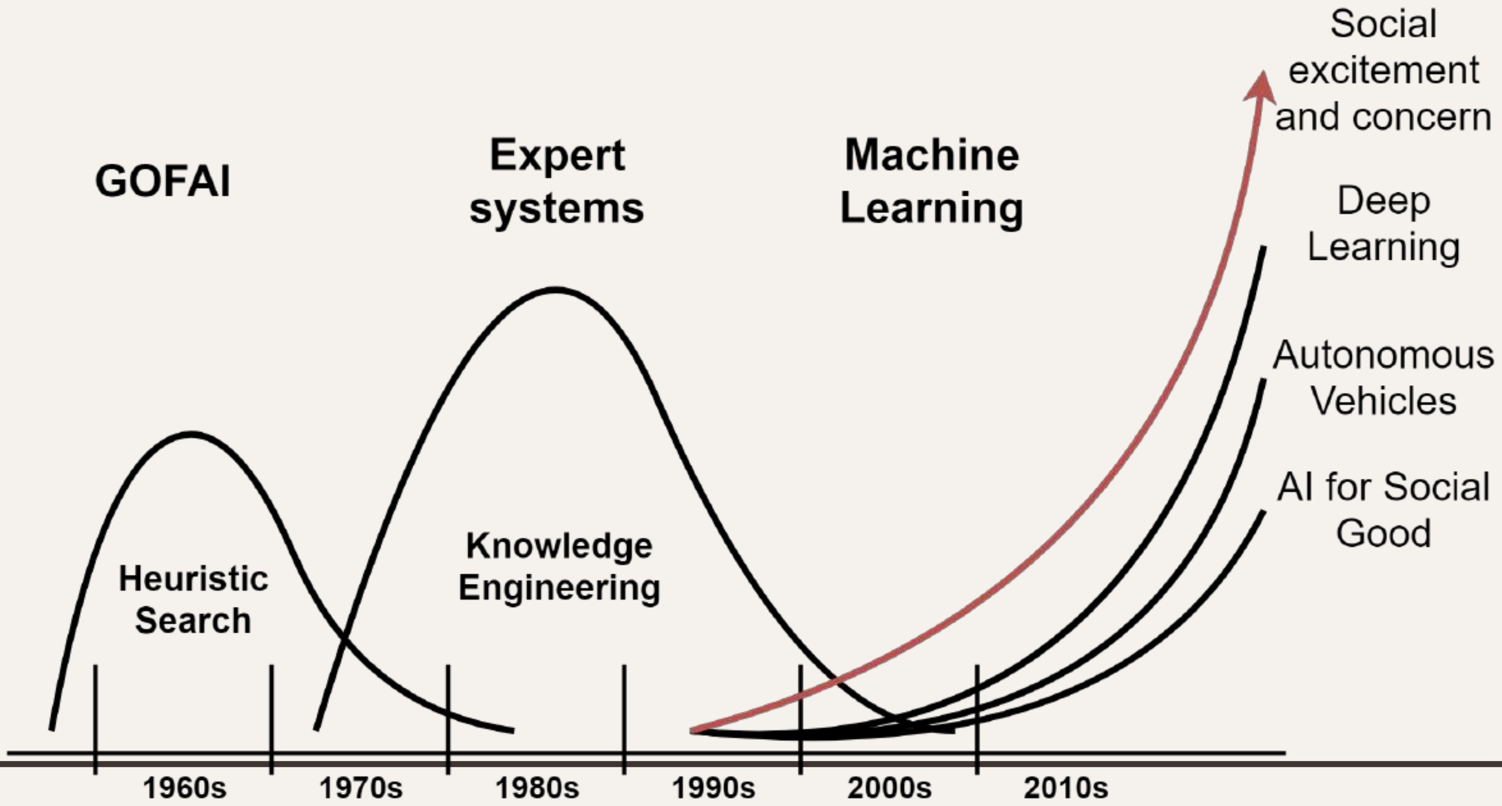
Machine Learning

Refresher

About Me

- Masters and PhD on Artificial Intelligence and Machine Learning
- Researcher at IT Aveiro
- Areas of interest: Artificial Intelligence, Machine Learning, text mining, stream mining, IoT, M2M

AI & ML

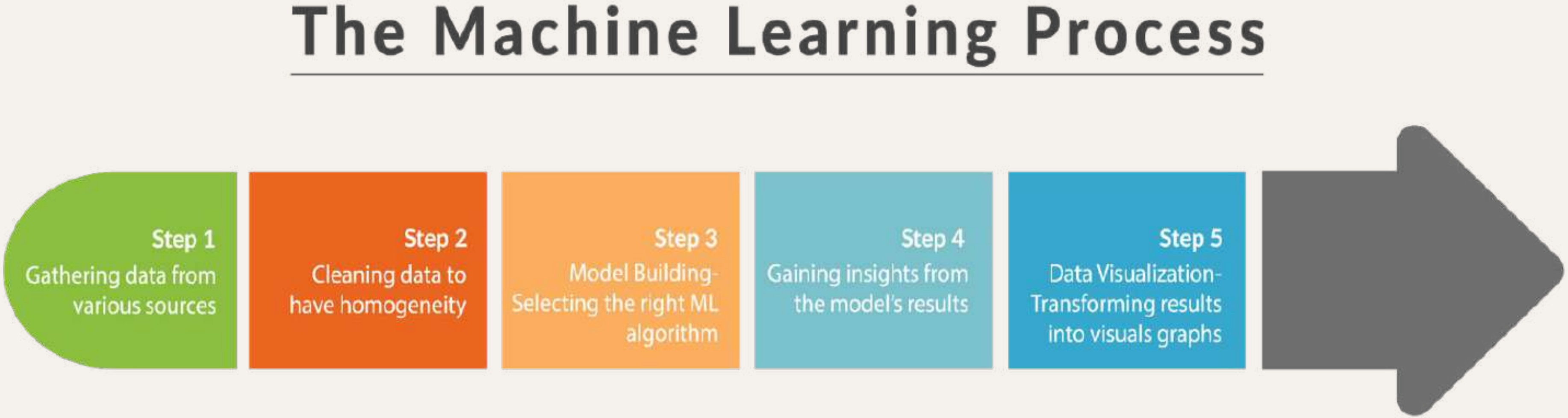

Social excitement and concern
GOFAI
Expert systems
Machine Learning
Deep Learning
Autonomous Vehicles
AI for Social Good
Heuristic Search
Knowledge Engineering
1960s
1970s
1980s
1990s
2000s
2010s

What is ML (Why should i Care)?

