

Week 1

Course. Introduction to Machine Learning

Theory 1. Introduction to ML

Dr. Maria Salamó Llorente

maria.salamo@ub.edu

Dept. Mathematics and Informatics,
Faculty of Mathematics and Informatics,
University of Barcelona (UB)



1. Introduction
2. Definitions
3. Basis of Machine Learning
4. Applications
5. History



1. Introduction

Overview

The word cloud illustrates the following concepts:

- Machine Learning**: The central theme, represented by large words like "machine", "algorithms", "learning", and "data".
- Data**: A significant component, with words like "data", "graphs", "complex", and "objects".
- Statistical Methods**: Represented by words such as "statistical", "classification", "clustering", "regression", and "inference".
- Theoretical Foundations**: Words like "theory", "behavior", "results", and "guarantees" are present.
- Computational Techniques**: Words like "algorithm", "complex", "patterns", "kernels", "online", and "sensor".
- Machine Components**: Words like "computer", "software", "hardware", and "network".
- Performance Metrics**: Words like "accuracy", "precision", "recall", and "f1 score".
- Modeling and Simulation**: Words like "model", "simulate", "forecast", and "predict".
- Optimization**: Words like "optimize", "minimize", and "maximize".
- Uncertainty and Risk**: Words like "risk", "uncertainty", and "robustness".
- Interpretability**: Words like "interpret", "explain", and "understand".
- Ethics and Bias**: Words like "bias", "fairness", and "equity".



Look at the video:

https://youtu.be/eWOf33_39Y

Why Study Machine Learning?

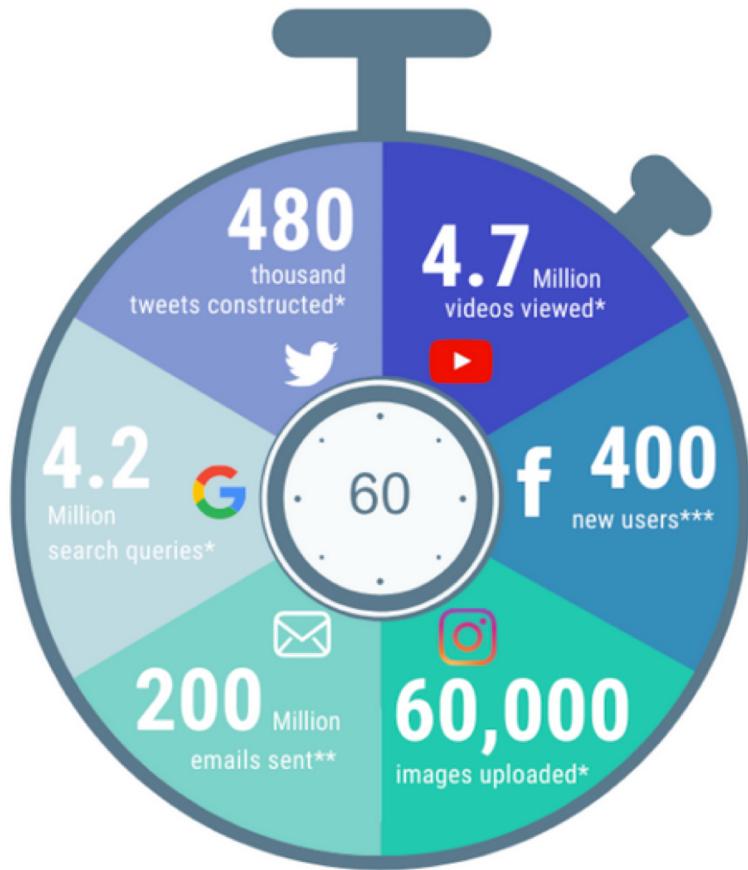
- Develop systems that are too difficult or expensive to construct manually because they require specific detailed skills or knowledge tuned to a specific task (***knowledge engineering bottleneck***)
- Develop systems that can automatically adapt and customize themselves to individual users (***personalization***)
 - Personalized news or mail filter
 - Personalized tutoring systems
- Discover new knowledge from large databases (***data mining***)
 - Market basket analysis (e.g. diapers and beer)
 - Medical text mining (e.g. migraines to calcium channel blockers to magnesium)

Why Study Machine Learning?

