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 determinate 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 programmatic esplicitamente**".

- ✓ Machine learning è uno strumento per la trasformazione dell'**Informazione in Conoscenza**. Gli schemi nascosti (pattern) e la conoscenza concernente un determinate problema possono essere utilizzati per predire eventi future ed eseguire tutti i tipi di decision making complessi.

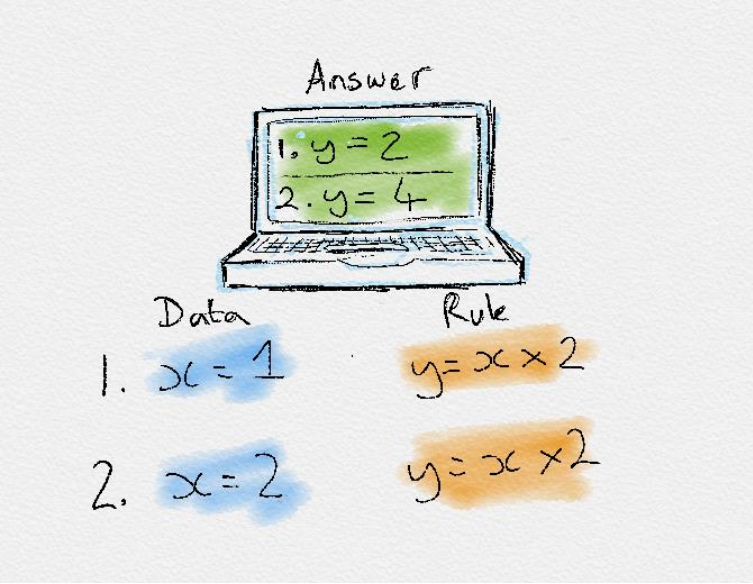
- ✓ Il Machine learning può anche essere definite come il processo per risolvere un problema pratico attraverso:
  - ✓ L'acquisizione di un dataset.
  - ✓ Costruendo algoritmicamente un modello statistic basato su quel dataset.

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

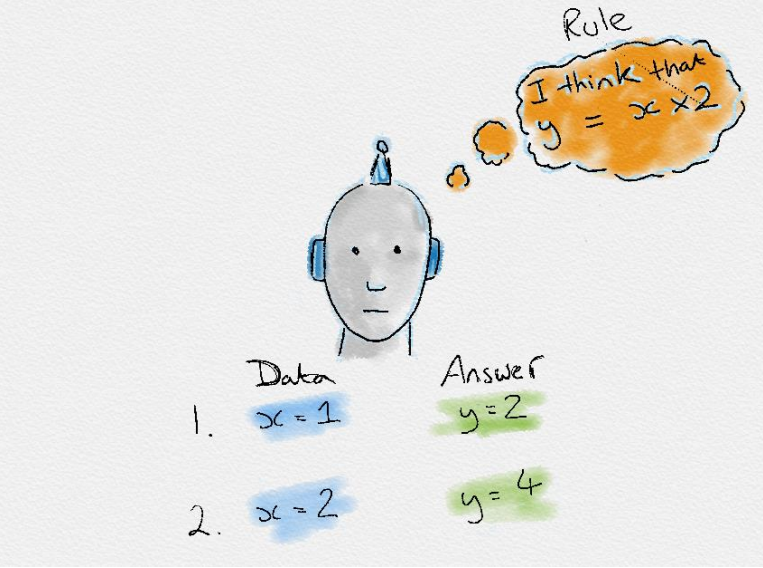
Programmazione Tradizionale vs Machine Learning

- ✓ 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.