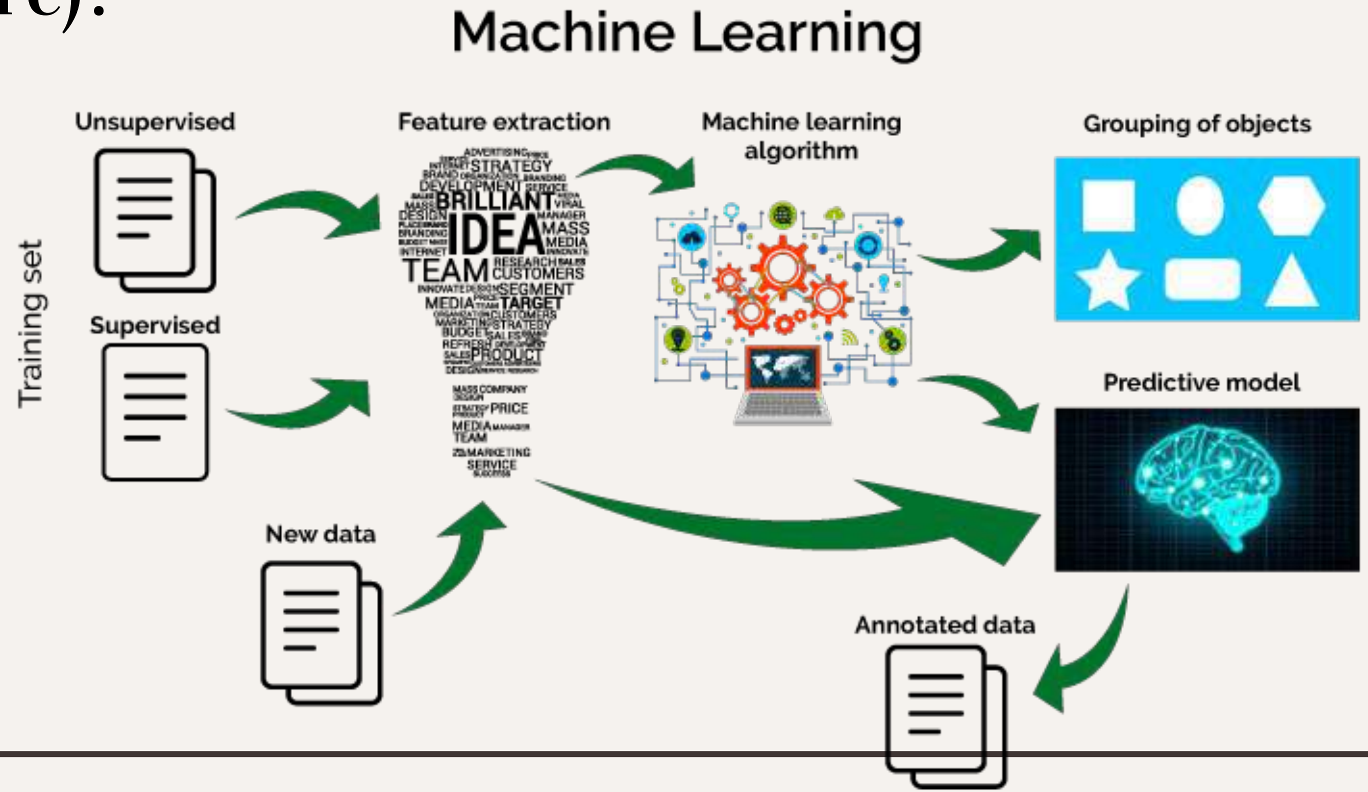
What does machine learning mean?

The term machine learning (abbreviated ML) refers to the capability of a machine to improve its own performance. It does so by using a statistical model to make decisions and incorporating the result of each new trial into that model. In essence, the machine is programmed to learn through trial and error.

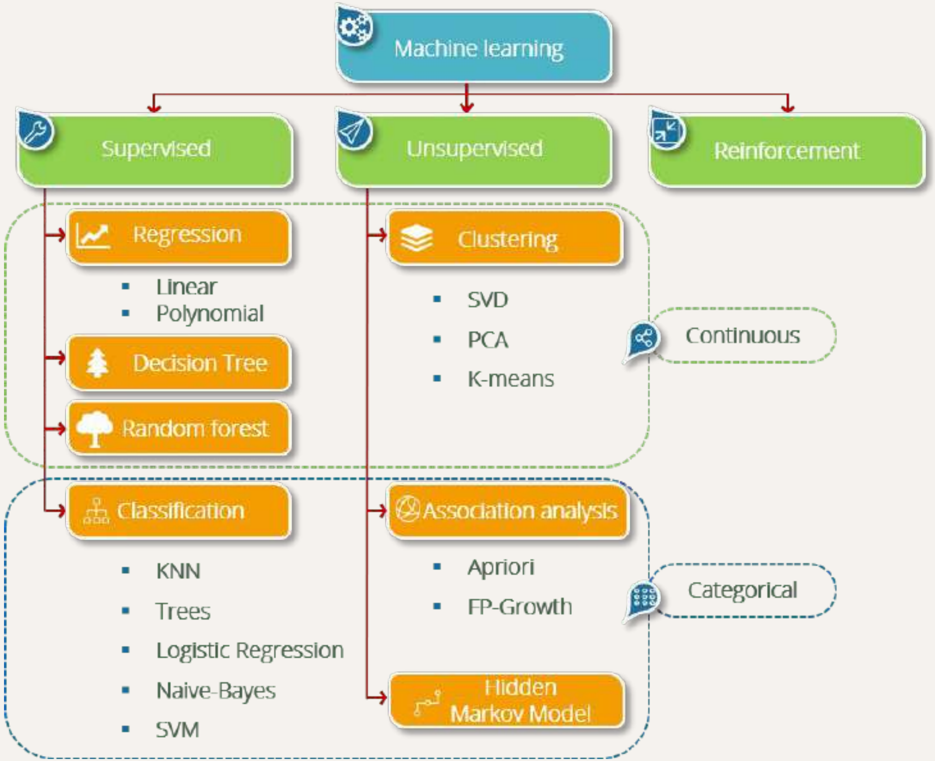
What is ML (Why should i Care)?



