

2. CASUS Annual Workshop

“Fusing Artificial Intelligence & Simulation”

Statistical methods for reduction of dimension of "fat" data sets



Dr. Malgorzata Bogdan is a Full Professor and the head of the Statistics group in the Department of Mathematics of University of Wroclaw and the Guest Professor in the Department of Statistics at Lund University. She received her PhD in Statistics from Wroclaw University of Science and Technology in 1996. She spent many years in USA working at the University of Washington and Purdue University and taking advantage of the Fulbright fellowship at Stanford University. She works on the development of novel statistical methods to deal with current challenges related to high dimensionality of modern data sets, deviations from standard probability assumptions and the presence of missing data. She also analyzes the properties of these methods through theoretical mathematical results and extensive simulation studies. The team of Dr. Bogdan cooperates with scientists from many fields and, apart from core statistical journals, publishes their research in genetic, medical and financial journals. Dr Bogdan published over 50 peer-reviewed articles and a monography on statistical gene detection in a Springer Series on Computational Biology. The methodology developed by Dr. Bogdan and her team is implemented in open source packages in R, Python and matlab.

Date: Friday, 9 October 2020

Time: 4:15 p.m. CET

Location: livestream link follows



Abstract:

Large data sets are now routinely collected in many fields of industry or science. Many of these data sets are "fat" – i.e. the number of measurements for one individual substantially exceeds the number of records in such data set. "Fat" data sets create opportunities for identifying new patterns but their analysis brings many statistical challenges. One of the most dangerous and often neglected phenomena is the problem of "false discoveries", which arise due to the random fluctuations in the data. Statistical methods for dimensionality reduction address this problem in a structured way and often provide some guarantees concerning the control over false discoveries. In this talk we will present several methods for dimensionality reduction developed by the Wrocław team in collaboration with colleagues from many research institutions in Europe and USA. We will cover some novel extensions of the Bayesian Information Criterion and of the Sorted L-One Penalized Estimator (SLOPE), both for supervised and unsupervised learning, some modifications of the knockoff methodology for controlling the False Discovery Rate (FDR) and the new method for unsupervised dimensionality reduction by the multiple latent component clustering.