

# Introduction to Machine Learning

*Francesco Pugliese, PhD*  
*neural1977@gmail.com*

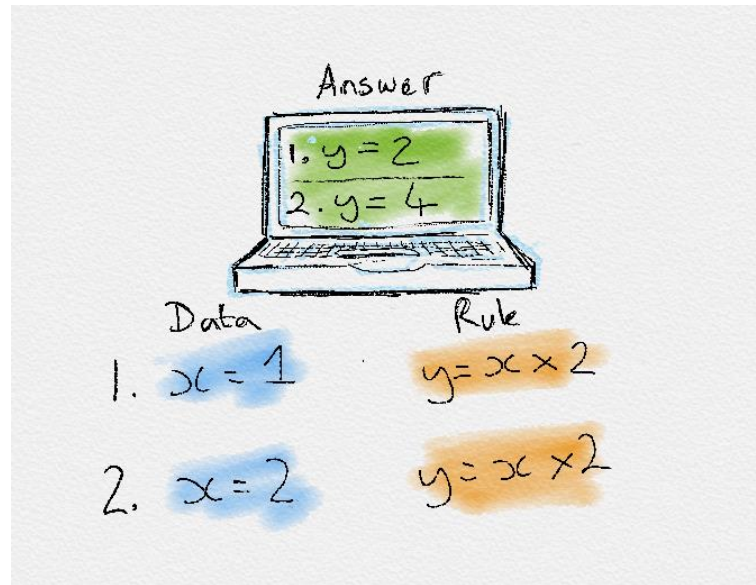
# Che cosa è il Machine Learning ?

- ✓ Machine learning è un ramo dell'Informatica che riguarda la costruzione di algoritmi, che si fondano su una collezione di dati riguardanti un determinato fenomeno.
- ✓ Il termine Machine Learning fu coniato da Arthur Samuel in 1959, un Americano pioniere nel campo dei videogiochi e dell'Intelligenza Artificiale che affermava: **"Il Machine Learning fornisce ai computer la capacità di apprendere senza essere programmaticamente esplicitamente"**.
- ✓ Machine learning è uno strumento per la trasformazione dell'**Informazione in Conoscenza**. Gli schemi nascosti (pattern) e la conoscenza concernente un determinato problema possono essere utilizzati per predire eventi future ed eseguire tutti i tipi di decision making complessi.
- ✓ Il Machine learning può anche essere definito come il processo per risolvere un problema pratico attraverso:
  - ✓ L'acquisizione di un dataset.
  - ✓ Costruendo algoritmicamente un modello statistico basato su quel dataset.

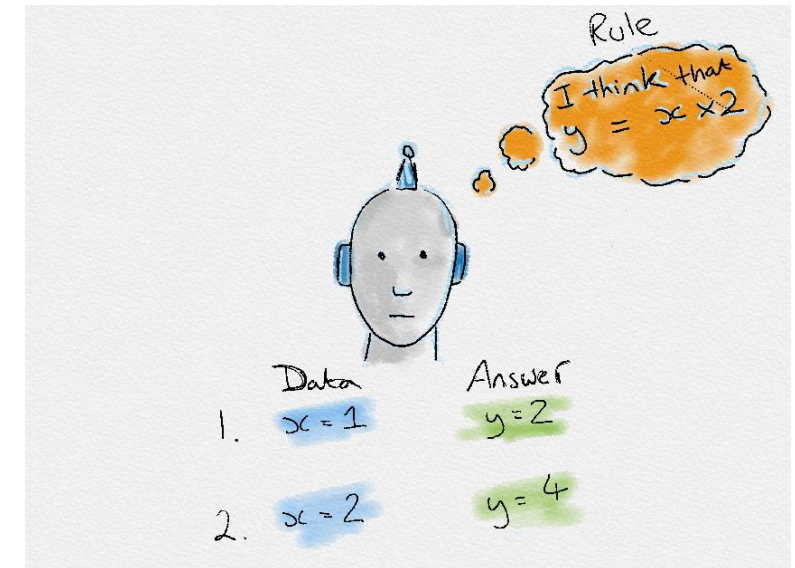
Quel modello statistico lo si assume essere in qualche modo capace di risolvere il problema pratico.

# Programmazione Tradizionale vs Machine Learning

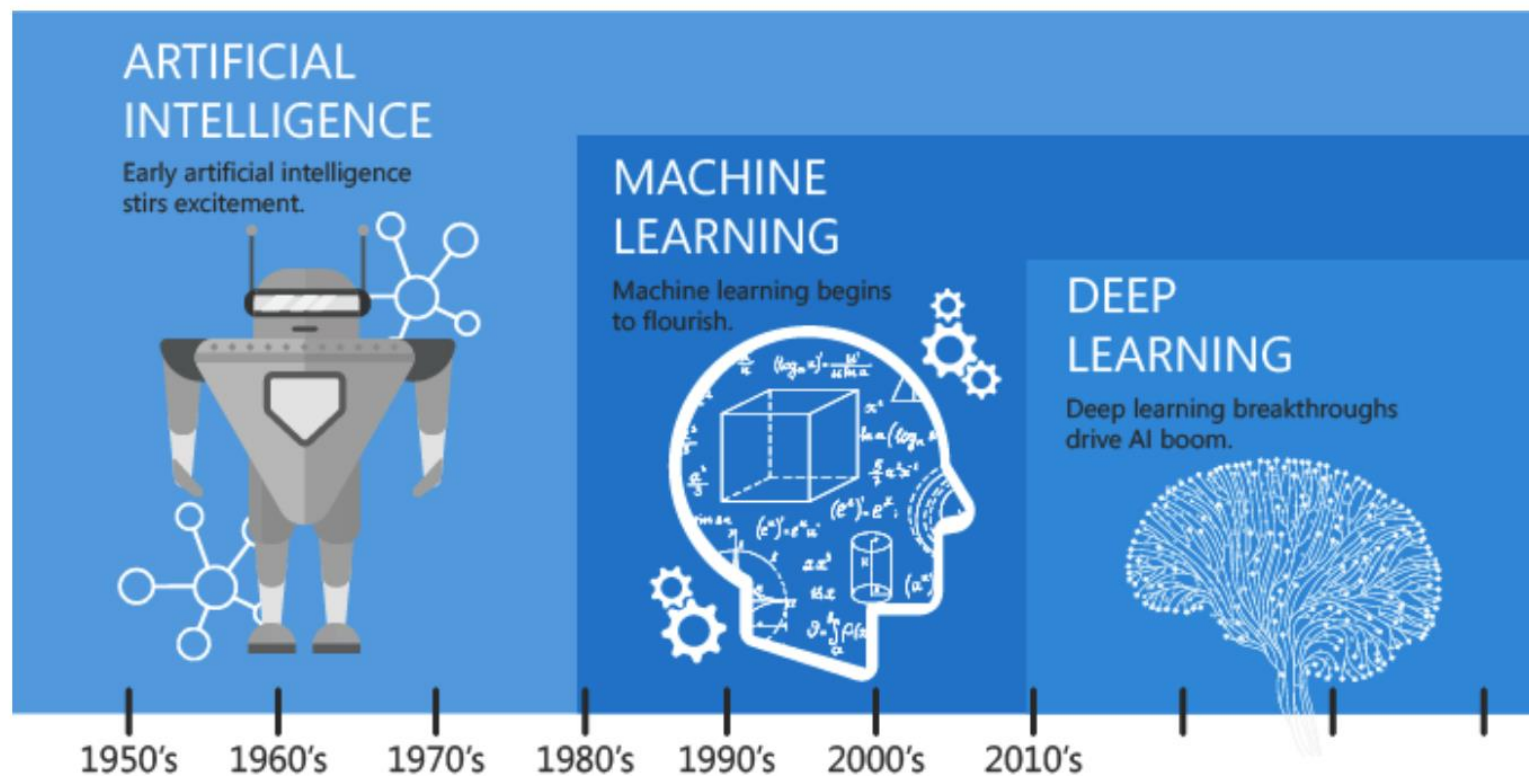
- ✓ Tradizionalmente, l'ingegneria del software combina le regole create dall'uomo con i dato al fine di creare delle risposte ad un problema specific. Invece il Machine Learning usa i dati e le risposte per scoprire le regole sottostanti ad un determinate problema.
- ✓ Per apprendere le regole che governano un fenomeno, le machine devono attraversare un processo di apprendimenti, provando differenti regole e cercando di capire come esse performano. Ecco perchè esso viene conosciuto come Machine Learning.



