

1. Record Nr.	UNINA9910633928903321
Autore	Alvo Mayer
Titolo	Statistical inference and machine learning for big data // Mayer Alvo
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2022] ©2022
ISBN	9783031067846 9783031067839
Descrizione fisica	1 online resource (442 pages)
Collana	Springer series in the data sciences
Disciplina	005.7
Soggetti	Big data Machine learning Mathematical statistics Dades massives Aprentatge automàtic Estadística matemàtica Llibres electrònics
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Intro -- Preface -- Acknowledgments -- Contents -- List of Acronyms -- List of Nomenclatures -- List of Figures -- List of Tables -- I. Introduction to Big Data -- 1. Examples of Big Data -- 1.1. Multivariate Data -- 1.2. Categorical Data -- 1.3. Environmental Data -- 1.4. Genetic Data -- 1.5. Time Series Data -- 1.6. Ranking Data -- 1.7. Social Network Data -- 1.8. Symbolic Data -- 1.9. Image Data -- II. Statistical Inference for Big Data -- 2. Basic Concepts in Probability -- 2.1. Pearson System of Distributions -- 2.2. Modes of Convergence -- 2.3. Multivariate Central Limit Theorem -- 2.4. Markov Chains -- 3. Basic Concepts in Statistics -- 3.1. Parametric Estimation -- 3.2. Hypothesis Testing -- 3.3. Classical Bayesian Statistics -- 4. Multivariate Methods -- 4.1. Matrix Algebra -- 4.2. Multivariate Analysis as a Generalization of Univariate Analysis -- 4.2.1. The General Linear Model -- 4.2.2. One Sample Problem -- 4.2.3. Two- Sample Problem -- 4.3. Structure in Multivariate Data Analysis -- 4.3.1.

Principal Component Analysis -- 4.3.2. Factor Analysis -- 4.3.3.
Canonical Correlation -- 4.3.4. Linear Discriminant Analysis -- 4.3.5.
Multidimensional Scaling -- 4.3.6. Copula Methods -- 5.
Nonparametric Statistics -- 5.1. Goodness-of-Fit Tests -- 5.2. Linear
Rank Statistics -- 5.3. U Statistics -- 5.4. Hoeffding's Combinatorial
Central Limit Theorem -- 5.5. Nonparametric Tests -- 5.5.1. One-
Sample Tests of Location -- 5.5.2. Confidence Interval for the Median
-- 5.5.3. Wilcoxon Signed Rank Test -- 5.6. Multi-Sample Tests --
5.6.1. Two-Sample Tests for Location -- 5.6.2. Multi-Sample Test for
Location -- 5.6.3. Tests for Dispersion -- 5.7. Compatibility -- 5.8.
Tests for Ordered Alternatives -- 5.9. A Unified Theory of Hypothesis
Testing -- 5.9.1. Umbrella Alternatives -- 5.9.2. Tests for Trend in
Proportions -- 5.10. Randomized Block Designs.
5.11. Density Estimation -- 5.11.1. Univariate Kernel Density
Estimation -- 5.11.2. The Rank Transform -- 5.11.3. Multivariate
Kernel Density Estimation -- 5.12. Spatial Data Analysis -- 5.12.1.
Spatial Prediction -- 5.12.2. Point Poisson Kriging of Areal Data --
5.13. Efficiency -- 5.13.1. Pitman Efficiency -- 5.13.2. Application of
Le Cam's Lemmas -- 5.14. Permutation Methods -- 6. Exponential
Tilting and Its Applications -- 6.1. Neyman Smooth Tests -- 6.2.
Smooth Models for Discrete Distributions -- 6.3. Rejection Sampling --
6.4. Tweedie's Formula: Univariate Case -- 6.5. Tweedie's Formula:
Multivariate Case -- 6.6. The Saddlepoint Approximation and Notions
of Information -- 7. Counting Data Analysis -- 7.1. Inference for
Generalized Linear Models -- 7.2. Inference for Contingency Tables --
7.3. Two-Way Ordered Classifications -- 7.4. Survival Analysis -- 7.4.1.
Kaplan-Meier Estimator -- 7.4.2. Modeling Survival Data -- 8. Time
Series Methods -- 8.1. Classical Methods of Analysis -- 8.2. State
Space Modeling -- 9. Estimating Equations -- 9.1. Composite
Likelihood -- 9.2. Empirical Likelihood -- 9.2.1. Application to One-
Sample Ranking Problems -- 9.2.2. Application to Two-Sample Ranking
Problems -- 10. Symbolic Data Analysis -- 10.1. Introduction -- 10.2.
Some Examples -- 10.3. Interval Data -- 10.3.1. Frequency -- 10.3.2.
Sample Mean and Sample Variance -- 10.3.3. Realization In SODAS --
10.4. Multi-nominal Data -- 10.4.1. Frequency -- 10.5. Symbolic
Regression -- 10.5.1. Symbolic Regression for Interval Data -- 10.5.2.
Symbolic Regression for Modal Data -- 10.5.3. Symbolic Regression in
SODAS -- 10.6. Cluster Analysis -- 10.7. Factor Analysis -- 10.8.
Factorial Discriminant Analysis -- 10.9. Application to Parkinson's
Disease -- 10.9.1. Data Processing -- 10.9.2. Result Analysis --
10.9.2.1. Viewer -- 10.9.2.2. Descriptive Statistics.
10.9.2.3. Symbolic Regression Analysis -- 10.9.2.4. Symbolic
Clustering -- 10.9.2.5. Principal Component Analysis -- 10.9.3.
Comparison with Classical Method -- 10.10. Application to
Cardiovascular Disease Analysis -- 10.10.1. Results of the Analysis --
10.10.2. Comparison with the Classical Method -- III. Machine Learning
for Big Data -- 11. Tools for Machine Learning -- 11.1. Regression
Models -- 11.2. Simple Linear Regression -- 11.2.1. Least Squares
Method -- 11.2.2. Statistical Inference on Regression Coefficients --
11.2.3. Verifying the Assumptions on the Error Terms -- 11.3. Multiple
Linear Regression -- 11.3.1. Multiple Linear Regression Model --
11.3.2. Normal Equations -- 11.3.3. Statistical Inference on Regression
Coefficients -- 11.3.4. Model Fit Evaluation -- 11.4. Regression in
Machine Learning -- 11.4.1. Optimization for Linear Regression in
Machine Learning -- 11.4.1.1. Gradient Descent -- 11.4.1.2. Feature
Standardization -- 11.4.1.3. Computing Cost on a Test Set -- 11.5.
Classification Models -- 11.5.1. Logistic Regression -- 11.5.1.1.
Optimization with Maximal Likelihood for Logistic Regression --

