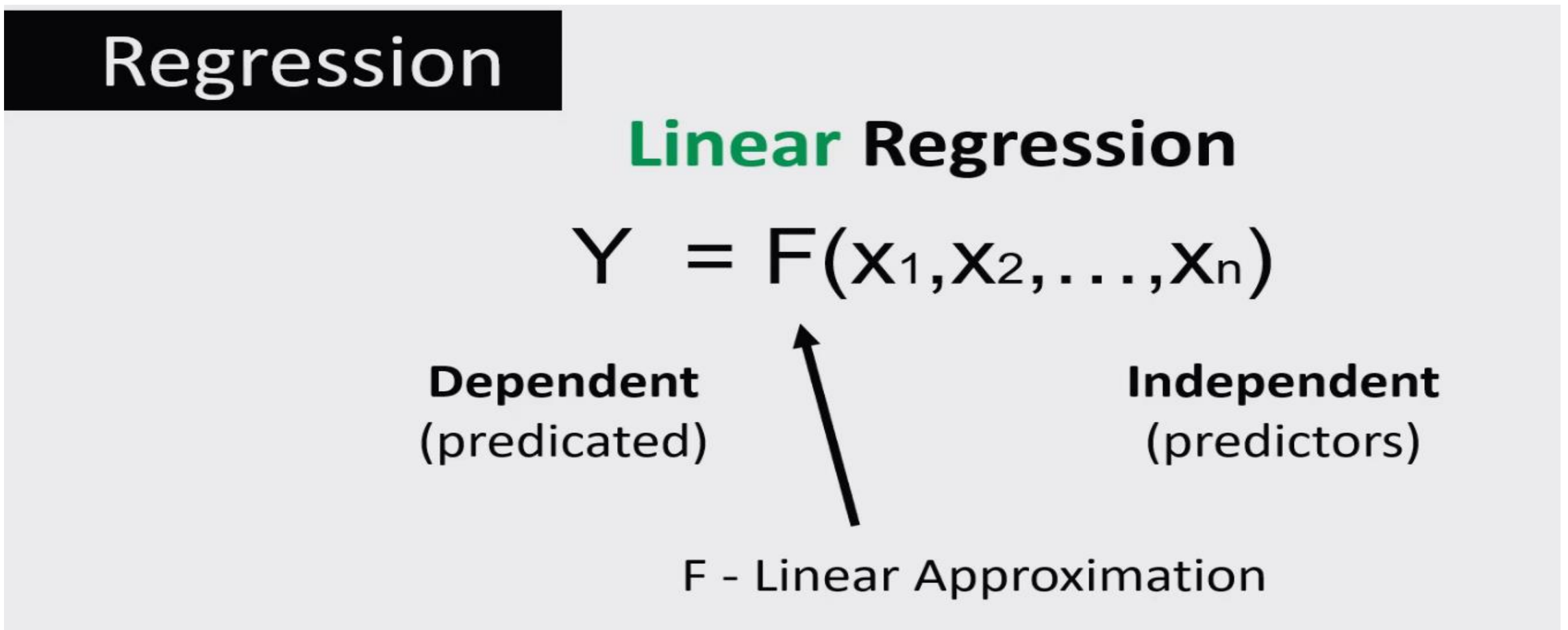
APPRENDIMENTO SUPERVISIONATO

REGRESSIONE



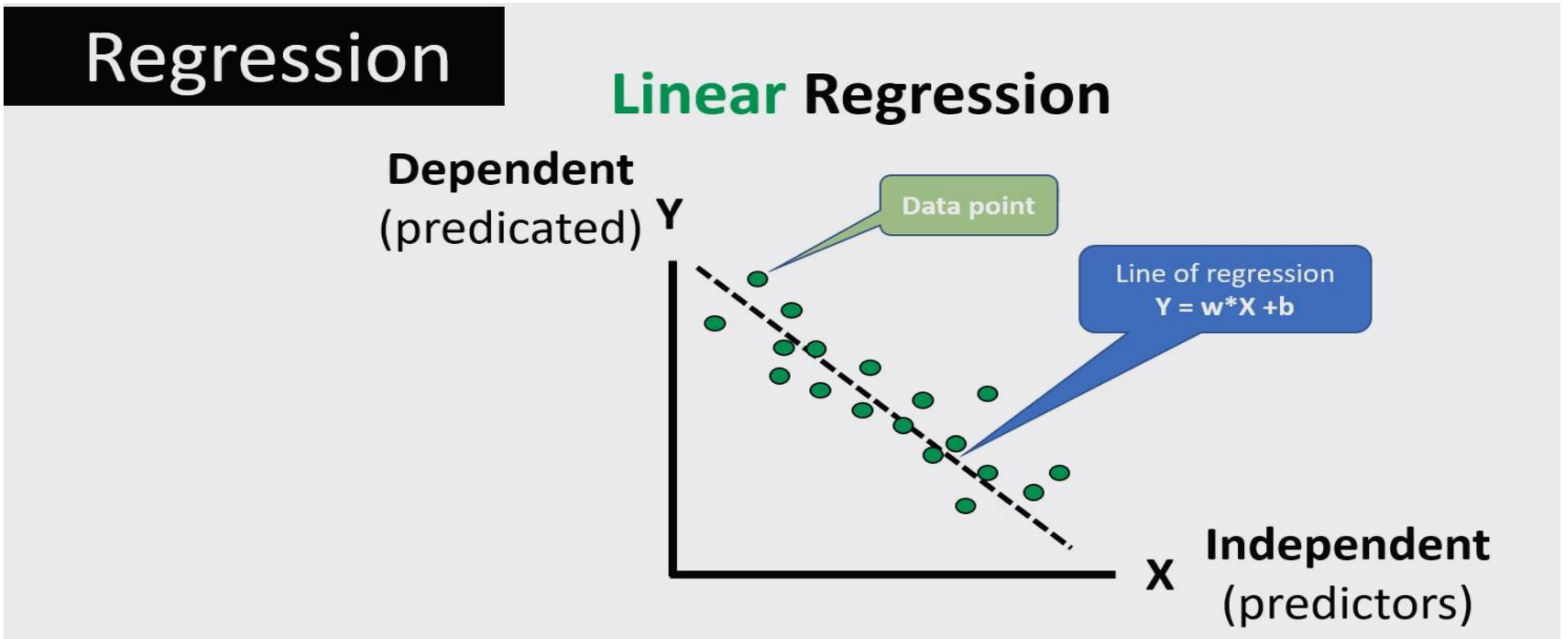
APPRENDIMENTO SUPERVISIONATO

REGRESSIONE



APPRENDIMENTO SUPERVISIONATO

REGRESSIONE



APPRENDIMENTO SUPERVISIONATO

REGRESSIONE

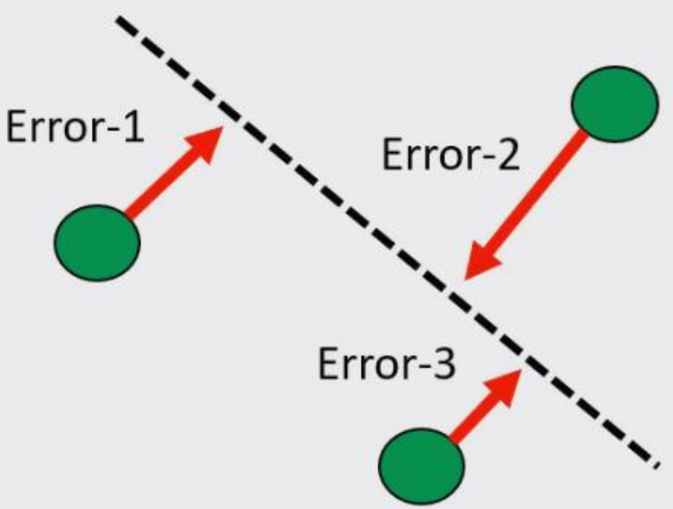
Regression

Linear Regression

Optimization using a Cost Function

Minimum(MSE)

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

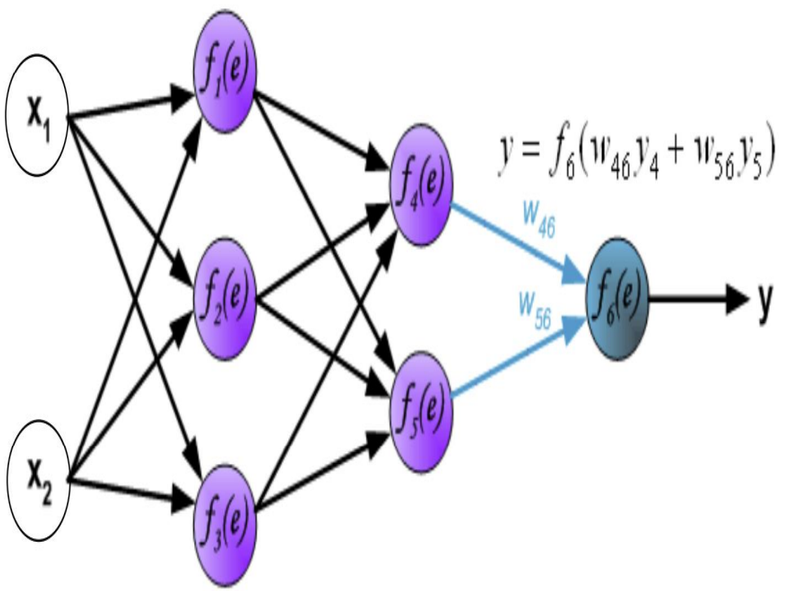


ALGORITMI DI MACHINE LEARNING

APPRENDIMENTO (TRAINING)

In genere, il comportamento di un algoritmo di Machine Learning è regolato da un set di parametri Θ (es. i pesi delle connessioni in una rete neurale). L'apprendimento consiste nel determinare il valore ottimo Θ^* di questi parametri.

Dato un training set $Train$ e un insieme di parametri, la funzione obiettivo $f(Train, \Theta)$ può indicare:



- l'ottimalità della soluzione (da massimizzare).

$$\Theta^* = argmax_\Theta\, f(Train, \Theta)$$

- oppure l'errore o perdita (loss-function) da minimizzare.

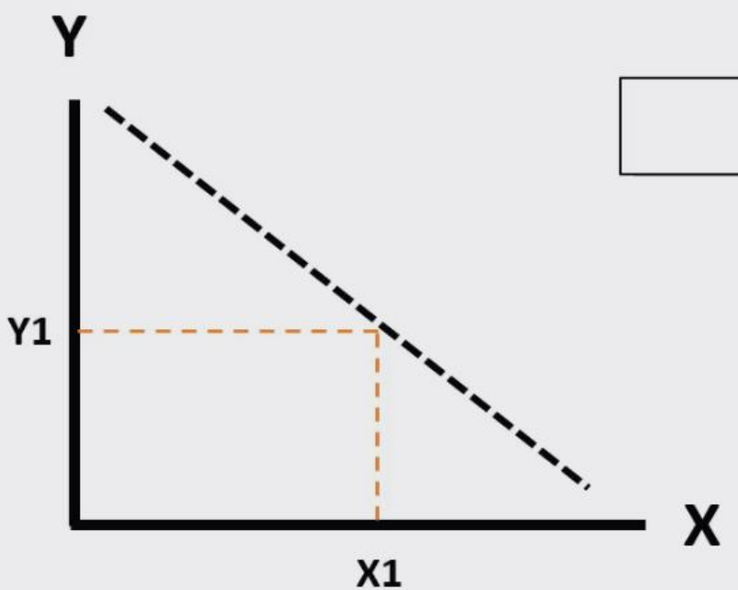
$$\Theta^* = argmin_\Theta\, f(Train, \Theta)$$

APPRENDIMENTO SUPERVISIONATO

REGRESSIONE

Regression

Linear Regression



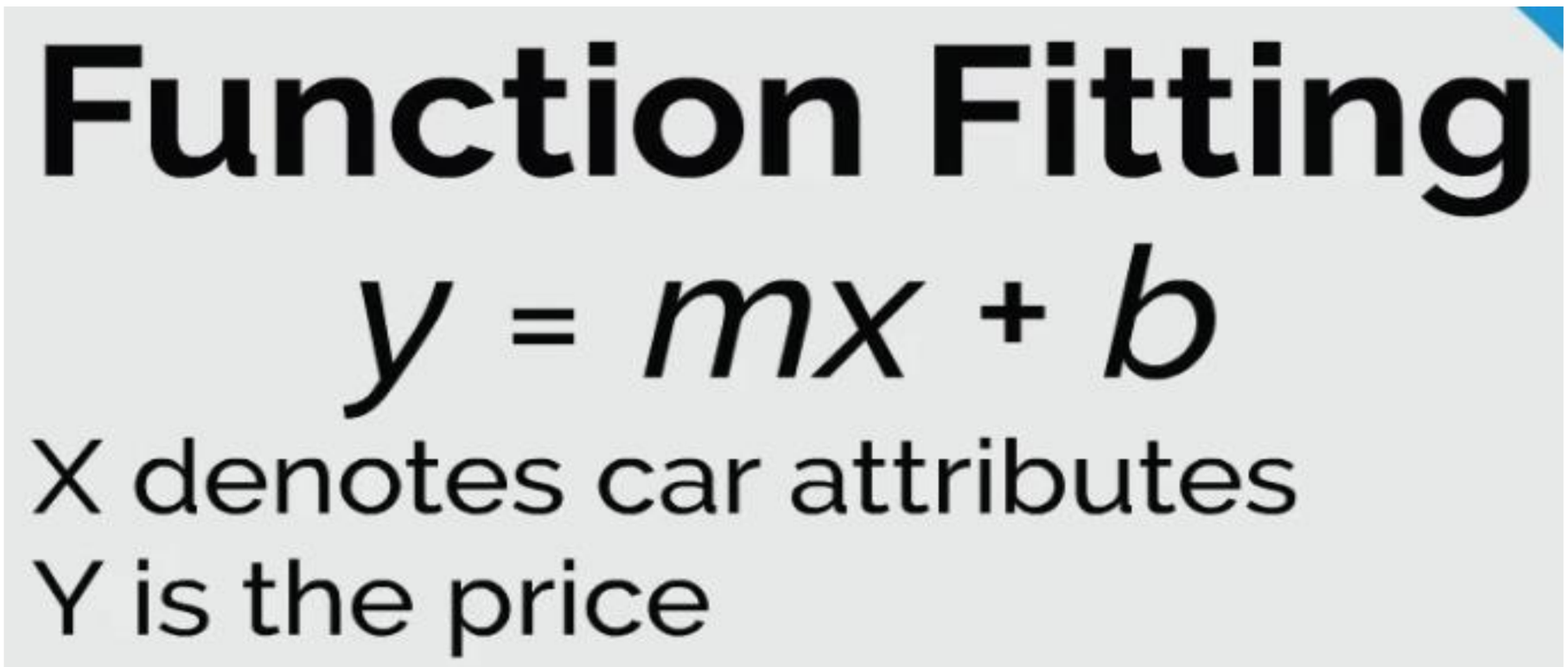
$$y = w[0]*x[0]+w[1]*x[1] +\ldots+ w[i]*x[i]+b$$