What is ML (Why should i Care)?



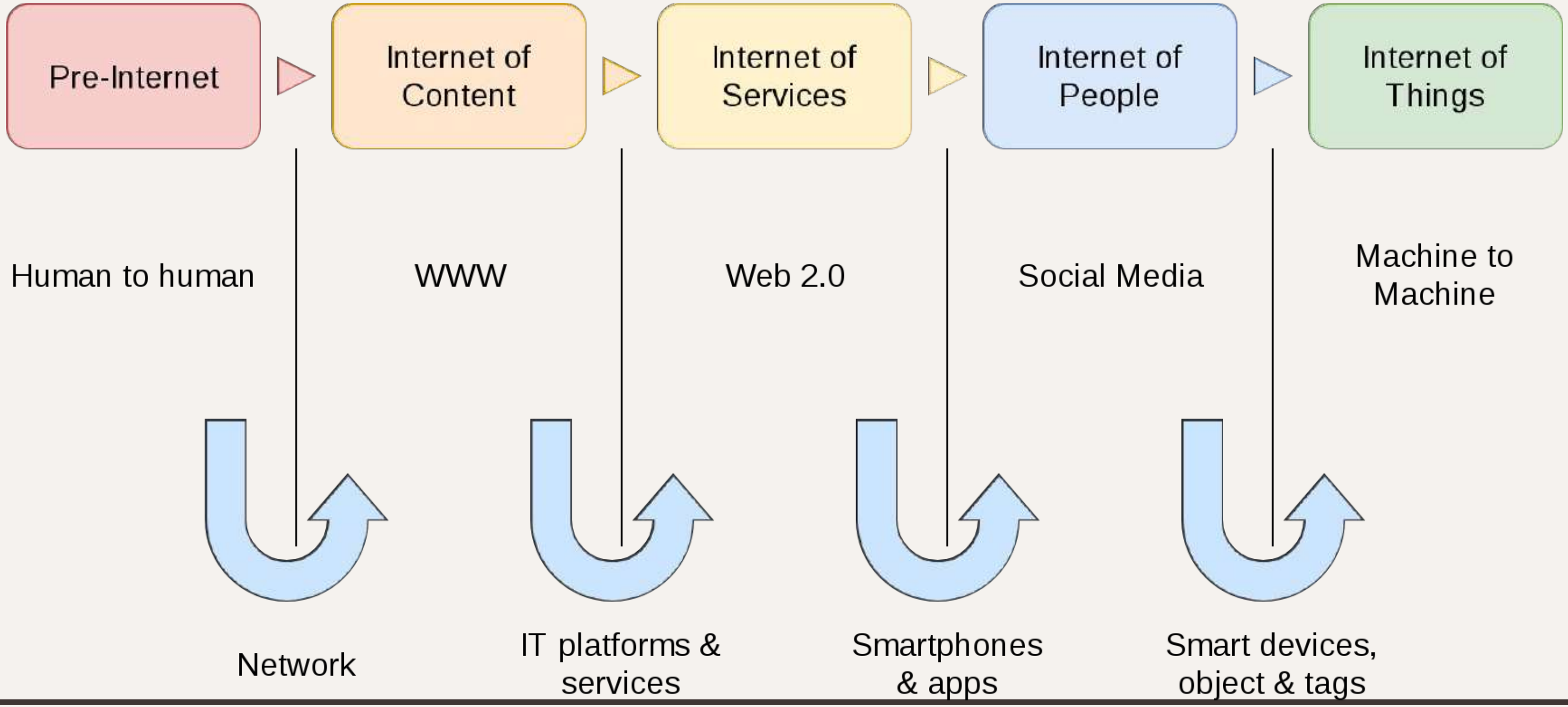
What is ML (Why should i Care)?



What is ML?

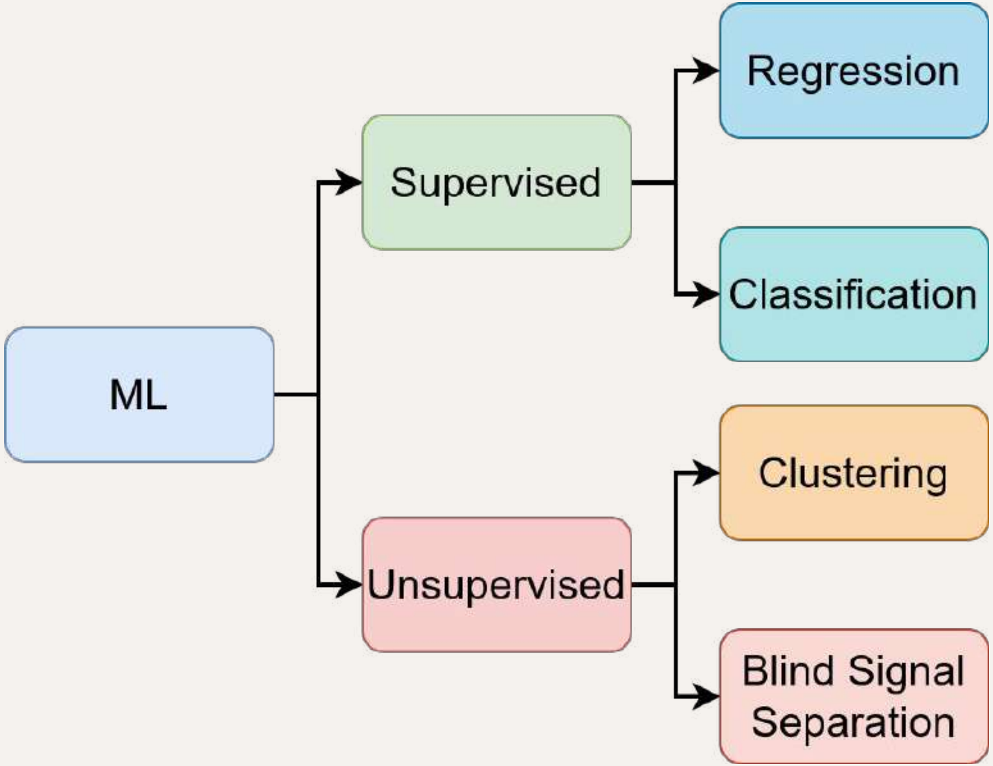
- A body of knowledge related with learning methods for machines (computers)
- Research area
- Opportunities for something useful