11.5.1.2. Statistical Inference -- 11.5.2. Logistic Regression for Binary Classification -- 11.5.2.1. Kullback-Leibler Divergence -- 11.5.3. Logistic Regression with Multiple Response Classes -- 11.5.4. Regularization for Regression Models in Machine Learning -- 11.5.4.1. Ridge Regression -- 11.5.4.2. Lasso Regression -- 11.5.4.3. The Choice of Regularization Method -- 11.5.5. Support Vector Machines (SVM) -- 11.5.5.1. Introduction -- 11.5.5.2. Finding the Optimal Hyperplane -- 11.5.5.3. SVM for Nonlinearly Separable Data Sets -- 11.5.5.4. Illustrating SVM -- 12. Neural Networks -- 12.1. Feed-Forward Networks -- 12.1.1. Motivation -- 12.1.2. Introduction to Neural Networks -- 12.1.3. Building a Deep Feed-Forward Network. 12.1.4. Learning in Deep Networks -- 12.1.4.1. Quantitative Model -- 12.1.4.2. Binary Classification Model -- 12.1.5. Generalization -- 12.1.5.1. A Machine Learning Approach to Generalization -- 12.2. Recurrent Neural Networks -- 12.2.1. Building a Recurrent Neural Network -- 12.2.2. Learning in Recurrent Networks -- 12.2.3. Most Common Design Structures of RNNs -- 12.2.4. Deep RNN -- 12.2.5. Bidirectional RNN -- 12.2.6. Long-Term Dependencies and LSTM RNN -- 12.2.7. Reduction for Exploding Gradients -- 12.3. Convolution Neural Networks -- 12.3.1. Convolution Operator for Arrays -- 12.3.1.1. Properties of the Convolution Operator -- 12.3.2. Convolution Layers -- 12.3.3. Pooling Layers -- 12.4. Text Analytics -- 12.4.1. Introduction -- 12.4.2. General Architecture -- IV. Computational Methods for Statistical Inference -- 13. Bayesian Computation Methods -- 13.1. Data Augmentation Methods -- 13.2. Metropolis-Hastings Algorithm -- 13.3. Gibbs Sampling -- 13.4. EM Algorithm -- 13.4.1. Application to Ranking -- 13.4.2. Extension to Several Populations -- 13.5. Variational Bayesian Methods -- 13.5.1. Optimization of the Variational Distribution -- 13.6. Bayesian Nonparametric Methods -- 13.6.1. Dirichlet Prior -- 13.6.2. The Poisson-Dirichlet Prior -- 13.6.3. Simulation of Bayesian Posterior Distributions -- 13.6.4. Other Applications -- Bibliography -- Index.
