

Part D Regression

Part D Regression Methods and Data Mining

28 Measures of Influence and Sensitivity in Linear Regression

Daniel Peña, Getafe (Madrid), Spain

29 Logistic Regression Tree Analysis

Wei-Yin Loh, Madison, USA

30 Tree-Based Methods and Their Applications

Nan Lin, St. Louis, USA
Douglas Noe, Champaign, USA
Xuming He, Champaign, USA

31 Image Registration and Unknown Coordinate Systems

Ted Chang, Charlottesville, USA

32 Statistical Genetics for Genomic Data Analysis

Jae K. Lee, Charlottesville, USA

33 Statistical Methodologies for Analyzing Genomic Data

Fenghai Duan, Omaha, USA
Heping Zhang, New Haven, USA

34 Statistical Methods in Proteomics

Weichuan Yu, New Haven, USA
Baolin Wu, Minneapolis, USA
Tao Huang, New Haven, USA
Xiaoye Li, New Heaven, USA
Kenneth Williams, New Haven, USA
Hongyu Zhao, New Haven, USA

35 Radial Basis Functions for Data Mining

Miyoung Shin, Daegu, Republic of Korea
Amrit L. Goel, Syracuse, USA

36 Data Mining Methods and Applications

Kwok-Leung Tsui, Atlanta, USA
Victoria Chen, Arlington, USA
Wei Jiang, Hoboken, USA
Y. Alp Aslandogan, Arlington, USA

Part D focuses on regression methods and data mining. The first chapter in this part, Chapt. 28, describes various diagnostic procedures for detecting single and multiple outliers and influential observations in linear regression. It also discusses procedures for detecting high-leverage outliers in large, high-dimensional data sets. Chapter 29 gives an overview of various logistic regression methods for fitting models to a binary-valued response variable and introduces the idea of a logistic regression tree based on a recursive partitioning algorithm to fit a linear logistic regression model for solving large, complex data sets. Chapter 30 introduces the basic structure of tree-based methods for constructing trees for both classification and regression problems by recursively partitioning a learning sample over its input variable space. It also compares classification and regression trees to multivariate adaptive regression splines, neural networks and support-vector machines. Chapter 31 presents the concept of a generalization of least-squares estimation (LSE), called M estimators, to solve the statistical problems involving unknown coordinate systems and image registration problems. This chapter also discusses in detail the differences between the LSE and M estimators and presents the statistical properties of M estimates for spherical regression.

The following three chapters focus on the statistical analysis of genomic and proteomics data. Chapter 32 provides an overview of the emerging statistical con-

cepts of statistical genetics, which are commonly used to analyze microarray gene-expression data, and further introduces recent statistical testing methods, such as significance analysis of microarray and local pooled-error tests, as well as supervised-learning discovery tools. Chapter 33 describe several statistical methods, such as the empirical Bayesian approach, significance analysis of microarray, support-vector machines, and tree- and forest-based classification, for analyzing genomic data and their applications in biochemical and genetic research. Chapter 34 discusses two proteomics statistical techniques, disease biomarker discovery and protein/peptide identification, and their applications in both the biological and medical research for analyzing mass-spectrometry data. The next two chapters focus on data mining and its applications. Chapter 35 describes the radical basis-function model architecture and its applications in bio-informatics and biomedical engineering and also describes the four algorithms commonly used for its design: clustering, orthogonal least squares, regularization, and gradient descent, while Chapt. 36 presents the basic principles of data-mining methodologies in databases, including knowledge discovery, supervised learning, software, the classification problem, neural networks, and association rules, and discusses several popular data-mining methods with applications in industry and business practice.