The Time is Ripe

- Many basic effective and **efficient algorithms** available
- Large amounts of ***on-line data*** available
- Large amounts of **computational resources** available

WHAT HAPPENS ONLINE
EVERY MINUTE?



The learning machine

Arthur Samuel (1959): *Machine Learning: Field of study that gives the computer the ability to learn without being explicitly programmed*



Look at the video:

<https://youtu.be/6tzt64XKNyQ>





2. Definitions

What is Learning?

- Webster's definition of “**to learn**”
“To gain **knowledge** or **understanding** of, or **skill** in **by study, instruction or experience**”
 - Learning a set of new facts
 - Learning HOW to do something
 - Improving ability of something already learned
- Herbert Simon: “Learning is any process by which a system improves performance from experience”

Machine Learning Definition by Tom Mitchell (1998)

- Study of algorithms that
 - at some task T
 - improve their performance P
 - based on experience E

well-defined learning task: $\langle T, P, E \rangle$



Look at the video: From minute 0:55 to minute 7:53

<https://youtu.be/m4NlfvrRCdg?list=PLI-BBnDxtUt1hLXmlwu27P22bTi6VwMkN>

Examples

T: Playing checkers

P: Percentage of games won against an arbitrary opponent

E: Playing practice games against itself

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

Exercise

“A computer program is said to *learn from experience E with respect to some task T and some performance measure P*, if its performance on T, as measured by P, improves with experience E.”

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is T, P, E in these sentences?

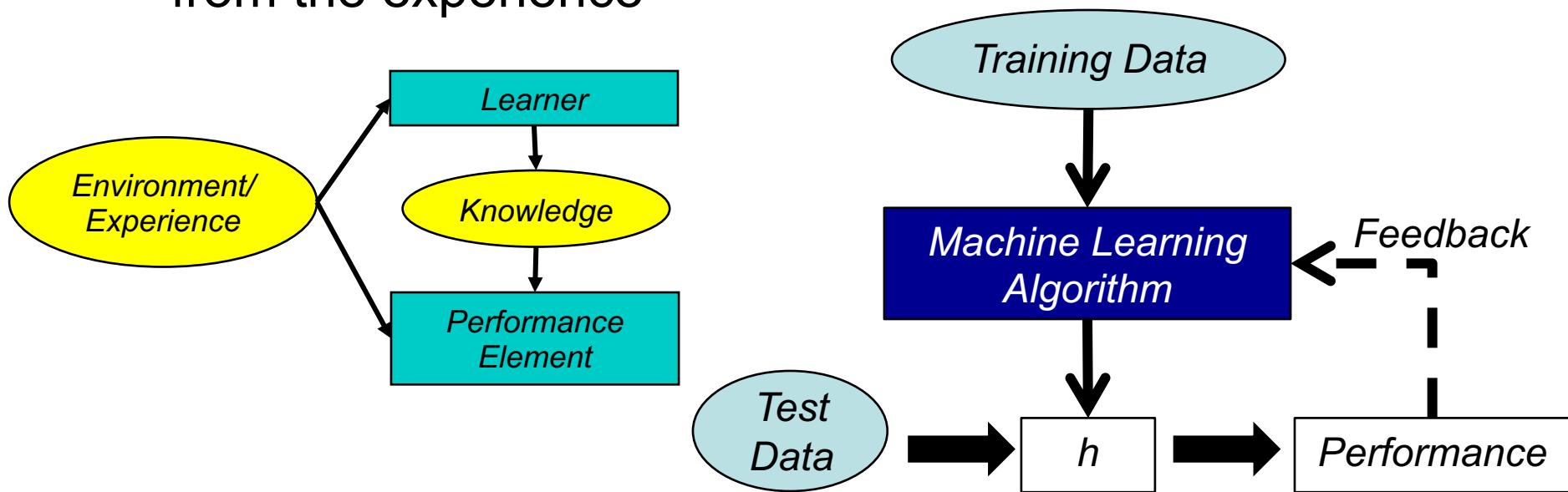
- [] Watching you label emails as spam or not spam
- [] The number (or fraction) of emails correctly classified as spam/not spam
- [] Classifying emails as spam or not spam



