MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

České vysoké učení technické v Praze
Fakulta jaderná a fyzikálně inženýrská
Břehová 7
115 19 Praha 1

Opponent's Report on Dissertation Thesis

General Framework for Classification at the Top

Author: Ing. Václav Mácha

Supervisors: doc. Ing. Václav Šmídl, Ph.D., Mgr. Lukáš Adam, Ph.D.

Institution: Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Mathematics

The thesis is devoted to problems of binary classification at the top of a samples, which is an actual and important topic in the field of mathematical optimization related to modern statistical and machine-learning methods.

In the first chapter, we can find general formulations for binary classification and the measures of performance for the binary classifiers, which enable to optimize their overall performance. Chapter 2 is focused on general optimization framework for classification at the top. It is shown that many important problems are related to this framework, namely ranking problems, accuracy at the top, and the Neyman-Pearson problem. New formations are proposed and relations between them are studied. These formulations enable to use linear as well as (general) nonlinear classifiers. In Chapter 3, basic properties (convexity, differentiability and stability) of the linear

Katedra pravděpodobnosti a matematické statistiky
Sokolovská 49/83, 186 75 Praha 8
kpms@mff.cuni.cz
95155 3287

classification models are investigated with a special focus on conditions which guarantee optimality of the obtained solutions of optimization models. Chapter 4 is devoted to dual formulations of the considered models. By applying these reformulations, two-special classes of the models are identified. Moreover, the reformulations enable to include nonlinearities into the classification rules. The chapter ends with an efficient algorithm to solve the optimization problems. In Chapter 5, nonlinear classification rules and models are elaborated which cover, e.g., neural networks. The candidate has shown that the formulations are not decomposable since the decision threshold is always a function of all classification scores. Therefore, the stochastic gradient descent algorithm is modified to solve this case. Chapter 6 contains results of many numerical experiments including image recognition, steganalysis, and malware detection. Specific settings of the rules and their parameters are deeply elaborated.

To summarize, the thesis brings original results which contribute significantly to the fields of classification rule models, mathematical optimization and numerical methods. The mathematical methods used to derive new results are original and in most cases highly non-trivial. I consider the goals of the work fulfilled. I would like to appreciate that the codes are made freely available using the candidate's GitHub repository.

The thesis is written very precisely. I found almost no typos or ambiguities. The candidate works correctly with the sources, they are properly cited in the text and his contribution is clearly specified.

The candidate has already published 2 papers (as a coauthor) in *Optimization Methods and Software*, which is a leading international journal of the field of mathematical optimization and numerical methods with strictly positive WOS Impact Factor. Another paper is prepared and submitted to a journal. This is very good publication record.

Based on the above facts, **I recommend the submitted thesis for defense.**

July 21, 2023, Prague

Doc. RNDr. Martin Branda, Ph.D.
Charles University, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics