

1. Record Nr.	UNINA9910140840803321
Autore	Dziuda Darius M
Titolo	Data mining for genomics and proteomics : analysis of gene and protein expression data // Darius M. Dzuida
Pubbl/distr/stampa	Hoboken, N.J., : Wiley, c2010
ISBN	1-282-70757-4 9786612707575 0-470-59341-5 0-470-59340-7
Descrizione fisica	1 online resource (348 p.)
Collana	Wiley Series on Methods and Applications in Data Mining ; ; v.1
Disciplina	572.8/6
Soggetti	Genomics - Data processing Proteomics - Data processing Data mining
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	DATA MINING FOR GENOMICS AND PROTEOMICS; CONTENTS; PREFACE; ACKNOWLEDGMENTS; 1 INTRODUCTION; 1.1 Basic Terminology; 1.1.1 The Central Dogma of Molecular Biology; 1.1.2 Genome; 1.1.3 Proteome; 1.1.4 DNA (Deoxyribonucleic Acid); 1.1.5 RNA (Ribonucleic Acid); 1.1.6 mRNA (messenger RNA); 1.1.7 Genetic Code; 1.1.8 Gene; 1.1.9 Gene Expression and the Gene Expression Level; 1.1.10 Protein; 1.2 Overlapping Areas of Research; 1.2.1 Genomics; 1.2.2 Proteomics; 1.2.3 Bioinformatics; 1.2.4 Transcriptomics and Other -omics . . . ; 1.2.5 Data Mining; 2 BASIC ANALYSIS OF GENE EXPRESSION MICROARRAY DATA 2.1 Introduction 2.2 Microarray Technology; 2.2.1 Spotted Microarrays; 2.2.2 Affymetrix GeneChip(®) Microarrays; 2.2.3 Bead-Based Microarrays; 2.3 Low-Level Preprocessing of Affymetrix Microarrays; 2.3.1 MAS5; 2.3.2 RMA; 2.3.3 GCRMA; 2.3.4 PLIER; 2.4 Public Repositories of Microarray Data; 2.4.1 Microarray Gene Expression Data Society (MGED) Standards; 2.4.2 Public Databases; 2.4.2.1 Gene Expression Omnibus (GEO); 2.4.2.2 ArrayExpress; 2.5 Gene Expression Matrix; 2.5.1 Elements of Gene Expression Microarray Data Analysis;

2.6 Additional Preprocessing, Quality Assessment, and Filtering
2.6.1 Quality Assessment
2.6.2 Filtering; 2.7 Basic Exploratory Data Analysis; 2.7.1 t Test; 2.7.1.1 t Test for Equal Variances; 2.7.1.2 t Test for Unequal Variances; 2.7.2 ANOVA F Test; 2.7.3 SAM t Statistic; 2.7.4 Limma; 2.7.5 Adjustment for Multiple Comparisons; 2.7.5.1 Single-Step Bonferroni Procedure; 2.7.5.2 Single-Step Sidak Procedure; 2.7.5.3 Step-Down Holm Procedure; 2.7.5.4 Step-Up Benjamini and Hochberg Procedure; 2.7.5.5 Permutation Based Multiplicity Adjustment; 2.8 Unsupervised Learning (Taxonomy-Related Analysis); 2.8.1 Cluster Analysis
2.8.1.1 Measures of Similarity or Distance
2.8.1.2 K-Means Clustering;
2.8.1.3 Hierarchical Clustering; 2.8.1.4 Two-Way Clustering and Related Methods; 2.8.2 Principal Component Analysis; 2.8.3 Self-Organizing Maps; Exercises; 3 BIOMARKER DISCOVERY AND CLASSIFICATION; 3.1 Overview; 3.1.1 Gene Expression Matrix . . . Again; 3.1.2 Biomarker Discovery; 3.1.3 Classification Systems; 3.1.3.1 Parametric and Nonparametric Learning Algorithms; 3.1.3.2 Terms Associated with Common Assumptions Underlying Parametric Learning Algorithms; 3.1.3.3 Visualization of Classification Results
3.1.4 Validation of the Classification Model
3.1.4.1 Reclassification;
3.1.4.2 Leave-One-Out and K-Fold Cross-Validation; 3.1.4.3 External and Internal Cross-Validation; 3.1.4.4 Holdout Method of Validation; 3.1.4.5 Ensemble-Based Validation (Using Out-of-Bag Samples); 3.1.4.6 Validation on an Independent Data Set; 3.1.5 Reporting Validation Results; 3.1.5.1 Binary Classifiers; 3.1.5.2 Multiclass Classifiers; 3.1.6 Identifying Biological Processes Underlying the Class Differentiation; 3.2 Feature Selection; 3.2.1 Introduction; 3.2.2 Univariate Versus Multivariate Approaches
3.2.3 Supervised Versus Unsupervised Methods

Sommario/riassunto

Data Mining for Genomics and Proteomics uses pragmatic examples and a complete case study to demonstrate step-by-step how biomedical studies can be used to maximize the chance of extracting new and useful biomedical knowledge from data. It is an excellent resource for students and professionals involved with gene or protein expression data in a variety of settings.
