

Capítol 1: Machine Learning

Aina Palacios

Aina Palacios

- Enginyera de Telecomunicacions especialitzada en Audiovisuals
- Màster en Tecnologies Avançades especialitzada en deep learning en Multimèdia!
- Experiència en programació web i machine learning.
- Mentora a IT Academy de **DataScience**



<https://www.linkedin.com/in/ainapc/>



ainaPali#2617

Objectius

- Aprendre conceptes bàsics sobre l'Intel·ligència artificial
- Veure perquè s'utilitza
- Bases matemàtiques i dades
- Machine Learning
- Crear el nostre entorn d'execució per aplicar Intel·ligència artificial
- Deep Learning
- Practicar els conceptes

SO LITTLE TIME

SO MUCH TO LEARN....

memegenerator.net

Intel·ligència artificial

“Part informàtica dedicada al desenvolupament d'algorismes que permet a una màquina prendre decisions intel·ligents o, si més no, comportar-se com si tingués una intel·ligència semblant a la humana”

Esta inteligencia artificial ayuda a los radiólogos a diagnosticar el cáncer de mama

RAQUEL HOLGADO / NOTICIA / 17.07.2022 - 08:00H



- El software creado por una startup alemana detecta la enfermedad y evita que a los sanitarios se les "pase por alto" en una radiografía.

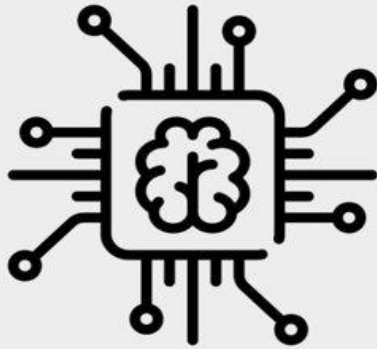
Experto cree que la inteligencia artificial ayudará a "objetivar la subjetividad" dentro de "unos 10 años"

Google despide a uno de sus ingenieros por afirmar que su Inteligencia Artificial "es capaz de sentir"

Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve It

ARTIFICIAL INTELLIGENCE

Any technique which enables computers to mimic human behavior



MACHINE LEARNING

AI techniques that give computers the ability to learn without being explicitly programmed to do so



DEEP LEARNING

A subset of ML which make the computation of multi-layer neural networks feasible



1950s 1960s 1970s 1980s 1990s 2000s 2010s

Machine Learning

Algoritmes que analitzen les dades, aprenen d'elles i fan decisions respecte al que han après

2

Deep Learning

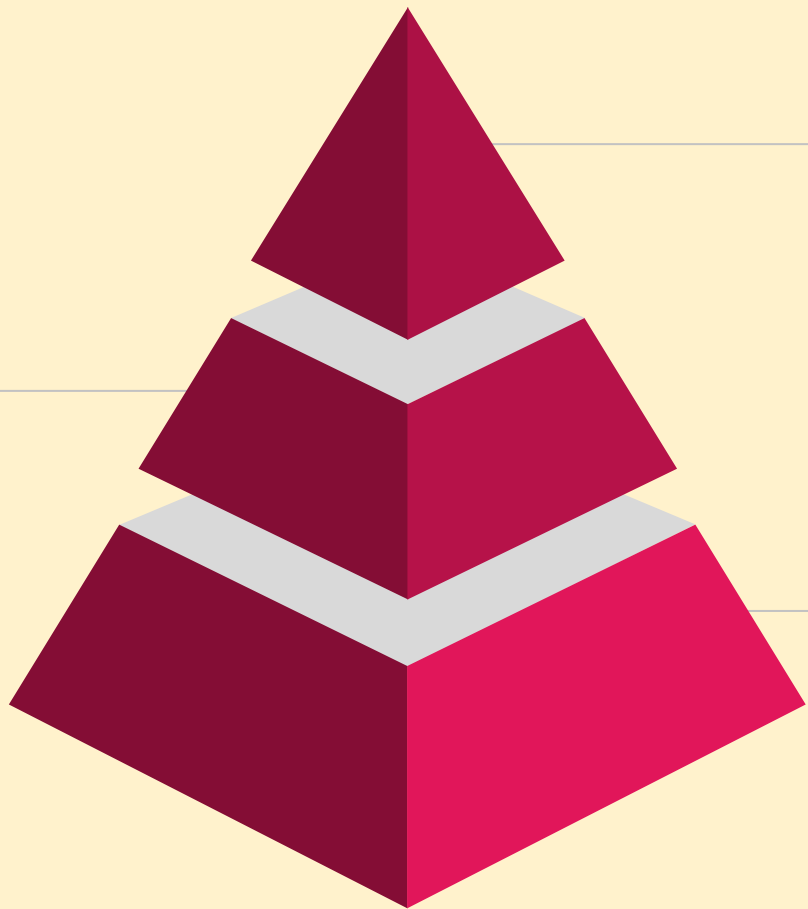
1

Nova evolució del Machine Learning basat en la forma que aprenen els humans. Basat en Neural Networks

Intel·ligència artificial

3

Tot allò que engloba comportaments intel·ligents.