3. Basis of Machine Learning

Designing a Learning System

- STEPS:
 - Choose the training experience
 - Choose exactly what is to be learned, i.e. the **target function**.
 - Choose how to represent the target function
 - Choose a learning algorithm to infer the target function from the experience





- What is the **task**?
 - Classification
 - Regression
 - Problem solving / planning / control
- How to evaluate **performance**?
 - Classification accuracy (or error)
 - Solution correctness
 - Solution quality (length, efficiency)
 - Speed of performance
- How to represent **experience**?
 - Neuron, case, tree, etc.

Classification tasks

- Assign object/event to one of a given finite set of categories.
 - Medical diagnosis
 - Credit card applications or transactions
 - Fraud detection in e-commerce
 - Worm detection in network packets
 - Spam filtering in email
 - Recommend books, movies, music, or jokes
 - Financial investments
 - DNA sequences
 - Spoken words
 - Handwritten letters
 - Astronomical images

- Performing actions in an environment in order to achieve a goal.
 - Solving calculus problems
 - Playing checkers, chess, or backgammon
 - Balancing a pole
 - Driving a car or a jeep
 - Flying a plane, helicopter, or rocket
 - Controlling an elevator
 - Controlling a character in a video game
 - Controlling a mobile robot

- **Models of intelligence**
 - Cognitive Psychology: The process of human learning
 - Neurobiology: The brain, the neuron
- **Knowledge/concepts**
 - Cognitive Psychology: What is a concept? How to represent them? Is there an explanation about how we represent them?
 - Mathematical Logic: How concepts can be combined and manipulated?
 - Statistics: What mathematical model represents them?
 - Information theory: How can we code them?

Related Disciplines

- Artificial Intelligence
- Data Mining
- Probability and Statistics
- Information theory
- Numerical optimization
- Computational complexity theory
- Control theory (adaptive)
- Psychology (developmental, cognitive)
- Neurobiology
- Linguistics
- Philosophy