**VS**



# Una breve storia del Machine Learning



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

Image: Linked In | Machine Learning vs Deep learning

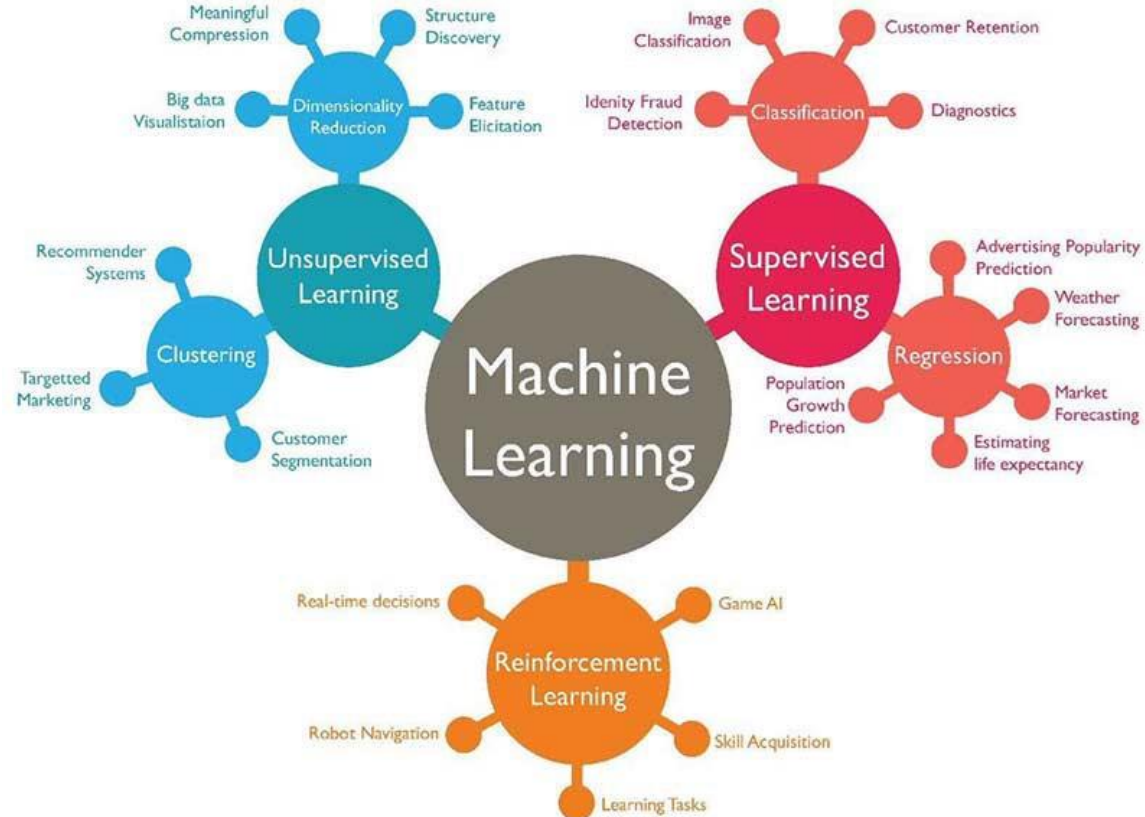
# Tipi di Machine Learning

- ✓ Esistono 4 tipi di Machine Learning ampoa,emte riconosciuti:
  - ✓ Supervised learning
  - ✓ Unsupervised learning
  - ✓ Semi-supervised learning
  - ✓ Reinforcement learning
- ✓ Ciascuna forma di Machine Learning presenta diversi approcci, ma tutti seguono lo stesso processo e la stessa teoria sottostante.
- ✓ Il **No Free Lunch theorem** è famoso nel Machine Learning. Questo teorema afferma che non esiste un singolo algoritmo che lavora bene per tutti i tipi di task. Ciascun task che cerchi di risolvere ha le sue proprie idiosincrasie, ovvero le proprie specificità ed avversità a determinati modelli del Machine Learning. Quindi, ci sono molti algoritmi e approcci che si adattano a ciascun specificità e stranezza di un singolo problema.