x[i] – input features

w[i], b – model parameters

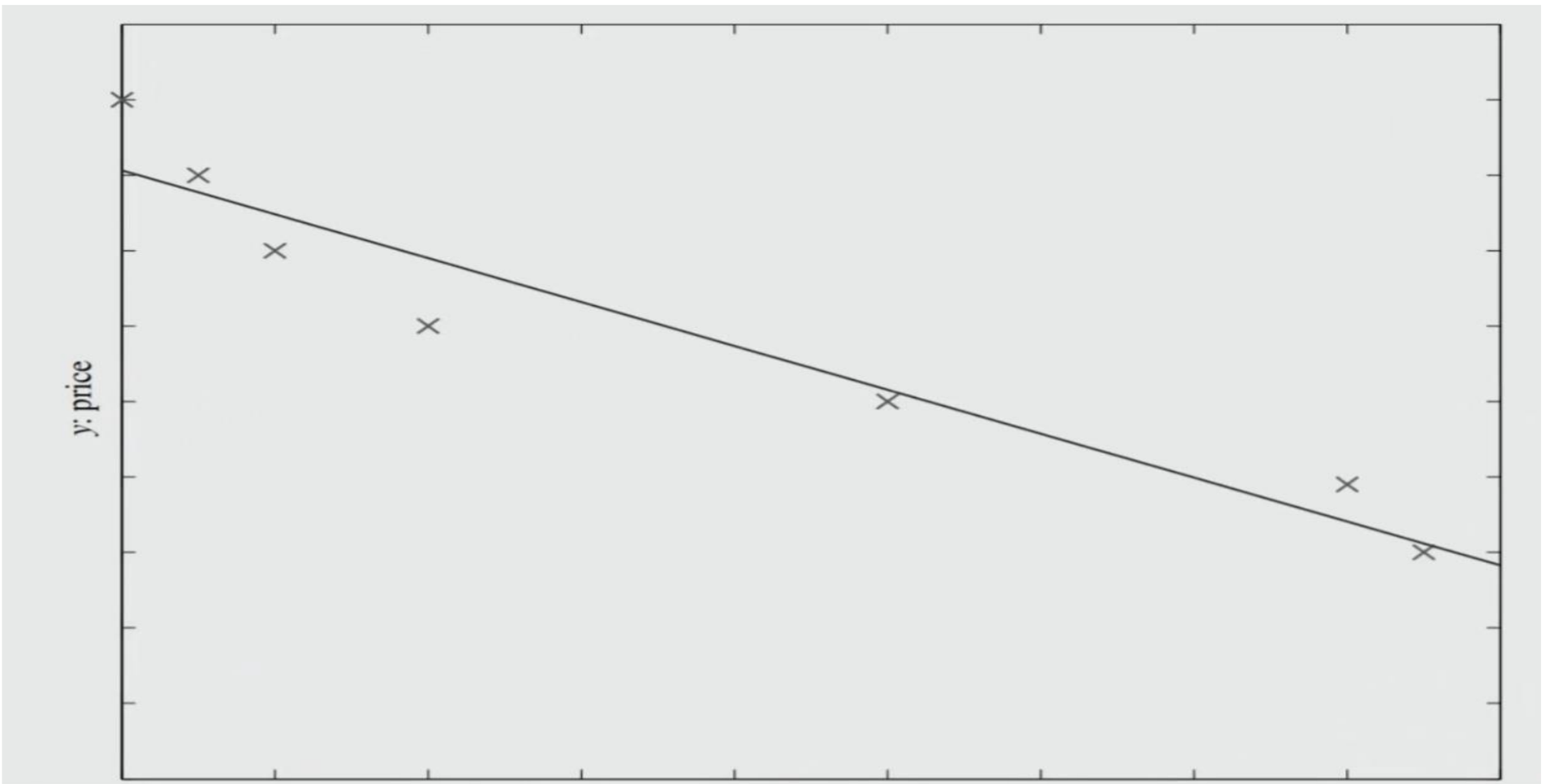
APPRENDIMENTO SUPERVISIONATO

REGRESSIONE



APPRENDIMENTO SUPERVISIONATO

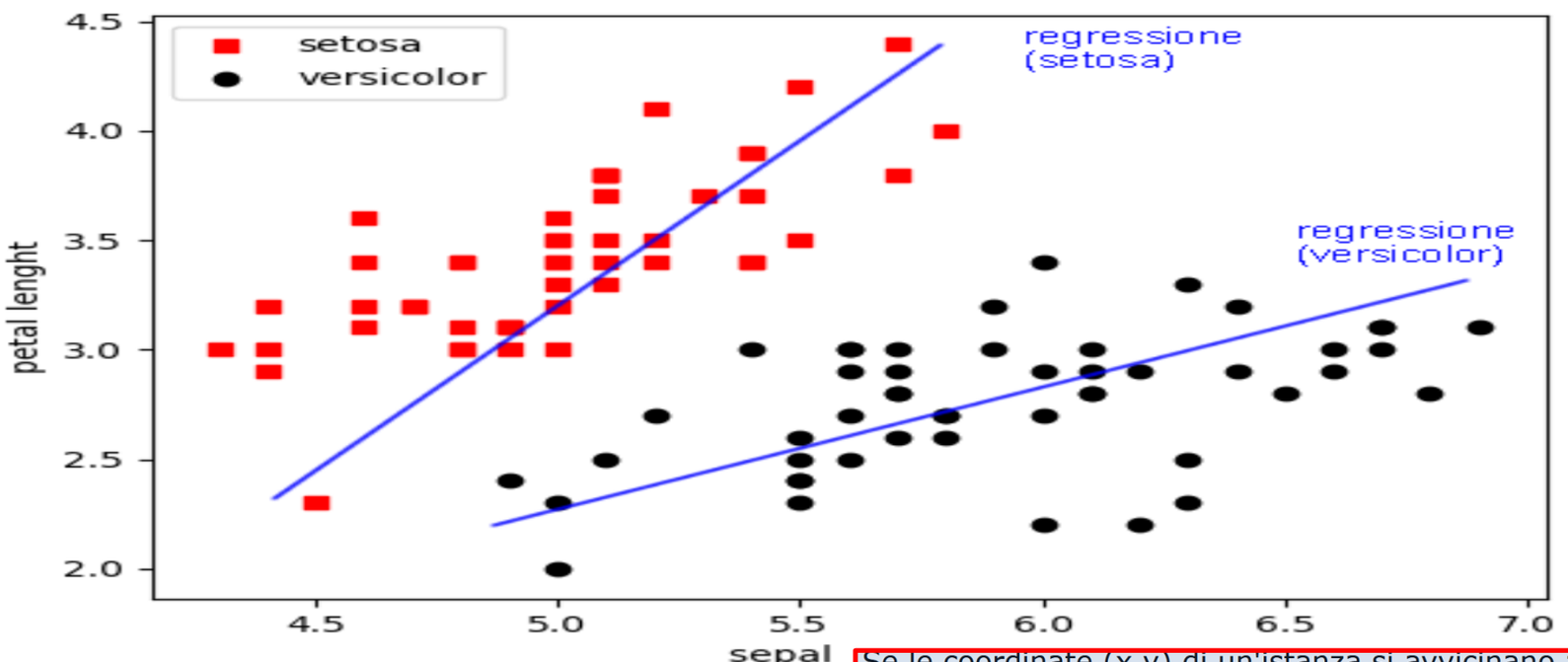
REGRESSIONE



APPRENDIMENTO SUPERVISIONATO

REGRESSIONE

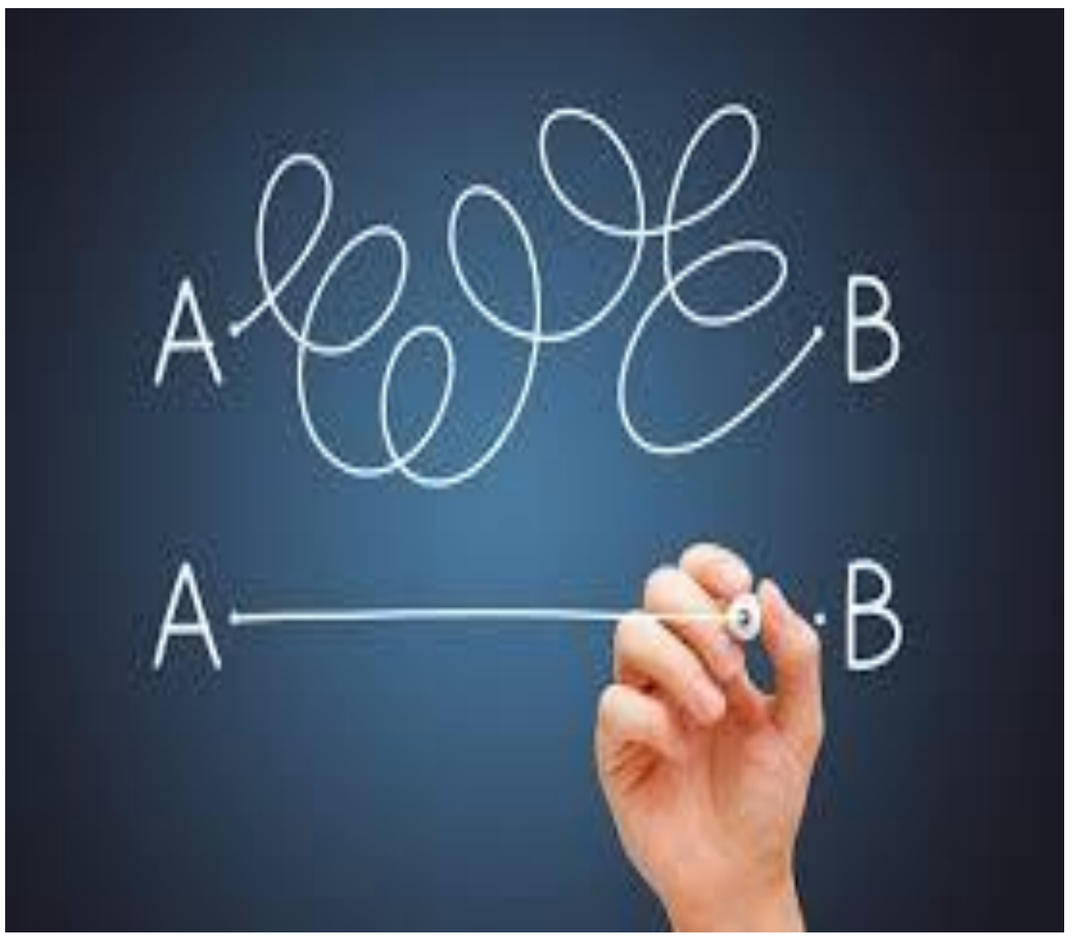
Il risultato finale è una linea retta che minimizza la distanza tra gli N esempi dell'insieme di training che appartengono alla stessa categoria.



Se le coordinate (x,y) di un'istanza si avvicinano alla funzione di regressione f(x), l'istanza viene classificata con l'etichetta Z.

Una volta trovata, la **funzione di classificazione** può essere utilizzata per valutare istanze diverse dall'insieme di training.


E=mc2
R=∫p(t)dt


A
B
A
B

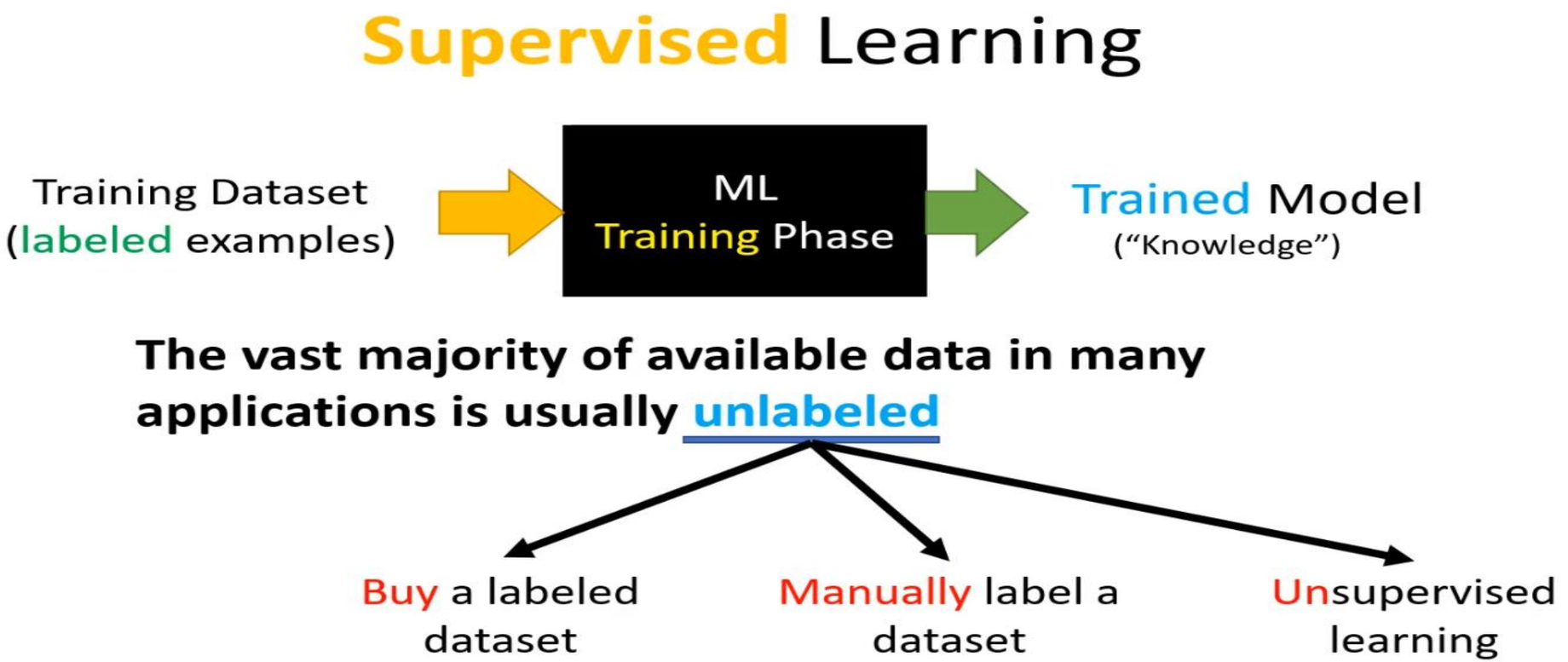
ALGORITMI DI MACHINE LEARNING

APPRENDIMENTO (TRAINING)

- Supervisionato (Supervised): sono note le classi dei patterr utilizzati per l'addestramento.
  - *il training set è etichettato*.
    - Non Supervisionato (Unsupervised): non sono note le classi dei pattern utilizzati per l'addestramento.
      - *il training set non è etichettato*.

- It is often easier to obtain unlabeled data – from a lab instrument or a computer – than labeled data, which can require human intervention

APPRENDIMENTO NON SUPERVISIONATO

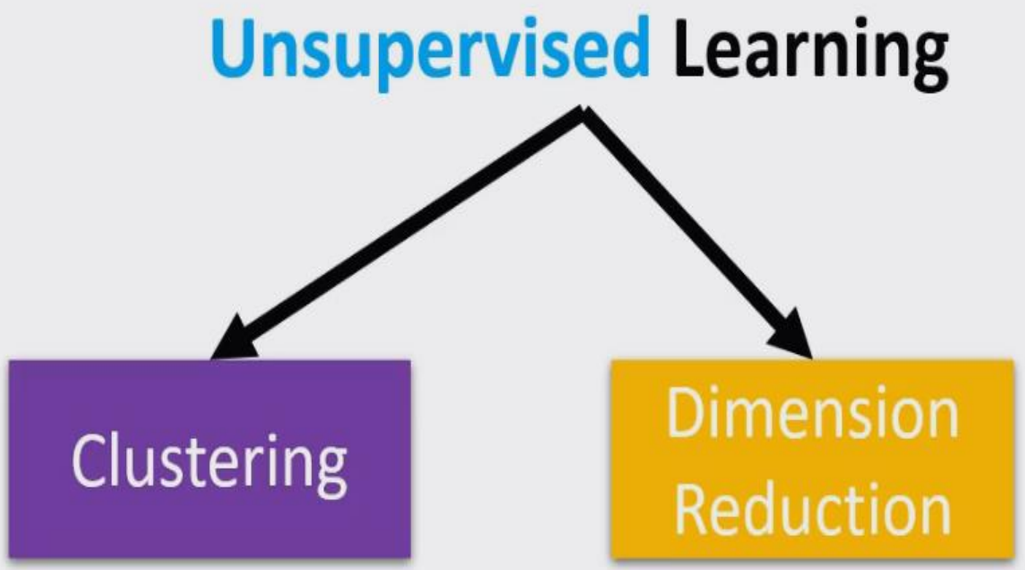


APPRENDIMENTO NON SUPERVISIONATO

Overview

Learning without a “teacher” supervising the learning process

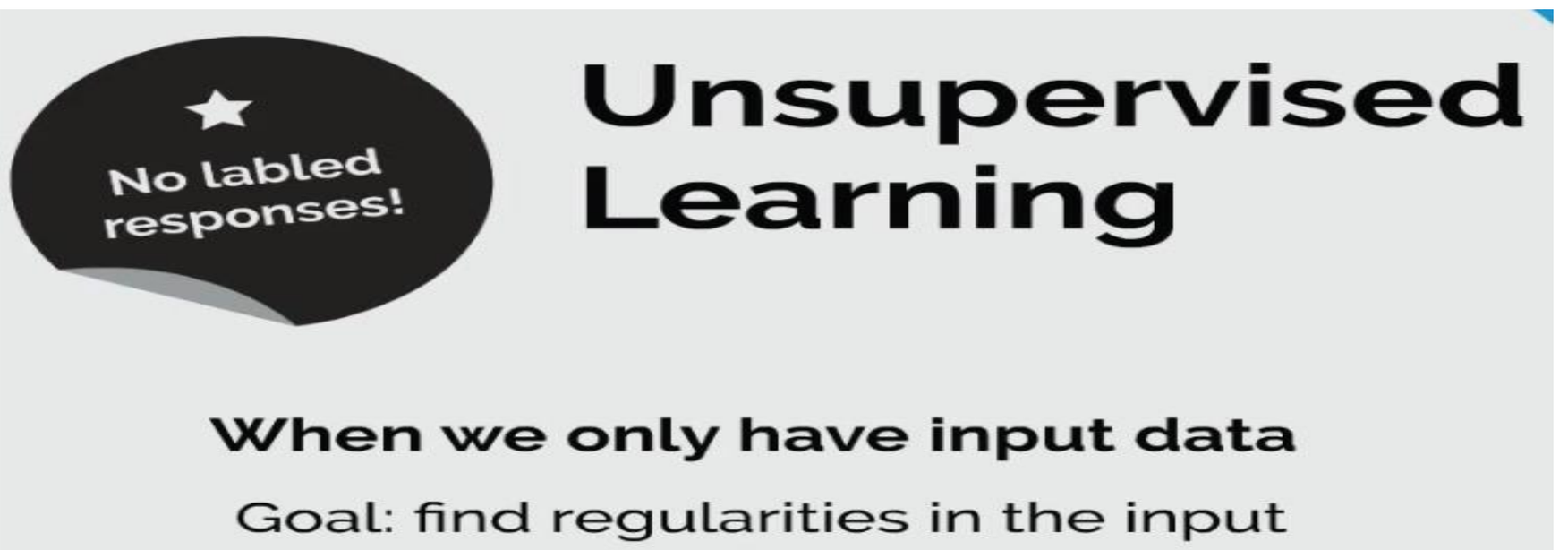
Identify automatically meaningful patterns in unlabeled data



ALGORITMI DI MACHINE LEARNING

APPRENDIMENTO (TRAINING)

- Non Supervisionato (Unsupervised): non sono note le classi dei pattern utilizzati per l'addestramento.
  - *il training set non è etichettato*.



APPRENDIMENTO NON SUPERVISIONATO

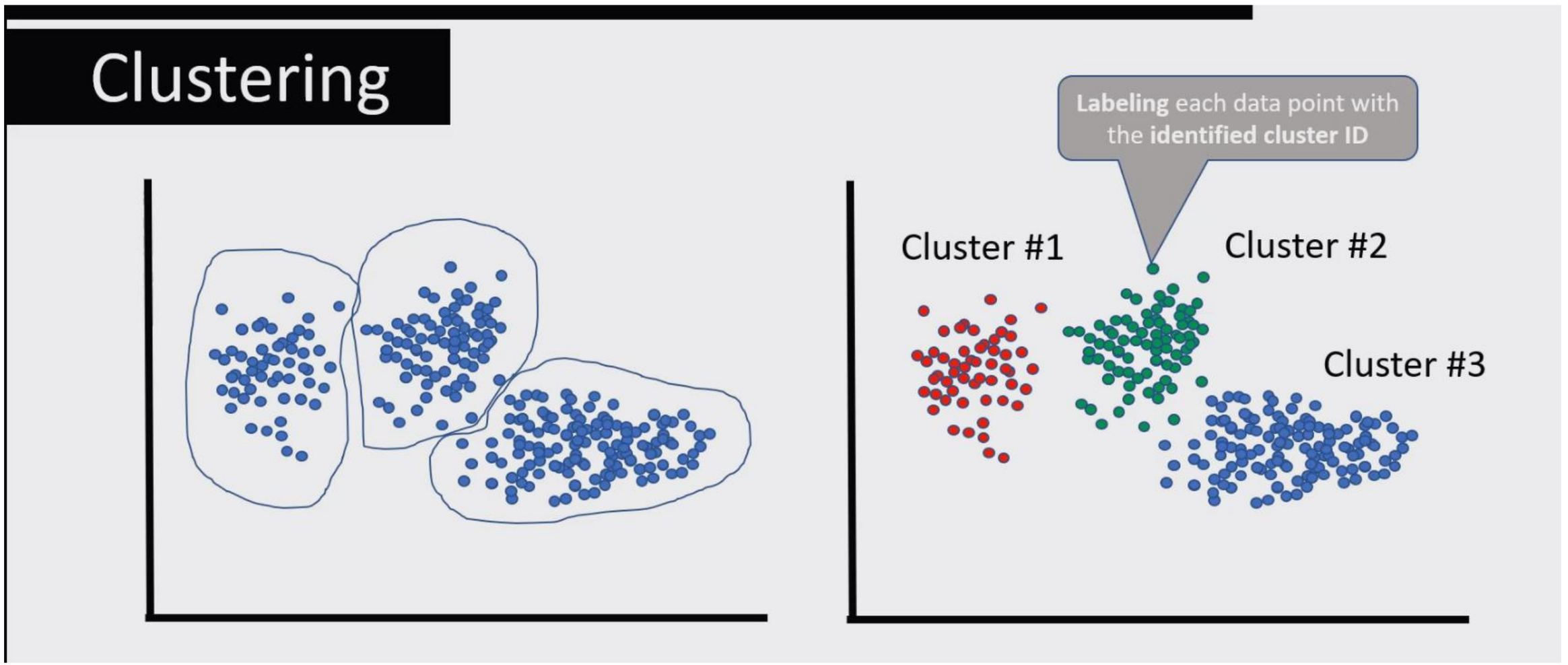
Clustering

What is Clustering?

