ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

BA/MA/SA/Group/Lab:

Learning (Mining) from Non-i.i.d. Examples

Machine learning is the study of how to let machines improve at a certain task from seeing training data. In supervised learning, this task is to predict the target value of unseen examples based on the features of these examples. Statistical learning theory studies fundamental questions such as how many training examples are needed to guarantee a good prediction performance. In the past, learning theory has been a source of inspiration for several efficient and accurate learning algotithms.

Most algorithms and analysis in statistical learning theory rely on the important assumption that all data points are identically and independently distributed (i.i.d.). This assumption works for propositional data, i.e., data involving unconnected observations.

However, this i.i.d. assumption becomes unrealistic when the examples are interdependent. A relational database can be represented with a graph with entities and the relations between them. Existing paradigms to learn from this type of relational data violate the i.i.d. assumption in several ways. For example, some paradigms collect networked training examples sharing common objects, and then apply a classic supervised learning algorithm assuming i.i.d. data. Ignoring these dependencies may be harmful for learning and generalization, but only very few statistically sound results have been established.

In this project, we would like to devise accurate algorithms to learn from non-i.i.d. data and/or show statistically sound guarantees.

Requirements: Creativity and an interest in machine learning algorithms and theory. Being familiar with probability&statistics and/or graph theory may be an advantage.

Interested? Please contact us for more details!

Contacts

- Yuyi Wang: yuyi.wang@tik.ee.ethz.ch, ETZ G94
- Thomas Ulrich: thomas.ulrich@tik.ee.ethz.ch, ETZ G63