Why Should You Care?

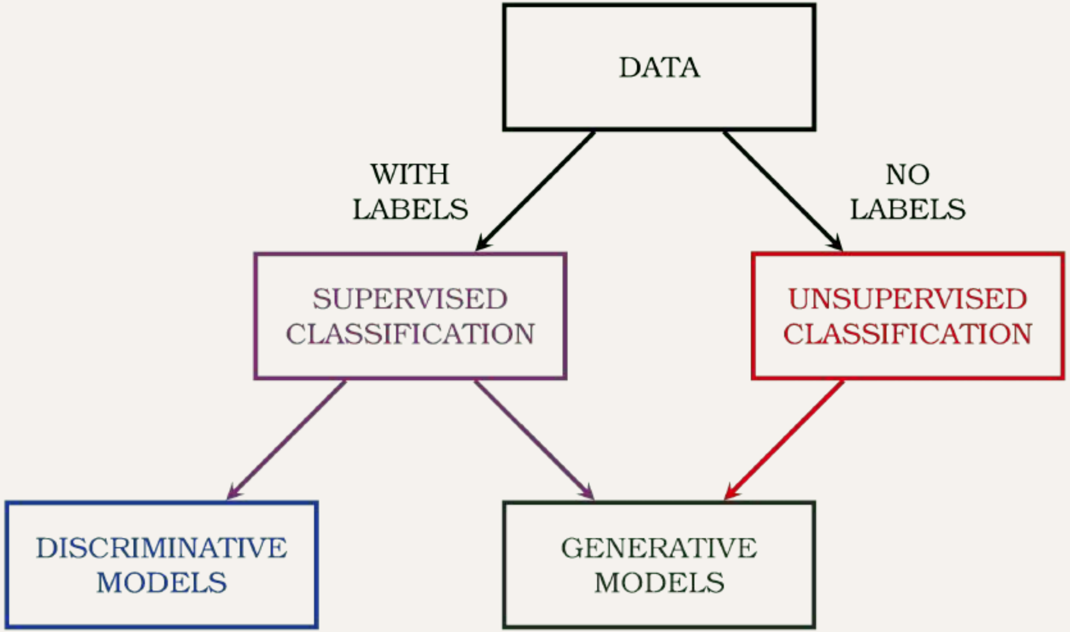


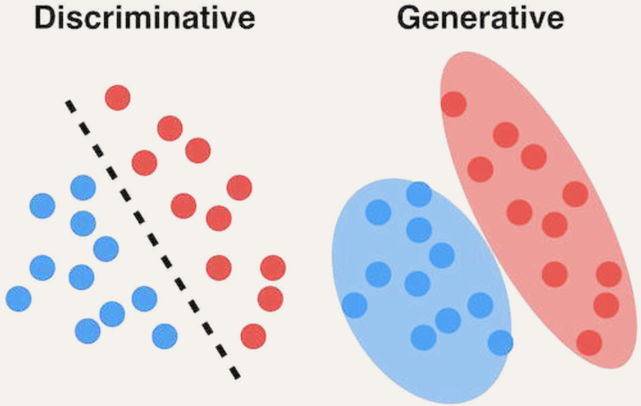
Taxonomy

Taxonomies...


ML
Supervised
Unsupervised
Regression
Classification
Clustering
Blind Signal Separation

Taxonomies...


DATA
WITH LABELS
NO LABELS
SUPERVISED CLASSIFICATION
UNSUPERVISED CLASSIFICATION
DISCRIMINATIVE MODELS
GENERATIVE MODELS


Discriminative
Generative

Taxonomies...

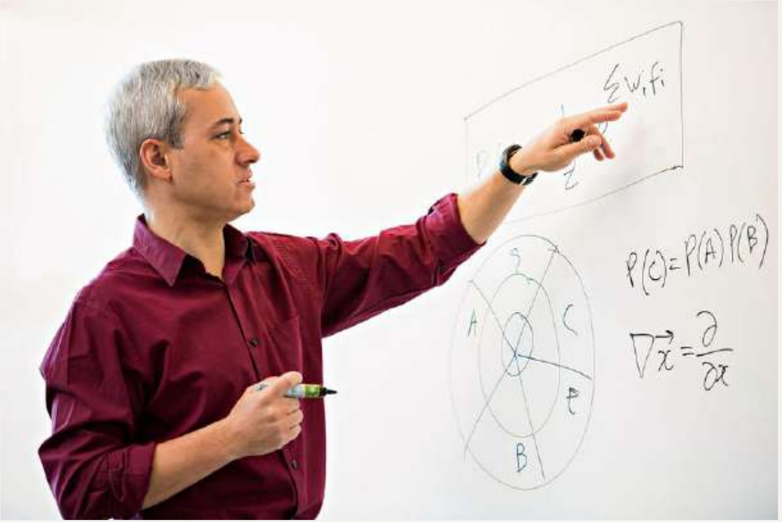
Induction symbolic reasoning

Neural Networks connections modelled on brain's neurons

Evolutionary algorithms learn from random generations (genetic algorithm)