Clustering is the task of identifying similar instances with **shared attributes** in a dataset and **group them together into clusters**

The **output** of the algorithm would be a set of **"labels"** assigning each data point to one of the identified clusters

APPRENDIMENTO NON SUPERVISIONATO



APPRENDIMENTO NON SUPERVISIONATO

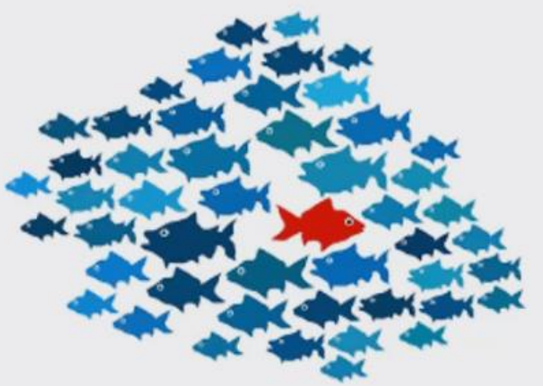
Clustering

Use Cases for Clustering

Customer Segmentation

Anomaly/Outlier Detection

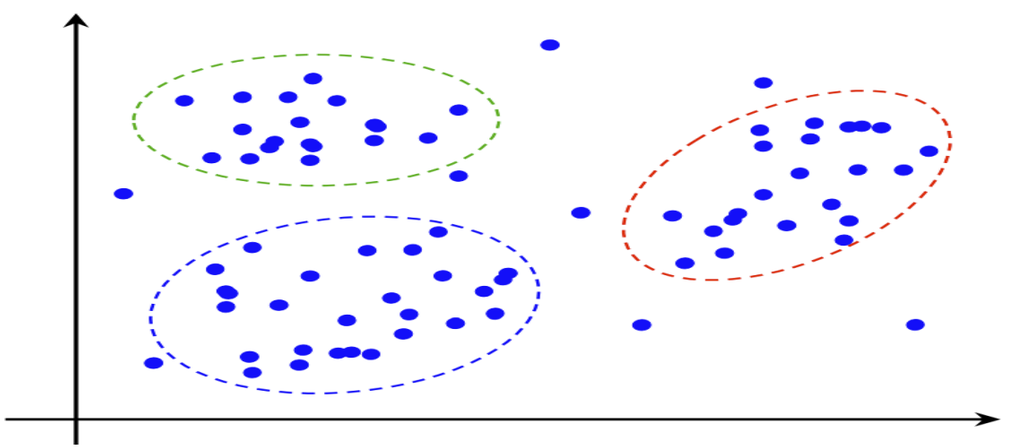
Semi-supervised Learning



ALGORITMI DI MACHINE LEARNING

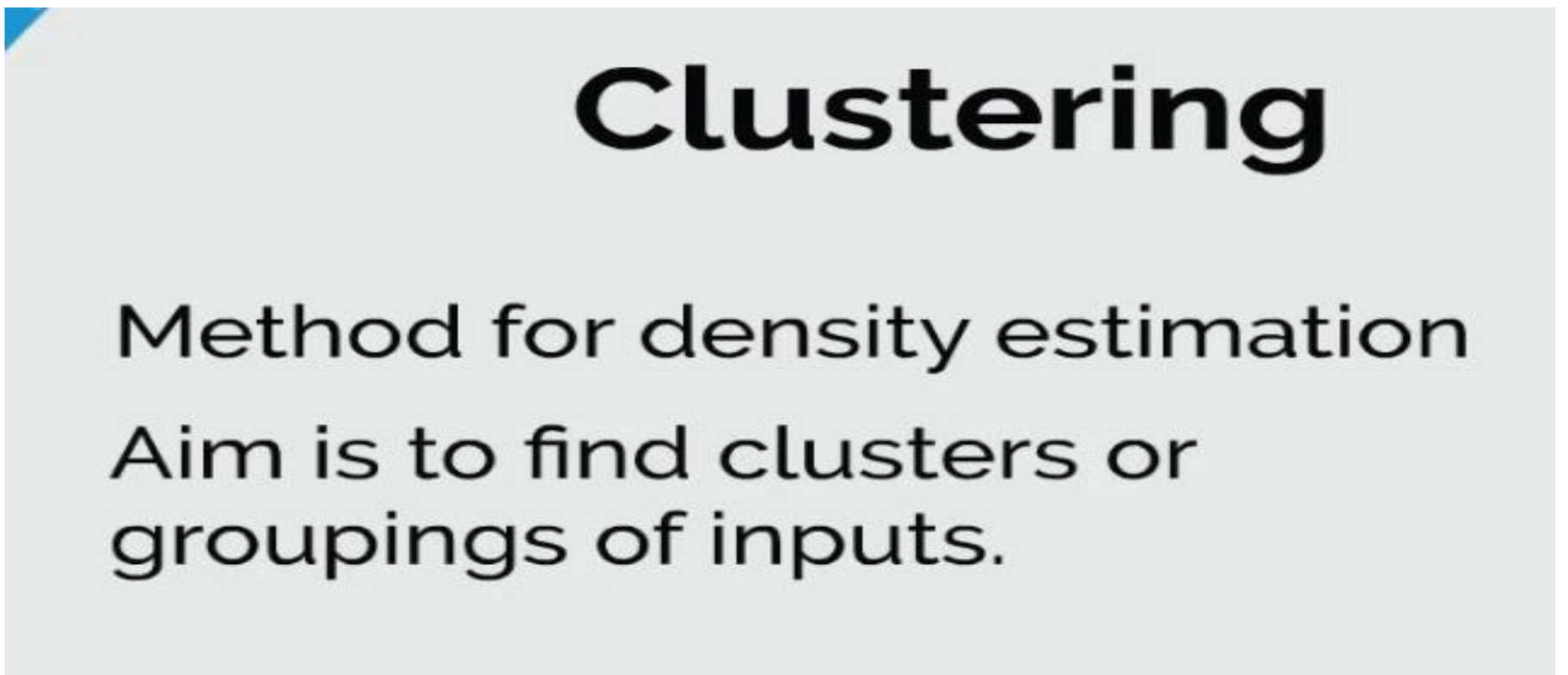
TRAINING NON SUPERVISIONATO - CLUSTERING

- Clustering: individua gruppi (cluster) di pattern con caratteristiche simili.
  - Le classi del problema non sono note e i pattern non etichettati → la natura non supervisionata del problema lo rende più complesso della classificazione.
  - Spesso nemmeno il numero di cluster è noto a priori
  - I cluster individuati nell'apprendimento possono essere poi utilizzati come classi.



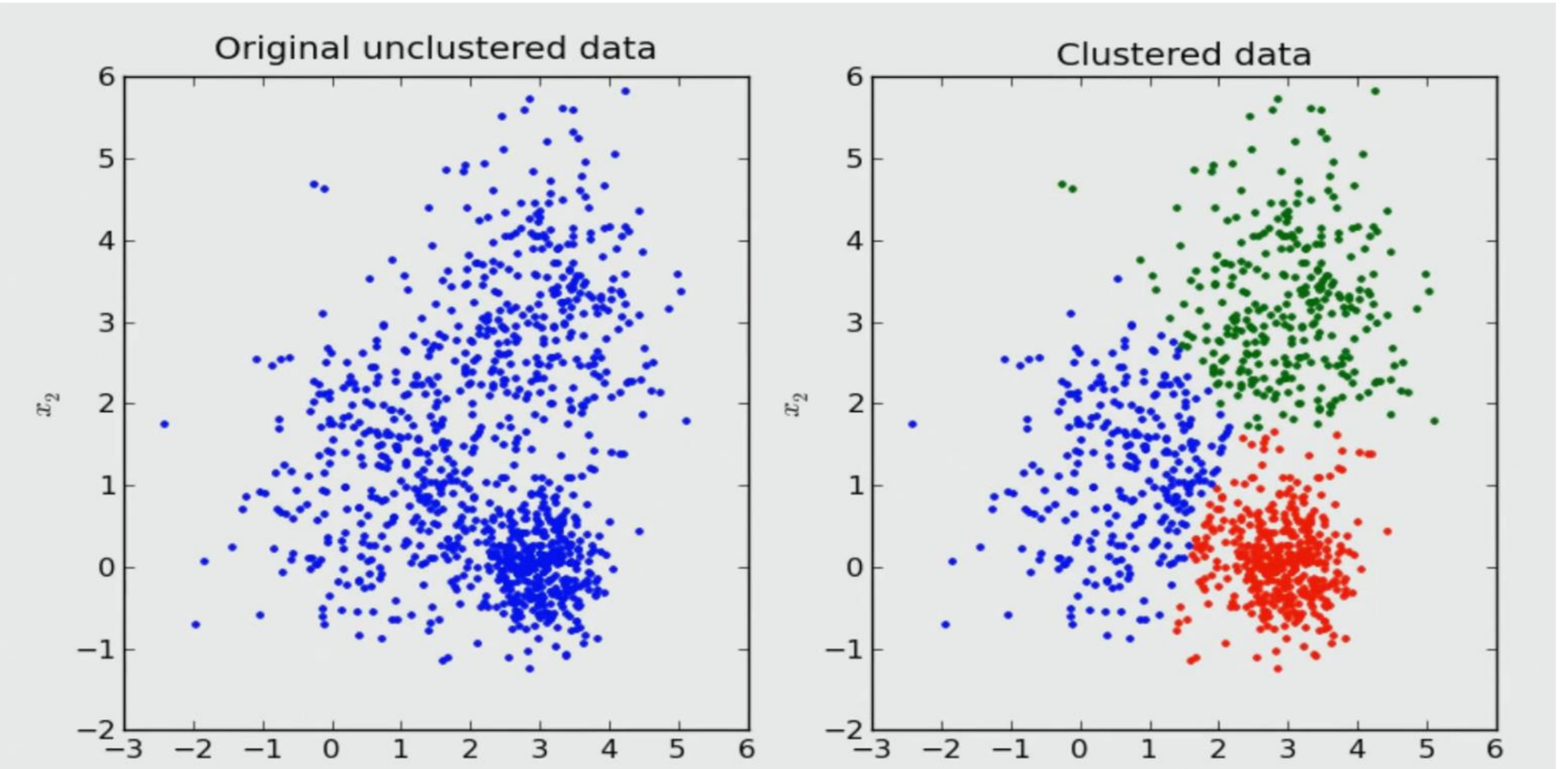
ALGORITMI DI MACHINE LEARNING

TRAINING NON SUPERVISIONATO - CLUSTERING



ALGORITMI DI MACHINE LEARNING

TRAINING NON SUPERVISIONATO - CLUSTERING



ABBIAMO SOLO DUE VARIABILI

ALGORITMI DI MACHINE LEARNING

TIPOLOGIE DI TRAINING

- Batch: l'addestramento è effettuato una sola volta su un training set dato.
  - una volta terminato il training, il sistema passa in «working mode» e non è in grado di apprendere ulteriormente.
  - Attualmente, la maggior parte dei sistemi di machine learning opera in questo modo.

- Incrementale: a seguito dell'addestramento iniziale, sono possibili ulteriori sessioni di addestramento.
  - Scenari: Sequenze di Batch, Unsupervised Tuning.
  - Rischio: Catastrofic Forgetting (il sistema dimentica quello che ha appreso in precedenza).

- Naturale: addestramento continuo (per tutta la vita)
  - Addestramento attivo in working mode.

ALGORITMI DI MACHINE LEARNING

ACCURATEZZA RISULTATI

In un problema di Classificazione, l'accuratezza di classificazione [0…100%] è la percentuale di pattern correttamente classificati. L'errore di classificazione è il complemento.

$$Accuratezza = \frac{pattern\ correttamente\ classificati}{pattern\ classificati}$$

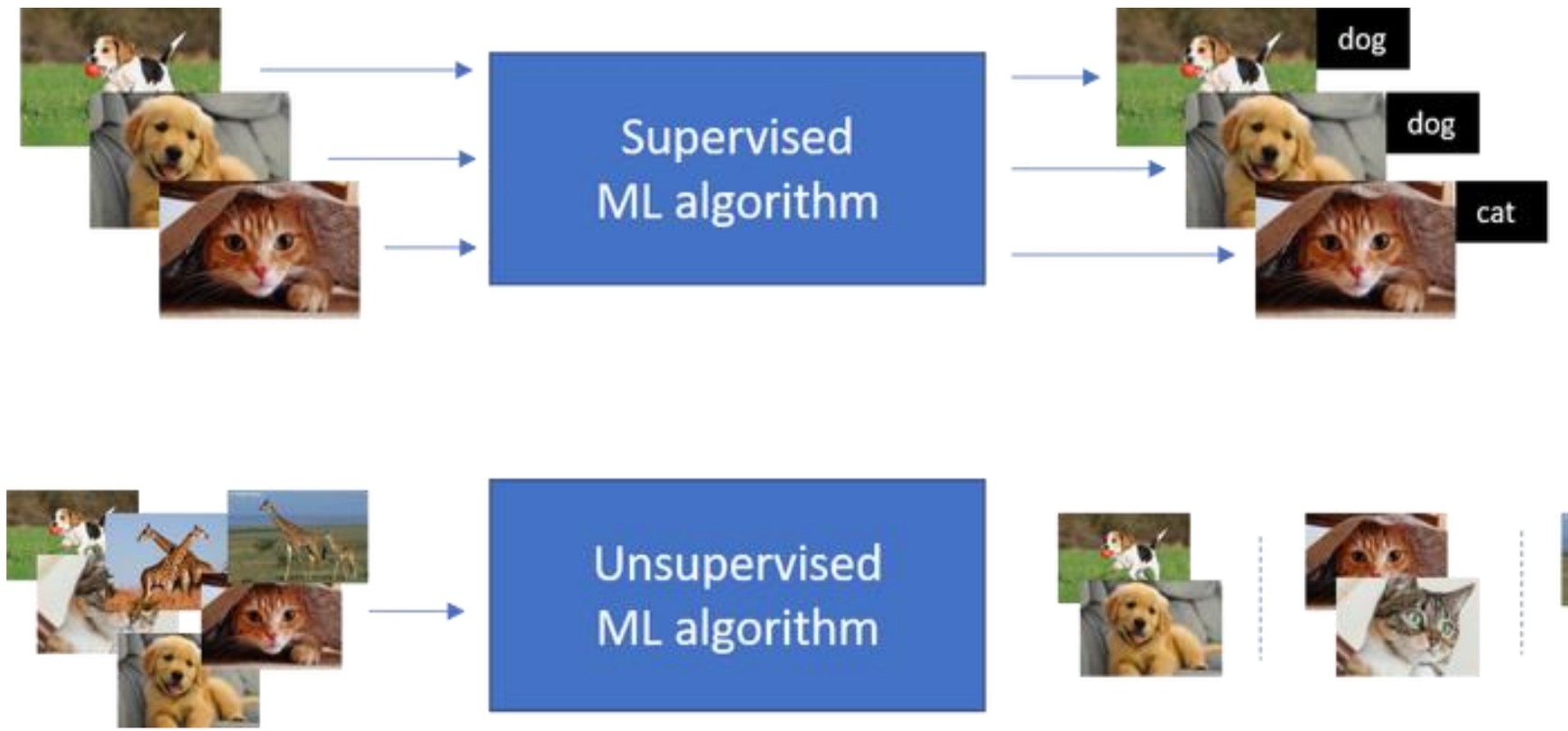
$$Errore = 100\% - Accuratezza$$

Nei problemi di Regressione, si valuta in genere l'RMSE (Root Mean Squared Error) ovvero la radice della media dei quadrati degli scostamenti tra valore vero e valore predetto.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1..N}(pred_i - true_i)^2}$$

ALGORITMI DI MACHINE LEARNING

ESTREMA SINTESI



ALGORITMI DI MACHINE LEARNING

ESTREMA SINTESI

ALGORITMI DI MACHINE LEARNING

MEMENTO

- Non utilizzate approcci di Machine Learning per problemi sui quali non avete a disposizione sufficienti esempi per il Training e il Test.
- Collezionare esempi (ed etichettarli) può richiedere ingenti sforzi, a meno che non siate in grado di reperire i pattern in rete, e/o non possiate pagare qualcuno per collezionarli/etichettarli al posto vostro (*es. Crowdsourcing via Amazon Mechanical Turk per ImageNet*).
- Collezionate pattern rappresentativi del problema da risolvere e distribuiteli adeguatamente tra Train, Valid e Test.

Tool per il Machine Learning

Nel corso degli anni ricercatori, sviluppatori indipendenti e imprese hanno sviluppato numerosi tool software (librerie, framework, simulatori), gran parte dei quali open-source.

La scelta del tool (e relativo linguaggio di programmazione) dipende dagli obiettivi del progetto e dalla preferenze dello sviluppatore.

Tra i tool più noti, ricordiamo:

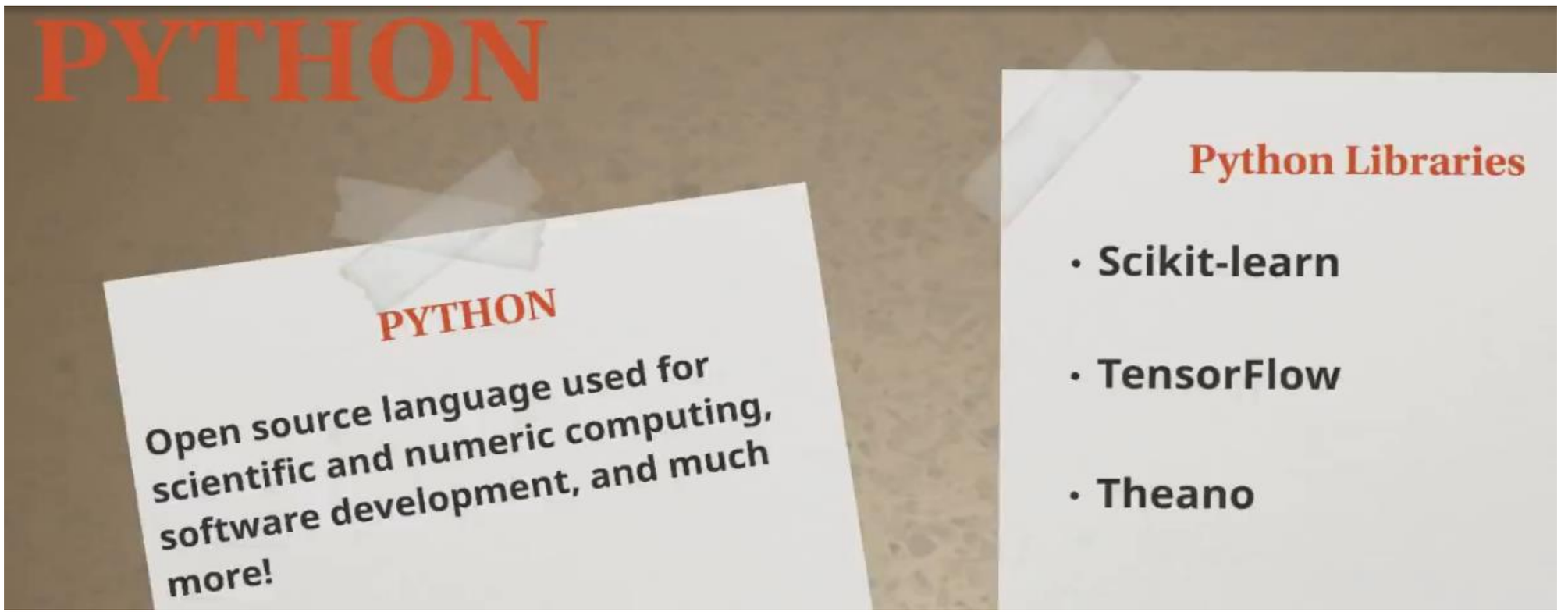
- Scikit-learn* (Python). *General Purpose*
- OpenCV (C++). *Molto utilizzato in ambito Visione Artificiale*
- Weka (Java). *Molto utilizzato in ambito Data Mining*
- R and Caret. *Dalla Statistica al Machine Learning*

Per un elenco più dettagliato:

https://github.com/josephmisiti/awesome-machine-learning.

SOFTWARE PER MACHINE LEARNING

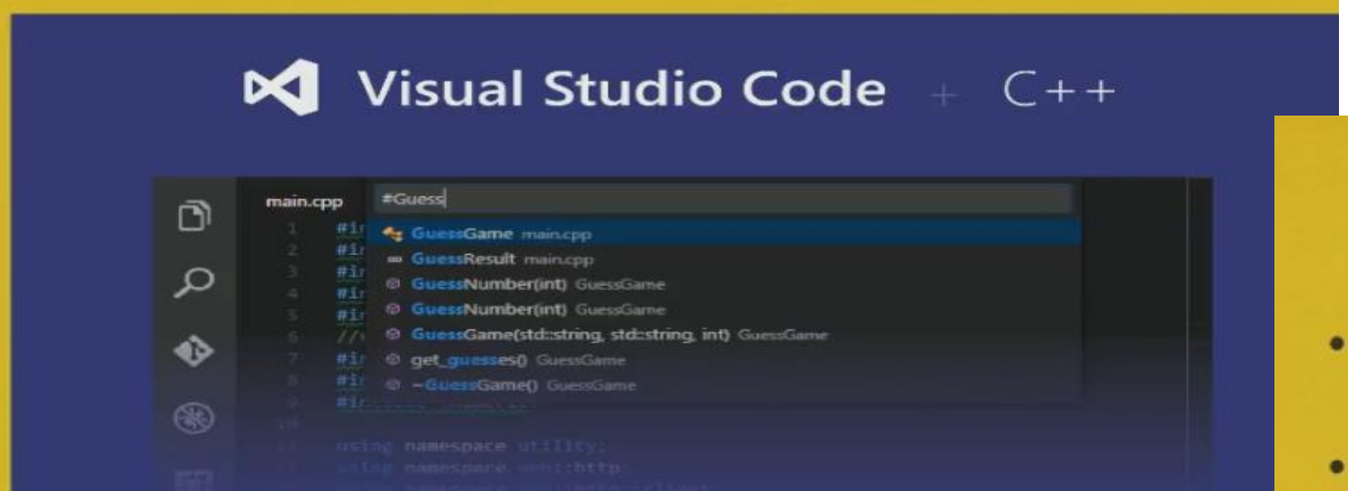
SOFTWARE PER MACHINE LEARNING



SOFTWARE PER MACHINE LEARNING

C++

General-purpose programming language with imperative, object-oriented, and generic programming features.



