NMST611 *Advanced Statistics Seminar*

Anomaly detection using data depth: multivariate case

Pavlo Mozharovskyi
Télécom Paris

October 19, 2022
15:40, Praktikum KPMS

Abstract

Anomaly detection is a branch of machine learning and data analysis which aims at identifying observations that exhibit abnormal behaviour. Be it measurement errors, disease development, severe weather, production quality default(s) (items) or failed equipment, financial frauds or crisis events, their on-time identification, isolation and explanation constitute an important task in almost any branch of industry and science. By providing a robust ordering, data depth — statistical function that measures belongingness of any point of the space to a data set — becomes a particularly useful tool for detection of anomalies. Already known for its theoretical properties, data depth has undergone substantial computational developments in the last decade and particularly recent years, which has made it applicable for contemporary-sized problems of data analysis and machine learning.

In this article, data depth is studied as an efficient anomaly detection tool, assigning abnormality labels to observations with lower depth values, in a multivariate setting. Practical questions of necessity and reasonability of invariances and shape of the depth function, its robustness and computational complexity, choice of the threshold are discussed. Illustrations include use-cases that underline advantageous behaviour of data depth in various settings.