

## DESCRIPTION FOR THE GENERAL PUBLIC

The major goal of this project is the development of new methods of analysis of Big Data and investigation of their statistical properties. Such data now occur in many fields of science and industry and need to be properly analyzed for the effective extraction of the information. The major problem with the analysis of such data is setting the appropriate thresholds for signal detection. If this problem is inadequately addressed the researchers often end up with a large number of false discoveries or with the insufficient power of detection of true patterns. The errors of both types largely reduce the predictive abilities of the estimated models and decrease the respective profit. Specifically, in the context of gene detection, which is the main practical focus of our research, a false discovery needs to be verified by expensive biological experiments, while the omission of important gene might reduce the chance of development of appropriate treatments.

In our research we plan to theoretically investigate the properties of the new convex optimization method, called SLOPE (Sorted L-One Penalized Estimation), recently developed by the author of this project and collaborators from Stanford University, and to extend the scope of applications of this method. The method is computationally feasible and satisfies strict requirements concerning the control of the false discovery rate, while preserving a high power of identification of true patterns. Theoretical investigation of the properties of SLOPE will allow for a better understanding of this method, developing its optimal versions and will set honest practical limits of its applicability. Further, we expect that the control of the false discovery rate provided by SLOPE will also allow to build statistical models with superior predictive properties. We also plan to develop new Bayesian versions of SLOPE, which will allow to use the prior information and to estimate uncertainty of the results of statistical inference. Apart from the development of new optimal versions of SLOPE we also plan to derive precise statistical criteria for the comparison of statistical methods in settings when the notions of the true and false discovery are not entirely clear. This will allow for a unified comparison of different statistical methods and will prevent misinterpreting of the results of the statistical experiments. The final product of our research will be the set of publicly available statistical packages for the analysis of Big Data, with the implementations of statistical methods developed in the course of this project.