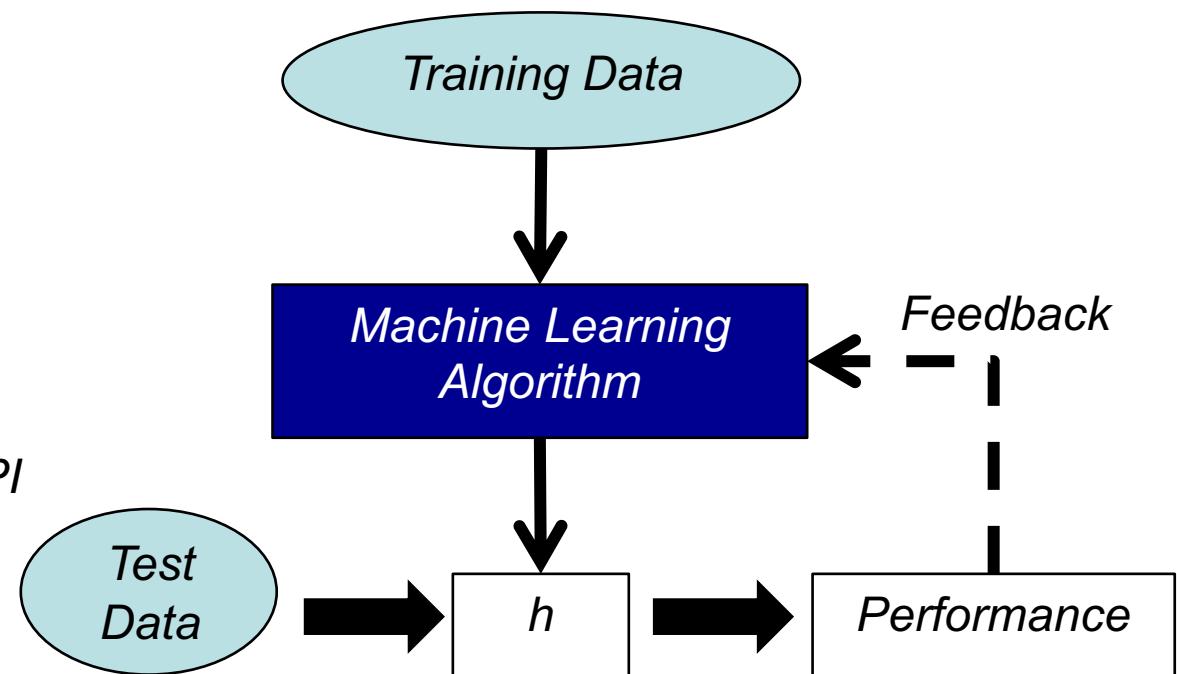
The machine learning process

- Data collection and Preparation
- Feature Selection
- Algorithm Choice
- Parameter and model selection
- Training
- Evaluation



**Look at the video:
from 0:00 to 45:00**

<https://youtu.be/mbyG85GZ0PI>



- **Supervised learning**

Given $D = \{\mathbf{X}_i, \mathbf{Y}_i\}$, learn $f(\cdot) : \mathbf{Y}_i = f(\mathbf{X}_i)$, s.t. $D^{\text{new}} = \{\mathbf{X}_j\} \Rightarrow \{\mathbf{Y}_j\}$

- **Unsupervised learning**

Given $D = \{\mathbf{X}_i\}$, learn $f(\cdot) : \mathbf{Y}_i = f(\mathbf{X}_i)$, s.t. $D^{\text{new}} = \{\mathbf{X}_j\} \Rightarrow \{\mathbf{Y}_j\}$

- **Semi-supervised learning**

- **Others:** Reinforcement Learning, Active learning, Recommender Systems

Introduction to Machine Learning

Unsupervised Learning

Cluster Analysis

Factor Analysis

Visualization

K-Means ,
EM

PCA, ICA

Self
Organized
Maps (SOM) ,
Multi-
Dimensional
Scaling

Lazy
Learning
(K-NN, IBL,
CBR)

Overfitting,
model
selection and
feature
selection

Kernel
Learning

Ensemble
Learning
(NN, Trees,
Adaboost)

Perceptron,
SVM

Supervised Learning

Non Linear Decision

Linear
Decision

Decision
Learning
Theory

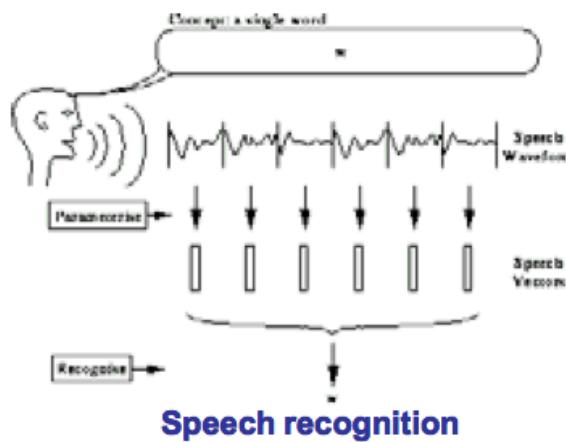
Basic
concepts of
Decision
Learning
Theory