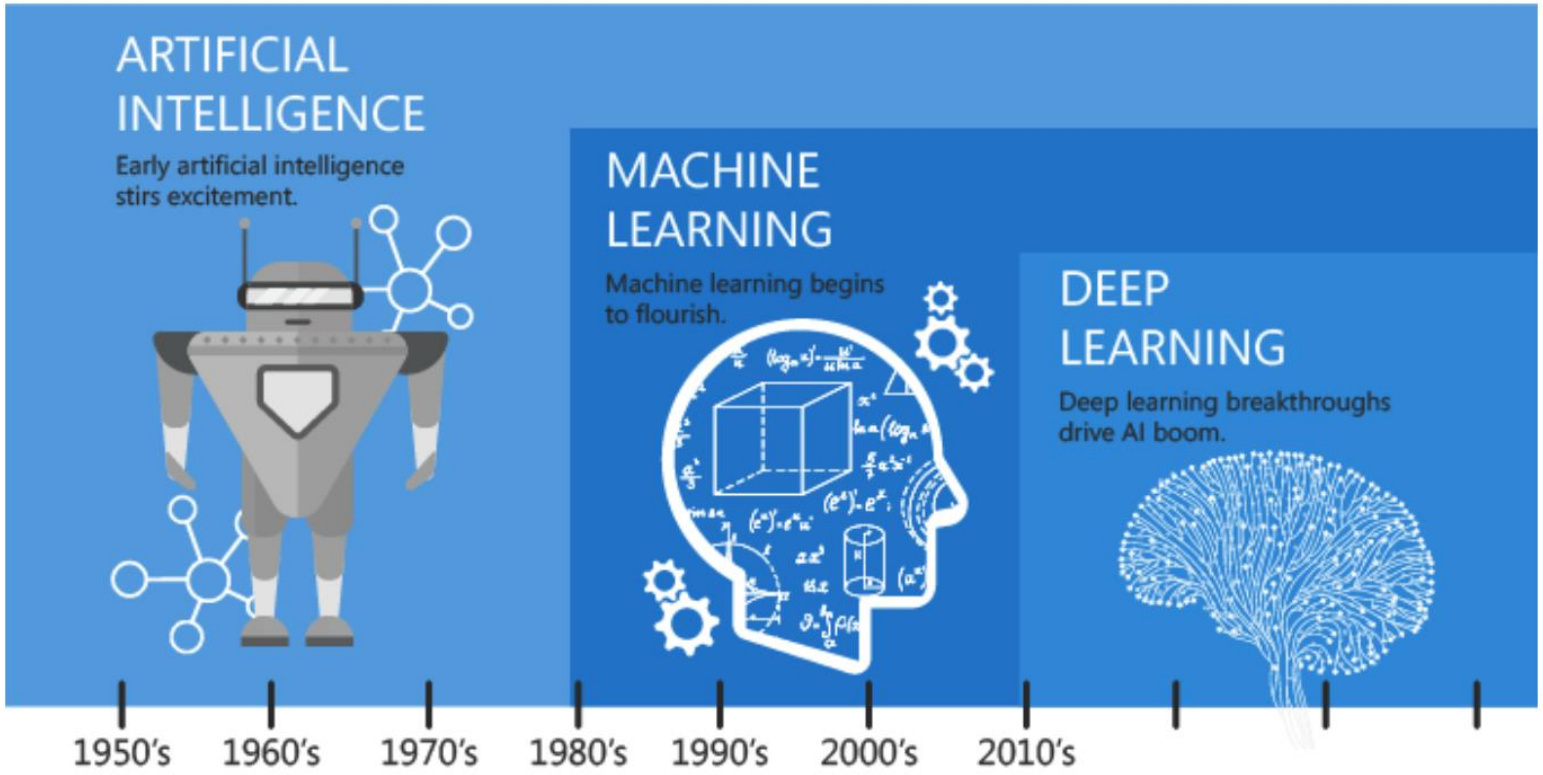
- ✓ Per apprendere le regole che governano un fenomeno, le machine devono attraversare un processo di apprendimenti, provando differenti regole e cercando di capire come esse performano. Ecco perchè esso viene conosciuto come Machine Learning.