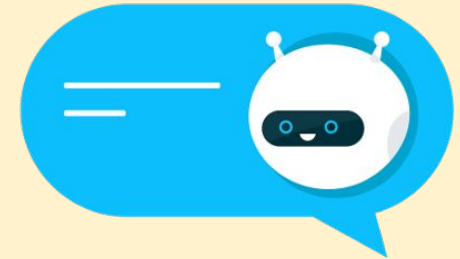


Symbolic AI -> Conegut com a GOF AI (Good old fashioned AI), es basa en el concepte que la intel·ligència és extrapolable del cos biològic, i que la informació pot ser entesa com a símbols. Symbolic AI tracta d'aportar coneixement humà i comportament directament als programes computacionals.

Basades en if/then rules.

- ChatBots
- Automatització

Moltes vegades aquest tipus de pràctica s'utilitza d'acompanyament en el Machine Learning



Machine Learning

Traditional use of computers

Dins de l'Intel·ligència Artificial, tenim el **Machine Learning**.

Consisteix en la utilització d'algoritmes per donar la capacitat a les màquines d'identificar patrons en un conjunt de dades històriques amb l'objectiu de proporcionar decisions automàtiques intel·ligents.

Podem trobar molts tipus de **models** per a cada tipus de problema a resoldre.



Machine learning at training phase



Classical Machine Learning

Task Driven

Supervised Learning

(Pre Categorized Data)

Classification

(Divide the socks by Color)

Eg. Identity
Fraud Detection

Regression

(Divide the Ties by Length)

Eg. Market
Forecasting

Unsupervised Learning

(Unlabelled Data)

Clustering

(Divide by Similarity)

Eg. Targeted
Marketing

Association

(Identify Sequences)

Eg. Customer
Recommendation

Dimensionality Reduction

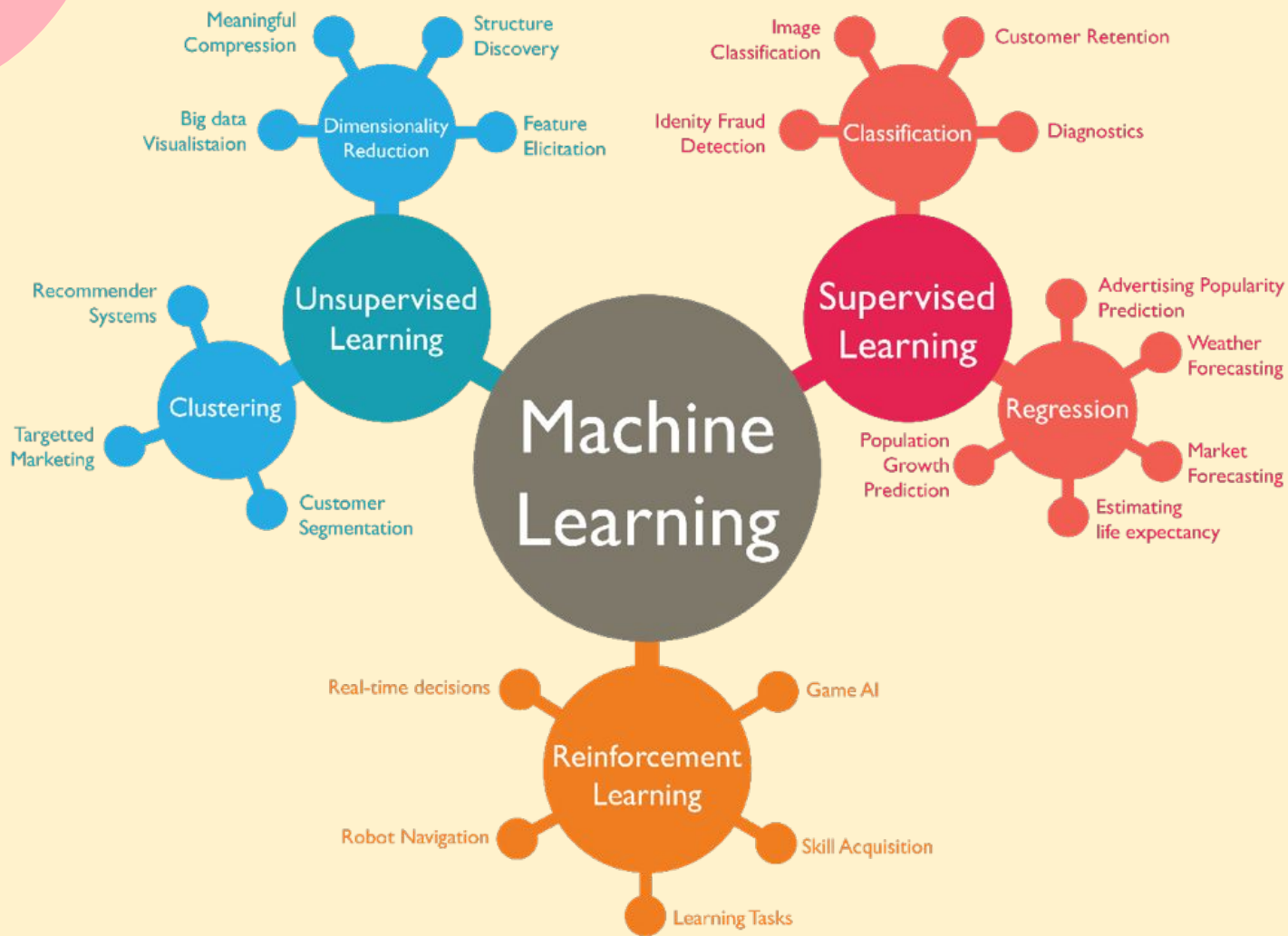
(Wider Dependencies)