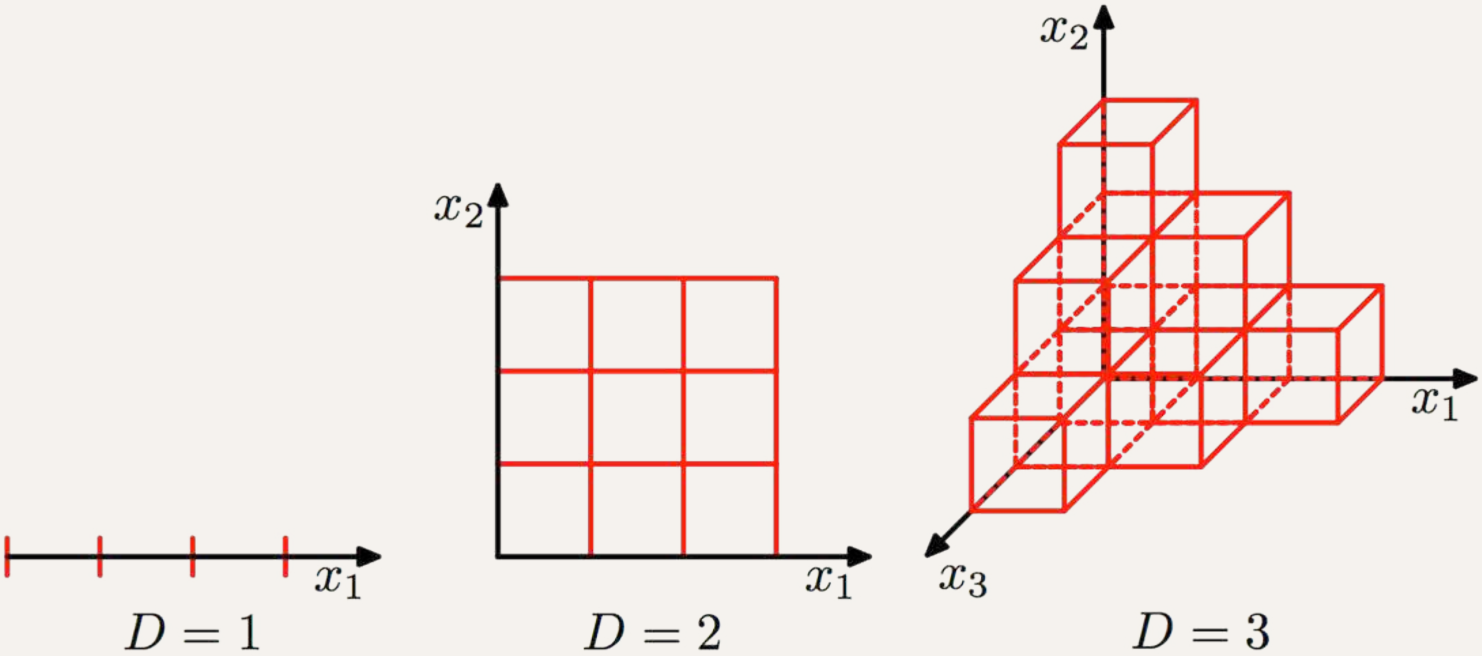
Bayesian inference probabilistic models based on bayes' theorem

Analogy learns by finding similar examples



Limitations

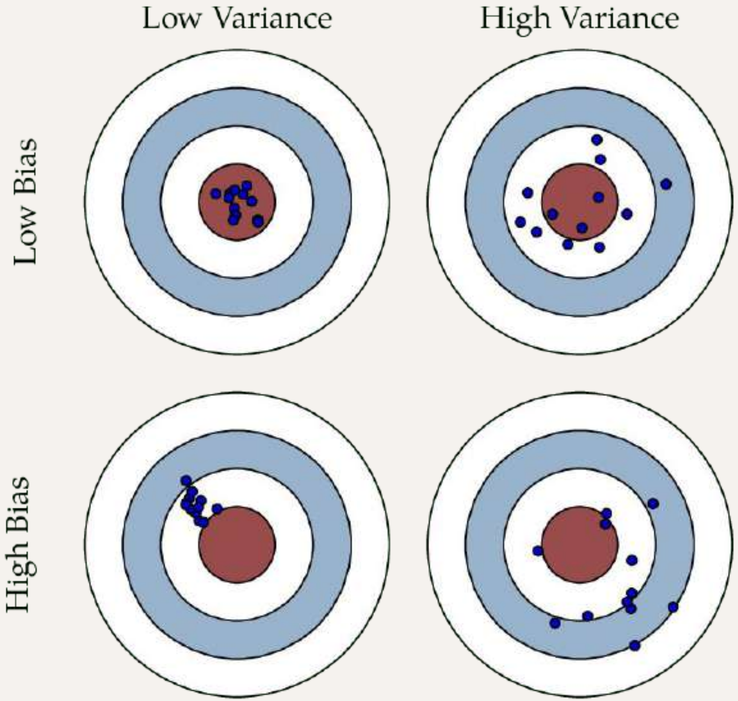
Limitations...



Limitations...

- Our model is a simplification of reality
- Simplification is based on assumptions (model bias)
- Assumptions fail in certain situations

Bias and Variance



Terminology

Terminology

Dataset: organized set of examples, typically composed of features and labels

Feature: single property of an example (input variable)

Label: classification category of an example (output variable)

Example: single instance of a dataset

Aprendizagem Aplicada à Segurança

Mário Antunes

September 22, 2023

University of Aveiro

Table of Contents

SPAM

- The term “spam” is internet slang that refers to unsolicited commercial email (UCE).
- The first reported case of spam occurred in 1898, when the New York Times reported unsolicited messages circulating in association with an old swindle.
- The term “spam” was coined in 1994, based on a now-legendary Monty Python’s Flying Circus sketch, where a crowd of Vikings sings progressively louder choruses of “SPAM! SPAM! SPAM!”