Common Libraries

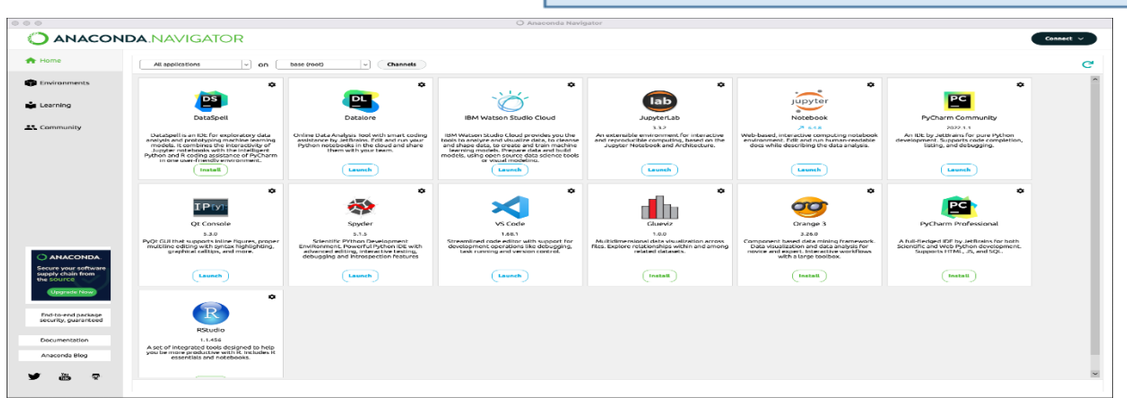
- Mlpack
- Shark
- Shogun

ALGORITMO DI MACHINE LEARNING

APPLICAZIONE PRATICA

Il sistema di ML utilizza quale *DATA set* quanto contenuto in un file, prodotto da sistemi di Cyber Defence,
posti a perimetro di una rete di un sistema aziendale e successivamente rielaborato da personale umano.
Il file di dati reali, denominato file A, è stato prodotto in formato .csv e contiene circa 14.000 righe, ciascuna descrivente un malware.
Utilizzeremo un CLASSIFICATORE SUPERVISIONATO
Ciascun *malware* è descritto da 79 features ripartite in (dettaglio nella grafica a colori):

- Indicazione delle 17 librerie che il malware utilizza per interagire con il Sistema Operativo;
- Indicazione di 37 API che vengono chiamate dal malware
- Estensione del file in cui è inserito il malware: sono considerate 11 diverse estensioni;
- Nome del malware: ne sono raccolti 14 diverse tipologie.



python™

ALGORITMO DI MACHINE LEARNING

APPLICAZIONE PRATICA

- Indicazione delle 17 librerie che il malware utilizza per interagire con il Sistema Operativo;
- Indicazione di 37 API che vengono chiamate dal malware
- Estensione del file in cui è inserito il malware: sono considerate 11 diverse estensioni;
- Nome del malware: ne sono raccolti 14 diverse tipologie

```
Keys of malware:
Index([ appexception', 'apicall', 'codeinjection', 'dll-loaded', 'doc_summary',
       'exploitcode', 'file', 'folder', 'heapspraying', 'hiddenproc',
       'malicious-alert', 'mutex', 'network', 'process', 'regkey', 'thread',
       'wmiquery', 'CheckRemoteDebuggerPresent', 'ClipboardFormatListener',
       'ClipboardSequenceNumber', 'CLSIDFromString', 'CryptAcquireContext',
       'EncryptMessage', 'EnumWindows', 'ExitProcess', 'FindWindowEx',
       'FindWindow', 'GetClipboardData', 'GetComputerName',
       'GetComputerNameEx', 'GetDesktopWindow', 'GetForegroundWindow',
       'GetLocalTime', 'GetSystemDefaultLangID', 'GetSystemDirectory',
       'GetSystemTime', 'GetTokenInformation', 'GetVersionEx',
       'GetVolumeNameForVolumeMountPoint', 'IcmpSendEcho', 'IsDebuggerPresent',
       'MessageBox', 'NtAdjustPrivilegesToken', 'RegisterRawInputDevices',
       'SetClipboardData', 'SetClipboardViewer', 'SetProcessDEPPolicy',
       'SetTimer', 'SetWindowsHookEx', 'ShellExecute', 'Sleep', 'SleepEx',
       'StartService', 'SystemTimeToFileTime', 'xls', 'doc', 'exe', 'xlsx',
       'zip', 'xlsm', 'docx', 'xlsb', 'pif', 'scr', 'altra_estensione',
       'agent_tesla', 'carrotbat', 'icedid', 'asyncrat', 'ave_maria', 'emotet',
       'formbook', 'gozi', 'lokibot', 'nanocore', 'netsky', 'remcos', 'qakbot',
       'autoit'],
      dtype='object')
(14257, 79)
```

ALGORITMO DI MACHINE LEARNING

APPLICAZIONE PRATICA

In [2]:
```
from sklearn import datasets
from sklearn.svm import SVC
from sklearn.multiclass import OneVsOneClassifier
```

In [3]:
```
#IMPORTAZIONE FILE CSV ESTERNO, NON COMPRESO IN SKLEARN
data_malware = pd.read_csv("malspam_def14k_new.csv")
```

In [4]:
```
#COMANDO PER IMPORTARE DA SKLEARN LA FUNZIONE PER SPLITTARE IL CAMPIONE
from sklearn.model_selection import train_test_split
```

In [5]:
```
#CREO I SET DI TEST E TRAIN AL 80%
train,test=train_test_split(data_malware,test_size=0.2)
```

In[2]: importo il modello del classificatore;

In[3]: importo il file di dati in formato .csv;

In[4]: importo la funzione per splittare i dati nelle due componenti: Training e Test.

In[5]: tramite la funzione importata al punto In[4] creo i due set di dati, Training e Test, secondo ripartizione che prevede 80% dei dati da utilizzare per il training e il restante 20% per i test

ALGORITMO DI MACHINE LEARNING

APPLICAZIONE PRATICA

In[12]: in questa parte viene effettuata la fase di addestramento del classificatore utilizzando i dati *xtrain* e *ytrain* ottenuti dalla funzione di *splitting* precedentemente descritta;

Out[12]: rappresenta *l'output* del comando precedente ossia il classificatore addestrato;

```
In [8]: xtrain, xtest, ytrain, ytest = train_test_split(x,y, test_size = 0.15)
```

```
In [9]: svc = SVC()
```

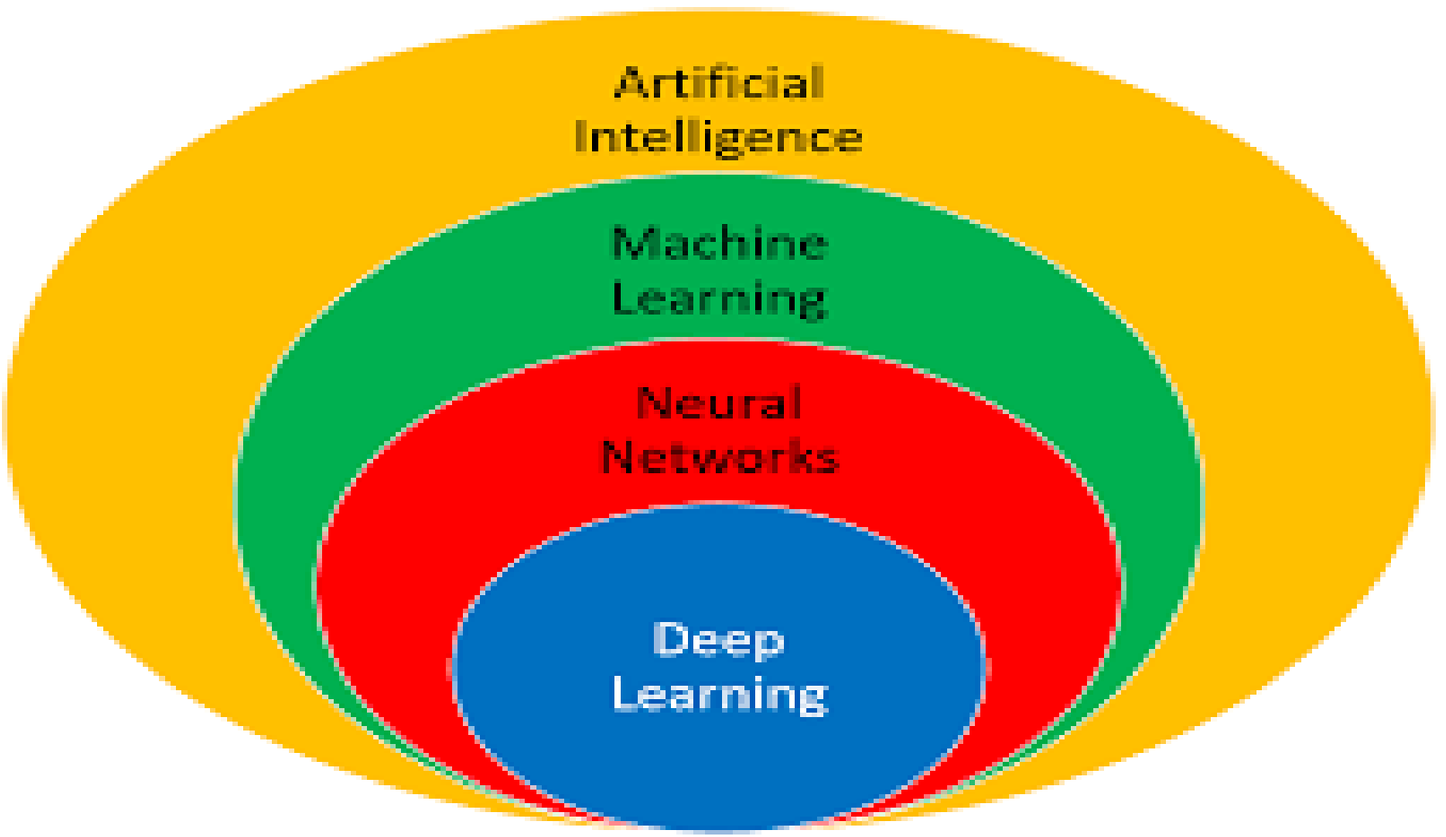
```
In [10]: o_vs_o = OneVsOneClassifier(svc)
```

```
In [12]: o_vs_o.fit(xtrain, ytrain)

Out[12]: OneVsOneClassifier(estimator=SVC())
```

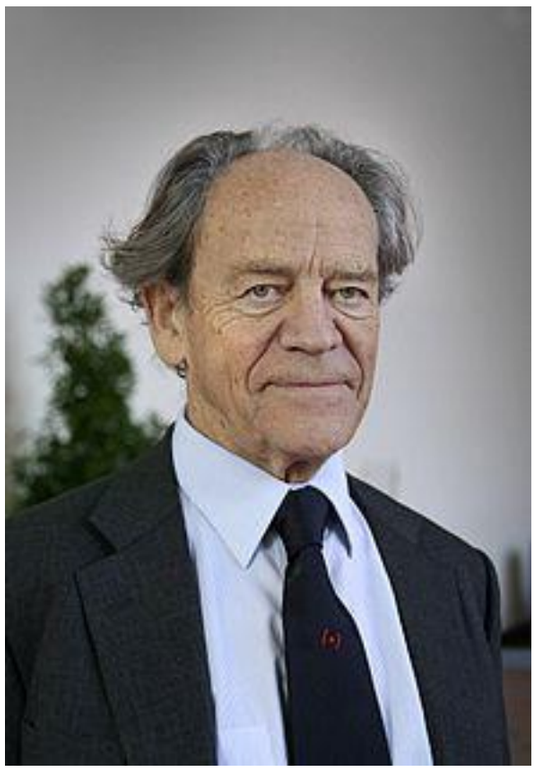
```
In [13]: ypred = o_vs_o.predict(xtest)
```

```
In [15]: # Utilizzo il modulo metrics per il calcolo, per poi stampare il risultato
         print("ACCURACY OF THE MODEL: ", metrics.accuracy_score(ytest, ypred))

ACCURACY OF THE MODEL:  0.9859747545582047
```

