



Institute for Data Sciences and Engineering
COLUMBIA UNIVERSITY

SEMINAR EVENT

DATE: MON, NOV. 11

TIME: 10:00 A.M.

LOCATION:
INTERSCHOOL LAB, 750
SCHAPIRO CEPSR



SPEAKER: Joachim M. Buhmann

*Computer Science Department, Machine Learning Laboratory,
ETH Zurich*

TITLE: What is the Information Content of an Algorithm?

DATE: MON, NOV. 11

TIME: 10:00 A.M.

LOCATION:
INTERSCHOOL LAB, 750
SCHAPIRO CEPSR

Title:

What is the Information Content of an Algorithm?

Abstract:

Algorithms are exposed to randomness in the input or noise during the computation. How well can they preserve the information in the data w.r.t. the output space? Algorithms especially in Machine Learning are required to show robustness to input fluctuations or randomization during execution. This talk elaborates a new framework to measure the "informativeness" of algorithmic procedures and their "stability" against noise. An algorithm is considered to be a noisy channel which is characterized by a generalization capacity (GC). The generalization capacity objectively ranks different algorithms for the same data processing task based on the bit rate of their respective capacities. The problem of grouping data is used to demonstrate this validation principle for clustering algorithms, e.g. k-means, pairwise clustering, normalized cut, adaptive ratio cut and dominant set clustering. Our new validation approach selects the most informative clustering algorithm, which filters out the maximal number of stable, task-related bits relative to the underlying hypothesis class. The concept also enables us to measure how many bits are extracted by sorting algorithms when the input and thereby the pairwise comparisons are subject to fluctuations.

Bio:

Joachim M. Buhmann leads the Machine Learning Laboratory in the Department of Computer Science at ETH Zurich. He has been a full professor of Information Science and Engineering since October 2003. He studied physics at the Technical University Munich and obtained his PhD in Theoretical Physics. As postdoc and research assistant professor, he spent 1988-92 at the University of Southern California, Los Angeles, and the Lawrence Livermore National Laboratory. He held a professorship for applied computer science at the University of Bonn, Germany from 1992 to 2003. His research interests spans the areas of pattern recognition and data analysis, including machine learning, statistical learning theory and information theory. Application areas of his research include image analysis, medical imaging, acoustic processing and bioinformatics. Currently, he serves as president of the German Pattern Recognition Society.

500 WEST 120TH STREET
NEW YORK, NY 10027

PHONE: 212-854-5660
IDSE.COLUMBIA.EDU