

1. Record Nr.	UNINA9910350247903321
Autore	Shimura Shuichi
Titolo	High-dimensional Microarray Data Analysis : Cancer Gene Diagnosis and Malignancy Indexes by Microarray // by Shuichi Shimura
Pubbl/distr/stampa	Singapore : , : Springer Nature Singapore : , : Imprint : Springer, , 2019
ISBN	981-13-5998-9
Edizione	[1st ed. 2019.]
Descrizione fisica	1 online resource (XXV, 419 p. 261 illus., 130 illus. in color.)
Disciplina	519.5
Soggetti	Biometry Statistics Social sciences - Statistical methods Biostatistics Statistical Theory and Methods Statistics in Social Sciences, Humanities, Law, Education, Behavioral Sciences, Public Policy
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	1 New Theory of Discriminant Analysis and Cancer Gene Analysis -- 2 Overview of Cancer Gene Diagnosis by RIP and Revised LP-OLDF -- 3 Cancer Gene Diagnosis of Alon Microarray -- 4 Further Examinations of SMs---Defect of Revised LP-OLDF and Correlations of Genes -- 5 Cancer Gene Diagnosis of Golub et al. Microarray -- 6 Cancer Gene Diagnosis of Shipp et al. Microarray -- 7 Cancer Gene Diagnosis of Singh et al. Microarray -- 8 Cancer Gene Diagnosis of Tian et al. Microarray -- 9 Cancer Gene Diagnosis of Chiaretti et al. Microarray -- 10 LINGO Programs of Cancer Gene Analysis -- Index.
Sommario/riassunto	This book shows how to decompose high-dimensional microarrays into small subspaces (Small Matryoshkas, SMs), statistically analyze them, and perform cancer gene diagnosis. The information is useful for genetic experts, anyone who analyzes genetic data, and students to use as practical textbooks. Discriminant analysis is the best approach for microarray consisting of normal and cancer classes. Microarrays are linearly separable data (LSD, Fact 3). However, because most linear discriminant function (LDF) cannot discriminate LSD theoretically and

error rates are high, no one had discovered Fact 3 until now. Hard-margin SVM (H-SVM) and Revised IP-OLDF (RIP) can find Fact3 easily. LSD has the Matryoshka structure and is easily decomposed into many SMs (Fact 4). Because all SMs are small samples and LSD, statistical methods analyze SMs easily. However, useful results cannot be obtained. On the other hand, H-SVM and RIP can discriminate two classes in SM entirely. RatioSV is the ratio of SV distance and discriminant range. The maximum RatioSVs of six microarrays is over 11.67%. This fact shows that SV separates two classes by window width (11.67%). Such easy discrimination has been unresolved since 1970. The reason is revealed by facts presented here, so this book can be read and enjoyed like a mystery novel. Many studies point out that it is difficult to separate signal and noise in a high-dimensional gene space. However, the definition of the signal is not clear. Convincing evidence is presented that LSD is a signal. Statistical analysis of the genes contained in the SM cannot provide useful information, but it shows that the discriminant score (DS) discriminated by RIP or H-SVM is easily LSD. For example, the Alon microarray has 2,000 genes which can be divided into 66 SMs. If 66 DSs are used as variables, the result is a 66-dimensional data. These signal data can be analyzed to find malignancy indicators by principal component analysis and cluster analysis.