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

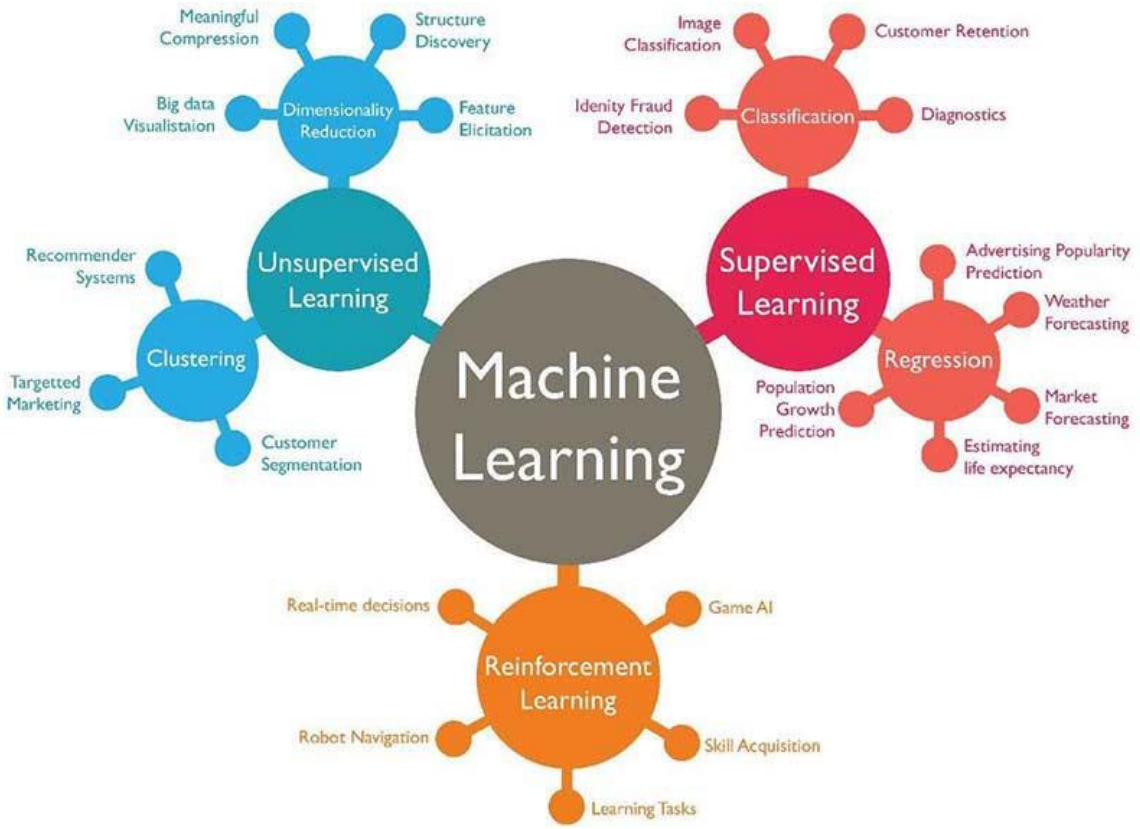
Tipi di Machine Learning

- ✓ 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.