DEEP LEARNING

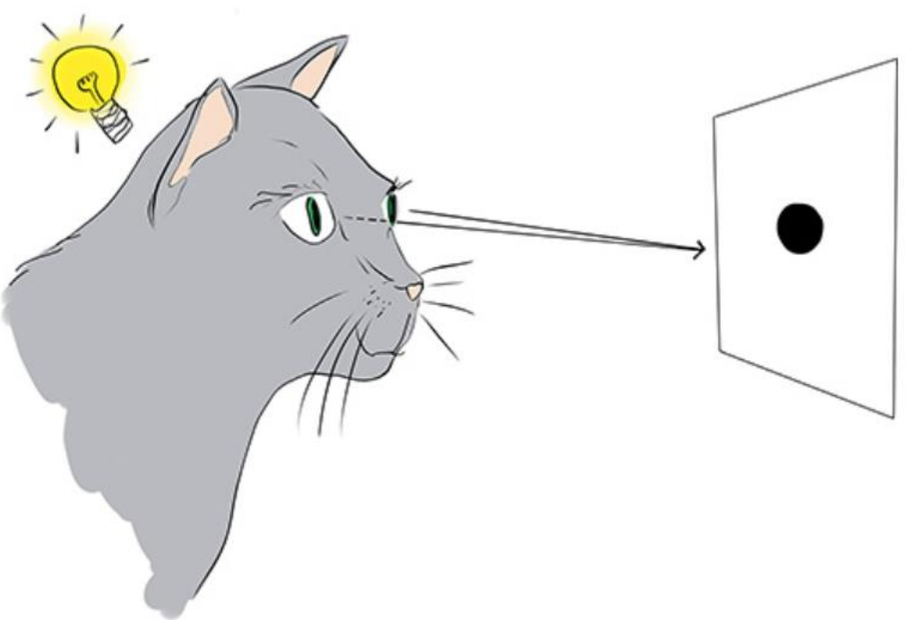


DEEP LEARNING

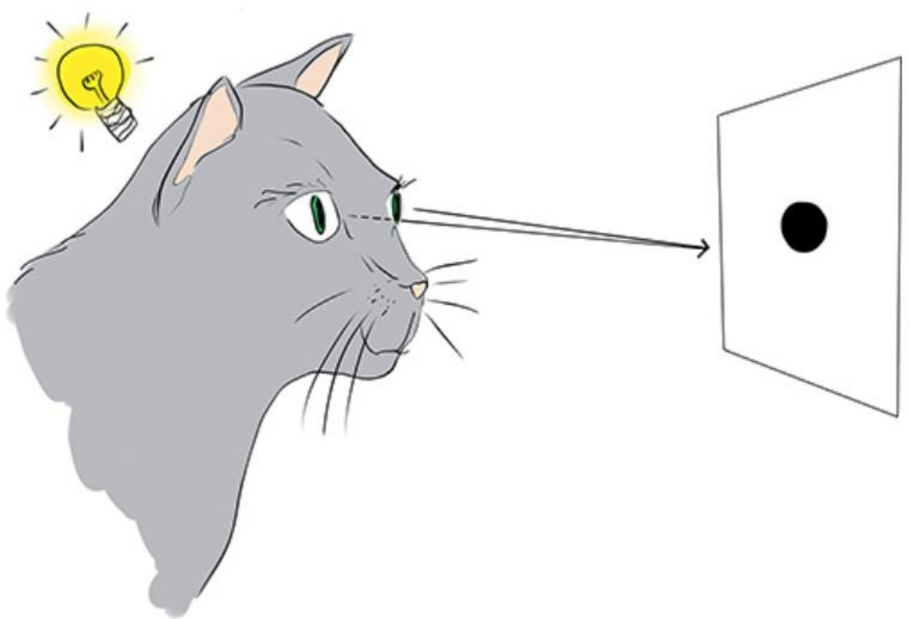


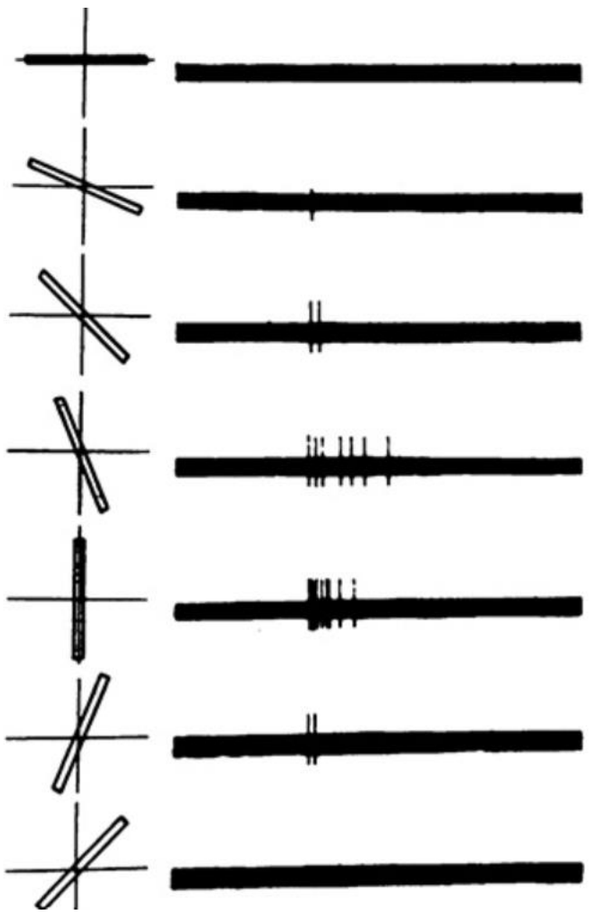
Torsten Wiesel

David Hubel

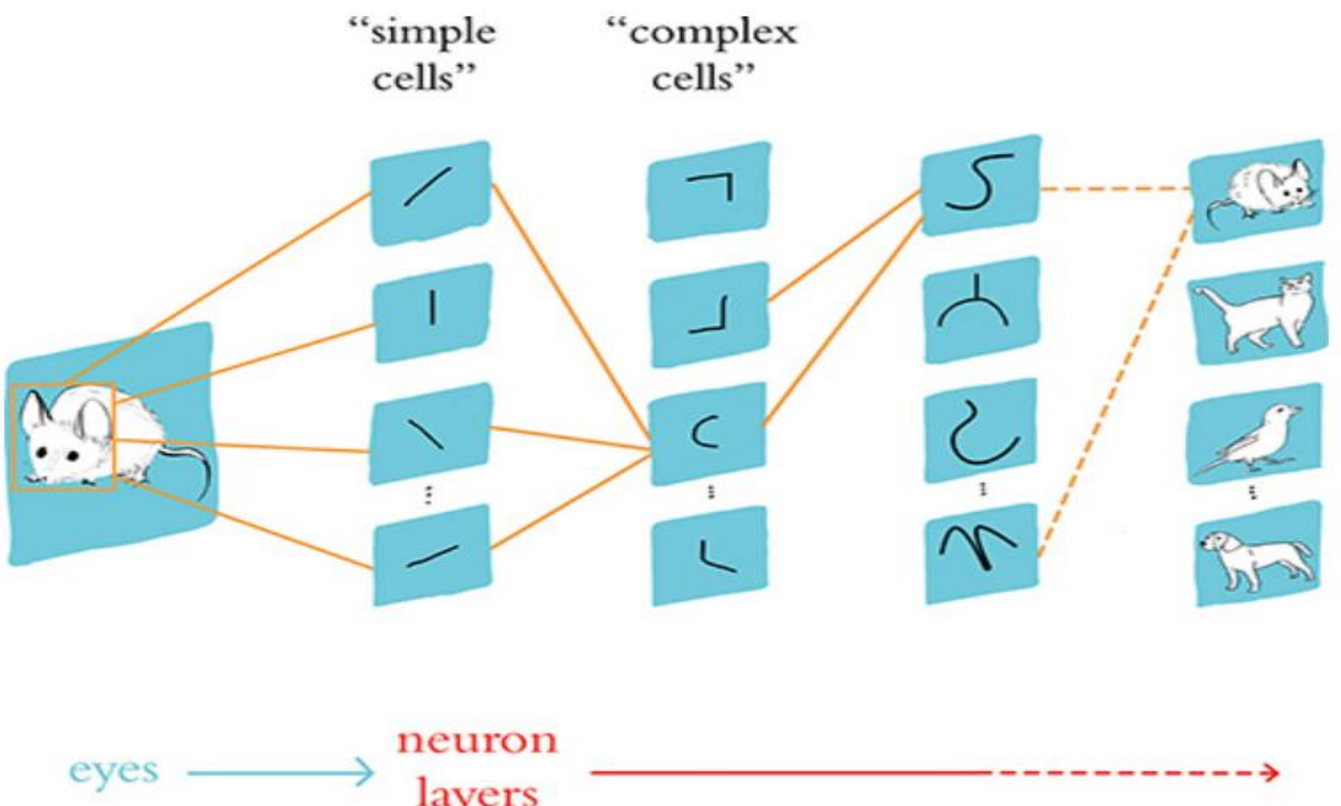


DEEP LEARNING





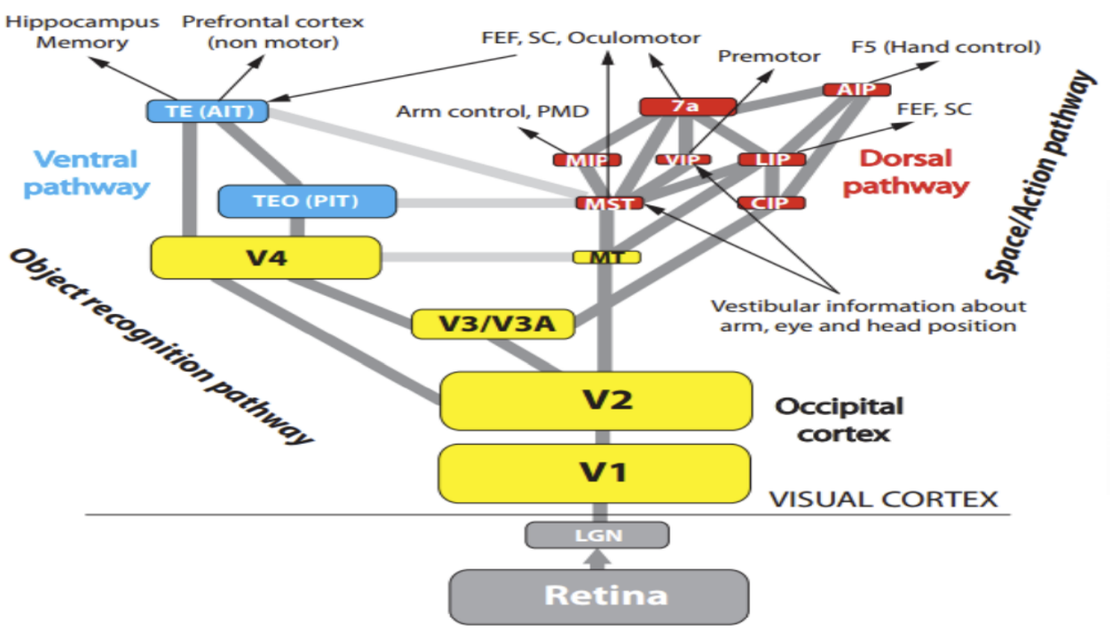
DEEP LEARNING



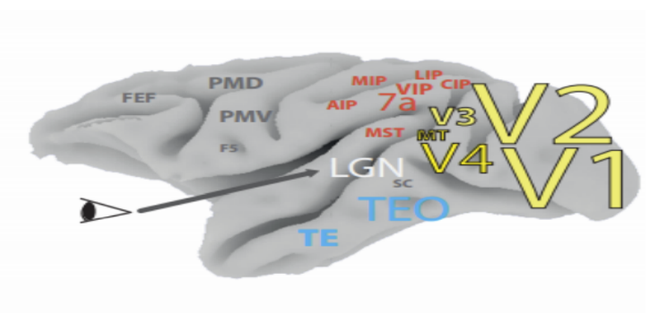
Perchè deep?

Con il termine DNN (Deep Neural Network) si denotano reti «profonde» composte da molti livelli (almeno 2 hidden) organizzati gerarchicamente.

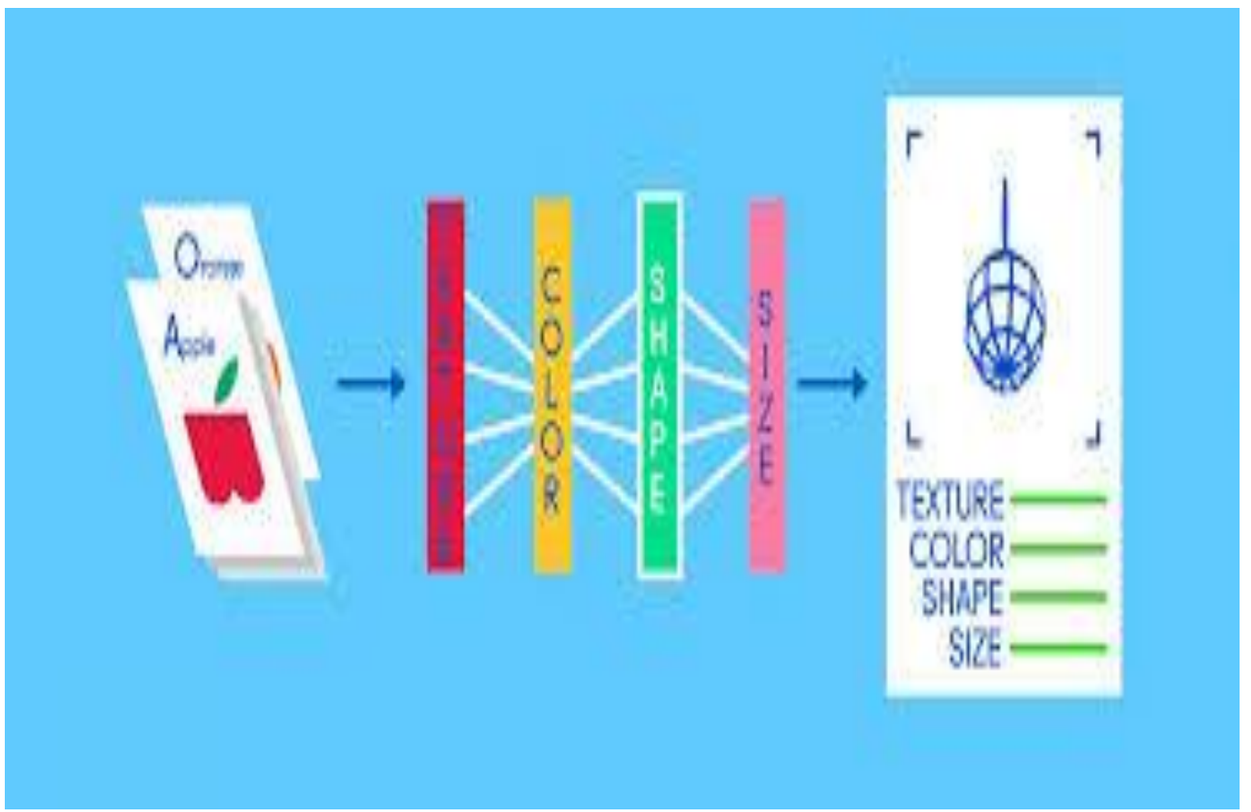
- Il nostro sistema visivo opera su una gerarchia di livelli (deep):



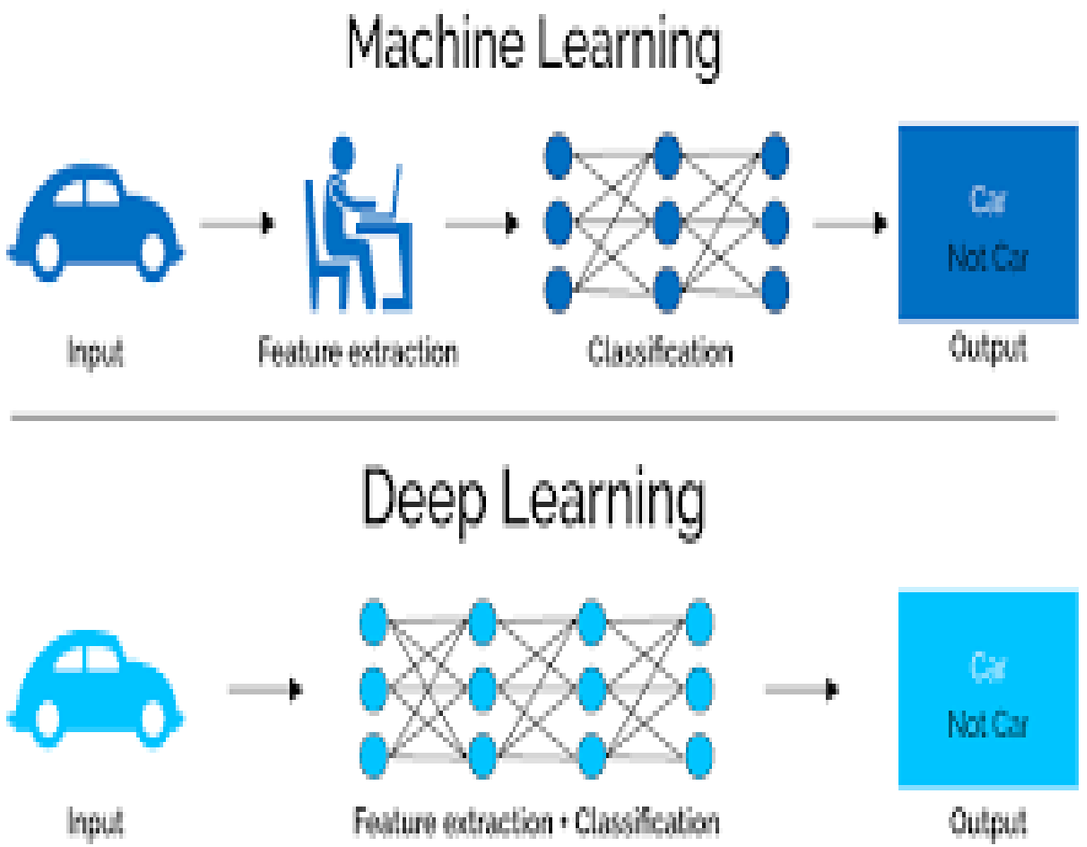
[Kruger et al. 2013]

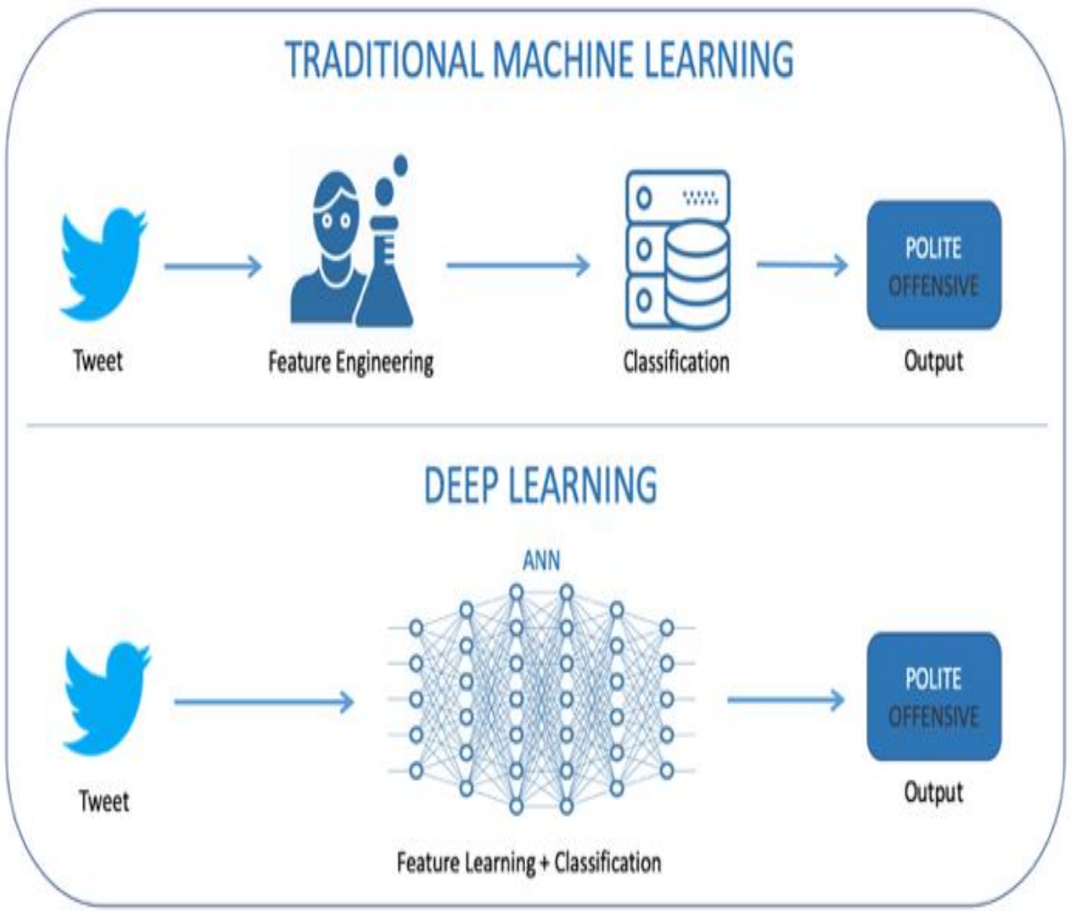


- L'organizzazione gerarchica consente di condividere e riusare informazioni (un po' come la programmazione strutturata). Lungo la gerarchia è possibile selezionare feature specifiche e scartare dettagli inutili (al fine di massimizzare l'invarianza).

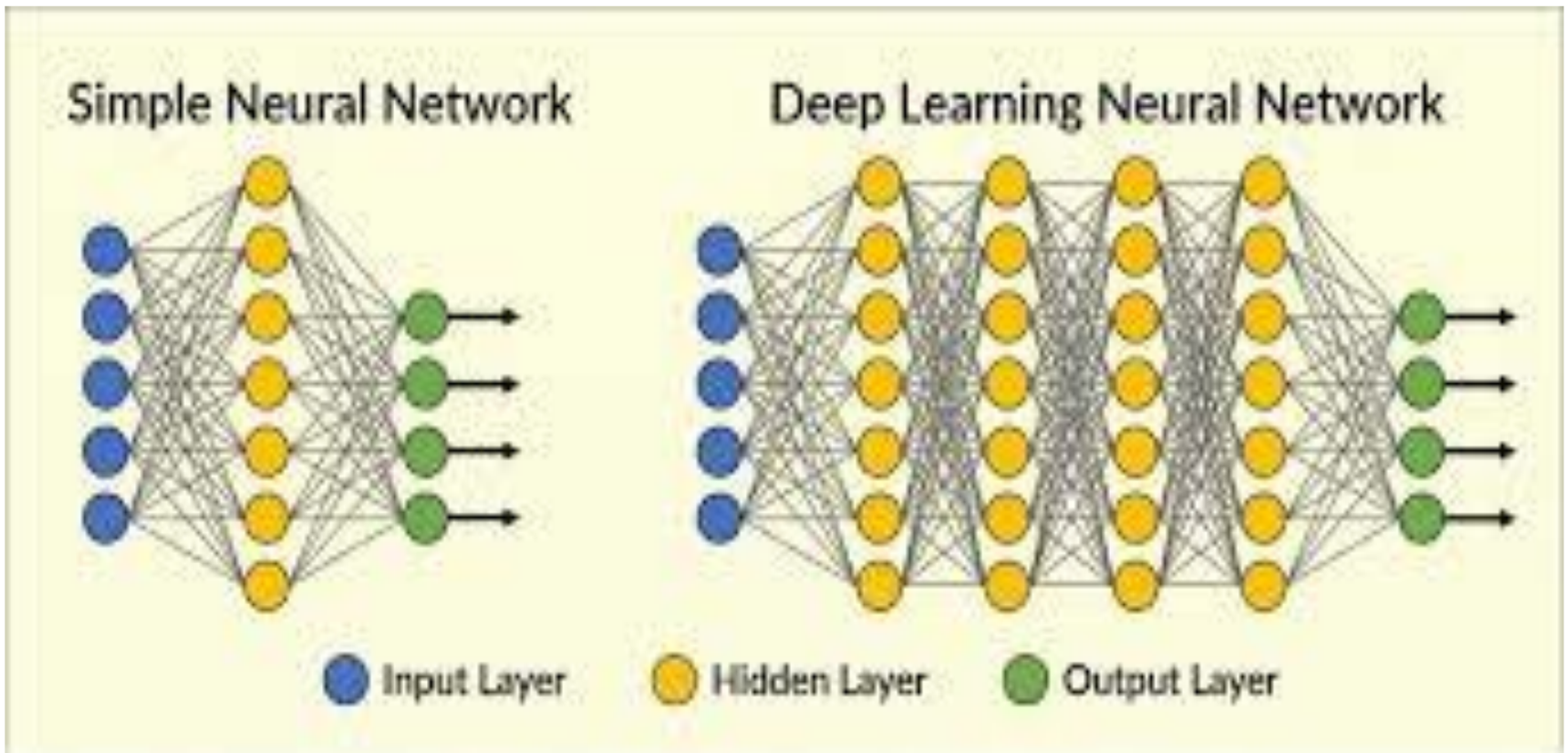


MACHINE LEARNING vs DEEP LEARNING





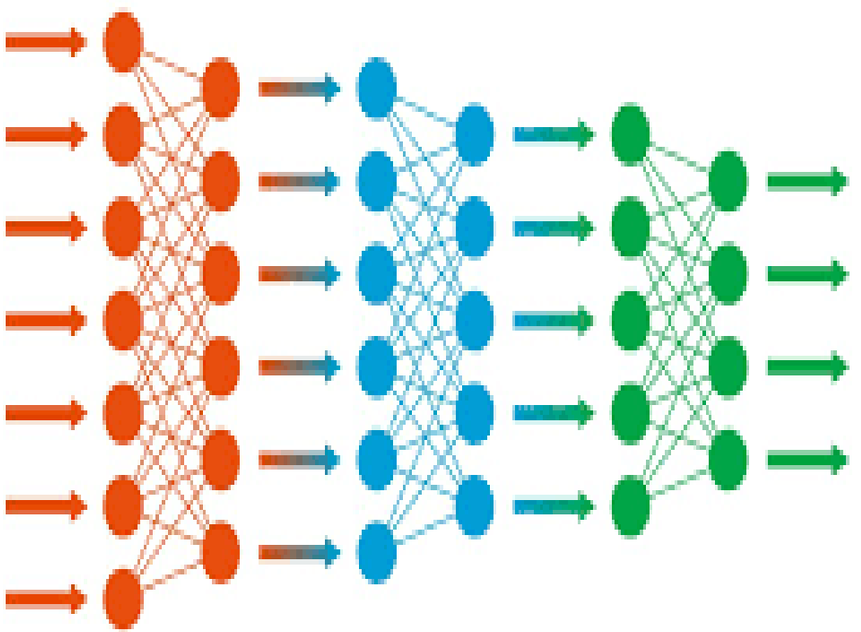
DEEP LEARNING



DEEP LEARNING

ma quanto deep?