# No Free Lunch Theorem

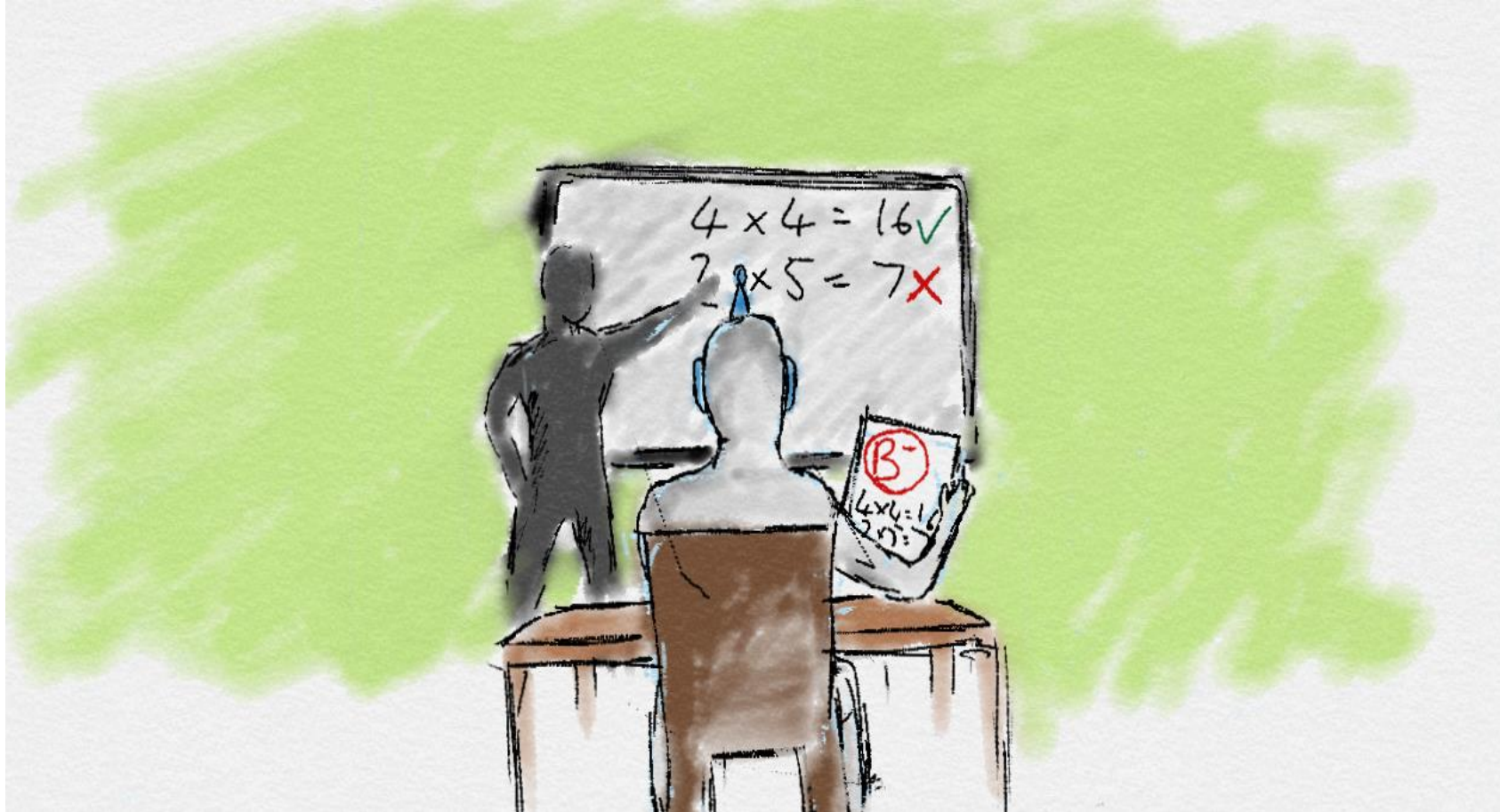
- ✓ Il **No Free Lunch Theorem (NFL o NFLT)** è spesso utilizzato nel campo dell'ottimizzazione e del Machine Learning, ed afferma che tutti gli algoritmi di ottimizzazione performano ugualmente bene quando è calcolata la media delle loro performance su tutti i possibili problemi che possono esistere.
- ✓ Questo implica che non esiste il miglior algoritmo di ottimizzazione in assoluto. E quindi che non esiste il miglior algoritmo di Machine Learning in assoluto per i problemi di modellazione predittiva come la classificazione o la regressione.
- ✓ Dunque su tutto lo spazio di tutti i possibili problemi, ogni tecnica di ottimizzazione performerà mediamente bene come ogni altra tecnica (inclusa la Random Search), e quindi la prima implicazione è che sia per problemi di ottimizzazione statica e dipendente dal tempo la performance media di ogni possibile coppia di algoritmi attraverso tutti i problemi è esattamente identica.

