

Contents

1	Introduction to Machine Learning Part Two	1
1	Summary	1
1.1	Machine Learning Is Convenient	1
1.2	Limitations	1
1.3	Wonderful Methods	1
1.4	Important Methods Reviewed in Volume 1	2
1.5	Important Methods Reviewed in Volume 2	2
2	Introduction, What Is so Convenient About Machine Learning	2
3	Limitations of Machine Learning, Advantages of Traditional Statistical Methods	3
4	Machine Learning and Megadata	4
5	The Machine Learning Methods Already Reviewed in the First Volume of This Publication	4
6	The Basic Models for Machine Learning	5
7	The Machine Learning Methods to Be Reviewed in the Second Volume of This Publication	5
8	Conclusions	6
	References	7
2	Two-Stage Least Squares	9
1	Summary	9
1.1	Background	9
1.2	Objective	9
1.3	Methods	9
1.4	Result	9
1.5	Conclusions	10
2	Introduction	10
3	Example Path Analysis	10
4	Multistage Least Squares Method	13

5	Discussion	14
6	Conclusions	15
	References	15
3	Multiple Imputations	17
1	Summary	17
1.1	Background	17
1.2	Objective and Methods	17
1.3	Results	17
1.4	Conclusions	18
2	Introduction	18
3	Current Methods for Missing Data Imputation	19
4	Example	19
5	Discussion	23
6	Conclusions	23
	References	26
4	Bhattacharya Analysis	27
1	Summary	27
1.1	Background	27
1.2	Objective	27
1.3	Methods	27
1.4	Results	27
1.5	Conclusions	28
2	Introduction	28
3	Unmasking Normal Values	28
4	Improving the p-Values of Data Testing	31
5	Objectively Searching Subsets in the Data	33
6	Discussion	36
7	Conclusions	37
	References	38
5	Quality-of-Life (QOL) Assessments with Odds Ratios	39
1	Summary	39
1.1	Background	39
1.2	Objective	39
1.3	Methods and Results	39
1.4	Conclusions	40
2	Introduction, Lack of Sensitivity of QOL-Assessments	40
3	Real Data Example	40
4	Odds Ratio Analysis	41
5	Discussion	43
6	Conclusions	43
	References	44

6	Logistic Regression for Assessing Novel Diagnostic Tests	
	Against Control	45
1	Summary	45
1.1	Background	45
1.2	Objective	45
1.3	Methods	45
1.4	Results	45
1.5	Conclusions	46
2	Introduction	46
3	Example	47
4	Comparison of the Tests with Binary Logistic Regression	47
5	Comparison of the Tests with Concordance (c)-Statistics	49
6	Discussion	50
6.1	Conclusions	52
7	Conclusions	52
	References	52
7	Validating Surrogate Endpoints	53
1	Summary	53
1.1	Background	53
1.2	Objective	53
1.3	Methods	53
1.4	Results	54
2	Introduction	54
3	Some Terminology	55
4	Validating Surrogate Markers Using 95 % Confidence Intervals	56
5	Validating Surrogate Endpoints Using Regression Modeling	58
6	Calculating the Required Sample Size in a Trial with Surrogate Endpoints	60
7	Discussion	62
8	Conclusions	63
	References	64
8	Two-Dimensional Clustering	65
1	Summary	65
1.1	Background	65
1.2	Objective	65
1.3	Methods	65
1.4	Results	65
1.5	Conclusions	66
2	Introduction	66
3	Three Clustering Methodologies	67
4	Example	68
4.1	Hierarchical Cluster Analysis	68
4.2	K-means Cluster Analysis	71
4.3	Density-Based Clustering	72

5	Discussion	73
6	Conclusions	74
	References	74
9	Multidimensional Clustering	77
1	Summary	77
1.1	Background	77
1.2	Objective	77
1.3	Methods	77
1.4	Results	77
1.5	Conclusions	78
2	Introduction	78
3	Three Clustering Methodologies, as Available in SPSS Statistical Software	79
4	Example	80
4.1	Hierarchical Clustering	80
4.2	K-means Clustering	83
4.3	Two-Step Clustering	85
5	Discussion	86
6	Conclusions	90
	References	90
10	Anomaly Detection	93
1	Summary	93
1.1	Background	93
1.2	Objective	93
1.3	Methods	93
1.4	Results	93
1.5	Conclusions	94
2	Introduction	94
3	Identifying Outliers Using BIRCH Clustering, Some Theory	95
4	Simulated Data Example	95
5	Real Data Example	99
6	Discussion	100
7	Conclusions	102
	References	103
11	Association Rule Analysis	105
1	Summary	105
1.1	Background	105
1.2	Objective	105
1.3	Methods	105
1.4	Results	105
1.5	Conclusions	106
2	Introduction	106
3	Some Theory	107