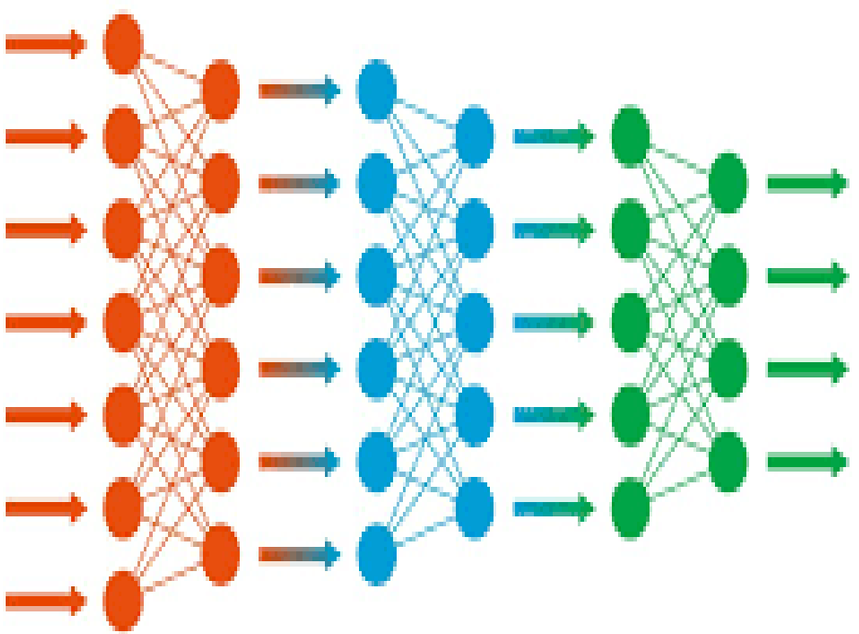
- Le DNN oggi maggiormente utilizzate consistono di un numero di livelli compreso tra 7 e 50.
  - Reti più profonde (100 livelli e oltre) hanno dimostrato di poter garantire prestazioni leggermente migliori, a discapito però dell'efficienza.



DEEP LEARNING

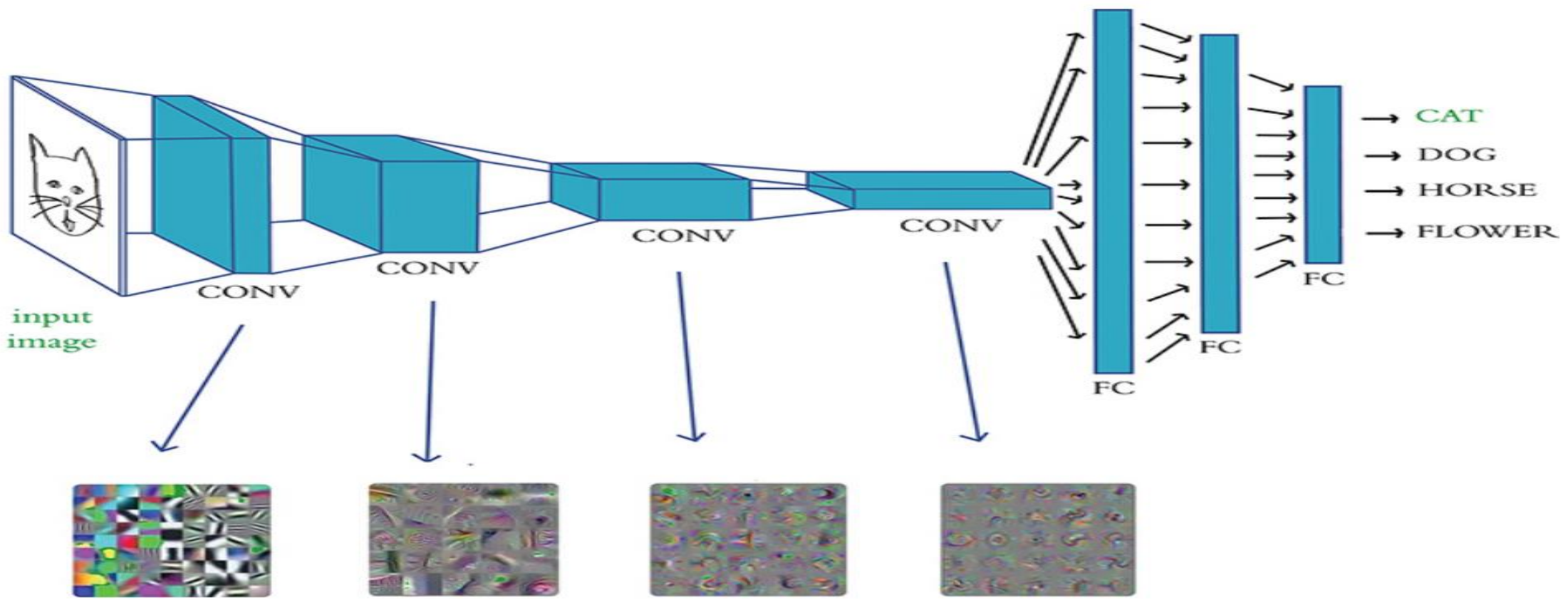
Livelli e Complessità

- La profondità (numero di livelli) è solo uno dei fattori di complessità. Numero di neuroni, di connessioni e di pesi caratterizzano altresì la complessità di una DNN.
- Maggiore è il numero di pesi (ovvero di parametri da apprendere) maggiore è la complessità del training. Al tempo stesso un elevato numero di neuroni (e connessioni) rende forward e back propagation più costosi, poiché aumenta il numero (G-Ops) di operazioni necessarie.
  - AlexNet: 8 livelli, 650K neuroni e 60M parametri
  - VGG-16: 16 livelli, 15M neuroni e 140M parametri
  - Corteccia umana: 10^{11} neuroni e 10^{14} sinapsi

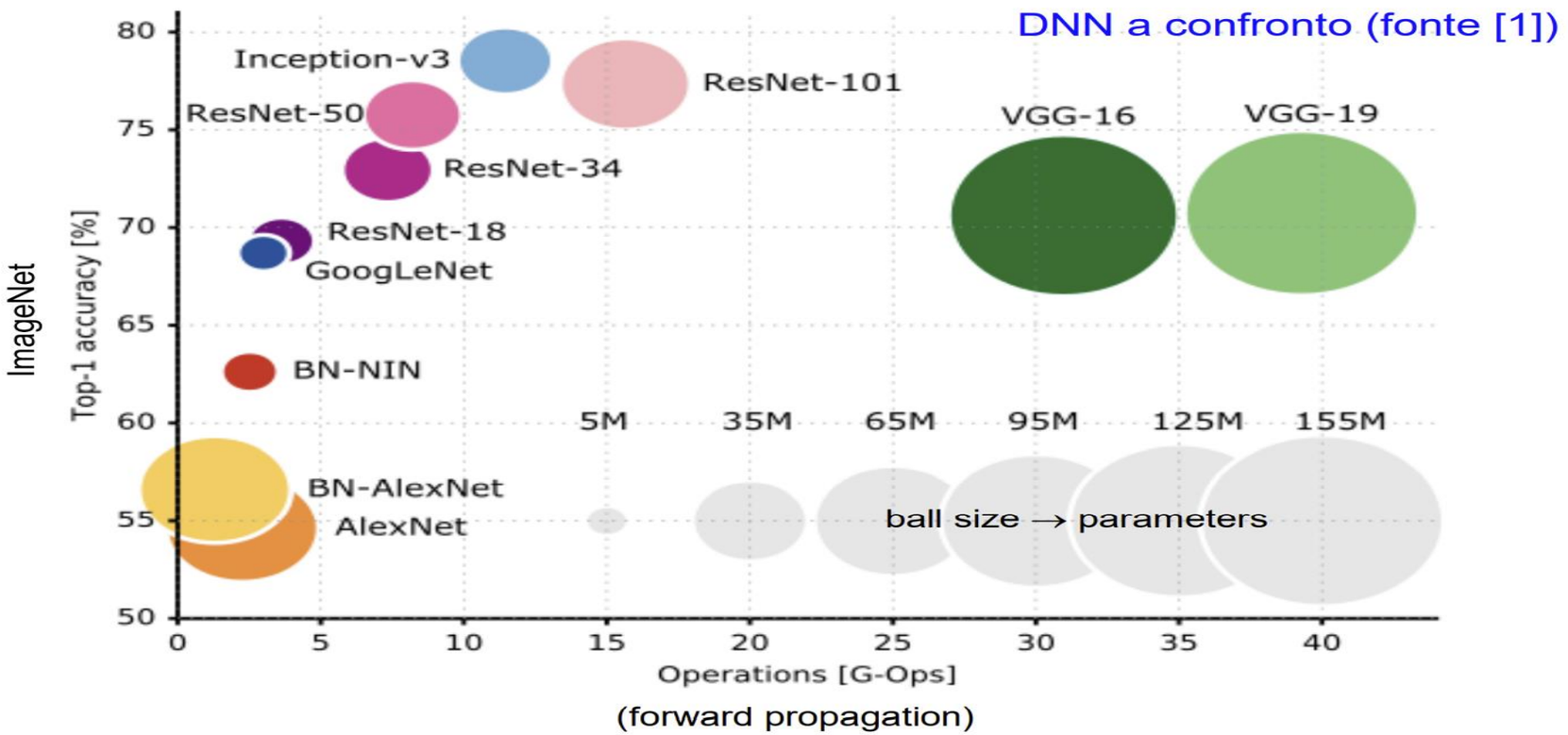


DEEP LEARNING

ALEXA NET



DEEP LEARNING



[1] Canziani et al. 2016, *An Analysis of Deep Neural Network Models for Practical Applications*

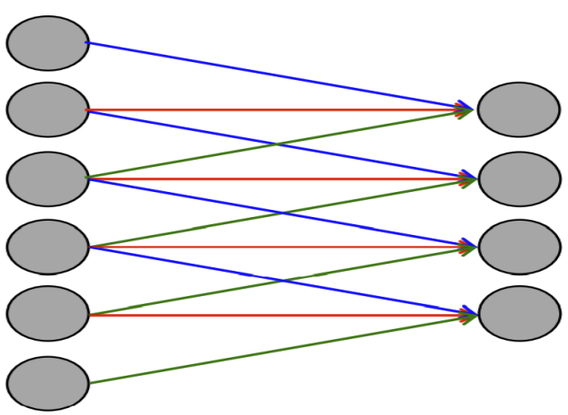
Principali tipologie di DNN

- Modelli feedforward «discriminativi» per la classificazione (o regressione) con training prevalentemente supervisionato:
  - CNN - Convolutional Neural Network (o ConvNet)
  - FC DNN - Fully Connected DNN (MLP con almeno due livelli hidden)
  - HTM - Hierarchical Temporal Memory
- Training non superivisionato (modelli «generativi» addestrati a ricostruire l'input, utili per pre-training di altri modelli e per produrre feature salienti):
  - Stacked (de-noising) Auto-Encoders
  - RBM - Restricted Boltzmann Machine
  - DBN - Deep Belief Networks

DA MULTILAYER PERCEPTRON (MLP) A CONVOLUTIONAL NEURAL NETWORK

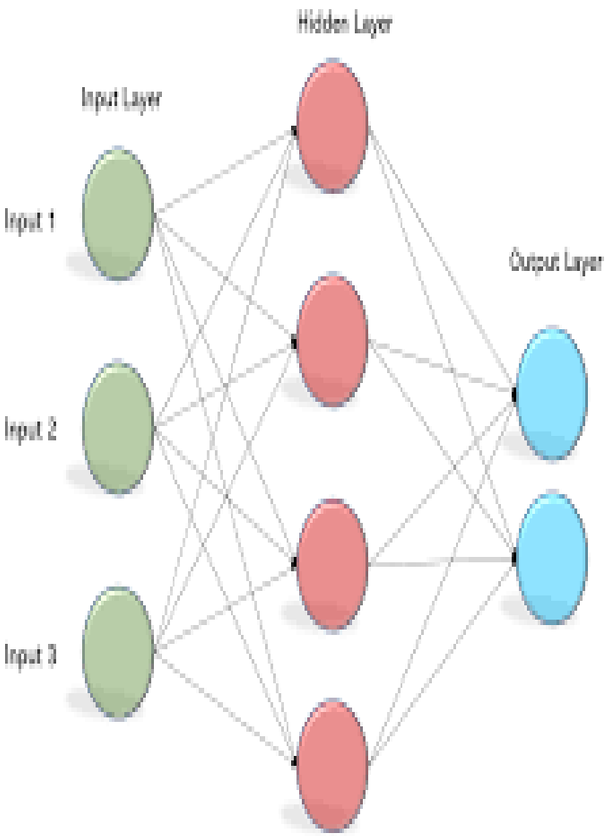
Convolutional Neural Networks (CNN) introdotte da LeCun et al., a partire dal 1998. Le principali differenze rispetto a MLP:

- processing locale: i neuroni sono connessi solo localmente ai neuroni del livello precedente. Ogni neurone esegue quindi un'elaborazione locale. Forte riduzione numero di connessioni.
- pesi condivisi: i pesi sono condivisi a gruppi. Neuroni diversi dello stesso livello eseguono lo stesso tipo di elaborazione su porzioni diverse dell'input. Forte riduzione numero di pesi.



Esempio: ciascuno dei 4 neuroni a destra è connesso solo a 3 neuroni del livello precedente. I pesi sono condivisi (stesso colore stesso peso). In totale 12 connessioni e 3 pesi contro le 24 connessioni + 24 pesi di una equivalente porzione di MLP.

MLP



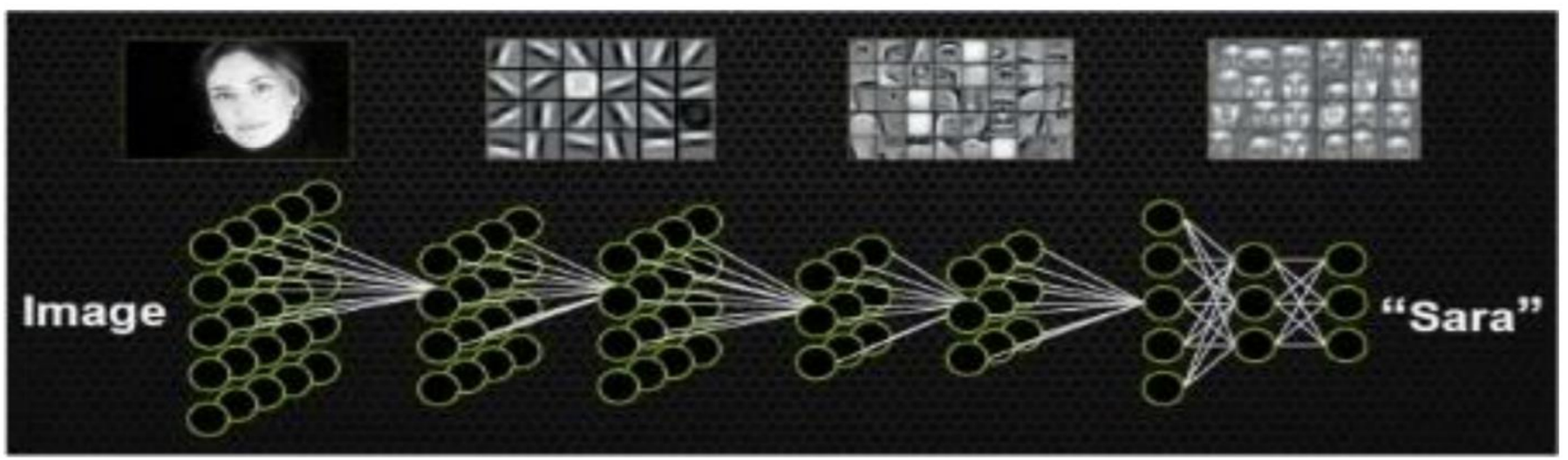
CONVOLUTIONAL NEURAL NETWORK

CNN: Architettura

Esplicitamente progettate per processare immagini, per le quali elaborazione locale, pesi condivisi, e pooling non solo semplificano il modello, ma lo rendono più efficace rispetto a modelli fully connected. Possono essere utilizzate anche per altri tipi di pattern (es. speech).

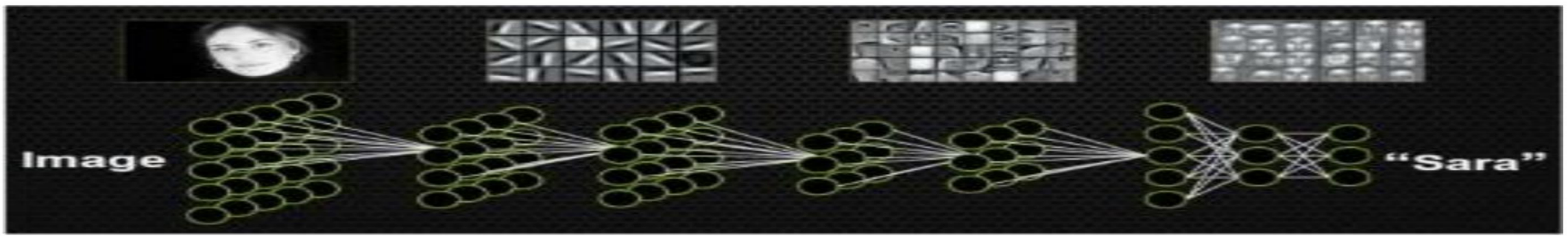
CONVOLUTIONAL NEURAL NETWORK

- Architettura: una CNN è composta da una gerarchia di livelli. Il livello di input è direttamente collegato ai pixel dell'immagine, gli ultimi livelli sono generalmente fully-connected e operano come un classificatore MLP, mentre nei livelli intermedi si utilizzano connessioni locali e pesi condivisi.



CONVOLUTIONAL NEURAL NETWORK

- Architettura: una CNN è composta da una gerarchia di livelli. Il livello di input è direttamente collegato ai pixel dell'immagine, gli ultimi livelli sono generalmente fully-connected e operano come un classificatore MLP, mentre nei livelli intermedi si utilizzano connessioni locali e pesi condivisi.