Eg. Big Data
Visualization

Obj: Predications & Predictive Models

Pattern/ Structure Recognition

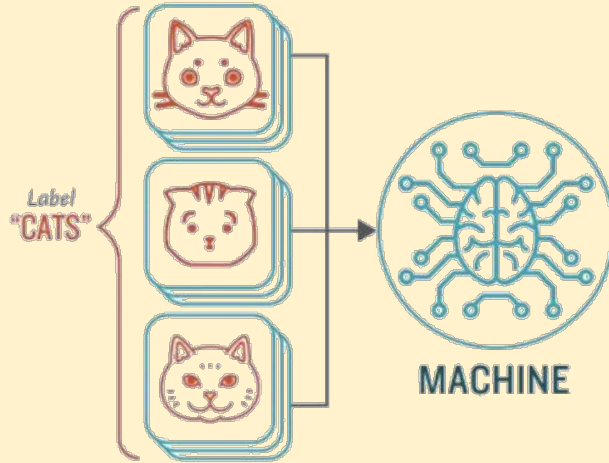




How Supervised Machine Learning Works

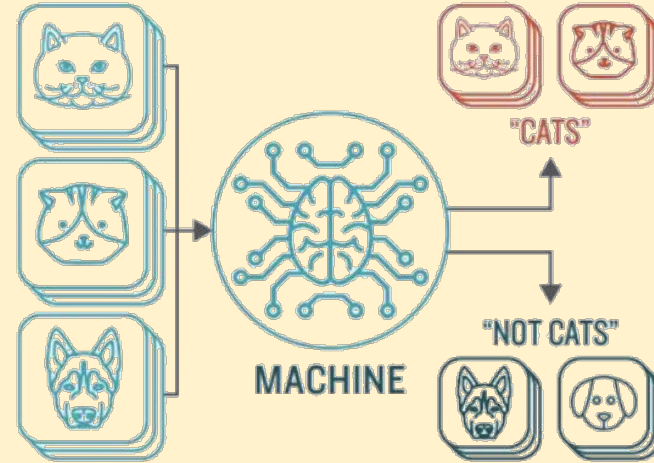
STEP 1

Provide the machine learning algorithm categorized or "labeled" input and output data from to learn

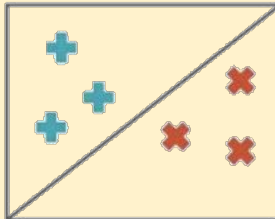


STEP 2

Feed the machine new, unlabeled information to see if it tags new data appropriately. If not, continue refining the algorithm