SPAM


1
1898
THE SPANISH PRISONER
The New York Times reports of unsolicited messages circulating in association with an old swindle.
2
Early 1900s
POST MAIL
Advertisement based on unsolicited content has been mailed to our doors by Post Mail services for over a century!
3
1978
ARPANET
The first reported case of email spam is attributed to Digital Equipment Corporation and circulated to 400 ARPANET users.
4
Mid 1990s
THE EMAIL EPIDEMIC
A growing fraction of emails is spam. Platforms and ISPs start investing in spam filtering techniques.
5
1995
SEARCH ENGINES
Web content spam and link farms are common forms of spamdexing, the manipulation of Web search result ranking.
6
2000s
SOCIAL NETWORKS
The rise of Facebook, Twitter, and Reddit leads to new opportunities for spammers to reach billions of Social Web users.
7
Early 2000s
FAKE REVIEWS
Giants of e-commerce like Amazon and Alibaba fight the manipulation of product popularity by opinion spam.
8
Mid 2000s
PHISHING
Social engineering and disguise may allow attackers to trick victims into revealing sensitive information. Ransomware are used to extort funds from the victims.
9
Early 2010s
SOCIAL BOTS
Millions of accounts operated by software populate social media to carry out nefarious spam campaigns.
10
Mid 2010s
FALSE NEWS
Spam Websites are created to deliberately propagate false news related to politics, public health, and social issues.
2018+
AI SPAM
Systems based on AI can manipulate reality, producing indistinguishable alternatives. AIs can also be target of manipulation and spam to elicit behaviors of the AI system or of its users.

SPAM

Dear Sir,
I am prince [illegible] from Nigeria. Your help would be very appreciated.
I want to transfer all of my fortune outside if Nigeria due to a frozen account,
If you could be so kind and transfer small sum of 3 500 USD to my account,
I would be able to unfreeze my account and transfer my money outside of
Nigeria. To repay your kindness, I will send 1 000 000 USD to your account.

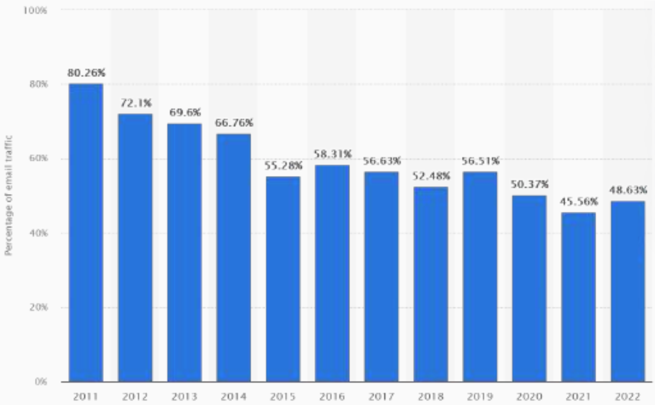
Please contact me to proceed

Prince [illegible]

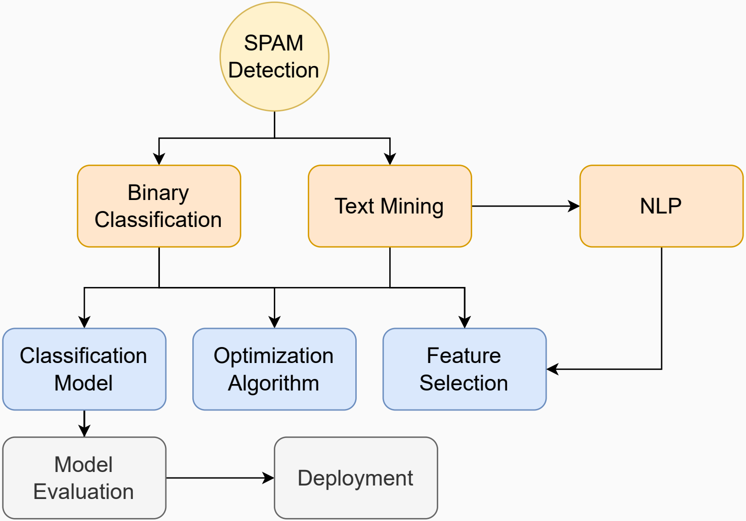
Fight against SPAM

- *Huge* list of `https://en.wikipedia.org/wiki/Anti-spam_techniques`
- From common sense to *Bayesian spam filtering*
- Unfortunately it is a costly battle

Fight against SPAM

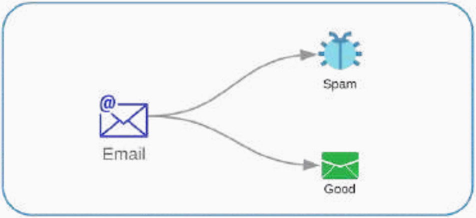


SPAM Detection



Binary Classification

- Binary classification is the task of classifying the elements of a set into two groups (each called class) on the basis of a classification rule.
- For this application one message can either be spam or ham.



Binary classification

- Text mining is the process of deriving high-quality information from text.
- Combines concepts from Machine Learning, Linguistic and statistical analysis.
- In this area we will explore the methods used to rank words/tokens and the BoW model.

Bag of Words (Bow) model

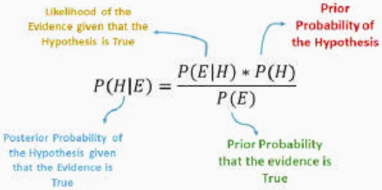
	the	red	dog	cat	eats	food
1. the red dog	1	1	1	0	0	0
2. cat eats dog	0	0	1	1	1	0
3. dog eats food	0	0	1	0	1	1
4. red cat eats	0	1	0	1	1	0

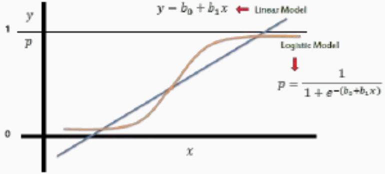
Natural Language Processing (NLP)

- NLP gives the computers the ability to understand text.
- Combines *Sintax* and *Semantic* into the analysis.
- One famous exemples are the Large Language Models (LLMs) that power OpenAI Chat GPT.

Classification Model

- SPAM detection is “considered” a toy example.
- As such, we will explore two of the simples learning models: Naive Bayes and Logistic Regression.