4	Examples	108
5	Discussion	111
6	Conclusions	112
	References	112
12	Multidimensional Scaling	115
1	Summary	115
1.1	Background	115
1.2	Objective	115
1.3	Methods	115
1.4	Results	116
1.5	Conclusions	116
2	Introduction	116
3	Some Theory	117
3.1	Proximity Scaling	117
3.2	Preference Scaling	119
4	Data Analysis Using SPSS	119
4.1	Proximity Scaling	119
4.2	Preference Scaling	123
5	Discussion	126
6	Conclusions	127
	References	127
13	Correspondence Analysis	129
1	Summary	129
1.1	Background	129
1.2	Objective	129
1.3	Methods	129
1.4	Results	129
1.5	Conclusions	130
2	Introduction	130
3	Some Theory	131
4	Example	133
5	Discussion	135
6	Conclusions	135
	Appendix	136
	References	137
14	Multivariate Analysis of Time Series	139
1	Summary	139
1.1	Background	139
1.2	Objective	139
1.3	Methods	139
1.4	Results	139
1.5	Conclusions	140
2	Introduction	140

3	Some Theory	141
4	Example	146
5	Discussion	146
6	Conclusions	149
	Appendix	149
	References	152
15	Support Vector Machines	155
1	Summary	155
1.1	Background	155
1.2	Objective and Methods	155
1.3	Results	155
1.4	Conclusions	156
2	Introduction	156
3	Example Data	156
4	Methodological Background of Support Vector Machines	157
5	Results of Fitting a Support Vector Machine to Our Example Data...	160
6	Conclusions	160
	References	161
16	Bayesian Networks	163
1	Summary	163
1.1	Background	163
1.2	Objective	163
1.3	Methods	163
1.4	Results	163
1.5	Conclusions	164
2	Introduction	164
3	Example Data	164
4	Methodological Background of Bayesian Networks	165
5	Results of Fitting a Bayesian Network to Our Example Data	168
6	Discussion	169
7	Conclusions	170
	References	170
17	Protein and DNA Sequence Mining	171
1	Summary	171
1.1	Background	171
1.2	Objective	171
1.3	Methods	171
1.4	Results	171
1.5	Conclusions	172
2	Introduction	172
3	Gene Expression Analysis	173
4	Sequence Similarity Searching	176
4.1	Example 1	180

4.2	Example 2	181
4.3	Example 3	182
4.4	Example 4	182
5	Discussion	183
6	Conclusions	184
	References	184
18	Continuous Sequential Techniques	187
1	Summary	187
1.1	Background	187
1.2	Objective	187
1.3	Methods	187
1.4	Results	187
1.5	Conclusions	188
2	Introduction	188
3	Parallel and Triangular Sequential Sets	189
4	Data Example of a Parallel Sequential Set	190
5	Triangular Sequential Sets	192
6	Discussion	193
7	Conclusions	194
	References	194
19	Discrete Wavelet Analysis	195
1	Summary	195
1.1	Background	195
1.2	Objective	195
1.3	Methods	195
1.4	Results	196
1.5	Conclusions	196
2	Introduction	196
3	Some Theory	197
4	Clinical Relevance of Discrete Wavelet Analysis	199
5	First Example, a Single Heart Beat's Electrocardiogram	199
6	Second Example, Trends and Patterns in a Patient's Blood Glucose Levels	200
7	Third Example, Health Institutions' Ongoing Consumption Patterns	202
8	Discussion	204
9	Conclusions	205
	References	205
20	Machine Learning and Common Sense	207
1	Summary	207
1.1	Background and Objective	207
1.2	Methods and Results	207
1.3	Conclusions	207

2	Introduction	207
3	The Performance of a Validated Test in Everyday Practice	208
4	Real Data Example	209
5	Discussion	210
6	Conclusions	211
	References	211
	Statistical Tables	213
	Index	219



<http://www.springer.com/978-94-007-6886-4>

Machine Learning in Medicine

Part Two

Cleophas, T.J.; Zwinderman, A.H.

2013, XIV, 231 p.,

ISBN: 978-94-007-6886-4