# Tipi di Machine Learning





# Supervised learning





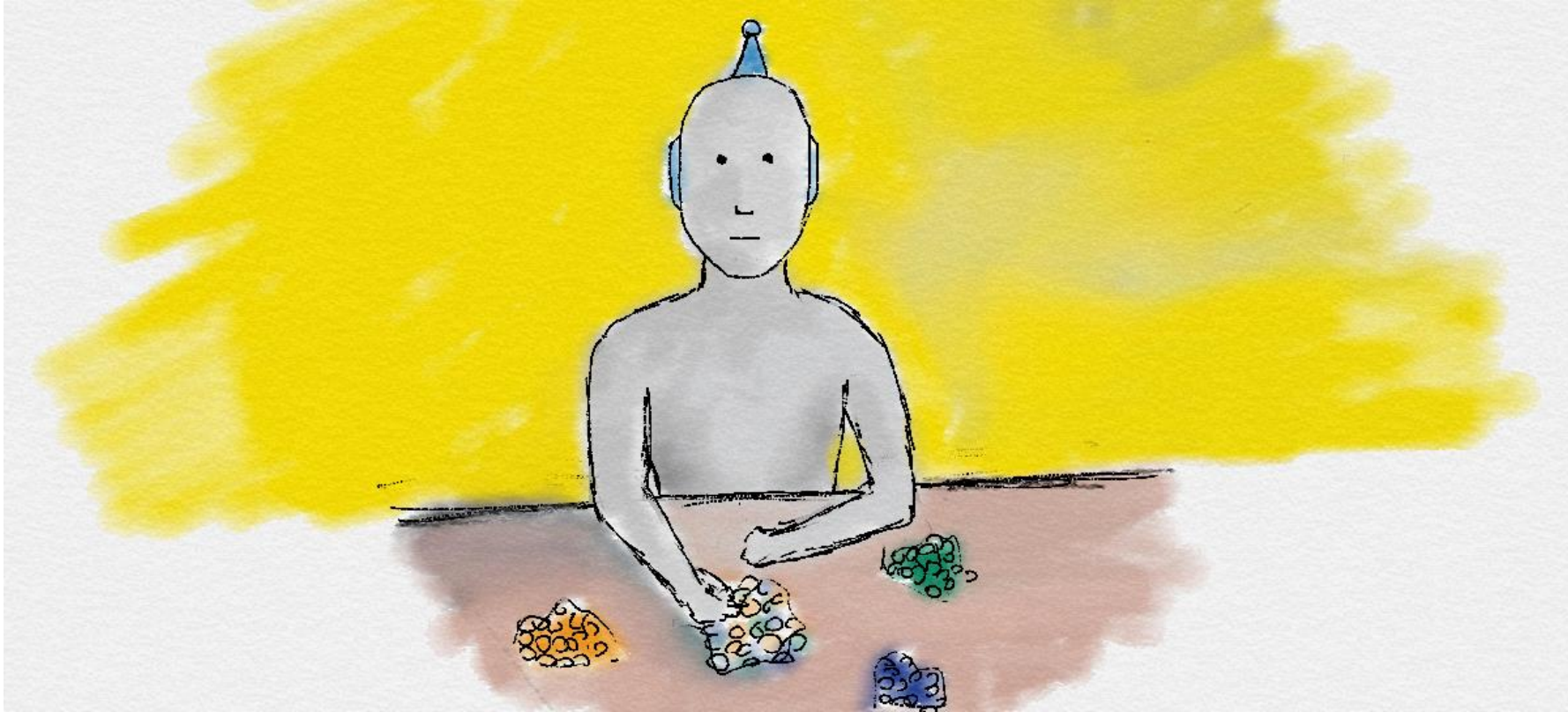
# Supervised learning

- ✓ Nel **Supervised Learning**, l'obiettivo è apprendere le regole di **mapping** (rules) tra un insieme di input ed un insieme di outputs.
- ✓ Per esempio, gli input potrebbero essere **previsione del tempo**, e gli output potrebbero essere il **numero di turisti** presso una spiaggia.
- ✓ L'obiettivo dell'**Apprendimento Supervisionato** è imparare il mapping che descrive le relazioni tra temperature e numero dei turisti.
- ✓ Un esempio di dati etichettati di input e output viene fornito durante il processo di apprendimento per insegnare al modello come comportarsi in corrispondenza di determinati input, da qui, **"apprendimento con supervisione"**.
- ✓ Nel caso della spiaggia per esempio, nuovi input potrebbero essere alimentati nell'algoritmo di Machine Learning che riguardano la previsione della temperature il quale fornirà una predizione futura del numero dei visitatori.

# Supervised learning

- ✓ Avere la capacità di adattarsi a nuovi input e fare predizioni è un element fondamentale per **la generalizzazione** nel machine learning. Durante l'addestramento, vogliamo **massimizzare la capacità di generalizzazione** del modello, in modo tale che il modello supervisionato definisca la relazione sottostante e generale.
- ✓ Se il modello è sovra-addestrato, possiamo causare over-fitting sugli esempi di training e il modello è incapace di adattarsi a nuovi dati provenienti dal mondo e mai visti prima.
- ✓ Un effetto collaterale è essere consapevoli che la supervision che forniamo introduce un **bias** nell'apprendimento. Il modello può solo imitare ciò che ha visto, pertanto è importante mostrargli esempi di training **affidabili e senza bias**.
- ✓ Inoltre, il supervised learning richiede un sacco di dati prima di poter apprendere.
- ✓ Ottenere abbastanza dati etichettati affidabili è spesso la parte più costosa e spesso più difficile del supervised learning. (Per questo i **dati sono definiti come il nuovo petrolio!**)

# Unsupervised learning



# Unsupervised learning

- ✓ Nell'unsupervised learning, ossia apprendimento non supervisionato, solo i dati di input sono forniti come dataset. Ovvero non ci sono **etichette** da prendere come esempi di output per il modello. Tuttavia è sorprendentemente utile sapere che c'è anche in questo caso la possibilità di trovare molti pattern (schemi) interessanti e complessi anche in presenza di dati senza etichette.
- ✓ Un esempio di apprendimento non supervisionato è nella vita reale e potrebbero essere **ordinare differenti monete di diverso colore in pile separate**. Nessuno penserebbe a come le state separando, tuttavia semplicemente guardando alle loro feature come i colori potresti osservare quali monete sono associate allo stesso colore e quali di esse vanno clusterizzate all'interno del Gruppo corretto.
- ✓ L'Unsupervised learning può essere molto **più difficile** del supervised learning, dal momento che la rimozione della supervisione può portare ad un problema che è diventato meno definito.
- ✓ Si parte da uno **stato pulito (clean slate)** con meno bias e ci si potrebbe anche trovare in uno nuovo, un miglior modo per risolvere il problema. Quindi, ecco perché l'unsupervised learning è anche conosciuto come **"scoperta della conoscenza" (knowledge discovery)**. L'apprendimento non supervisionato è molto utile quando si esegue una analisi dei dati **"Esplorativa" (EDA - Exploratory Data Analysis)**.

# Tecniche di Apprendimento non Supervisionato

- ✓ Alcuni tipi di apprendimento non supervisionato sono:
  - ✓ Clustering (Cluster Analysis)
  - ✓ Stima della Densità (Density estimation)
  - ✓ Riduzione della Dimensionalità (Dimensionality reduction)
  - ✓ Modelli a Variabili Latenti (Latent variable models)
  - ✓ Individuazione delle Anomalie (Anomaly detection)
- ✓ Tecniche più complesse di apprendimento non supervisionato implicano l'uso delle reti neurali come gli **Auto-encoders** e le **Deep Belief Networks**