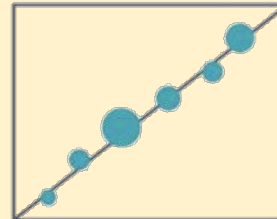


TYPES OF PROBLEMS TO WHICH IT'S SUITED



CLASSIFICATION

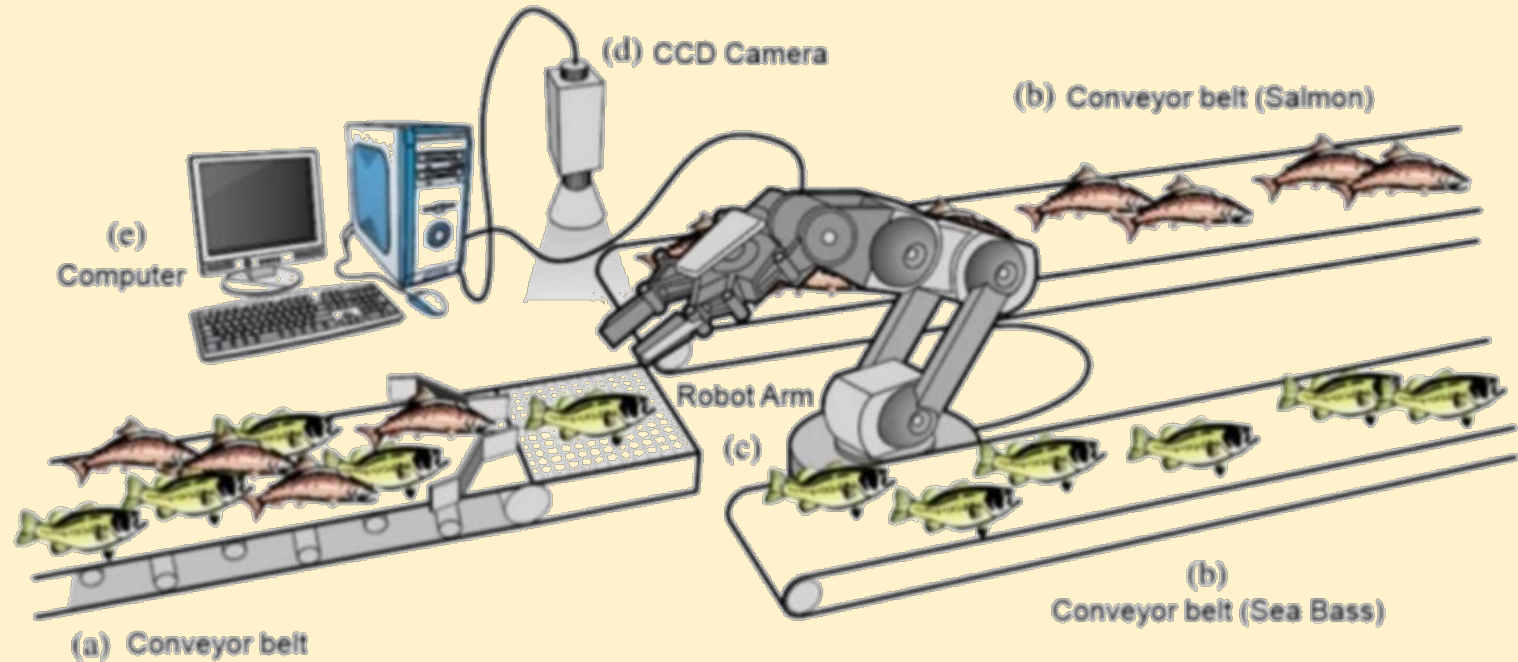
Sorting items into categories



REGRESSION

Identifying real values (dollars, weight, etc.)

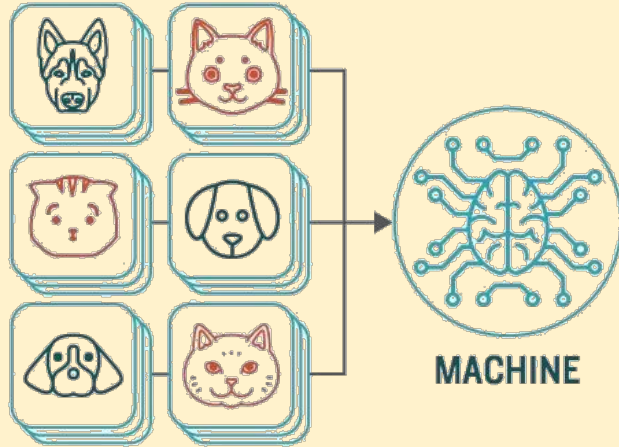
Features are extracted the CCD image (e.g. brightness, length, width) and a computer takes decisions that are executed by the robot.



How **Unsupervised** Machine Learning Works

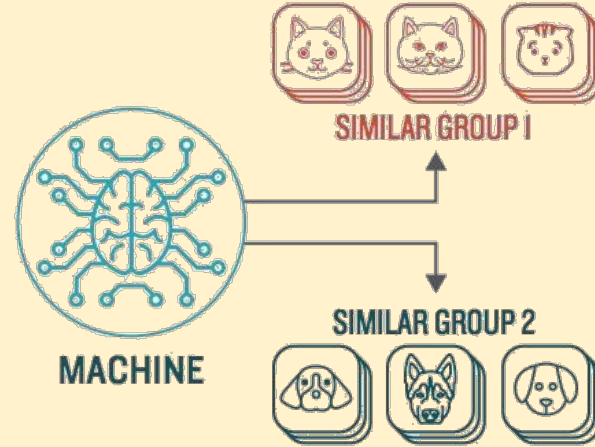
STEP 1

Provide the machine learning algorithm uncategorized, unlabeled input data to see what patterns it finds

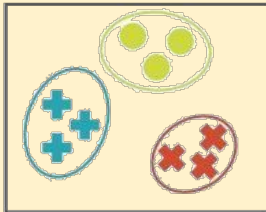


STEP 2

Observe and learn from the patterns the machine identifies



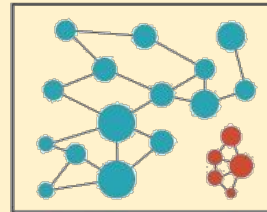
TYPES OF PROBLEMS TO WHICH IT'S SUITED



CLUSTERING

Identifying similarities in groups

For Example: Are there patterns in the data to indicate certain patients will respond better to this treatment than others?



