DATA SCIENCE SERIES

STATISTICAL FOUNDATIONS OF DATA SCIENCE

JIANQING FAN
RUNZE LI
CUN-HUI ZHANG
HUI ZOU

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Contents

Statistical Foundations of Data Science gives a thorough introduction to commonly used statistical models and contemporary statistical machine learning techniques and algorithms, along with their mathematical insights and statistical theories. It serves as a graduate-level textbook and a research monograph on high-dimensional statistics, sparsity and covariance learning, machine learning, and statistical inference. It includes ample exercises that involve both theoretical studies as well as empirical applications.

The book begins by introducing the stylized features of big data and their impacts on statistical analysis. It then introduces multiple linear regression and expands the techniques of model building via nonparametric regression and kernel tricks. It provides a comprehensive account of sparsity explorations and model selections for multiple regression, generalized linear models, quantile regression, robust regression, hazards regression, among others. High-dimensional inference is also thoroughly addressed and so is feature screening. The book also provides a comprehensive account of high-dimensional covariance estimation, learning latent factors and hidden structures, as well as their applications to statistical estimation, inference, prediction and machine learning problems. It also introduces thoroughly statistical machine learning theory and methods for classification, clustering, and prediction. These include CART, random forests, boosting, support vector machines, clustering algorithms, sparse PCA, and deep learning.

The authors are international authorities and leaders on the presented topics. All are fellows of the Institute of Mathematical Statistics and the American Statistical Association.

Jianqing Fan is Frederick L. Moore Professor, Princeton University. He is co-editing *Journal of Business and Economics Statistics* and was the co-editor of *The Annals of Statistics, Probability Theory and Related Fields,* and *Journal of Econometrics* and has been recognized by the 2000 COPSS Presidents' Award, AAAS Fellow, Guggenheim Fellow, Guy medal in silver, Noether Senior Scholar Award, and Academician of Academia Sinica.

Runze Li is Elberly family chair professor and AAAS fellow, Pennsylvania State University, and was co-editor of *The Annals of Statistics*.

Cun-Hui Zhang is distinguished professor, Rutgers University and was co-editor of *Statistical Science*.

Hui Zou is professor, University of Minnesota and was action editor of *Journal of Machine Learning Research*.

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com
CRC Press titles are available as eBook ed

ID 20 1001 6752
ISBN 9781466510845

STATISTICS
ISBN: 978-1-4665-1084-5
90000
9 781466 510845