# Tecniche di Apprendimento non Supervisionato

- ✓ **Dimensionality Reduction:** si tratta di una tecnica di trasformazione dei dati fra uno spazio ad alta dimensionalità verso uno spazio a bassa dimensionalità in modo che la rappresentazione a bassa dimensionalità mantenga comunque alcune proprietà significative dei dati originali. Lavorando su spazi ad alta dimensionalità che possono essere non desiderabili per varie ragioni, i dati grezzi sono spesso sparsi come conseguenza del fatto che analizzare i dati può essere spesso **computazionalmente intrattabile** (difficile da controllare e gestire). I metodi più comuni sono la **PCA** che cerca una combinazione di feature che catturano bene la varianza delle feature originali (es. decomposizione di segnali in componenti con la matrix factorization), le proiezioni casuali (**Random Projections**) che forniscono parecchi strumenti per la riduzione dei dati attraverso proiezioni casuali, e la **Feature Agglomeration** che applica un **Clustering Gerarchico** per raggruppare insieme feature che si comportano allo stesso modo.
- ✓ **Latent Variable Models:** che mettono in relazione un insieme di variabili osservate (anche chiamate variabili manifeste) ad un insieme di variabili latenti, ossia variabili che non possono essere direttamente osservate, ma piuttosto inferite attraverso modelli matematici delle variabili osservabili. Si assume che la risposta sugli indicatori o variabili manifeste sono il risultato della posizione degli individui sulle variabili latenti. Esistono diversi tipi di latent variable model a seconda del fatto che le variabili manifeste e le variabili latenti siano categoriche o continue. Nella **Factor Analysis** e la **Latent Trait Analysis**, sia le variabili latenti sono continue e normalmente distribuite (secondo la normale) mentre nella **Latent Profile Analysis** e nella **Latent Class Analysis** le variabili hanno una distribuzione multinomiale, che è una generalizzazione della distribuzione binomiale.

# Tecniche di Apprendimento non Supervisionato

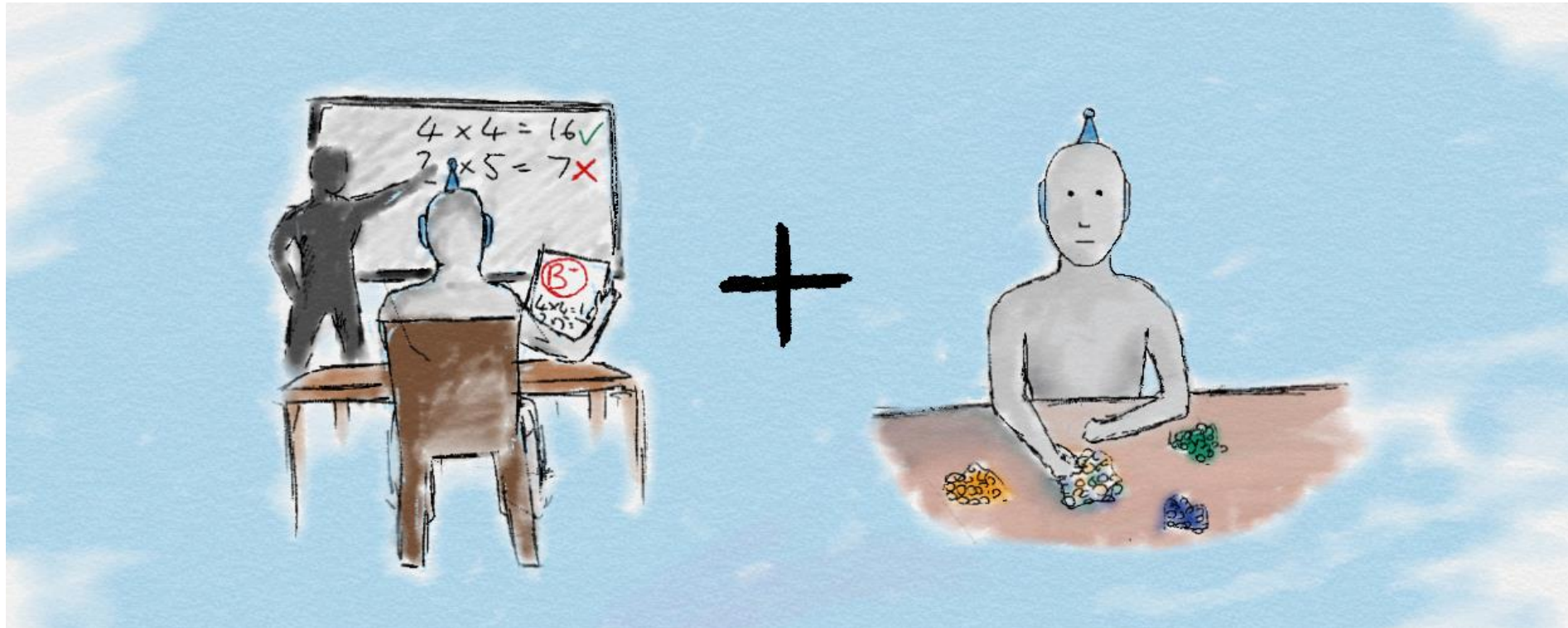
- ✓ **Clustering:** Il clustering (anche detta cluster analysis) consiste in un insieme di metodi per raggruppare oggetti in classi omogenee. Un cluster è un insieme di oggetti che presentano tra loro delle similarità, ma che per contro presentano dissimilarità con oggetti di altri cluster. E' un task principale della EDA (Exploratory Data Analysis) e una comune tecnica per la data analysis statistica, nel pattern recognition, nell'information retrieval, bioinformatica, ecc. Non si tratta di un algoritmo specific, ma il task generale da risolvere con vari algoritmi che differiscono sostanzialmente nella loro comprensione di cosa costituisce un cluster e come trovarli efficientemente. Secondo Scikit-Learn (libreria di machine learning di Python) esistono i seguenti algoritmi di clustering: **K-Means, Affinity propagation, Mean-shift, Spectral clustering, Ward hierarchical clustering, Agglomerative clustering, DBSCAN, OPTICS, Gaussian Mixture, Birch, Bisecting K-Means.**
- ✓ **Density Estimation:** si tratta di un metodo non parametrico utilizzato per il riconoscimento di pattern e per la classificazione attraverso una stima di densità degli spazi metrici, o spazio delle feature. E' una tecnica tra unsupervised learning, feature engineering e data modelling. Alcuni degli algoritmi più popolari sono i modelli a mistura come Gaussian Mixture e gli approcci basati su vicinanza come il kernel density estimate. Per esempio un istogramma è una semplice visualizzazione di dati ma un problem principale è la scelta del binning che può avere un effetto sproporzionato sulla visualizzazione risultante. Il risultato è una stima smooth della densità che è derivate dai dati, e le funzioni come un modello non-parametrico potente della distribuzione dei punti.



