

1. Record Nr.	UNINA9911019548903321
Autore	Giudici Paolo
Titolo	Applied data mining for business and industry / / Paolo Giudici, Silvia Figini
Pubbl/distr/stampa	Hoboken, NJ, : John Wiley, 2009
ISBN	9786612123696 9781282123694 1282123696 9780470058862 0470058862 9780470745830 0470745835 9780470745823 0470745827
Edizione	[2nd ed.]
Descrizione fisica	1 online resource (259 p.)
Altri autori (Persone)	FiginiSilvia
Disciplina	005.74068 006.312
Soggetti	Data mining Business - Data processing Commercial statistics
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	Applied Data Mining for Business and Industry; Contents; 1 Introduction; Part I Methodology; 2 Organisation of the data; 2.1 Statistical units and statistical variables; 2.2 Data matrices and their transformations; 2.3 Complex data structures; 2.4 Summary; 3 Summary statistics; 3.1 Univariate exploratory analysis; 3.1.1 Measures of location; 3.1.2 Measures of variability; 3.1.3 Measures of heterogeneity; 3.1.4 Measures of concentration; 3.1.5 Measures of asymmetry; 3.1.6 Measures of kurtosis; 3.2 Bivariate exploratory analysis of quantitative data 3.3 Multivariate exploratory analysis of quantitative data3.4 Multivariate exploratory analysis of qualitative data; 3.4.1

Independence and association; 3.4.2 Distance measures; 3.4.3 Dependency measures; 3.4.4 Model-based measures; 3.5 Reduction of dimensionality; 3.5.1 Interpretation of the principal components; 3.6 Further reading; 4 Model specification; 4.1 Measures of distance; 4.1.1 Euclidean distance; 4.1.2 Similarity measures; 4.1.3 Multidimensional scaling; 4.2 Cluster analysis; 4.2.1 Hierarchical methods; 4.2.2 Evaluation of hierarchical methods; 4.2.3 Non-hierarchical methods; 4.3 Linear regression; 4.3.1 Bivariate linear regression; 4.3.2 Properties of the residuals; 4.3.3 Goodness of fit; 4.3.4 Multiple linear regression; 4.4 Logistic regression; 4.4.1 Interpretation of logistic regression; 4.4.2 Discriminant analysis; 4.5 Tree models; 4.5.1 Division criteria; 4.5.2 Pruning; 4.6 Neural networks; 4.6.1 Architecture of a neural network; 4.6.2 The multilayer perceptron; 4.6.3 Kohonen networks; 4.7 Nearest-neighbour models; 4.8 Local models; 4.8.1 Association rules; 4.8.2 Retrieval by content; 4.9 Uncertainty measures and inference; 4.9.1 Probability; 4.9.2 Statistical models; 4.9.3 Statistical inference; 4.10 Non-parametric modelling; 4.11 The normal linear model; 4.11.1 Main inferential results; 4.12 Generalised linear models; 4.12.1 The exponential family; 4.12.2 Definition of generalised linear models; 4.12.3 The logistic regression model; 4.13 Log-linear models; 4.13.1 Construction of a log-linear model; 4.13.2 Interpretation of a log-linear model; 4.13.3 Graphical log-linear models; 4.13.4 Log-linear model comparison; 4.14 Graphical models; 4.14.1 Symmetric graphical models; 4.14.2 Recursive graphical models; 4.14.3 Graphical models and neural networks; 4.15 Survival analysis models; 4.16 Further reading; 5 Model evaluation; 5.1 Criteria based on statistical tests; 5.1.1 Distance between statistical models; 5.1.2 Discrepancy of a statistical model; 5.1.3 Kullback-Leibler discrepancy; 5.2 Criteria based on scoring functions; 5.3 Bayesian criteria; 5.4 Computational criteria; 5.5 Criteria based on loss functions; 5.6 Further reading; Part II Business case studies; 6 Describing website visitors; 6.1 Objectives of the analysis; 6.2 Description of the data; 6.3 Exploratory analysis; 6.4 Model building; 6.4.1 Cluster analysis

Sommario/riassunto

The increasing availability of data in our current, information overloaded society has led to the need for valid tools for its modelling and analysis. Data mining and applied statistical methods are the appropriate tools to extract knowledge from such data. This book provides an accessible introduction to data mining methods in a consistent and application oriented statistical framework, using case studies drawn from real industry projects and highlighting the use of data mining methods in a variety of business applications. Introduces data mining methods and applications.
