Reviewer: Marlene Müller
Beuth University of Applied Sciences Berlin

Multivariate Nonparametric Regression and Visualization. With **R** and Applications to Finance

Jussi Klemelä
John Wiley & Sons, Hoboken, 2014.
ISBN 978-0-470-38442-8. 396 pp. USD 115 (Hardcover).
http://jklm.fi/regstruct/

This book on nonparametric regression can be seen as complementing "Smoothing of Multivariate Data" published in 2009 by the same author (Klemelä 2009). While the former book focuses on nonparametric density estimation in one or more dimensions the current publication now addresses a variety of regression-type models.

The subtitle of the book emphasizes the R software and applications to finance as particular aspects. The supplementary R code is however only covered in the appendix. After a short introduction, the book divides into the following parts:

- *Part I: Methods of Regression and Classification*

  This is the most extensive part with five chapters (about 275 pages), comprising a fundamental introduction to different regression models (e.g., linear models, kernel regression and semiparametric approaches). The application examples in the book are mainly directed towards finance. For that reason this part also introduces numerous terms from quantitative finance (e.g., copulas, risk measures, GARCH etc.).

- *Part II: Visualization*

  The two chapters in this part (about 50 pages) describe the visualization of data and the visualization of functions.

- *Appendix A: R Tutorial*

  This appendix (about 20 pages) is a collection of R code examples used to compute and visualize some of the estimation methods in the book. These examples employ R functions which are available in the R packages **denpro** (Klemelä 2015) and **regpro** (Klemelä 2016) on CRAN. The tutorial is also available from the author's homepage (http://jklm.fi/regstruct/tutorial.html).

The major *Part I* of the book concentrates on the theoretical foundations of nonparametric regression (classification issues are rather a byproduct). It is accompanied by some examples

and graphical illustrations. Chapters 2 and 3 involve application sections on volatility and quantile estimation in finance and on regression methods used for portfolio selection.

To name some of the contents of the book in more detail: The introductory Chapter 1 starts with an overview on regression and introduces topics such as conditional distributions and densities, random and fixed design in regression, classical mean regression and quantile regression, partial effects and derivative estimation, regression variants (like variance, covariance and correlation regression, time series, instrumental variables) and stochastic control. The subsequent sections consider discrete response variables, parametric distribution families, copula models, (a short excursion to) classification, examples from quantitative finance, and data transformations. A discussion on measuring the performance of estimators completes this general presentation.

The further chapters are concerned with specific classes of regression estimators. Chapter 2 introduces linear regression and related models, for example generalized linear models, varying coefficients, series estimators, ARCH, quantile estimation and linear classification. Chapter 3 is dedicated to kernel regression and discusses in particular the regressogram, the Nadaraya-Watson estimator and alternative estimators for fixed design, moving averages, nearest neighbor methods and median smoothing, state space smoothing and GARCH models. Chapter 4 concentrates on selected semiparametric (structural) models such as single index, additive and partially linear regression models. Finally Chapter 5 summarizes more complex models under the title "Empirical Risk Minimization". This comprises for example local polynomial estimation, local likelihood, support vector machines, stagewise modeling, projection pursuit, adaptive regressograms and CART (classification and regression trees).

I do like *Part I* of the book for a presentation that is focused on the main facts and is formulated in a mathematically precise way. It is, however, not really an introduction to the topics but rather a compendium or catalogue of methods. I would have also preferred to find more literature references in the text.

As already mentioned above, *Part II* deals with two main topics: visualization of multivariate data and visualization of multivariate curves. The two chapters repeat many well known techniques as in particular scatterplots, QQ plots, histogram and density estimates, one-dimensional functions, perspective surfaces and contour plots for two-dimensional functions. In addition, the book discusses some multivariate techniques such as projection pursuit, multidimensional scaling, graphical matrices, parallel coordinates, level sets, level set trees, volume functions and barycenter plots.

The software part of the book, *Appendix A* (http://jklm.fi/regstruct/tutorial.html), is written as a user tutorial mainly for the R package **regpro**. It aims to explain how to use the functions in that package and to replicate some of the estimation examples throughout the book. Unfortunately, the provided R code is not free of errors, in particular:

- The function `pcf.histo` used in A.1.3 and A.1.4 (and also `histo2data`, `rotation3d` in A.1.4) are not exported in the namespace of package **denpro**. (To circumvent this, one could modify their call to `denpro:::pcf.histo` etc., for example.)

- The code examples in A.3.3 and subsequently A.4.2 are missing the initialization of the vector `y`. This produces an R error if these codes are run independently from other code.

All kernel-based methods allow for the Gaussian and uniform kernels only. It should also

be remarked that *Appendix A* solely provides examples with generated data. I really would prefer a possibility to rerun all the examples within the book with R. My biggest concern however is, that there are no references to other existing R packages on non- or semiparametric regression estimation. Also, the package **regpro** is completely implemented in R (which means that it does not make use of code in C/C++/Fortran etc.). For that reason, running the examples from *Appendix A* might take quite some time if the sample sizes and grids are set to somewhat larger sizes. Thus, the R tutorial may be seen as a starting point for users who want to take their first steps into nonparametric function estimation. They are guided with (in most cases) clearly written code examples. For real-world applications (in particular in quantitative finance) a deeper insight into other existing R packages and R programming would be preferable.

To summarize: I do recommend this book to readers who are interested to learn about a large variety of non- and semiparametric estimators and who want to see some examples on their application in finance. This is not an introductory text for quantitative finance or R, the intended audience seems to be rather experienced statisticians and students with a good understanding of mathematical notation. The software (or R) part of the book is rather a prototypic implementation for illustrating some of the main concepts.

References

Klemelä J (2009). *Smoothing of Multivariate Data: Density Estimation and Visualization.* Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken. ISBN 978-0-470-29088-0.

Klemelä J (2015). **denpro**: *Visualization of Multivariate Functions, Sets, and Data.* R package version 0.9.2, URL https://CRAN.R-project.org/package=denpro.

Klemelä J (2016). **regpro**: *Nonparametric Regression.* R package version 0.1.1, URL https://CRAN.R-project.org/package=regpro.

Reviewer:
Marlene Müller
Beuth University of Applied Sciences Berlin
Department II — Mathematics, Physics and Chemistry
D-13353 Berlin, Germany
E-mail: marlene.mueller@beuth-hochschule.de
URL: http://prof.beuth-hochschule.de/mmueller/