# Tecniche di Apprendimento non Supervisionato

- ✓ **Anomaly Detection:** si identifica come il compito di riconoscere elementi rari, eventi o osservazioni che deviano significativamente dalla maggioranza dei dati e non si conformano ad una ben definita nozione di comportamento normale. Tali esempi possono emergere come elementi sospetti in quanto sembrano generate da un differente meccanismo o appaiono inconsistenti con il resto dell'insieme dei dati. L'anomaly detection trova applicazione in molti domini che includono la cyber sicurezza, la medicina, la computer vision, la statistica, le neuroscienze, le frodi finanziarie, ecc. Le anomalie vengono inizialmente ricercate per chiaro rigetto o omissione dai dati per supportare l'analisi statistica, per esempio, per computare la media o la deviazione standard. Le anomalie possono anche essere rimosse per migliorare le prestazioni di modelli predittivi come la regressione lineare e più recentemente la loro rimozione migliora le performance degli algoritmi di machine learning. Tuttavia, in molte applicazioni le anomalie stesse sono interessanti in quanto devono essere identificate e separate dal rumore o dagli outlier irrilevanti. Le anomalie possono essere distinte in :
  1. **anomalie puntuali:** si verificano quando un dato individuale è considerato come un'anomalia rispetto al resto dei dati
  2. **anomalie contestuali:** che sono dipendenti dal contesto, e si verificano quando i dati sono anomali se all'interno di uno specifico contesto
  3. **anomalie collettive:** che si verificano quando una collezione di istanze di dati è anomala rispetto all'intero dataset piuttosto che a singoli valori
- ✓ Alcuni degli algoritmi più popolari di Anomaly Detection sono: statistici (z-score, grubb's test), netid density based (KNN, local outlier factor), One-class (SVM), replicator neural networks, autoencoders, Bayesian networks, hmms, cluster analysis based outlier detection, fuzzy logic, ensemble con feature bagging, ecc.

# Apprendimento Semi-Supervisionato (Semi-supervised learning)



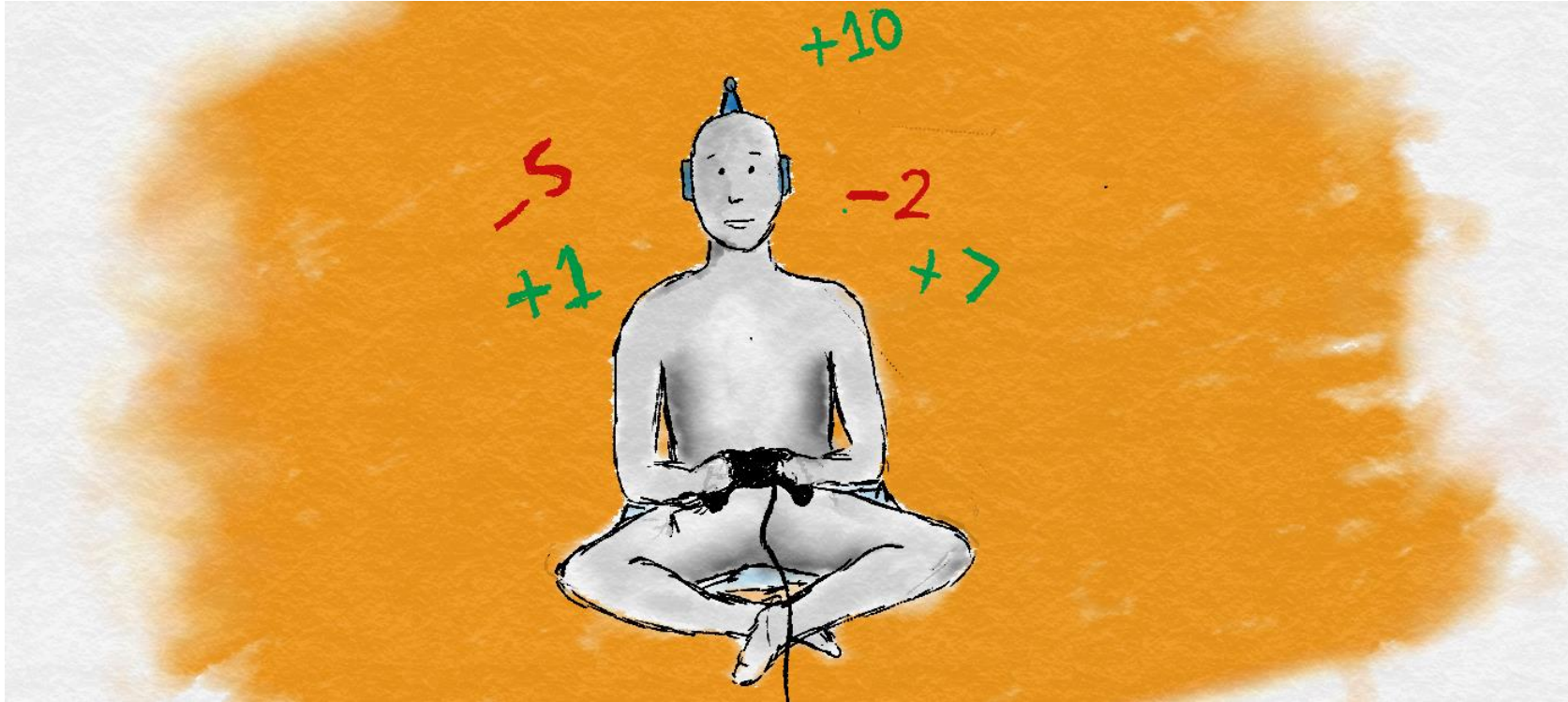
# Semi-supervised learning

- ✓ Il **Semi-supervised learning** o **apprendimento semi-supervisionato** è un **mix** tra approcci supervisionato e non supervisionato.
- ✓ Il processo di apprendimento non è strettamente supervisionato con etichette per ogni singolo input, ma neanche lasciamo l'algoritmo a briglie sciolte senza fornire nessun tipo di feedback.
- ✓ L'apprendimento semi-supervisionato si posiziona nella strada di mezzo. Miscelando insieme una piccola quantità di dati etichettati con una più grande quantità di dati non etichettati viene **ridotto il problema di non avere abbastanza dati etichettati**. Quindi, si aprono nuovi scenari che possono essere risolti nel machine learning.
- ✓ **Esempio:** Un perfetto esempio è nelle analisi mediche, come tac per cancro ai polmoni. Un esperto addestrato è necessario per etichettare il dataset e questo può diventare costoso sia economicamente che in termini di tempo. Invece, un esperto può etichettare solo un piccolo insieme di tac, e l'algoritmo di apprendimento semi-supervisionato potrebbe essere capace di sfruttare questo piccolo dataset per fare predizioni su un insieme più ampio.

