MR3445317 62G08 60D05 60E15 62G07 62G20 62H30 68T05 68T10
Biau, Gérard (F-PARIS6-NDM); **Devroye, Luc** [**Devroye, Luc P.**] (3-MGL-C)

★**Lectures on the nearest neighbor method.**
Springer Series in the Data Sciences.
Springer, Cham, 2015. *ix*+290 *pp. ISBN* 978-3-319-25386-2; 978-3-319-25388-6

As the authors state in the Preface, nearest neighbor methods are one of the main paradigms in machine learning, making the text a fitting start to the new Springer series. They express their intention "to bring the key statistical, probabilistic, combinatorial, and geometric ideas required in the analysis under one umbrella". No target audience is specified, but a casual glance through the book reveals that it should have a certain mathematical (around third-year level) but not necessarily statistical maturity—a welcome feature, given the diversity of backgrounds that an instructor wanting to use this book may face. Several open problems are suggested, which add to the interest of the book from the viewpoint of researchers.

In part, the book modifies, and sometimes strengthens, results that may be deemed classic (some of which originate from papers of the first author from the late 1970s and 1980s). Extensive and careful use of real analysis is made. The authors maintain a pleasant conversational tone and avoid digressing into major related topics such as concentration inequalities (page 141), Vapnik-Chervonenkis theory (page 162), and minimax theory (page 211)—certainly a justified decision.

The book has three parts: "Density estimation" (Chapters 1–7), "Regression estimation" (Chapters 8–16), and "Supervised classification" (Chapters 17–19). This is complemented by an appendix (Chapter 20) on topics from probability, geometry, and real analysis.

Chapter 1 gives basic material on order statistics and nearest neighbors, and shows the universal nature of uniform order statistics; versions of the notion of universality also appear later on. Chapter 2 introduces the nearest neighbor distance, with ties being broken by comparing indices. Chapter 3 defines the k-nearest neighbor density estimate of the unknown density f; weak and strong pointwise consistency are proven here, and uniform consistency in Chapter 4. Weighted k-nearest neighbor density estimates are the topic of Chapter 5. Local behavior is investigated in Chapter 6, under the assumption that $f(\mathbf{x})$ is estimated using only $\|\mathbf{X}_1-\mathbf{x}\|,\dots,\|\mathbf{X}_n-\mathbf{x}\|$ for the sample $\{\mathbf{X}_i\}$, allowing for the reduction to a one-dimensional problem. In Chapter 7, the variance and bias of the estimate proposed in [L. F. Kozachenko and N. N. Leonenko, Problemy Peredachi Informatsii **23** (1987), no. 2, 9–16; MR0908626] are investigated. Tools used are the powerful Efron-Stein inequality and Hardy-Littlewood maximal functions.

Turning to Part II, the brief Chapter 8 introduces the regression model as a weighted average of response values for nearest neighbors, where the tie issue is a messy complication. In Chapter 9, it is shown that the 1-nearest neighbor estimate is L^2 consistent only in the noiseless case. In Chapter 10, universal L^p consistency is proven with a theorem of C. J. Stone [Ann. Statist. **5** (1977), no. 4, 595–645; MR0443204], whose proof is given in full. Chapter 11 contains results on universal weak pointwise consistency, where the tie issue causes technical difficulties, and strong pointwise consistency. In Chapter 12, for proving strong uniform consistency, a uniform exponential tail condition on the modulus of the residual is assumed. After an interlude in Chapter 13 on advanced properties of uniform order statistics, the local rate of convergence is investigated in Chapter

14. Chapter 15 discusses noiseless estimation for fixed and diverging k. Chapter 16 gives a concise yet very engaging account of important issues regarding the choice of nearest neighbor estimate, with an oracle inequality and a nice example on feature (or variable) selection for the mean integrated squared error.

Part III on (supervised) classification begins, in Chapter 17, with a concise account of basic issues — mainly for the binary case — which apply not only to nearest-neighbor methods; in particular, it states the connection of classification to regression via plug-in of the regression estimate. Chapters 18 and 19 build on this last point, and give precise conditions on the sequence of weights to obtain universal consistency.

The examples interspersed throughout the text, and the parts or variations of results whose proofs are left to the reader, compensate at least partly for the lack of formal exercises. Misprints are few, and a list of them may be found on the webpage of the first author.

This book is a valuable addition to the literature on this topic and will be found to be useful by students, teachers, and researchers.

Christian Rau