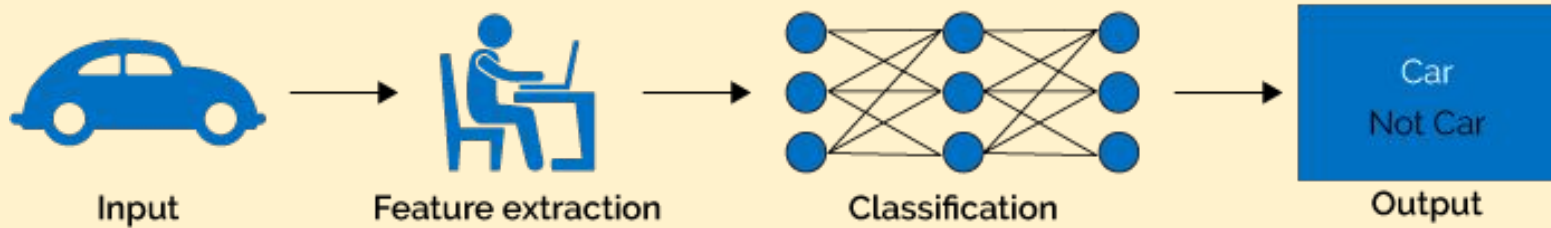
ANOMALY DETECTION

Identifying abnormalities in data

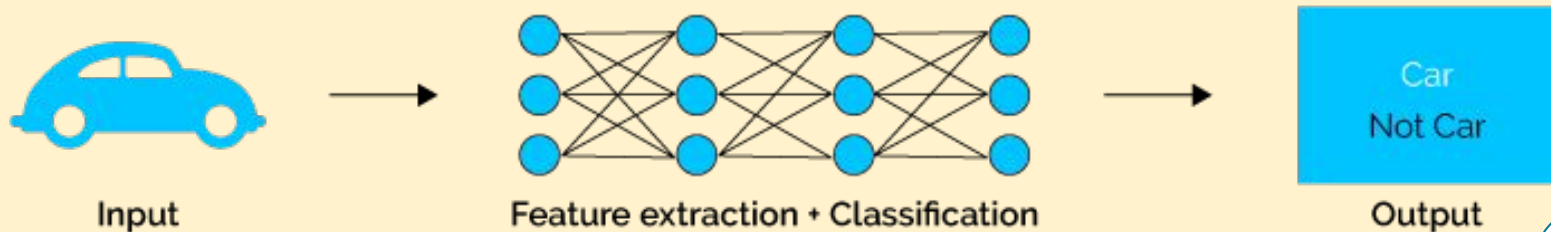
For Example: Is a hacker intruding in our network?

Deep Learning

Machine Learning



Deep Learning



El **Deep Learning** és una part del Machine Learning on el seu algoritme es basa a imitar el comportament de les Xarxes Neuronals humanes per poder efectuar Classificació, Regressió, Unsupervised i Reinforcement.

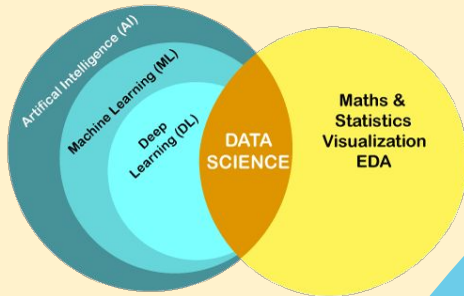
Aquest tipus de xarxes necessiten moltes dades i treballa molt bé en Computer Vision i NLP (Natural Language Processing)



Data Science

Data Science és un camp que estudia les dades i com extreure la informació d'aquestes.

Moltes vegades es dirà que Data Science no forma part de l'IA, però per aplicar-la bé, és necessari poder proporcionar al nostre model la informació de la millor manera possible. Així com un estudi i una neteja d'aquestes dades.



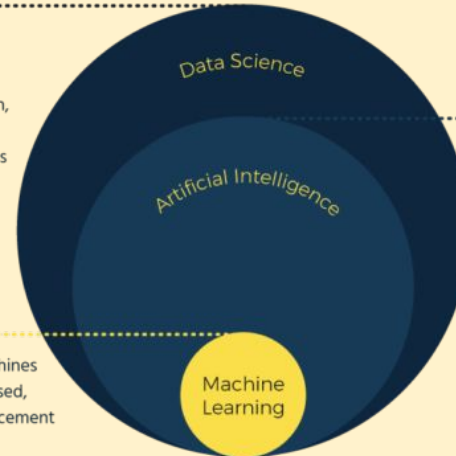
AI vs. Data Science vs. Machine Learning

Data Science

- Collection, preparation, and analysis of data
- Leverages AI/ML, research, industry expertise, and statistics to make business decisions

Machine Learning

- Algorithms that help machines improve through supervised, unsupervised, and reinforcement learning
- Subset of AI and Data Science tool



Artificial Intelligence

- Technology for machines to understand/interpret, learn, and make 'intelligent' decisions
- Includes Machine Learning among many other fields

1. Entorn de treball

Anaconda

[Link](#)



Jupyter Notebook

[Link](#)





Github

Per poder treballar amb repositoris, compartir codi i tenir una còpia de les teves tasques



SQL i BD

Necessitem conèixer com s'estructuren les bases de dades per poder seleccionar i gestionar les dades d'interès

2. Tipus de dades

Quan treballem en dades a IA, el que tenim són mostres, dades significatives d'alguna cosa en particular. Aquest conjunt de mostres formarà el nostre Data Set. Aquestes mostres poden venir de moltes formes diferents. A Python podem distingir entre aquest tipus de dades i més:

- **Text** -> Strings
- **Numèric** -> Integers, float, complex
- **Seqüències** -> List, tuple, range, array
- **Mapes/Objectes** -> Diccionaris python

Les nostres bases de dades poden contenir guardades les dades en aquests formats:

- **Numèriques** -> Boolean, Integer, Float, Nan...
- **Categòriques** -> Identifiers, Strings and Chars, Objects
- **Ordinal** -> Dades que poden ser mesurades entre elles (Notable, excel·lent, aprovat,...)
- **Imatges**
- **Text**



NumPy

Numpy

NumPy és una llibreria de Python especialitzada en el càlcul numèric i anàlisi de dades, especialment per a un gran volum de dades. Treballa molt més ràpid les dades que Python



pandas

Pandas

Pandas és una llibreria de Python especialitzada en el maneig i l'anàlisi d'estructures de dades. Defineix les estructures en NumPy. Conté molts **mètodes** que ens permeten treballar fàcilment