# Self-supervised learning

- ✓ Il **Self-supervised learning** o **apprendimento auto-supervisionato** è un metodo di machine learning che apprende da dato non etichettati. Può essere considerata in una posizione intermedia tra l'apprendimento supervisionato e l'apprendimento non supervisionato ed è basato quasi sempre sulle reti neurali artificiali.
- ✓ In questo metodo, la rete neurale apprendere in due step: nel primo, il compito viene risolto usando delle pseudo-label che aiutano ad inizializzare i pesi della rete. Nella seconda fase, il task viene eseguito in maniera supervisionata o non supervisionata.
- ✓ Nel Self-supervised learning il modello addestra se stesso per imparare una parte dell'input da un'altra parte dell'input, è anche conosciuto come predictive o pretext learning. In questo processo, un problema totalmente non supervisionato viene trasformato in un problema supervisionato attraverso una auto-generazione delle etichette.
- ✓ Per fare uso di una grande quantità di dati non etichettati, è necessario impostare i corretti obiettivi di apprendimento per ottenere la supervision dai dati stessi. Per esempio, in NLP, se abbiamo alcune parole, usando il self-supervised learning possiamo completare il resto della frase. Allo stesso modo in un video possiamo predire i frame passati o futuri sulla base dei dati video disponibili.
- ✓ Il Self-supervised learning usa la struttura dei dati per fare uso di una varietà di segnali di supervision attraverso grandi dataset, il tutto senza fare affidamento sulle etichette.

# Reinforcement learning



# Reinforcement learning

- ✓ Il Reinforcement learning o apprendimento per rinforzo, non fa uso di etichette, invece usa un **reward, ossia un premio per apprendere.**
- ✓ Se conosci la psicologia avrai senz'altro sentito parlare di reinforcement learning. Altrimenti, conoscerai il concetto dal modo in cui apprendiamo ogni giorno.
- ✓ In questo approccio, **i feedback positivi e negative occasionali vengono usati per rinforzare i comportamenti.**
- ✓ Pensala come nel caso di addestramento di un **cane, i buoni comportamenti vengono premiati con un good behaviours are rewarded with a un premio e diventano pertanto i più comuni. I brutti comportamenti vengono puniti e diventano meno comuni.** Questo comportamento motivato dal premio è la chiave dell'apprendimento per rinforzo.
- ✓ Questo è molto simile al modo in cui noi umano apprendiamo. Durante le nostre vite, riceviamo segnali positive e negative e costantemente apprendiamo da essi. Le reazioni chimiche nel nostro cervello sono uno dei modi in cui riceviamo questi segnali. Quando accade qualcosa di buono, i neuroni nel nostro cervello forniscono una scarica di neurotrasmettitori positivi come la dopamine che ci fanno sentire bene e diventiamo più propensi a ripetere quella determinate azione.

# Reinforcement learning





# Reinforcement learning

- ✓ Non abbiamo bisogno di una supervisione costante per imparare come nel caso dell'apprendimento supervisionato. Solo fornendo segnali di rinforzo occasionali, otteniamo un apprendimento efficiente.
- ✓ Una delle parti più eccitanti dell'Apprendimento con Rinforzo, è che si discosta dall' addestramento su dei dataset statico, invece usa dataset dinamici in ambienti ricchi di rumore. Questo conduce il Machine Learning ad essere più vicino all'apprendimento in stile umano. Il mondo è semplicemente il nostro ambiente pieno di rumore e di dati complessi.
- ✓ I giochi sono molto popolari nella ricerca dell'apprendimento per Rinforzo. Essi forniscono ambienti ideali e ricchi di dati. Un algoritmo di Apprendimento per Rinforzo semplicemente ha lo scopo di massimizzare il suo premio giocando il gioco più e più volte.
- ✓ Se si può suddividere il gioco come un problema con uno score frequente come rewards, è probabilmente facile adattarlo all'Apprendimento per Rinforzo.

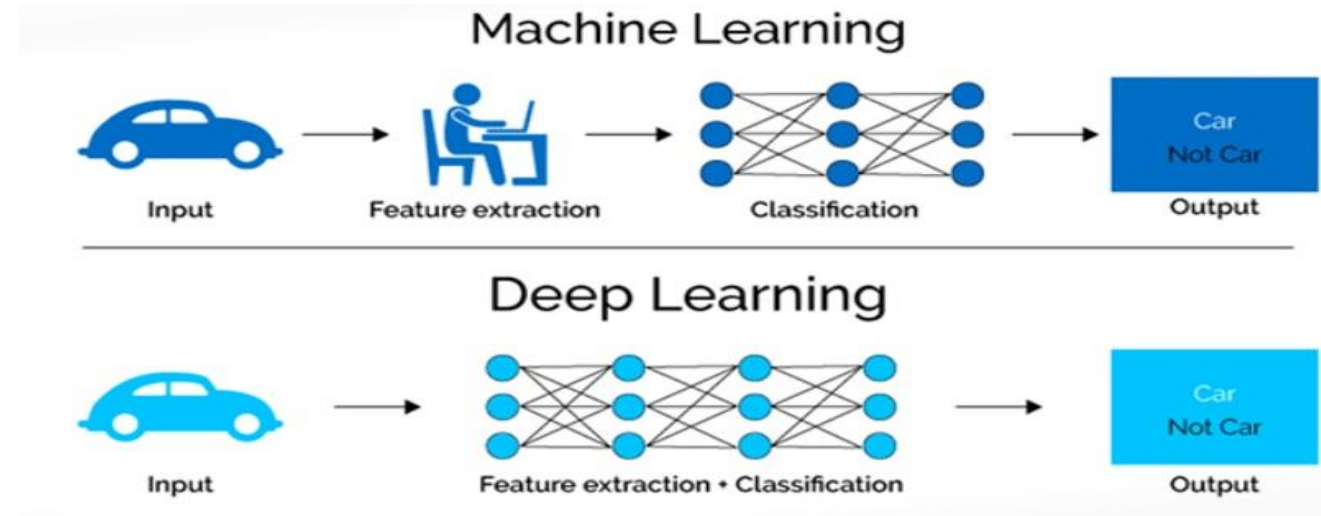
# Shallow vs Deep learning

- ✓ Un algoritmo di apprendimento "poco profondo" (shallow) impara i parametri del modello **direttamente dalle feature degli esempi di training.**
- ✓ La maggior parte degli algoritmi di apprendimento sono shallow.
- ✓ Le eccezioni note sono gli algoritmi di apprendimento di reti neurali, specificatamente quelle reti neurali che sono costruite **con più di uno strato** tra input e output.
- ✓ Tali reti neurali sono chiamate deep neural networks.
- ✓ Nell'apprendimento delle deep neural network (or, simply, deep learning), al contrario dello shallow learning, la maggior parte dei modelli di parametri non sono appresi dalle feature degli esempi di addestramento ma dagli output degli strati precedenti