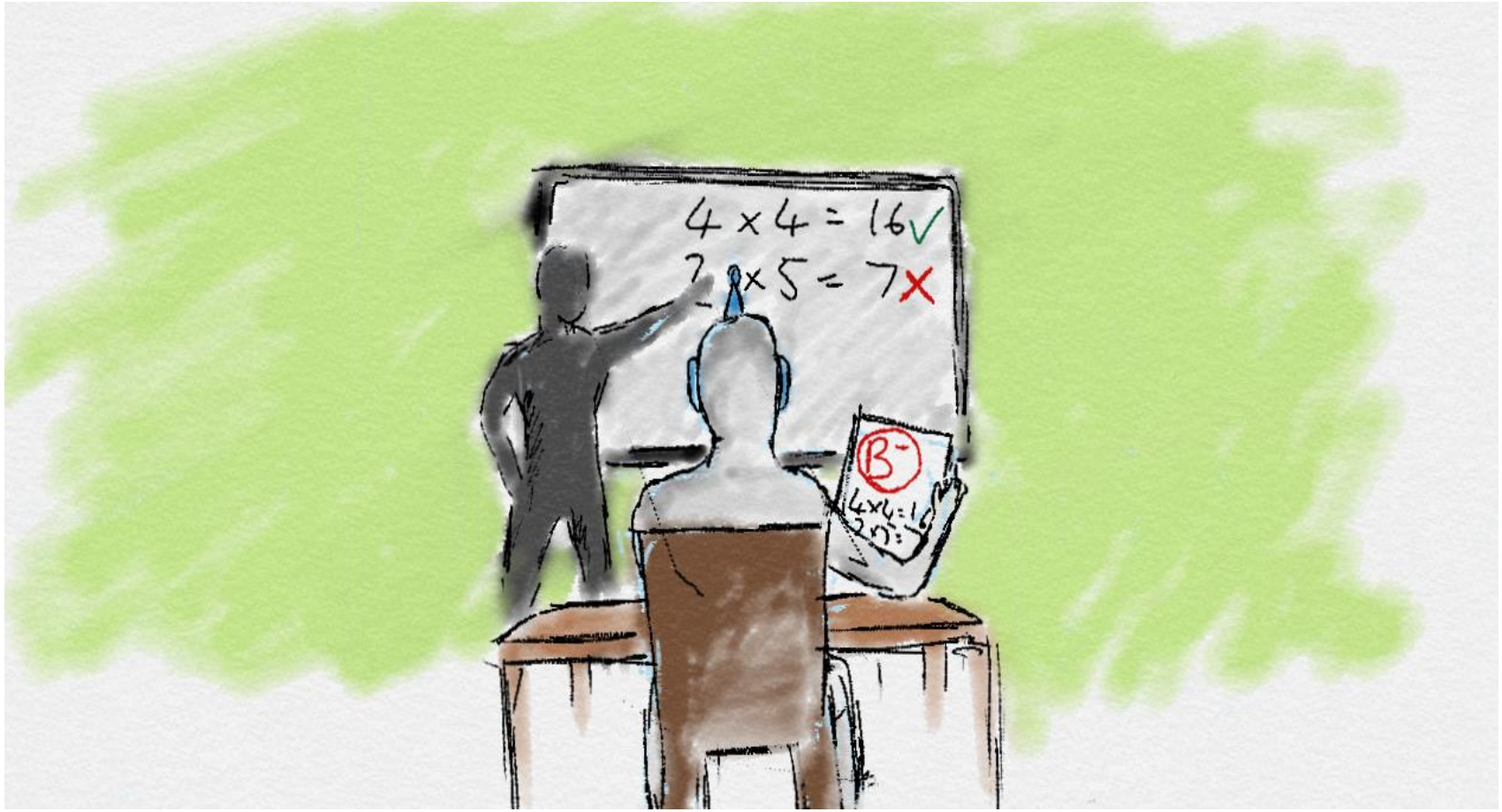
No Free Lunch Theorem

- ✓ 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 i assoluto. E quindi che non esiste il miglior algoritmo di Machine Learning in assoluto per i problem di modellazione predittiva come la classificazione o la regression.
- ✓ 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 corrispondeza 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 temperatire 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.

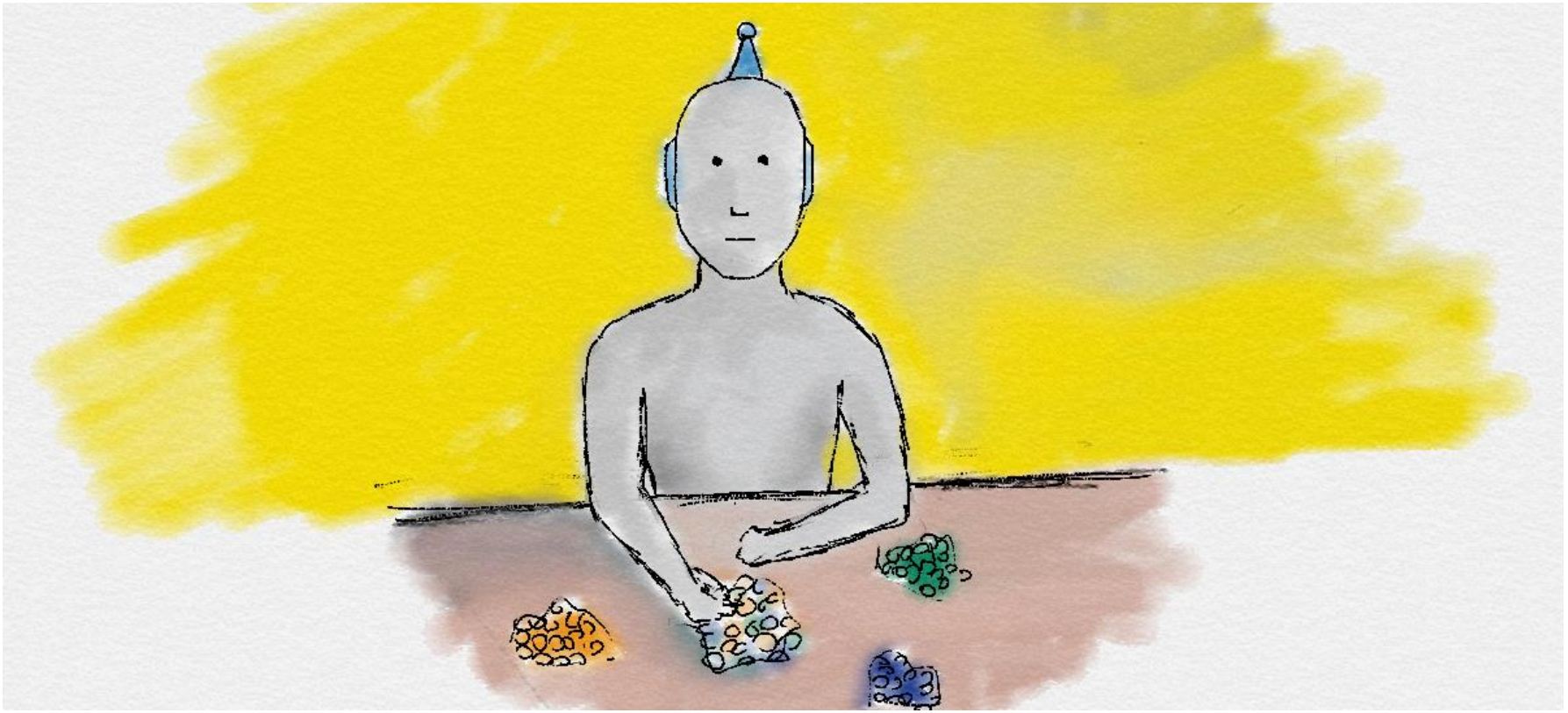
- ✓ Se il modello è sovra-addestrato, possiamo causare over-fitting sugli esempi di training e il modello è incapace di adattarsi a nuovi dati provenienti dal mondo e mai visti prima.