Bias/Variance,
VC dimension,
Practical
advice of how
to use
learning
algorithms

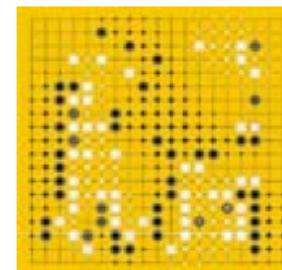
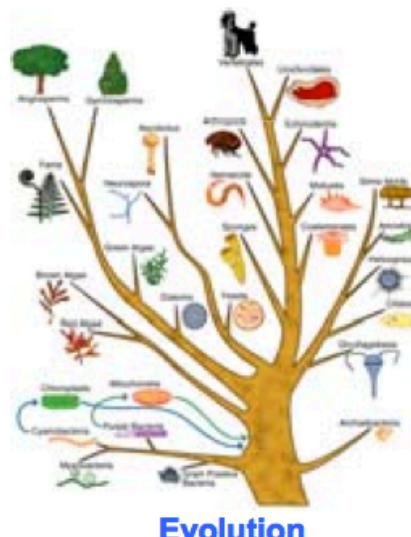
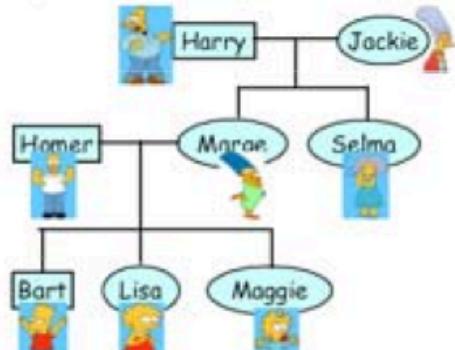


4. Applications

Applications



Computer vision



Games



Robotic control



Pedigree



Look at the video <https://youtu.be/ahRcGObyEZO>

- Machine translation
- Automatic Summarization
- Sentiment Analysis
- Question Answering

