1. Record Nr. UNINA9910254092303321

Autore	Berk Richard A
Titolo	Statistical Learning from a Regression Perspective / / by Richard A. Berk
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2016
ISBN	3-319-44048-9
Edizione	[2nd ed. 2016.]
Descrizione fisica	1 online resource (XXIII, 347 p. 120 illus., 91 illus. in color.)
Collana	Springer Texts in Statistics, , 1431-875X
Disciplina	519.2
Soggetti	Statistics Probabilities Public health Psychology—Methodology Psychological measurement Social sciences Statistical Theory and Methods Probability Theory and Stochastic Processes Statistics for Social Sciences, Humanities, Law Public Health Psychological Methods/Evaluation Methodology of the Social Sciences
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Statistical Learning as a Regression Problem -- Splines, Smoothers, and Kernels -- Classification and Regression Trees (CART) -- Bagging -- Random Forests -- Boosting -- Support Vector Machines -- Some Other Procedures Briefly -- Broader Implications and a Bit of Craft Lore.
Sommario/riassunto	This textbook considers statistical learning applications when interest centers on the conditional distribution of the response variable, given a set of predictors, and when it is important to characterize how the predictors are related to the response. As a first approximation, this can be seen as an extension of nonparametric regression. This fully revised new edition includes important developments over the past 8 years. Consistent with modern data analytics, it emphasizes that a

proper statistical learning data analysis derives from sound data collection, intelligent data management, appropriate statistical procedures, and an accessible interpretation of results. A continued emphasis on the implications for practice runs through the text. Among the statistical learning procedures examined are bagging, random forests, boosting, support vector machines and neural networks. Response variables may be quantitative or categorical. As in the first edition, a unifying theme is supervised learning that can be treated as a form of regression analysis. Key concepts and procedures are illustrated with real applications, especially those with practical implications. A principal instance is the need to explicitly take into account asymmetric costs in the fitting process. For example, in some situations false positives may be far less costly than false negatives. Also provided is helpful craft lore such as not automatically ceding data analysis decisions to a fitting algorithm. In many settings, subject-matter knowledge should trump formal fitting criteria. Yet another important message is to appreciate the limitation of one’s data and not apply statistical learning procedures that require more than the data can provide. The material is written for upper undergraduate level and graduate students in the social and life sciences and for researchers who want to apply statistical learning procedures to scientific and policy problems. The author uses this book in a course on modern regression for the social, behavioral, and biological sciences. Intuitive explanations and visual representations are prominent. All of the analyses included are done in R with code routinely provided.