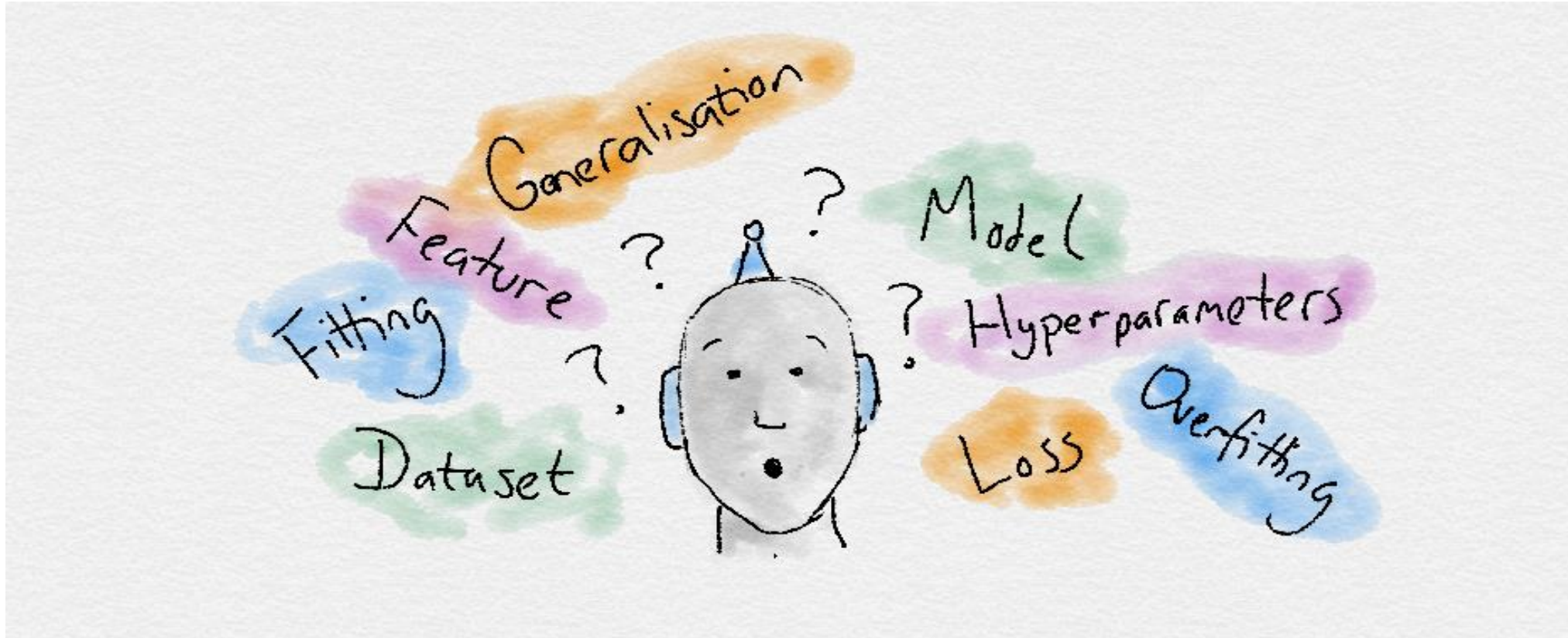
# Deep Learning is Machine Learning

Il Deep Learning si riferisce a quella classe di algoritmi che modellano automaticamente astrazioni di dati ad alto livello

- i. Qui 'modello' significa: definire, trovare, riconoscere e sfruttare
- ii. Qui 'automaticamente' significa: direttamente dai dati, senza affidarsi a feature specifiche al task e fatte a mano.



# Terminologie del Machine Learning



# Terminologie del Machine Learning

- ✓ **Dataset:** Un insieme di esempi di dati, che contengono caratteristiche importanti per risolvere i problemi.
- ✓ **Features:**
  - ✓ Pezzi Importanti di dati che aiutano a comprendere un problema.
  - ✓ Queste vengono inserite in un algoritmo di Machine Learning e lo aiutano ad apprendere.
- ✓ **Modello:**
  - ✓ La rappresentazione (modello interno) di un fenomeno che un algoritmo di Machine Learning è stato appreso. Essa impara questo dai dati ed è mostrato durante il training.
  - ✓ Il modello è l'output che si ottiene dopo aver addestrato un algoritmo.
  - ✓ Per esempio, un albero di decisione dovrebbe essere addestrato e produrre un modello di albero di decisione.

# Parametri vs Iperparametri

- ✓ Un iperparametro è una **proprietà** di un algoritmo di apprendimento, di solito (ma non sempre) che ha un valore numeric. Questo **influenza** il modo in cui un algoritmo lavora.
- ✓ Gli iperparametri non vengono appresi dall'algoritmo stesso dai dati. Essi devono **essere settati dagli analisti dei dati prima di avviare l'algoritmo di training**.
- ✓ I parametri sono **variabili che definiscono il modello** appreso dall'algoritmo di apprendimento.
- ✓ I parametri sono **direttamente modificati** dall'algoritmo di apprendimento basato sui dati di addestramento.
- ✓ L'obiettivo dell'apprendimento è quello di trovare tali valori dei parametri che rendono il modello **ottimale** in un certo senso.

# Bibliografia

Nolfi, S., Floreano, D., & Floreano, D. D. (2000). *Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines*. MIT press.

Pugliese, F., Acerbi, A., & Marocco, D. (2015). Emergence of leadership in a group of autonomous robots. *PloS one*, 10(9), e0137234.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2(1), 1-127.



Francesco **Pugliese**