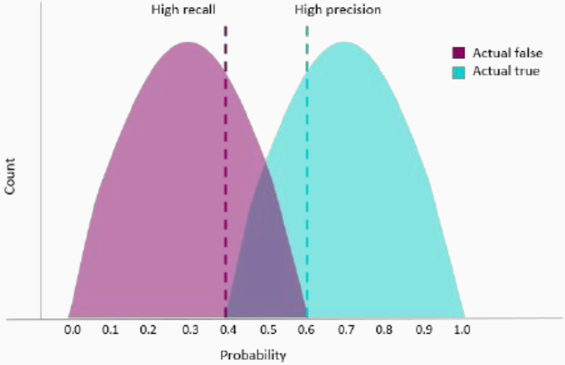


Model Evaluation

- Classification model can be evaluated using a confusing matrix
- The simplest methods to evaluate a model is throuhgh accuracy: $acc = \frac{TP+TN}{TN+TN+FP+FN}$

	Predicted Positive	Predicted Negative	
Actual Positive	**TP** *True Positive*	**FN** *False Negative*	Sensitivity $\frac{TP}{(TP+FN)}$
Actual Negative	**FP** *False Positive*	**TN** *True Negative*	Specificity $\frac{TN}{(TN+FP)}$
	Precision $\frac{TP}{(TP+FP)}$	Negative Predictive Value $\frac{TN}{(TN+FN)}$	Accuracy $\frac{TP+TN}{(TP+TN+FP+FN)}$

Model Evaluation



Aprendizagem Aplicada à Segurança

Mário Antunes

October 14, 2023

Universidade de Aveiro

Context

It is becoming difficult to identify Cybersecurity attacks. These attacks can originate internally due to malicious intent or negligent actions or externally by malware, target attacks, and APT (Advanced Persistent Threats).

But insider threats are more challenging and can cause more damage than external threats because they have already entered the network.

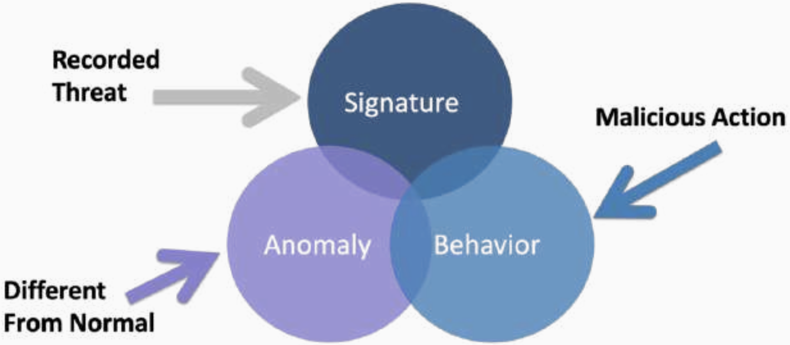
These activities present unknown threats and can steal, destroy or alter the assets.

Context

Earlier firewalls, web gateways, and some other intrusion prevention tools are enough to be secure, but now hackers and cyber attackers can bypass approximately all these defense systems.

Therefore with making these prevention systems strong, it is also equally essential to use detection. So that if hackers get into the network, the system should be able to detect their presence.

Context



Recorded
Threat
Signature
Malicious Action
Anomaly
Behavior
Different
From Normal

Context

Signature detection requires knowing what to look for and comparing hashes or other strings to identify a match. Signature detection is a common feature found within antivirus and IPS/IDS products.

Behavior detection looks for malicious or other known behavior characteristics and alarms the SOC when a match is made. An example is identifying port scanning or a file attempting to encrypt your hard drive, which is an indication of ransomware behavior. Antimalware and sandboxes are examples of tools that heavily leverage behavior detection capabilities.

Anomay detection it takes into consideration hot topics including big data, threat intelligence, and "zero-day" detection.

Anomaly Detection

Anomaly detection, also called outlier detection, is the identification of unexpected events, observations, or items that differ significantly from the norm:

- Anomalies in data occur only very rarely
- The features of data anomalies are significantly different from those of normal instances

What is an anomaly?

Generally speaking, an **anomaly** is something that differs from a norm: a deviation, an exception. In software engineering, by anomaly we understand a rare occurrence or event that doesn't fit into the pattern, and, therefore, seems suspicious. Some examples are:

- sudden burst or decrease in activity;
- error in the text logs;
- sudden rapid drop or increase in temperature.

Common reasons for outliers are:

- data preprocessing errors;
- noise;
- fraud;
- attacks.

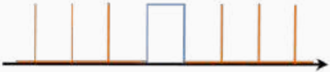
Types of Anomalies

Anomalies can be broadly categorized as:

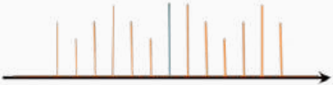
- Point anomalies: A single instance of data is anomalous if it's too far off from the rest.
- Contextual anomalies: The abnormality is context specific. This type of anomaly is common in time-series data.
- Collective anomalies: A set of data instances collectively helps in detecting anomalies.

Types of Anomalies

(a) Point Anomaly



(b) Collective Anomaly



(c) Contextual Anomaly

Anomaly Detection - Example #1

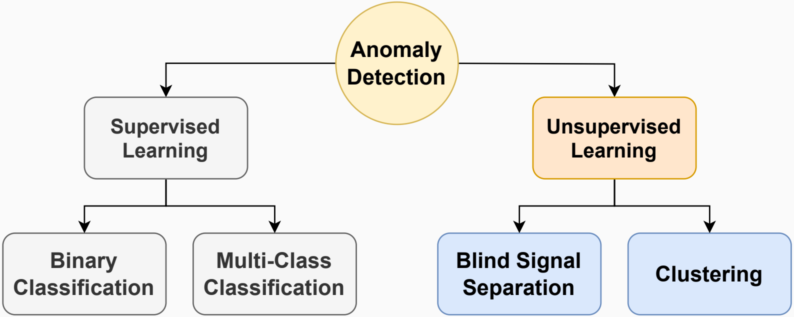
Network anomalies: Anomalies in network behavior deviate from what is normal, standard, or expected. To detect network anomalies, network owners must have a concept of expected or normal behavior. Detection of anomalies in network behavior demands the continuous monitoring of a network for unexpected trends or events.