- ✓ 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 fortiti come dataset. Ovvero non ci sono **etichette** da prendere come esempi di output per il modello. Tuttavia è soprendentemente 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 potrebber 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 definite.

- ✓ 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 perchè 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 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 tanenti siano categoriche o continue. Nella **Factor Analysis** e la **Latent Trait Analysis,** sia le variabili latent sono continue e normalmente distribuite (secondo la normale) mentre nella **Latent Profile Analysis** e nella **Latent Class Analysis** le variabili hanno una distibuzione multinomiale, che è una generalizzazione della distribuzione binomiale.

Tecniche di Apprendimento non Supervisionato

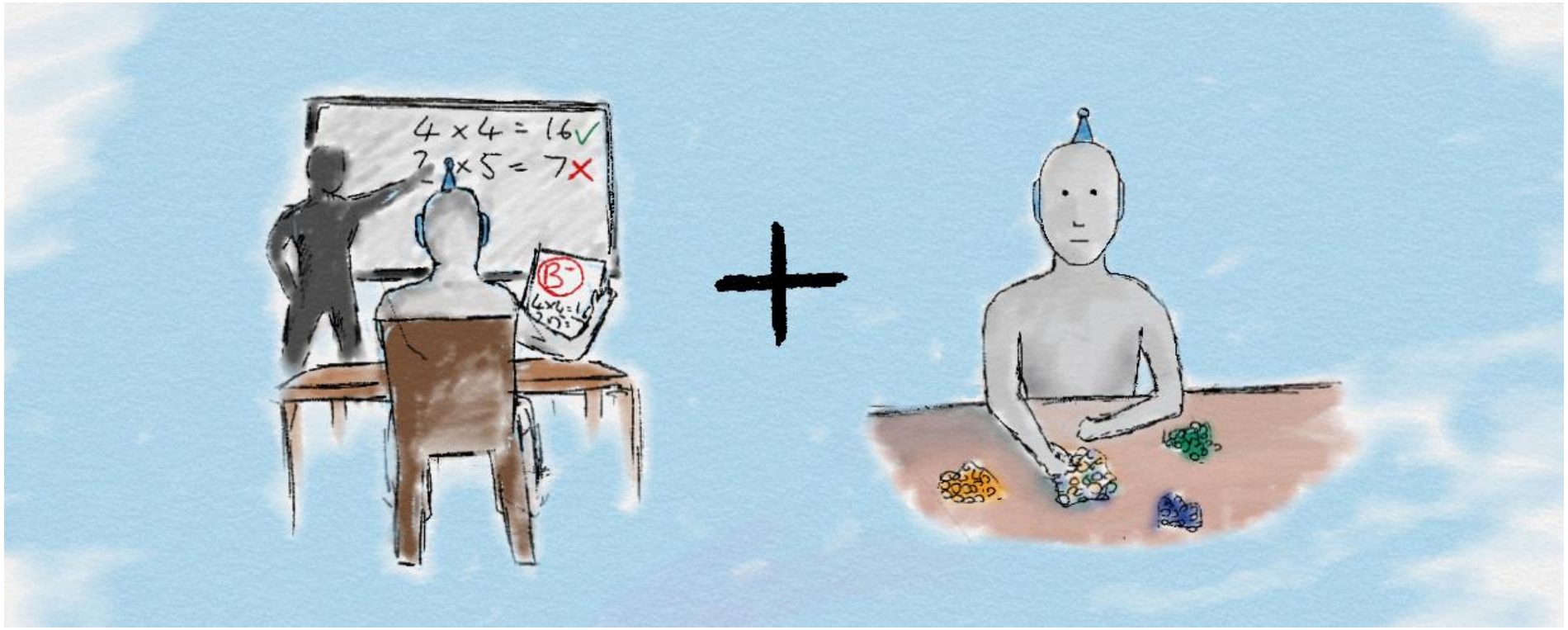
- ✓ **Clustering:** Il clustering (anche detta cluster analysis) consiste in un insieme di metodi per raggruppare oggetti in classi omogenee. Un cluster è un insieme di oggetti che presentano tra loro delle similarità, ma che per contro presentano dissimilarità con oggetti di altri cluster. E' un task principale della EDA (Exploratory Data Analysis) e una commune 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 definite 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 appllicazioni 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.

