

Preface

This book concerns the problem of data smoothing. There are many smoothing techniques, yet the kernel smoothing seems to be one of the most important and widely used ones. In this book, I focus on a well-known technique called kernel density estimation (KDE), which is an example of a nonparametric approach to data analysis.

During the last few decades, many books and papers devoted to this broad field have been published, so it seems that this area of knowledge is quite well understood and reached its maturity point. However, many (or even most) of the practical algorithms and solutions designed in the context of KDE are very time-consuming with quadratic computational complexity being a commonplace. This might be not problematic for situations, where datasets are not that big (at the level of hundreds of individual data points) but it already can be an obstacle for datasets containing thousands or more individual data points, especially in case of multivariate data. Progress in terms of theoretical results related to KDE does not go hand in hand with the development of fast and accurate algorithm for speeding up the required calculations in practical terms. In this sense, this book can be considered a valuable contribution to the field of KDE.

This book is a result of my research in the area of numerical and computational problems related to KDE, an interest that has been developing since ca. 2010. It should be viewed primarily as a research monograph and is intended both for those new to such topics as well as for more experienced readers. The first few chapters present a background material, describing the fundamental concepts related to the nonparametric density estimation, kernel density estimation, and bandwidth selection methods. The presented material is richly illustrated by numerical examples, using both toy and real datasets. The following chapters are devoted to the presentation of our own research on fast computation of kernel density estimators and bandwidth selection. The developed methods are based on the fast Fourier transform (FFT) algorithm that relies on a preliminary data transformation known as data binning. Certain results obtained by me on utilizing field-programmable gate arrays (FPGA) in the context of fast bandwidth selection are also included. FPGA devices are a not so common choice in terms of implementing purely numerical algorithms.

The proposed implementation can be seen as a preliminary study of practical usability of such FPGA-based applications. The monograph ends with a chapter presenting a number of applications related to KDE. The following example applications are given: discriminant analysis, cluster analysis, kernel regression, multivariate statistical process control, and flow cytometry.

I wish to express my deepest gratitude to Prof. Józef Korbicz, Ph.D., D.Sc., a corresponding member of the Polish Academy of Sciences for his continuing support since 2000 to this day and for motivating me to work hard on the problems at hand. I would also like to express my sincere thanks to Prof. Andrzej Obuchowicz, Ph.D., D.Sc., for his help at a time of particular need in my life. I also extend my thanks to Prof. Dariusz Uciński, Ph.D., D.Sc., and to Marcin Mrugalski, Ph.D., D.Sc., for their continuing kindness. Marek Sawerwain, Ph.D., helped me a lot in all programming tasks related to FPGA, and I thank him very much for this. My warm thanks also go to my mother, to my wife Edyta, and to my children Ola and Kuba for their love, patience, and support. Finally, I would like to express my sincere gratitude to my brother Jarosław because I truly believe that writing this book would not have been possible without his powerful and continuous assistance and support, both in professional and in family life.

Zielona Góra, Poland
September 2017

Artur Gramacki

Nonparametric Kernel Density Estimation and Its
Computational Aspects

Gramacki, A.

2018, XXIX, 176 p. 70 illus., Hardcover

ISBN: 978-3-319-71687-9