

Machine Learning

Federico Matteoni

A.A. 2021/22

Index

0.1	Introduction	2
-----	------------------------	---

0.1 Introduction

What is ML? Area of research combining aims of creating computers that could learn and powerful and adaptive statistical tools with rigorous foundation in computational science. Luxury or necessity? Growing availability and need for analysis of empirical data and difficult to provide intelligence and adaptivity by programming it. Change of paradigm.

Examples: spam classification, written text recognition. . . No or poor prior knowledge and rules for solving the problem, but easier to have a source of training experience.

ML is considered the latest general-purpose technology, capable of drastically affect pre-existing economic and social structures. And already has. The ultimate aim is to bring benefits to the people by solving big and small problems, accelerating human progress and empowering humans to add intelligence in any other science field.

Machine Learning We restrict to the computational framework: principles, methods and algorithms for learning and prediction, from experience. Building a model to be used for predictions. Common framework: infer a model or a **function** from a set of examples which allows the generalization (accurate response to new data).

When can we use ML? Be aware of the opportunity and awareness. ML is useful when there's no or poor theory surrounding the phenomenon, or uncertain, noisy or incomplete data which hinders formalization of solutions. The requests are: source of training experience (representative data) and a tolerance on the precision of results. The best examples are models to solve real-world problems that are difficult to be treated with traditional techniques: face and voice recognition (knowledge too difficult to formalize in an algorithm), predicting binding strength of molecules to proteins (not enough human knowledge) and personalized behavior, such as recommendation systems, scoring messages according to user preferences. . .

Definition The ML studies and proposes methods to build functions/hypothesis from examples of observed data that fits the known examples and able to generalize, with reasonable accuracy, for new data (according to verifiable results and under statistical and computational conditions and criteria.

Data Data **represents the available experience**. Representation problem: capturing the structure of the analyzed objects. Flat (attribute-value), structured. . . , categorical or continuous, missing data. . . **preprocessing**: variable scaling, encoding, selection. . .

Task The task defines the purpose of the application: knowledge that we want to achieve? which is the helpful nature of the result? what information are available?

Predictive task, classification and regression: function approximation

Descriptive task, cluster analysis and association rules: find subsets or groups of unclassified data.