3. Anàlisi estadístic

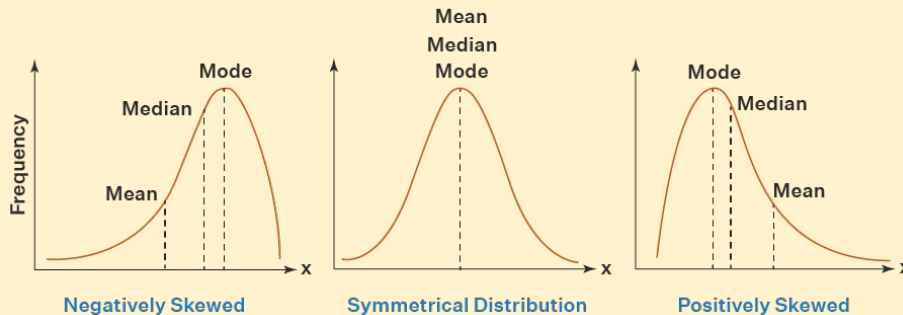
És necessari abans de posar-nos a aplicar l'intel·ligència artificial entendre les nostres dades. El nostre coneixement sobre aquestes pot influir molt a l'hora d'obtenir els millors resultats.

Imaginem que tenim unes dades numèriques i volem fer una anàlisi d'aquestes. Quines coses són necessàries entendre?

mean < median < mode

mean = median = mode

mean > median > mode



Median (Middle)

The number which is in the middle or the middle value.

11 7 11 18 9 7 6 23 7

6 7 7 7 9 11 11 18 23

Median: 9

Mode (Most)

The number that appears the most.

11 7 11 18 9 7 6 23 7

6 7 7 7 9 11 11 18 23

Mode: 7

Mean (Average)

The total of the numbers divided by how many numbers there are.

11 7 11 18 9 7 6 23 7

$11+7+11+18+9+7+6+23+7=99$

$99 / 9 = 11$

Mean: 11

Range (Difference)

The difference between the largest and the smallest number.

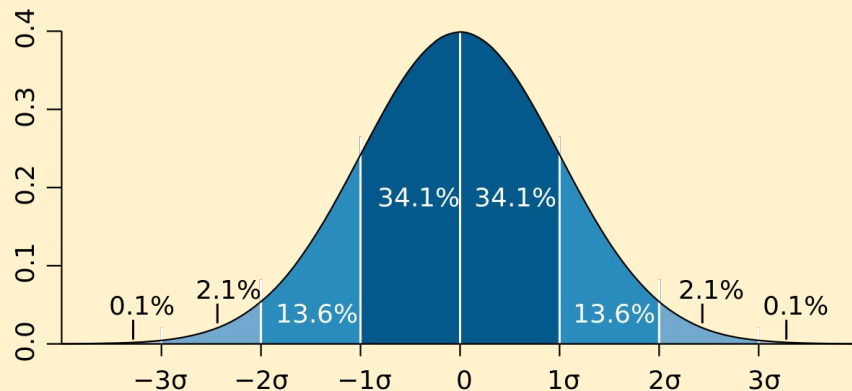
11 7 11 18 9 7 6 23 7

Large : 23 Small : 6

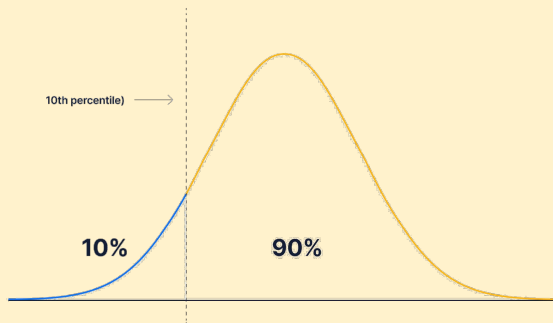
$23 - 6 = 17$

Range: 17

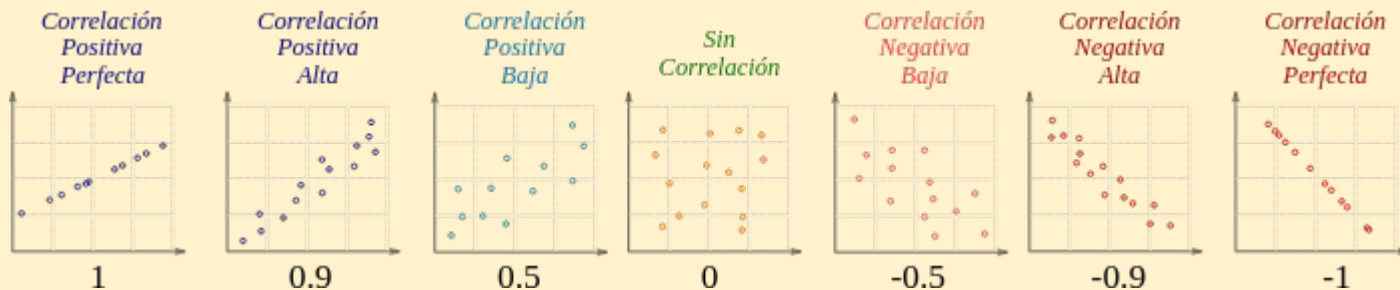
- **Desviació estàndard (std)** -> Número que descriu la distribució de les dades. Si és baix, vol dir que els valors estan a prop del valor mitjà, si és gran, vol dir que estan lluny de la mitjana.