Apprendimeno 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 piccolo 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 eseperto può etichettare solo un piccolo insieme di tac, e l'algortimo di apprendimento semi-supervisionato potrebbe essere capaci 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 considerate in una posizione intermedia tra l'apprendimento supervisionato e l'apprendimento non supervisionato ed è basato quasi sempre sulle reti neurali artificiali.

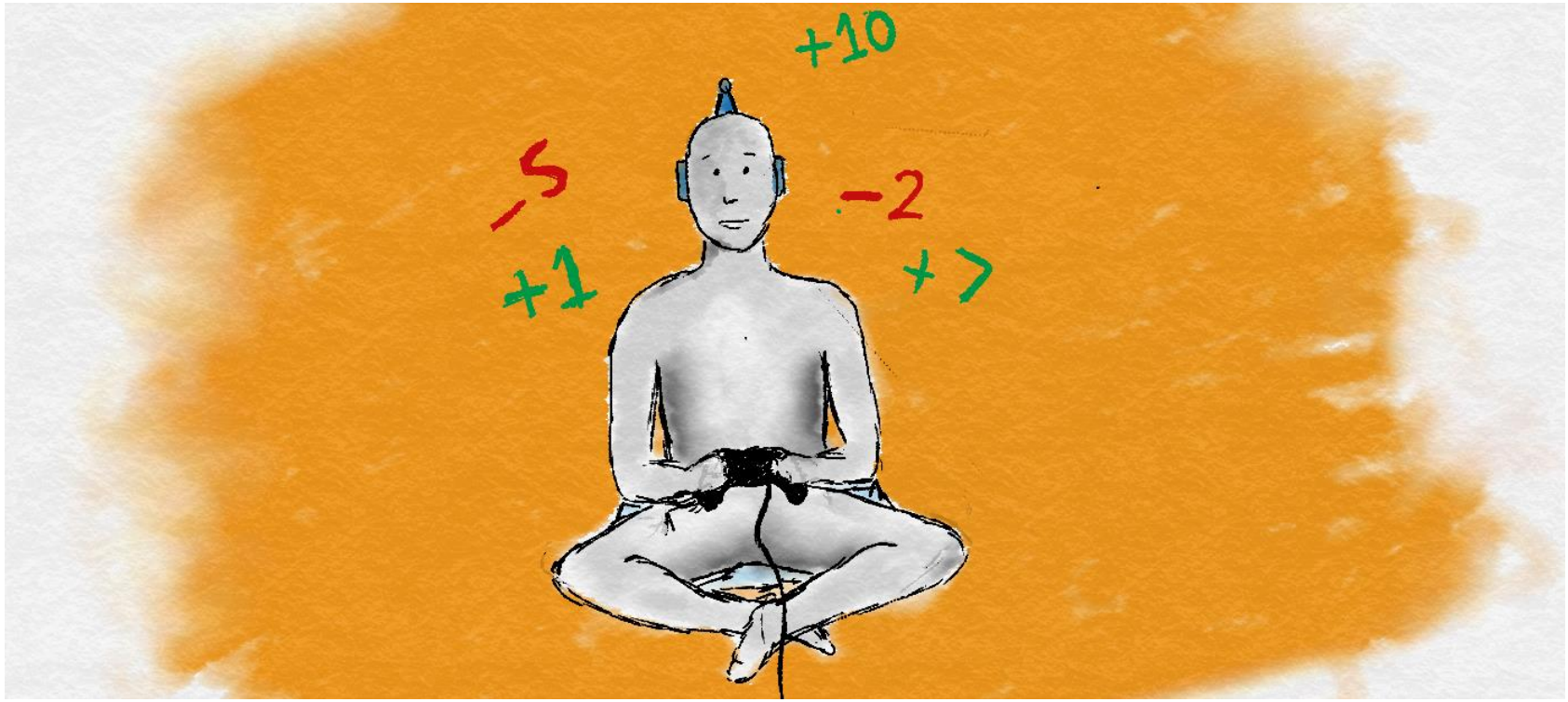
- ✓ In questo metodo, la rete neurale apprendere in due step: nel primo, il compito viene risolto usando delle pseudo-label che aiutano ad inizializzare i pesi della rete. Nella seconda fase, il task viene eseguito in maniera supervisionata o non supervisionata.

