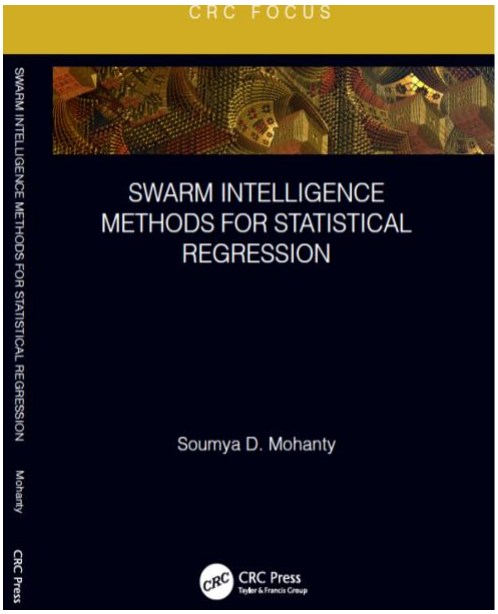
Swarm intelligence methods for statistical regression

BigDat 2019, Cambridge, UK.

Course summary

(**This document will be updated as needed)**

A core task in statistical analysis, especially in the era of Big Data, is the fitting of flexible, high-dimensional, and non-linear models to noisy data in order to capture meaningful patterns. This can often result in challenging non-linear and non-convex global optimization problems. The large data volume that must be handled in Big Data applications further increases the difficulty of these problems. ***Swarm Intelligence Methods for Statistical Regression*** describes methods from the field of computational swarm intelligence (SI), and how they can be used to overcome the optimization bottleneck encountered in statistical analysis.



Course textbook

"Swarm intelligence methods for statistical regression", Soumya D. Mohanty (Chapman and Hall/CRC Press, 2018). Available from the publisher's website as well as amazon.com.

Course highlights

· Overview of statistical data analysis and key results in stochastic optimization theory

· Focuses on methodology and results rather than formal proofs

· Reviews SI methods with a deeper focus on Particle Swarm Optimization (PSO)

· Practical tips and tricks for tuning PSO to extract good performance in real world data analysis challenges

Lecture 1

The importance of global optimization in statistical data analysis will be discussed. We will emphasize the fact that optimization plays as fundamental a role as probability theory in many modern applications of statistics, especially in big data, and that our failure in handling the optimization task often limits the type of models we can fit to observational data. Concrete real-world applications from the instructor's research in gravitational wave data analysis will be used to illustrate the benefits of overcoming the optimization bottleneck using swarm intelligence methods.

Lecture 2

High dimensional and flexible data models require stochastic optimization methods for fitting. Fundamental results on stochastic global optimization methods will be reviewed. Methods in evolutionary computation and swarm intelligence (SI) will be outlined, with a focus on particle swarm optimization (PSO), and strategies for applying them to statistical regression problems will be discussed.

Lecture 3

PSO will be reviewed in depth. We will consider some representative, but realistic, statistical regression problems and show step-by-step how to use PSO effectively in solving the associated optimization challenges.

Supplemental material

Please check the folders provided in this shared area for supplemental material. The contents of these folders are listed below and will be updated as time progresses.

SLIDES

This folder will contain the lecture slides. (Please check this folder closer to the start of BigDat19.)

READING

This folder contains pointers to reference material in addition to the course textbook.

CODES

This folder will contain Matlab codes and scripts related to the concrete statistical regression examples to be discussed in the course. (Please check this folder closer to the start of BigDat19.)

Instructor bio

Soumya D. Mohanty, Professor of Physics at the University of Texas Rio Grande Valley, completed his PhD degree in 1997 at the Inter-University Center for Astronomy and Astrophysics, India. He subsequently held post-doctoral positions at Northwestern University, Penn State, and the Max-Planck Institute for Gravitational Physics. He was also a visiting scholar with the LIGO project at Caltech. Mohanty's research has focused on solving some of the important data analysis challenges faced in Gravitational Wave (GW) astronomy across all observational frequency bands. These include non-parametric regression of very weak signals in noisy data, high-dimensional non-linear parametric regression, time series classification, and analysis of data from large heterogeneous sensor arrays. Mohanty’s work has been funded by grants from the Research Corporation, the U.S. National Science Foundation, and NASA.