

1. Record Nr.	UNINA9910349351303321
Autore	Komori Osamu
Titolo	Statistical Methods for Imbalanced Data in Ecological and Biological Studies // by Osamu Komori, Shinto Eguchi
Pubbl/distr/stampa	Tokyo : , : Springer Japan : , : Imprint : Springer, , 2019
ISBN	4-431-55570-6
Edizione	[1st ed. 2019.]
Descrizione fisica	1 online resource (63 pages)
Collana	JSS Research Series in Statistics, , 2364-0057
Disciplina	519.5
Soggetti	Statistics Biostatistics Statistics for Life Sciences, Medicine, Health Sciences Statistical Theory and Methods Statistics for Social Sciences, Humanities, Law
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	1. Imbalance Data -- 2. Weighted Logistic Regression -- 3. Beta-Maxent -- 4. Generalized-t Statistic -- 5. Machine Learning Methods for Imbalance Data.
Sommario/riassunto	This book presents a fresh, new approach in that it provides a comprehensive recent review of challenging problems caused by imbalanced data in prediction and classification, and also in that it introduces several of the latest statistical methods of dealing with these problems. The book discusses the property of the imbalance of data from two points of view. The first is quantitative imbalance, meaning that the sample size in one population highly outnumbers that in another population. It includes presence-only data as an extreme case, where the presence of a species is confirmed, whereas the information on its absence is uncertain, which is especially common in ecology in predicting habitat distribution. The second is qualitative imbalance, meaning that the data distribution of one population can be well specified whereas that of the other one shows a highly heterogeneous property. A typical case is the existence of outliers commonly observed in gene expression data, and another is heterogeneous characteristics often observed in a case group in case-control studies. The extension

of the logistic regression model, maxent, and AdaBoost for imbalanced data is discussed, providing a new framework for improvement of prediction, classification, and performance of variable selection. Weights functions introduced in the methods play an important role in alleviating the imbalance of data. This book also furnishes a new perspective on these problem and shows some applications of the recently developed statistical methods to real data sets.

---