- ✓ 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 passata o future 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

- ✓ Reinforcement learning doesn't use labels as such, and instead **uses rewards to learn**.
- ✓ If you're familiar with psychology, you'll have heard of reinforcement learning. If not, you'll already know the concept from how we learn in everyday life.
- ✓ In this approach, **occasional positive and negative feedback is used to reinforce behaviours**.
- ✓ Think of it like training a **dog**, **good behaviours are rewarded with a treat** and become more common. **Bad behaviours are punished and become less common**. This reward-motivated behaviour is key in reinforcement learning.
- ✓ This is very similar to how we as humans also learn. Throughout our lives, we receive positive and negative signals and constantly learn from them. The chemicals in our brain are one of many ways we get these signals. When something good happens, the neurons in our brains provide a hit of positive neurotransmitters such as dopamine which makes us feel good and we become more likely to repeat that specific action.

Reinforcement learning

Reinforcement learning

- ✓ We don't need constant supervision to learn like in supervised learning. By only giving the occasional reinforcement signals, we still learn very effectively.

- ✓ One of the most exciting parts of Reinforcement Learning is that is a **first step away from training on static datasets**, and instead of being able to use dynamic, noisy data-rich environments. This brings Machine Learning closer to a learning style used by humans. The world is simply our noisy, complex data-rich environment.

- ✓ Games are very popular in Reinforcement Learning research. They provide ideal data-rich environments. A Reinforcement Learning algorithm just aims to maximise its rewards by playing the game over and over again.

- ✓ If you can frame a problem with a frequent 'score' as a reward, it is likely to be suited to Reinforcement Learning.

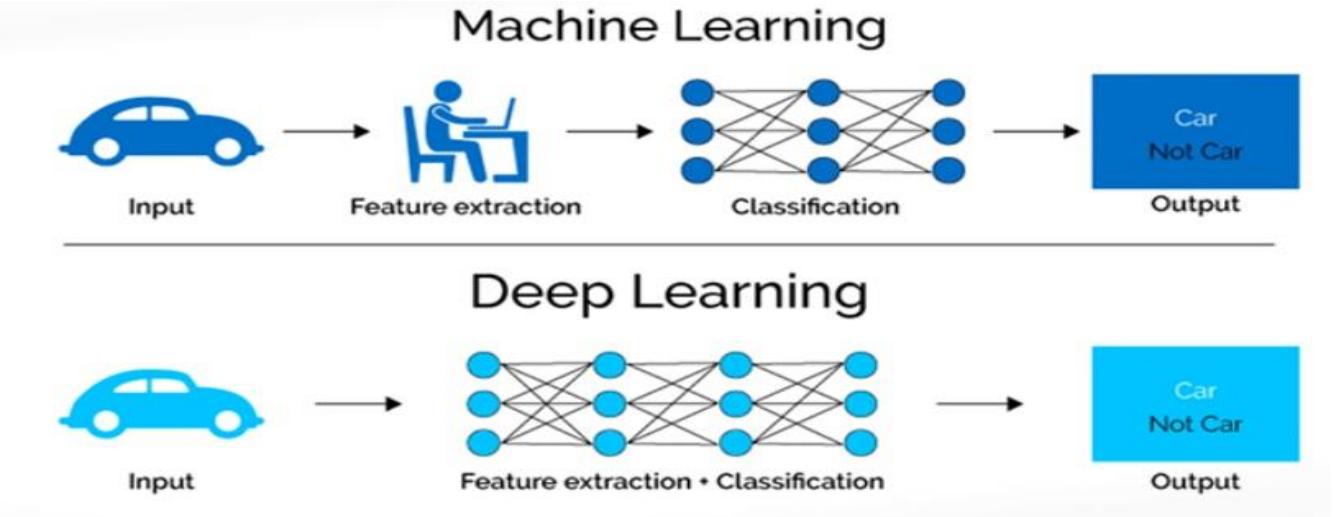
Shallow vs Deep learning

- ✓ A shallow learning algorithm learns the parameters of the model **directly from the features of the training examples**.
- ✓ Most supervised learning algorithms are shallow.
- ✓ The notorious exceptions are **neural network learning algorithms**, specifically those that build neural networks with **more than one layer** between input and output.
- ✓ Such neural networks are called deep neural networks.
- ✓ In deep neural network learning (or, simply, deep learning), contrary to shallow learning, most model parameters are learned not directly from the features of the training examples, but from the **outputs of the preceding layers**.
- ✓ More details about this when we talk about deep learning…