- Il campo visivo (receptive field) dei neuroni aumenta muovendosi verso l'alto nella gerarchia.
- Le connessioni locali e condivise fanno sì che i neuroni processino nello stesso modo porzioni diverse dell'immagine. Si tratta di un comportamento desiderato, in quando regioni diverse del campo visivo contengono lo stesso tipo di informazioni (bordi, spigoli, porzioni di oggetti, ecc.).
- Visualizzazione 3D di una CNN:
  http://www.cs.cmu.edu/~aharley/vis/

CONVOLUTIONAL NEURAL NETWORK

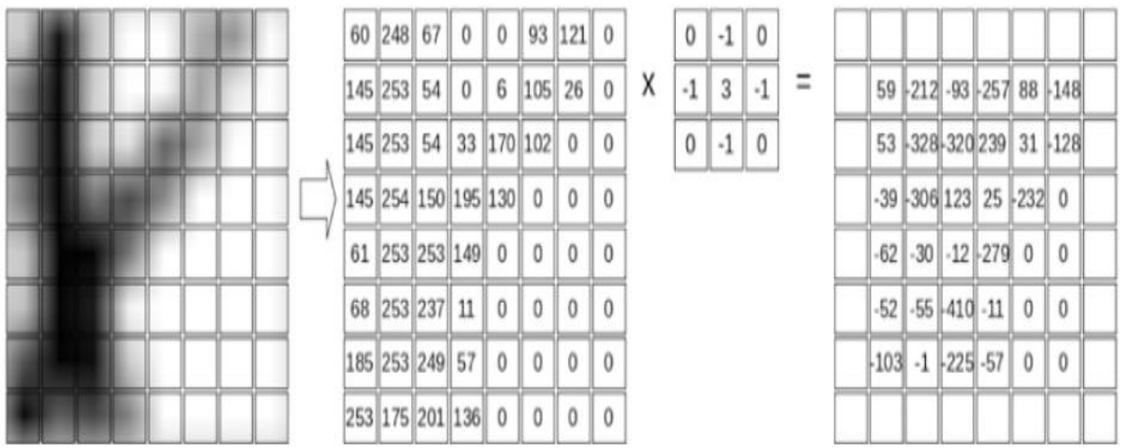


Figure 12-1: Convolving a kernel with an image

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 0 \end{bmatrix} \qquad \begin{bmatrix} 60 & 248 & 67 \\ 145 & 253 & 54 \\ 145 & 253 & 54 \end{bmatrix}$$

$$\begin{bmatrix} 60 & 248 & 67 \\ 145 & 253 & 54 \\ 145 & 253 & 54 \end{bmatrix} \times \begin{bmatrix} 0 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -248 & 0 \\ -145 & 759 & -54 \\ 0 & -253 & 0 \end{bmatrix}$$

0 + (–248) + 0 + (–145) + 759 + (–54) + 0 + (–253) + 0 = 59

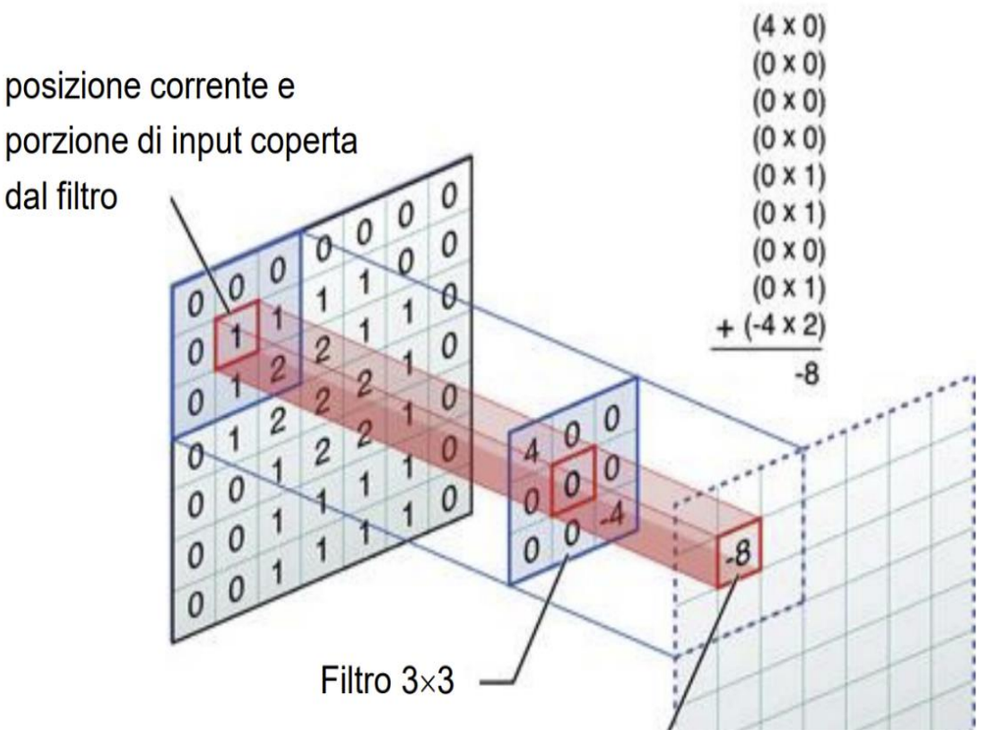
$$\begin{bmatrix} 248 & 67 & 0 \\ 253 & 54 & 0 \\ 253 & 54 & 33 \end{bmatrix} \times \begin{bmatrix} 0 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -67 & 0 \\ -253 & 162 & 0 \\ 0 & -54 & 0 \end{bmatrix}$$

212

CONVOLUTIONAL NEURAL NETWORK

- La convoluzione è una delle più importanti operazioni di image processing attraverso la quale si applicano filtri digitali.
- Un filtro digitale (un piccola maschera 2D di pesi) è fatta scorrere sulle diverse posizioni di input; per ogni posizione viene generato un valore di output, eseguendo il prodotto scalare tra la maschera e la porzione dell'input coperta (entrambi trattati come vettori).

CONVOLUTIONAL NEURAL NETWORK



: pesi del filtro in

rosso, porzione coperta in giallo

$1_{\times 1}$	$1_{\times 0}$	$1_{\times 1}$	0	0
$0_{\times 0}$	$1_{\times 1}$	$1_{\times 0}$	1	0
$0_{\times 1}$	$0_{\times 0}$	$1_{\times 1}$	1	1
0	0	1	1	0
0	1	1	0	0

4		

CONVOLUTIONAL NEURAL NETWORK

Esempio Filtri a immagine

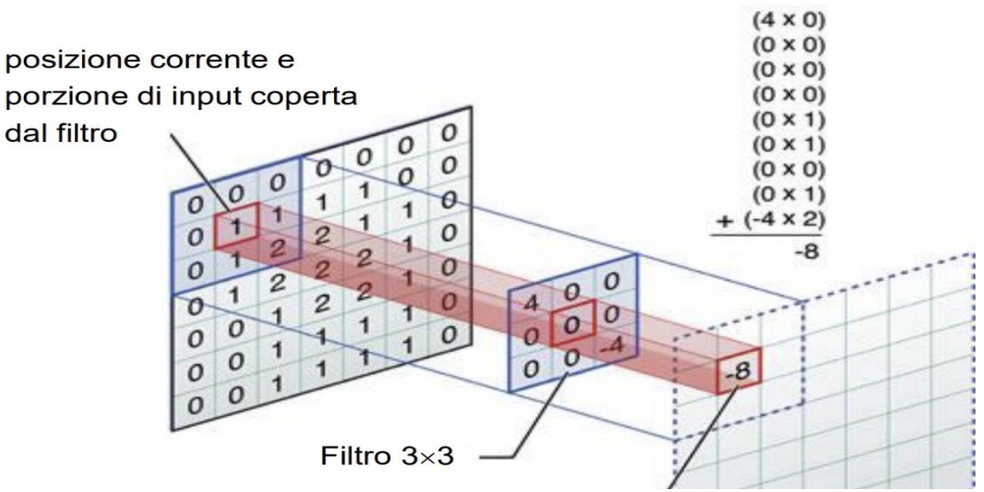


Immagine input

Filtro

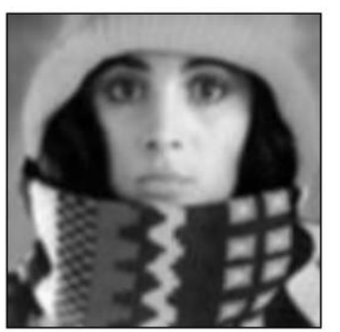
$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Immagine output

$1_{\times 1}$	$1_{\times 0}$	$1_{\times 1}$	0	0
$0_{\times 0}$	$1_{\times 1}$	$1_{\times 0}$	1	0
$0_{\times 1}$	$0_{\times 0}$	$1_{\times 1}$	1	1
0	0	1	1	0
0	1	1	0	0

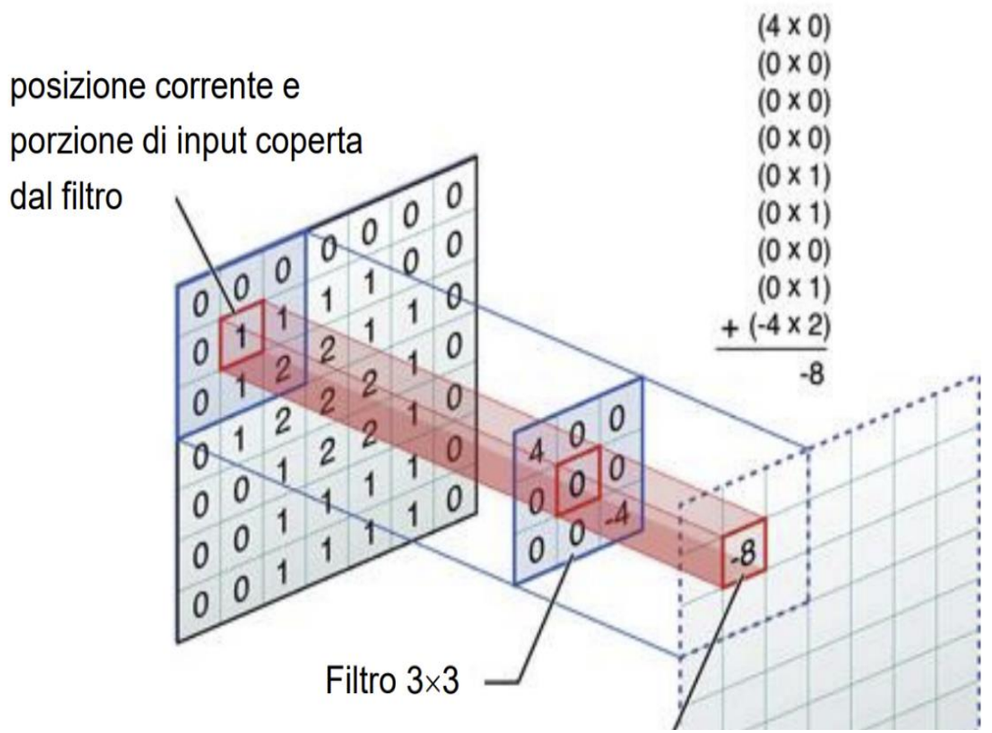
4		

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

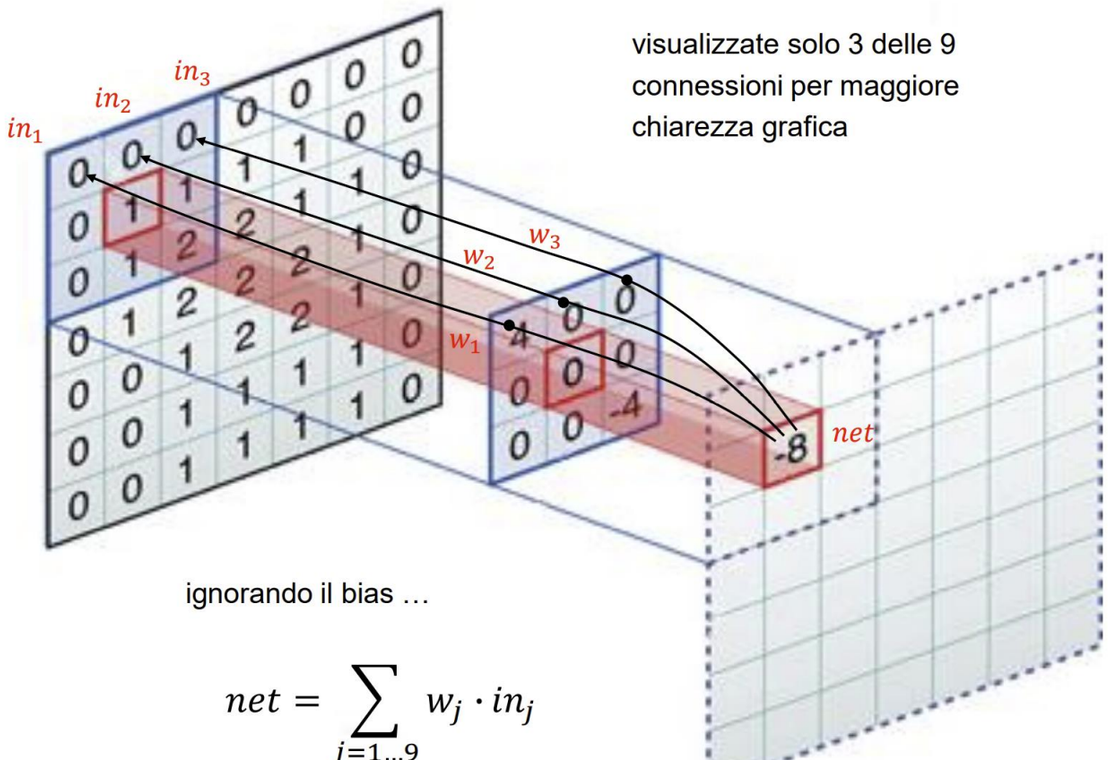


CONVOLUTIONAL NEURAL NETWORK

CONVOLUZIONE



OUT



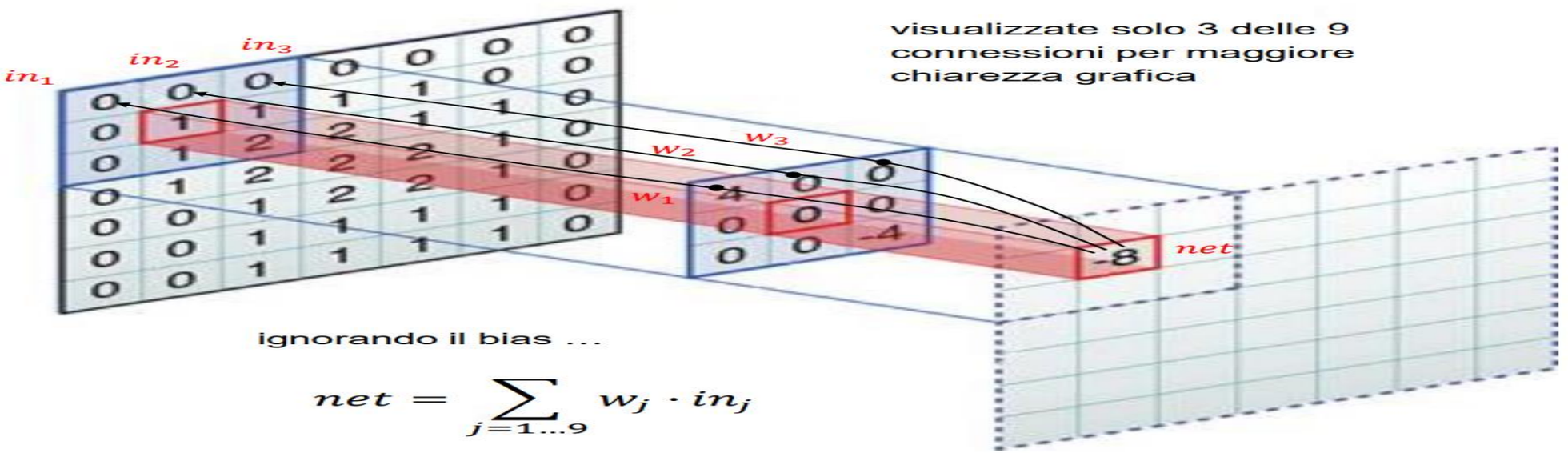
visualizzate solo 3 delle 9 connessioni per maggiore chiarezza grafica

ignorando il bias …

$$net = \sum_{j=1\ldots9} w_j \cdot in_j$$

CONVOLUTIONAL NEURAL NETWORK

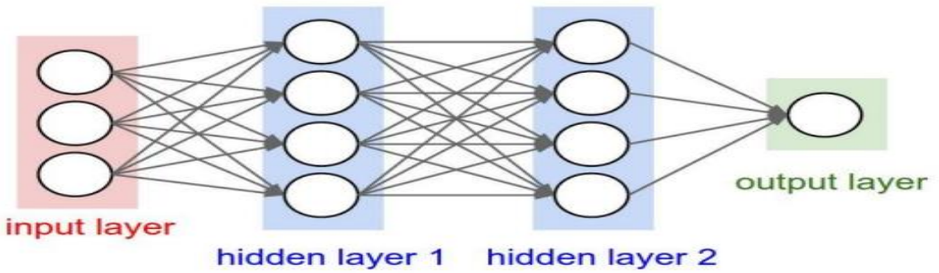
- Consideriamo i pixel come neuroni e le due immagini di input e di output come livelli successivi di una rete. Dato un filtro 3×3, se colleghiamo un neurone ai 9 neuroni che esso «copre» nel livello precedente, e utilizziamo i pesi del filtro come pesi delle connessioni w, notiamo che un classico neurone (di una MLP) esegue di fatto una convoluzione.



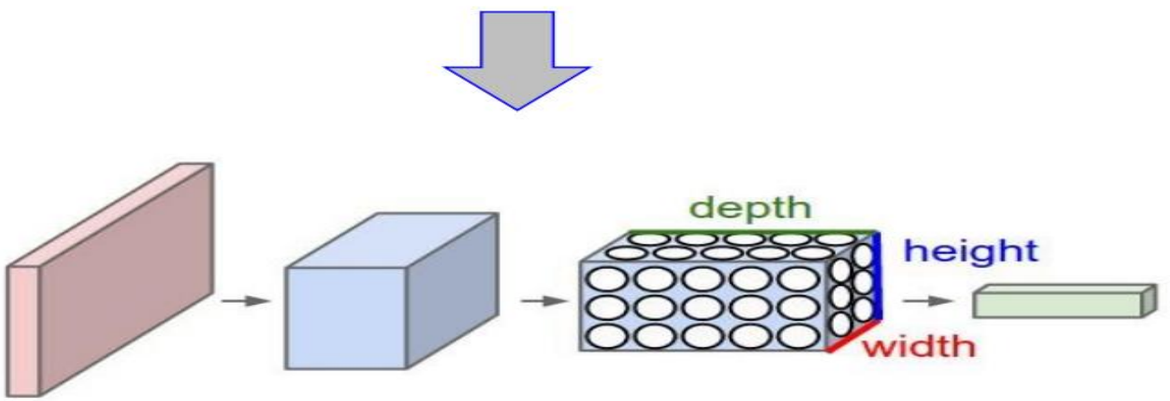
ignorando il bias ...

$$net = \sum_{j=1\ldots9} w_j \cdot in_j$$

CONVOLUTIONAL NEURAL NETWORK



MLP: organizzazione lineare dei neuroni nei livelli

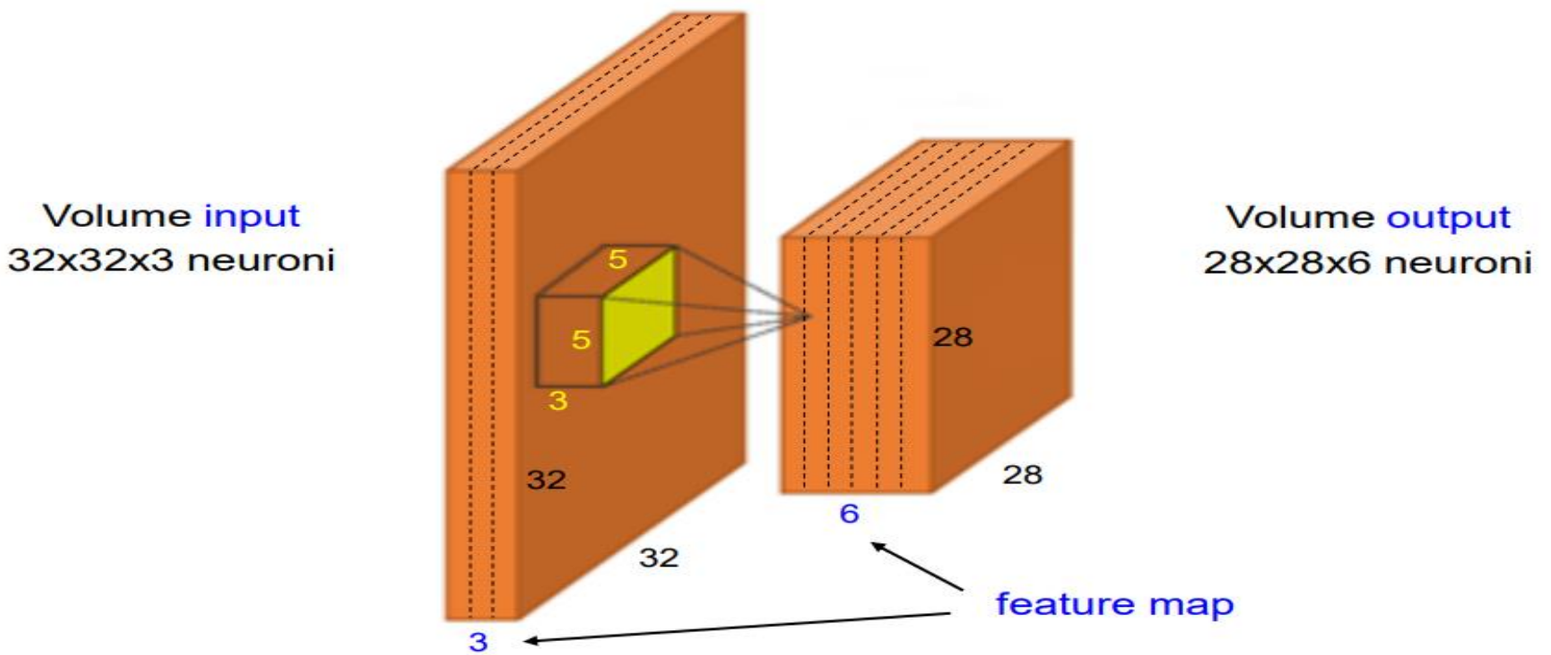


CNN: i livelli sono organizzati come griglie 3D di neuroni

sui piani width - height si conserva l'organizzazione spaziale «retinotipica» dell'immagine di input.