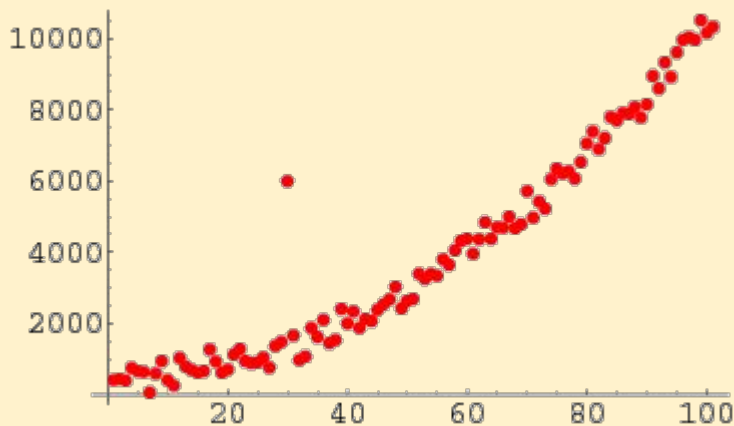
- **Percentil** -> Et dona el valor que descriu un percentatge donat de les mostres inferiors



- **Quantil** -> Divideix les dades en trossos de la mateixa probabilitat.
- **Correlació** -> Relació entre variables. Pot ser positiva o negativa



- **Outlier** -> Valor que s'allunya massa de la moda. Valor atípic.



Estadística descriptiva

Moltes vegades la representació visual de les dades ens dona informació clara sobre aquestes. Acompanyar les nostres dades amb gràfiques és clau per entendre-les.

Per visualitzar gràfiques a Python, podem utilitzar aquestes dues llibreries:

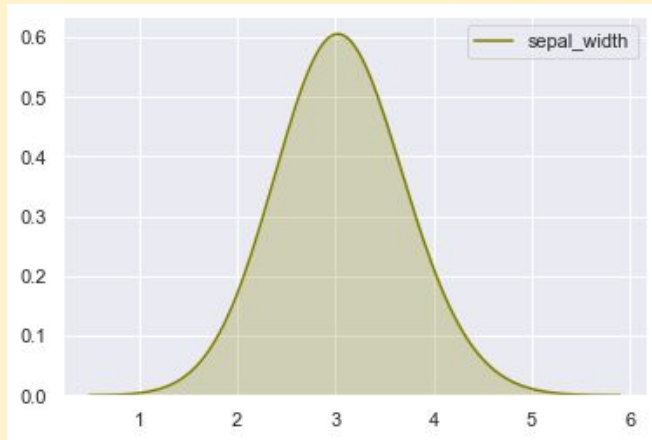
- [Seaborn](#)
- [Matplotlib](#)

Aquesta pàgina també ens dona un ventall de totes les gràfiques disponibles:

[Python Graph Gallery](#)

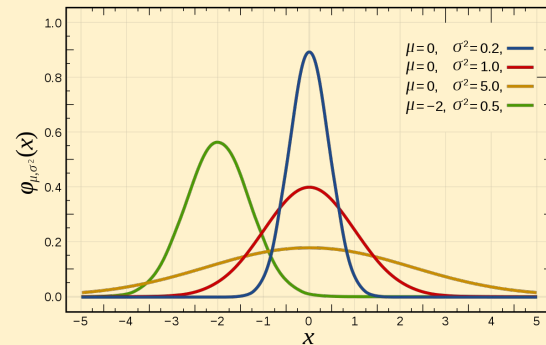
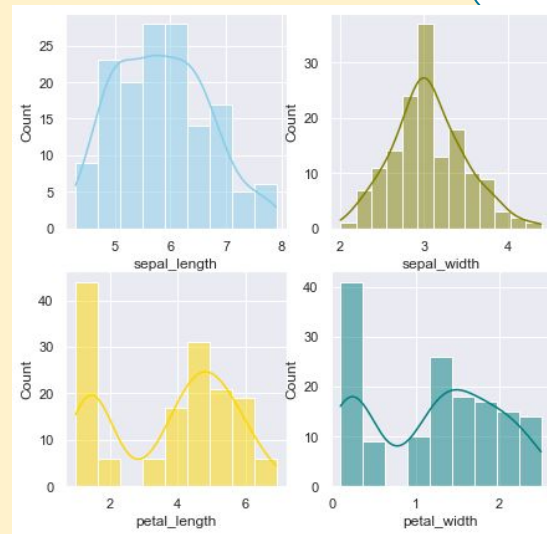
Les distribucions de les dades poden ser en forma:

- Continua: les dades tenen un nombre infinit de possibilitats (com per exemple la temperatura)
- Discreta: les dades tenen un nombre determinat de possibilitats (com per exemple el dia del teu naixement)