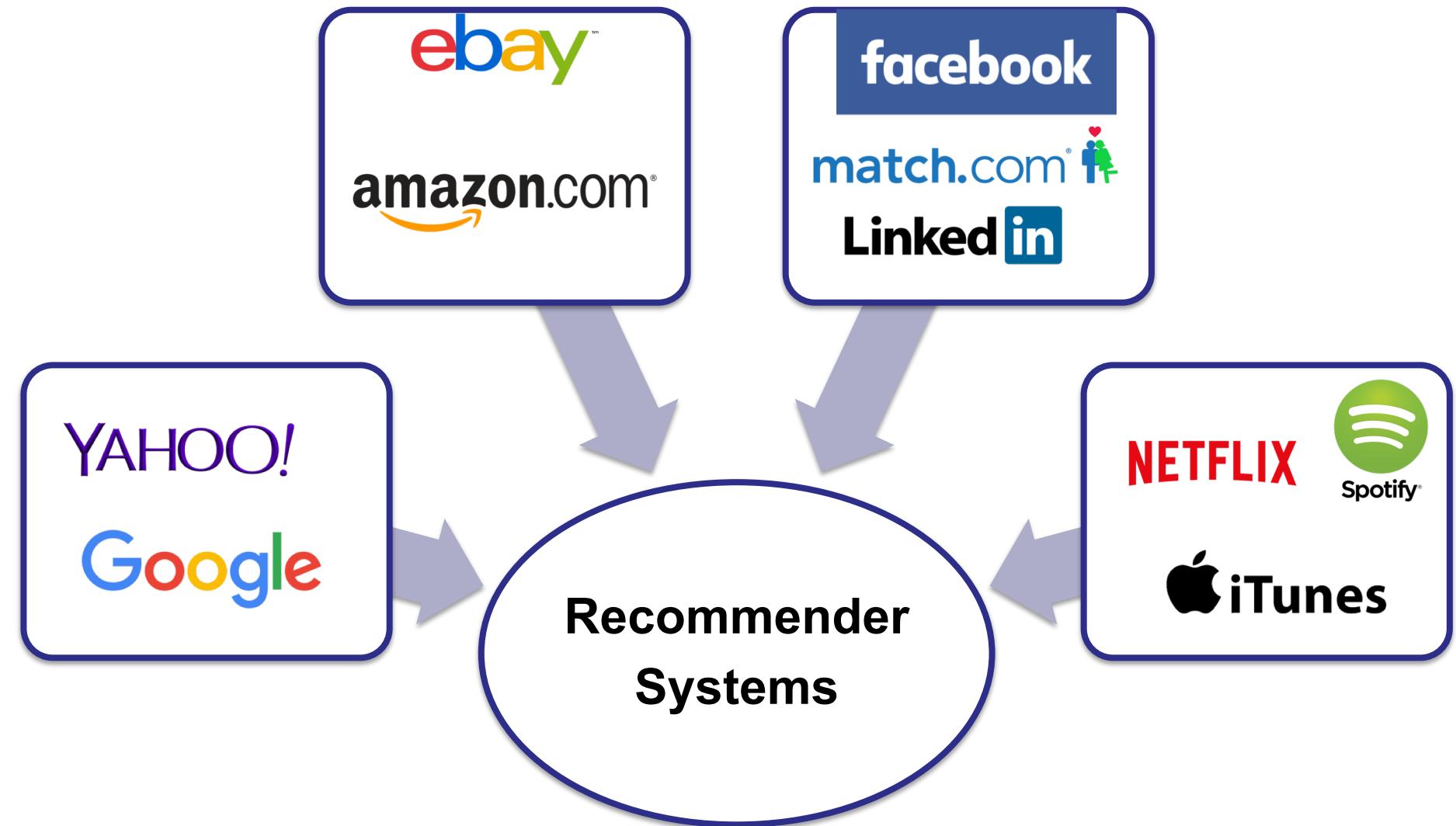


Look at the video

<https://youtu.be/WnzlbyTZsQY>

- Recommender systems are personalized **information systems** that provide recommendations: suggestions for items likely to be of use to a user (Burke, 2007)
- Recommendation as a prediction problem (Celma & Lamere, 2007)
 - Attempt to predict items that a user might be interested in
 - compute *similarity* between objects
 - user-user
 - item-item
 - form *predictions* based on the computed similarities

Use of Recommender Systems



5. History



TO READ: [HTTPS://WWW.DATAVERSITY.NET/A-BRIEF-HISTORY-OF-MACHINE-LEARNING/](https://www.dataversity.net/a-brief-history-of-machine-learning/)

History of Machine Learning

- **1950s**
 - Samuel's checker player
 - Selfridge's Pandemonium
- **1960s:**
 - Neural networks: Perceptron
 - Pattern recognition
 - Learning in the limit theory
 - Minsky and Papert prove limitations of Perceptron
- **1970s:**
 - Symbolic concept induction
 - Winston's arch learner
 - Expert systems and the knowledge acquisition bottleneck
 - Quinlan's ID3
 - Michalski's AQ and soybean diagnosis
 - Scientific discovery with BACON
 - Mathematical discovery with AM

History of Machine Learning

- **1980s:**
 - Advanced decision tree and rule learning
 - Explanation-based Learning (EBL)
 - Learning and planning and problem solving
 - Utility problem
 - Analogy
 - Cognitive architectures
 - Resurgence of neural networks (connectionism, backpropagation)
 - Valiant's PAC Learning Theory
 - Focus on experimental methodology
- **1990s:**
 - Data mining
 - Adaptive software agents and web applications
 - Text learning
 - Reinforcement learning (RL)
 - Inductive Logic Programming (ILP)
 - Ensembles: Bagging, Boosting, and Stacking
 - Bayes Net learning

History of Machine Learning

- **2000s**
 - Support vector machines
 - Kernel methods
 - Graphical models
 - Statistical relational learning
 - Transfer learning
 - Sequence labeling
 - Collective classification and structured outputs
 - Computer Systems Applications
 - Compilers
 - Debugging
 - Graphics
 - Security (intrusion, virus, and worm detection)
 - Email management
 - Personalized assistants that learn
 - Learning in robotics and vision



Week 1

Course. Introduction to Machine Learning

Theory 1. Introduction to ML

Dr. Maria Salamó Llorente

maria.salamo@ub.edu

Dept. Mathematics and Informatics,
Faculty of Mathematics and Informatics,
University of Barcelona (UB)