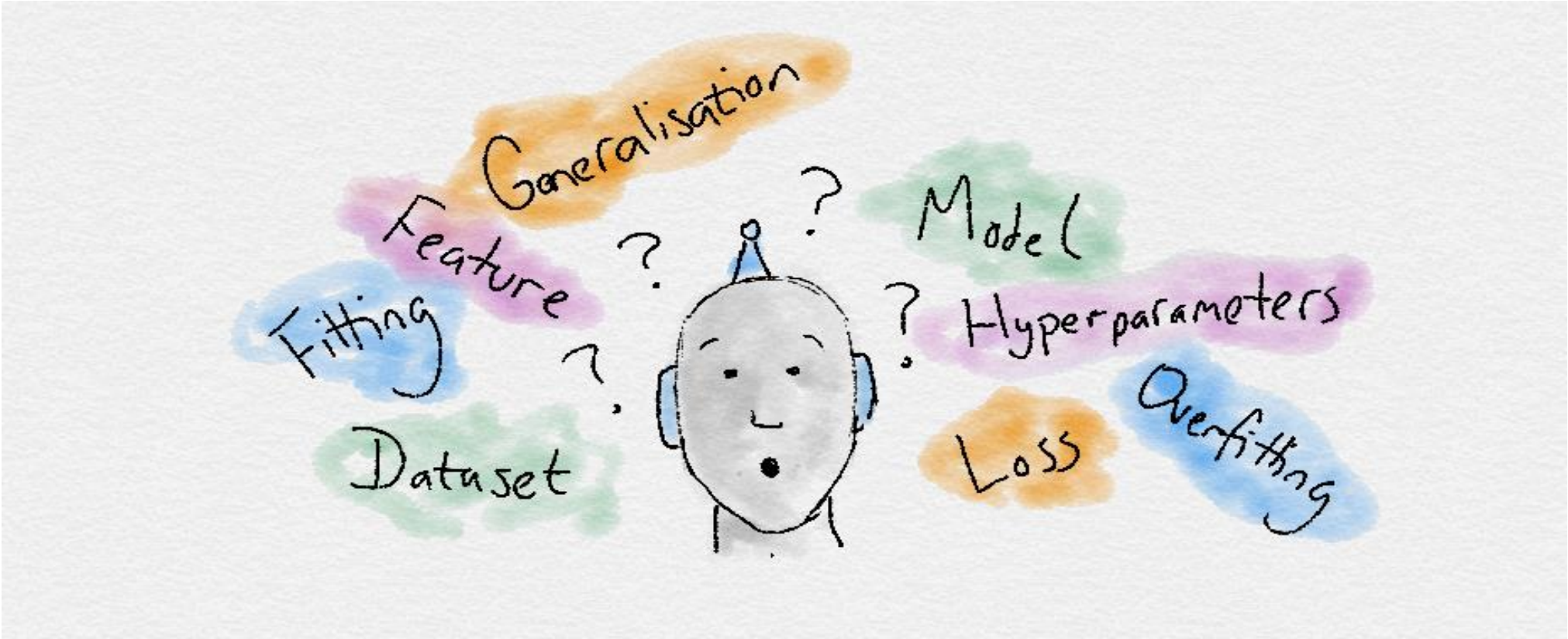
Deep Learning is Machine Learning

Deep Learning refers to algorithms that automatically ‘model’ high-level abstractions in data

i. here ‘model’ means: define, find, recognize and exploit
ii. here ‘automatically’ means: directly from data, without hinging upon handcrafted, task-specific features.



Machine Learning terminologies



Machine Learning terminologies

- ✓ **Dataset**: A set of data examples, that contain features important to solving the problem.

- ✓ **Features**:
  - ✓ Important pieces of data that help us understand a problem.
  - ✓ These are fed in to a Machine Learning algorithm to help it learn.

- ✓ **Model**:
  - ✓ The representation (internal model) of a phenomenon that a Machine Learning algorithm has learnt. It learns this from the data it is shown during training.
  - ✓ The model is the output you get after training an algorithm.
  - ✓ For example, a decision tree algorithm would be trained and produce a decision tree model.

Parameter vs Hyperparameters

- ✓ A hyperparameter is a **property** of a learning algorithm, usually (but not always) having a numerical value. That value **influences** the way the algorithm works.
- ✓ Hyperparameters aren't learned by the algorithm itself from data. They have to be **set by the data analyst before running the algorithm**. More details when we discuss how we tune Machine learning models in the next sections.
- ✓ Parameters are **variables that define the model** learned by the learning algorithm.
- ✓ Parameters are **directly modified** by the learning algorithm based on the training data.
- ✓ The goal of learning is to find such values of parameters that make the model **optimal** in a certain sense.

References

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**