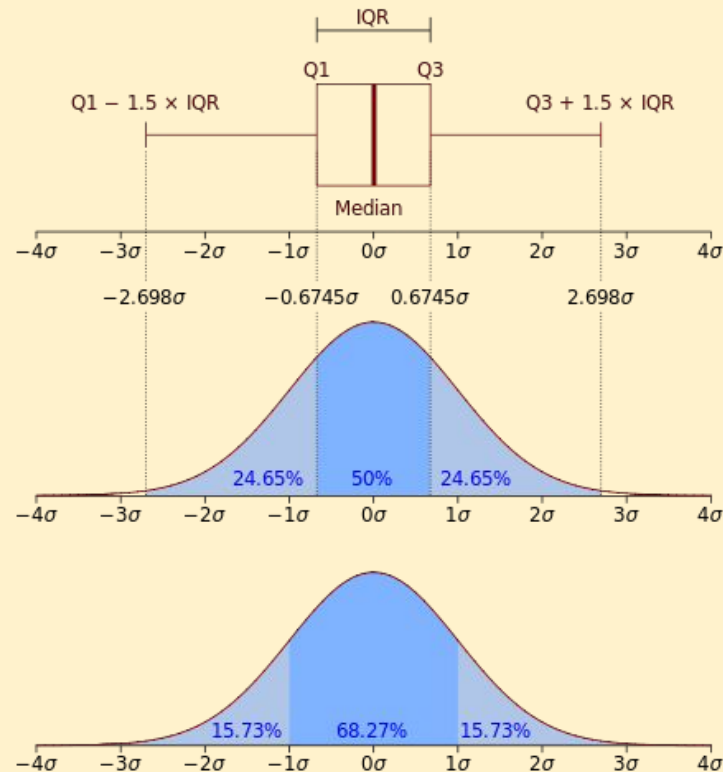
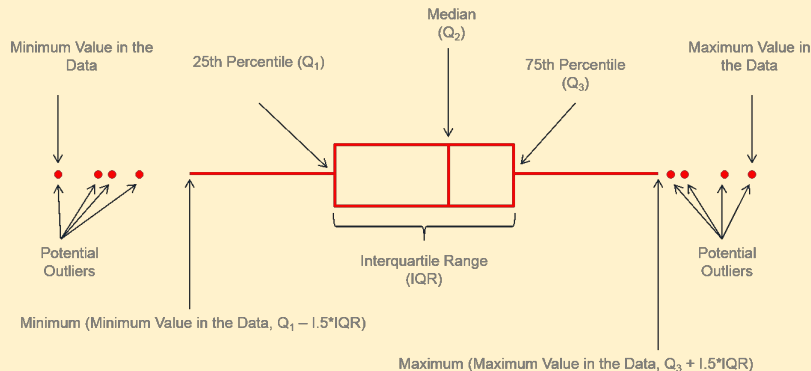
Anàlisi estadístic

- **Histograma o distribució** -> Representació de la mostra. En l'eix X, el valor. En l'eix Y, la freqüència d'aparició.
- **Normal Distribution** -> Distribució que sol aparèixer amb molta freqüència en dades reals. Forma acampanada. La mitjana i la mediana són molt properes.

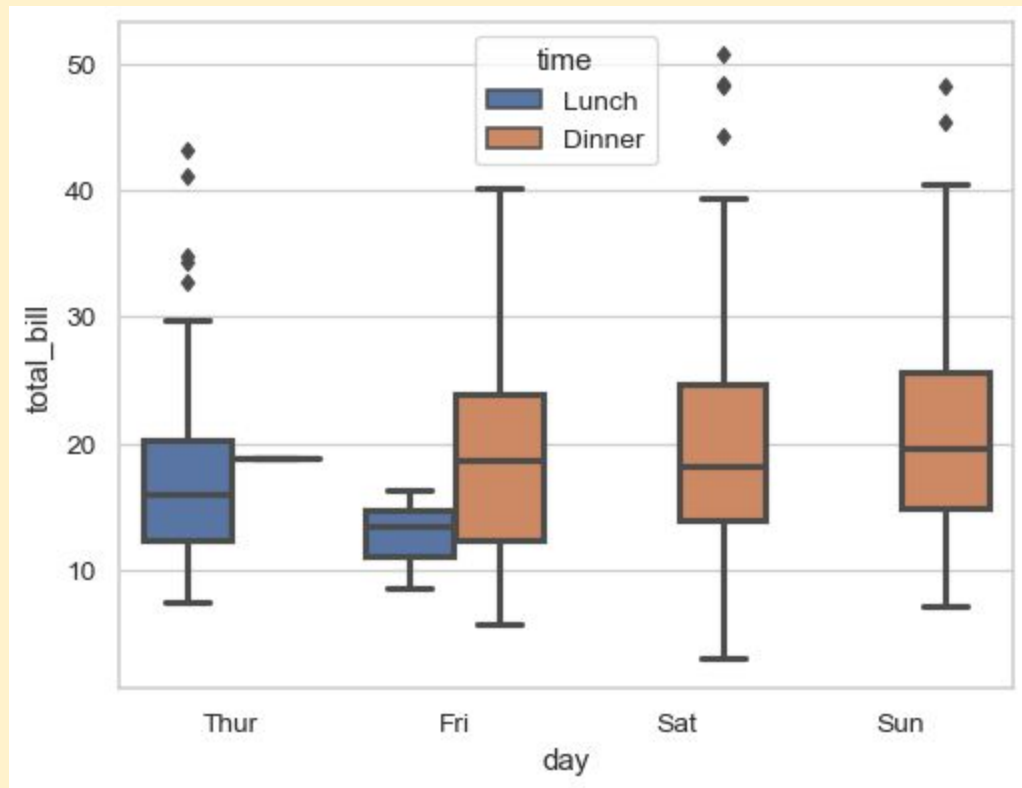


BoxPlot

Aquest tipus de representació de les dades ens dona un ràpid anàlisi de les dades mostrades. Ens dona informació sobre els quantils, la **media** i a més ens representa els outliers.



Anàlisi estadística



**Ja hem
acabat la
primera
part!**

Gràcies a tots!