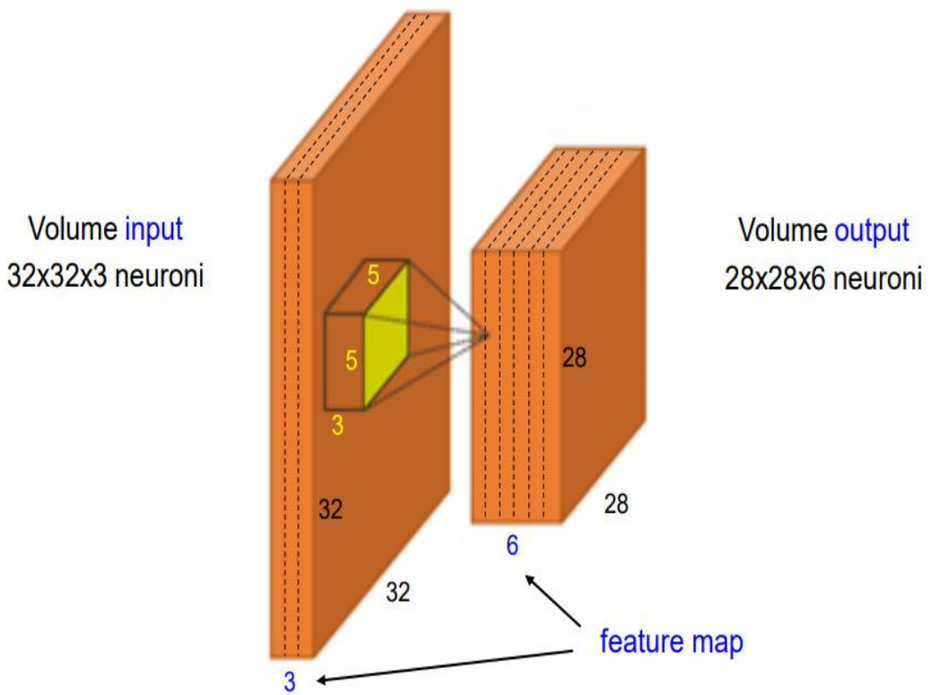
CONVOLUTIONAL NEURAL NETWORK – 3 D

- Il filtro opera su una porzione del volume di input. Nell'esempio ogni neurone del volume di output è connesso a 5x5x3=75 neuroni del livello precedente.



CONVOLUTIONAL NEURAL NETWORK – 3 D

- Il filtro opera su una porzione del volume di input. Nell'esempio ogni neurone del volume di output è connesso a 5x5x3= neuroni del livello precedente.



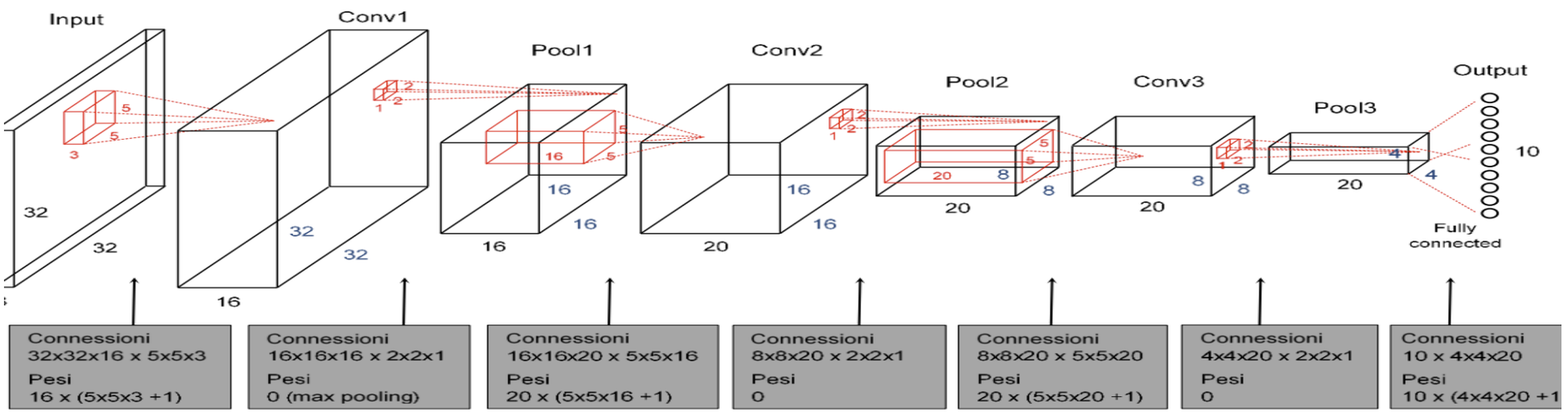
- Ciascuna «fetta» di neuroni (stessa depth) denota una feature map. Nell'esempio troviamo:
  - 3 feature map (dimensione 32x32) nel volume di input.
  - 6 feature map (dimensione 28x28) nel volume di output.
- I pesi sono condivisi a livello di feature map. I neuroni di una stessa feature map processano porzioni diverse del volume di input nello stesso modo. Ogni feature map può essere vista come il risultato di uno specifico filtraggio dell'input (filtro fisso).
- Nell'esempio il numero di connessioni tra i due livelli è (28x28x6) x (5x5x3) = 352800, ma il numero totale di pesi è 6 x (5x5x3 + 1) = 456. *In analoga porzione di MLP quanti pesi?*

CNN

Architettura

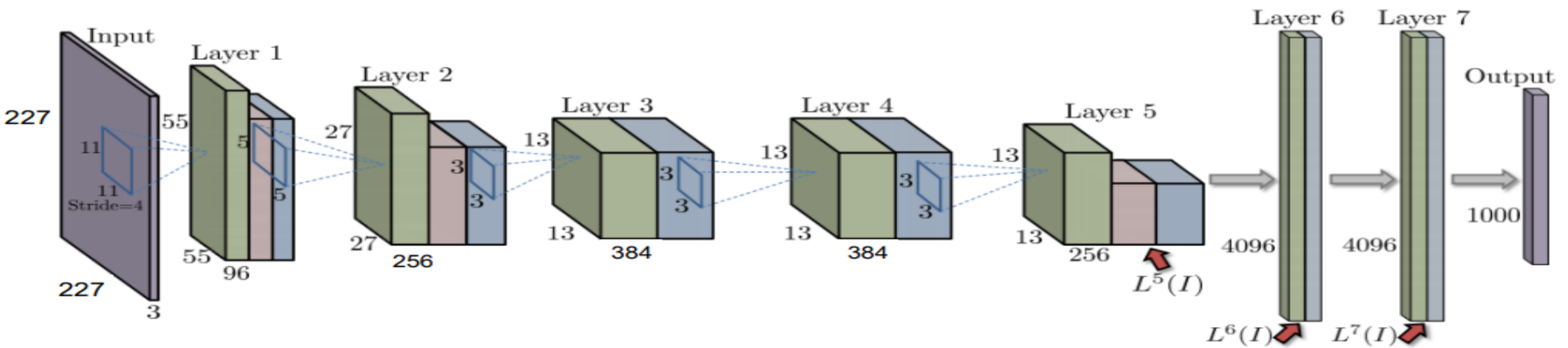
- Input: Immagini RGB 32x32x3;
- Conv1: Filtri:5x5, FeatureMaps:16, stride:1, pad:2, attivazione: Relu
- Pool1: Tipo: Max, Filtri 2x2, stride:2
- Conv2: Filtri:5x5, FeatureMaps:20, stride:1, pad:2, attivazione: Relu
- Pool2: Tipo: Max, Filtri 2x2, stride:2
- Conv3: Filtri:5x5, FeatureMaps:20, stride:1, pad:2, attivazione: Relu
- Pool3: Tipo: Max, Filtri 2x2, stride:2
- Output: Softmax, NumClassi: 10

Disegniamo la rete e calcoliamo neuroni sui livelli, connessioni e pesi



Neuroni totali: 31.562 (incluso livello input)
Connessioni totali: 3.942.784
Pesi totali: 22.466 (inclusi bias)

CNN

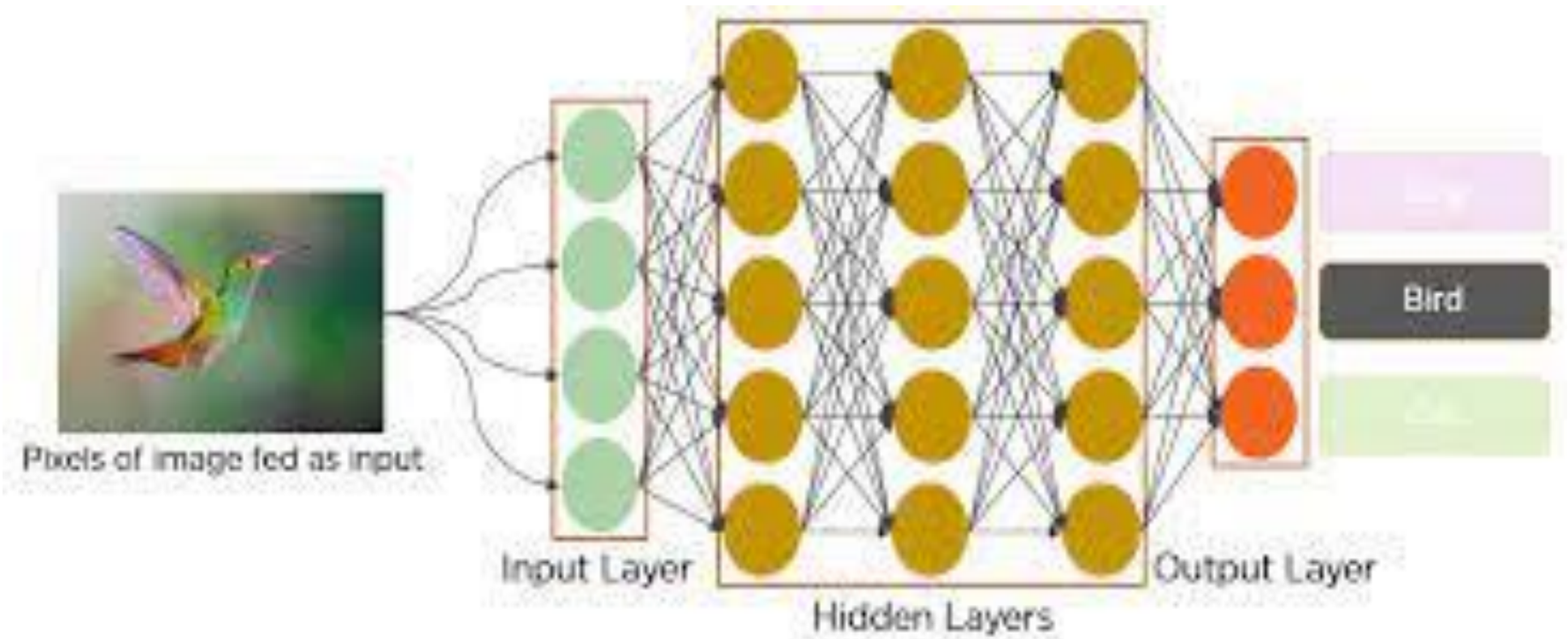


Codici colore

- Viola: Input (immagini 227x227x3) e Output (1000 classi di ImageNet)
- Verde: Convoluzione
- Rosa: Pooling (max)
- Blu: Relu activation

- Numero totale di parametri: 60M circa

CONVOLUTIONAL NEURAL NETWORK



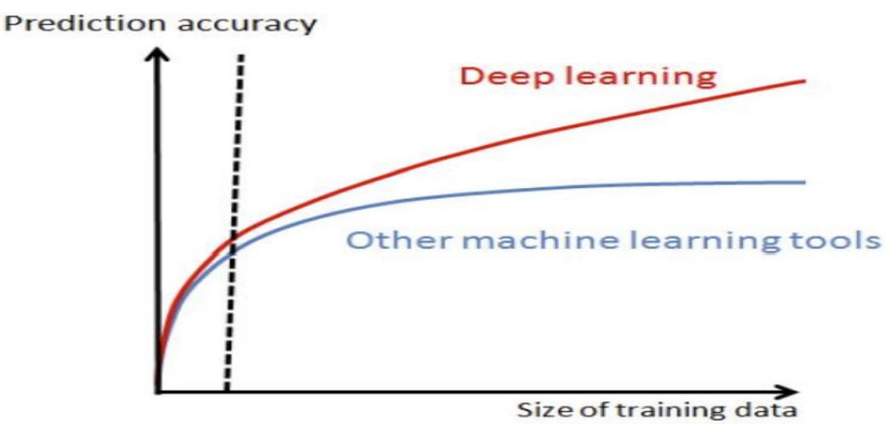
CONVOLUTIONAL NEURAL NETWORK

Ingredienti necessari

CNN ottengono già nel 1998 buone prestazioni in problemi di piccole dimensioni (es. riconoscimento caratteri, riconoscimento oggetti a bassa risoluzione), ma bisogna attendere il 2012 (AlexNet) per un radicale cambio di passo. AlexNet non introduce rilevanti innovazioni rispetto alle CNN di LeCun del 1998, ma alcune condizione al contorno sono nel frattempo cambiate:

- BigData: disponibilità di dataset etichettati di grandi dimensioni (es. ImageNet: milioni di immagini, decine di migliaia di classi).

La superiorità delle tecniche di deep learning rispetto ad altri approcci si manifesta quando sono disponibili grandi quantità di dati di training.



CONVOLUTIONAL NEURAL NETWORK

- GPU computing: il training di modelli complessi (profondi e con molti pesi e connessioni) richiede elevate potenze computazionali. La disponibilità di GPU con migliaia di core e GB di memoria interna ha consentito di ridurre drasticamente i tempi di training: da mesi a giorni.

Hardware per il training

- Il training di modelli complessi (profondi e con molti pesi e connessioni) su dataset di grandi dimensioni (es. ImageNet) richiede elevate potenze computazionali.
- La disponibilità di GPU con migliaia di core e GB di memoria interna è di fatto necessaria per contenere i tempi di training: da mesi a giorni/ore.

Nvidia Titan RTX (2.7K€)

- 4608 Core
- 24 GB RAM
- 16.3 TFLOPS (fp32)

CPU normalmente < 1 TFLOPS

CONVOLUTIONAL NEURAL NETWORK

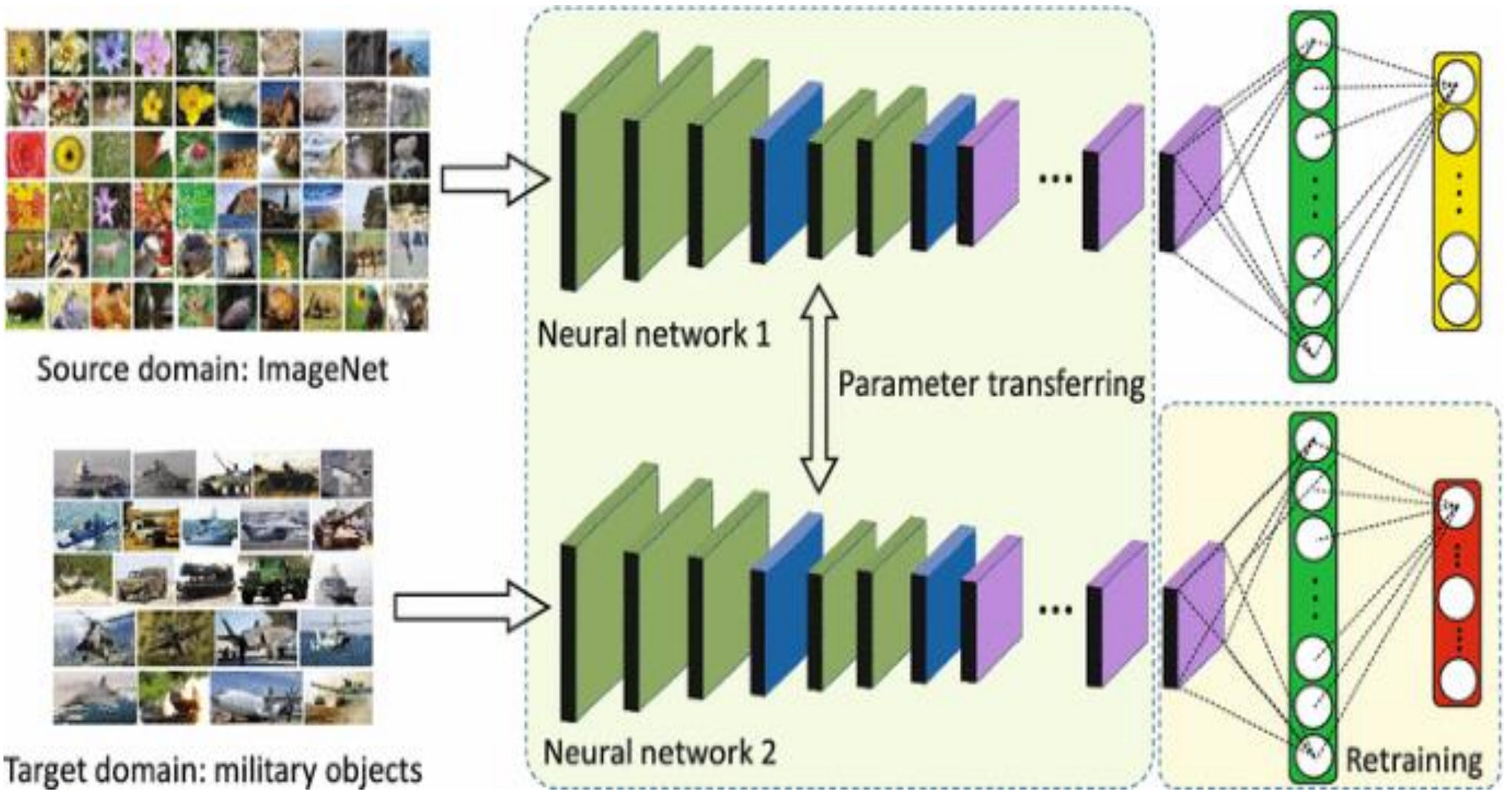
- L'addestramento può essere parallelizzato suddividendo il carico tra diverse GPU (i principali framework lo supportano in modo trasparente):
  - Workstation (es: NVIDIA DXG Station) con 4 GPU collegate tra loro (Nvlink) per massimizzare il trasferimento dei dati senza passare dal bus PCI-express.
  - GPU Cloud (es. Amazon, Google).
  - GPU-Cluster (es. DAVIDE del Cineca) con 180 GPU e una peak performance di circa 1 PFLOPS (1000 TFLOPS)

TOOL PER DEEP LEARNING

I principali framework per il deep learning (tutti con interfaccia Python) sono:

- Tensorflow* (Google). *Il più noto e diffuso*
- PyTorch (Facebook). *Molto apprezzato in ambito ricerca*
- Caffe (Berkeley). *Efficiente per Visione Artificiale*

APPLICAZIONI



APPLICAZIONI

APPLICAZIONI