Anomaly Detection - Example #2

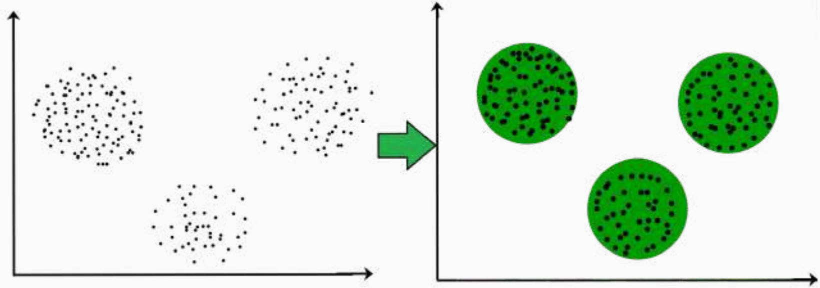
Application performance anomalies: These are simply anomalies detected by end-to-end application performance monitoring. These systems observe application function, collecting data on all problems, including supporting infrastructure and app dependencies. When anomalies are detected, rate limiting is triggered and admins are notified about the source of the issue with the problematic data.

Web application security anomalies: These include any other anomalous or suspicious web application behavior that might impact security such as CSS attacks or DDOS attacks.

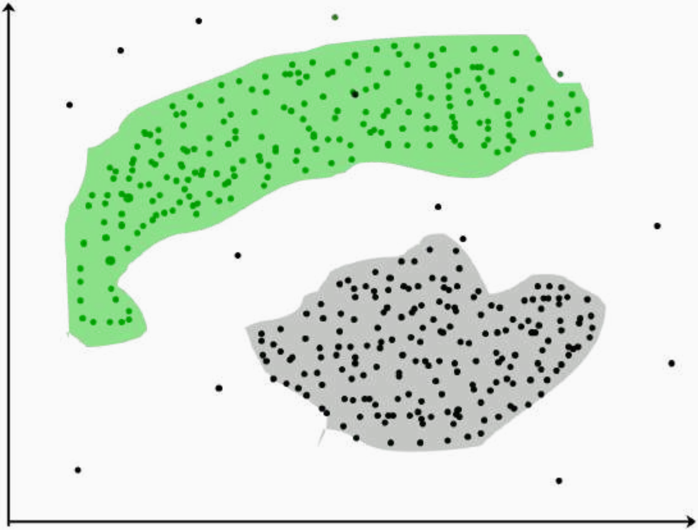
Anomaly Detection



Type of **unsupervised learning method**. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.



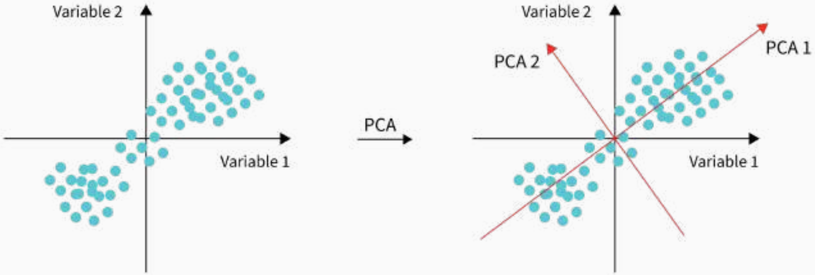
- **Density-Based Methods:** These methods consider the clusters as the dense region having some similarities and differences from the lower dense region of the space. These methods have good accuracy and the ability to merge two clusters.
- **Hierarchical Based Methods:** The clusters formed in this method form a tree-type structure based on the hierarchy. New clusters are formed using the previously formed one.
- **Partitioning Methods:** These methods partition the objects into k clusters and each partition forms one cluster. This method is used to optimize an objective criterion similarity function such as when the distance is a major parameter.



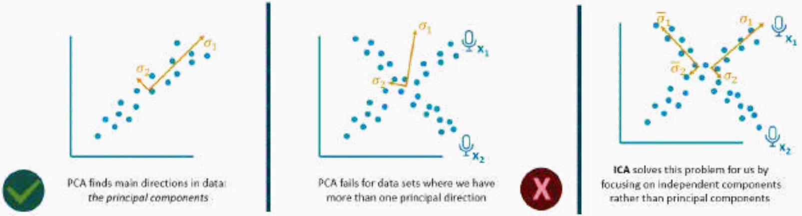
Blind Source Separation (BSS) refers to a problem where both the sources and the mixing methodology are unknown, only mixture signals are available for further separation process.

In several situations it is desirable to recover all individual sources from the mixed signal, or at least to segregate a particular source.

** Principal component analysis**, or PCA, is a statistical procedure that allows you to summarize the information content in large data tables by means of a smaller set of “summary indices” that can be more easily visualized and analyzed.



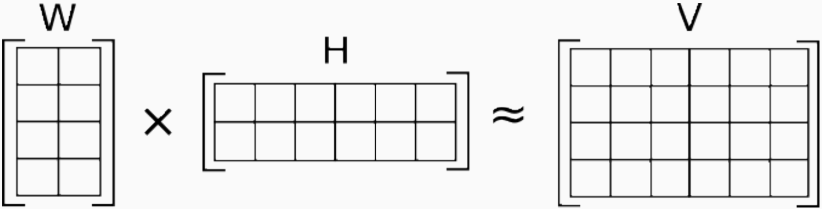
Independent Component Analysis (ICA) is a powerful technique in the field of data analysis that allows you to separate and identify the underlying independent sources in a multivariate data set.



Non-negative matrix factorization (NNMF) is a group of algorithms in multivariate analysis and linear algebra where a matrix V is factorized into two matrices W and H, with the property that all three matrices have no negative elements.

This non-negativity makes the resulting matrices easier to inspect. Also, in applications such as processing of audio spectrograms or muscular activity, non-negativity is inherent to the data being considered.

Since the problem is not exactly solvable in general, it is commonly approximated numerically.



Blind Source Separation: Anomaly Detection