

Contents

Preface	ix
1 Introduction	1
1.1 Categorical Data: Examples and Basic Concepts	1
1.2 Organization of This Book	5
1.3 Basic Components of Structured Regression	6
1.4 Classical Linear Regression	15
1.5 Exercises	27
2 Binary Regression: The Logit Model	29
2.1 Distribution Models for Binary Responses and Basic Concepts	29
2.2 Linking Response and Explanatory Variables	33
2.3 The Logit Model	37
2.4 The Origins of the Logistic Function and the Logit Model	48
2.5 Exercises	49
3 Generalized Linear Models	51
3.1 Basic Structure	51
3.2 Generalized Linear Models for Continuous Responses	53
3.3 GLMs for Discrete Responses	56
3.4 Further Concepts	60
3.5 Modeling of Grouped Data	62
3.6 Maximum Likelihood Estimation	63
3.7 Inference	67
3.8 Goodness-of-Fit for Grouped Observations	72
3.9 Computation of Maximum Likelihood Estimates	75
3.10 Hat Matrix for Generalized Linear Models	76
3.11 Quasi-Likelihood Modeling	78
3.12 Further Reading	79
3.13 Exercises	79
4 Modeling of Binary Data	81
4.1 Maximum Likelihood Estimation	82
4.2 Discrepancy between Data and Fit	87
4.3 Diagnostic Checks	93
4.4 Structuring the Linear Predictor	101
4.5 Comparing Non-Nested Models	113
4.6 Explanatory Value of Covariates	114
4.7 Further Reading	119
4.8 Exercises	120

5	Alternative Binary Regression Models	123
5.1	Alternative Links in Binary Regression	123
5.2	The Missing Link	130
5.3	Overdispersion	132
5.4	Conditional Likelihood	138
5.5	Further Reading	140
5.6	Exercises	140
6	Regularization and Variable Selection for Parametric Models	143
6.1	Classical Subset Selection	144
6.2	Regularization by Penalization	145
6.3	Boosting Methods	163
6.4	Simultaneous Selection of Link Function and Predictors	170
6.5	Categorical Predictors	173
6.6	Bayesian Approach	178
6.7	Further Reading	179
6.8	Exercises	179
7	Regression Analysis of Count Data	181
7.1	The Poisson Distribution	182
7.2	Poisson Regression Model	185
7.3	Inference for the Poisson Regression Model	186
7.4	Poisson Regression with an Offset	190
7.5	Poisson Regression with Overdispersion	192
7.6	Negative Binomial Model and Alternatives	194
7.7	Zero-Inflated Counts	198
7.8	Hurdle Models	200
7.9	Further Reading	203
7.10	Exercises	204
8	Multinomial Response Models	207
8.1	The Multinomial Distribution	209
8.2	The Multinomial Logit Model	210
8.3	Multinomial Model as Random Utility Model	215
8.4	Structuring the Predictor	215
8.5	Logit Model as Multivariate Generalized Linear Model	217
8.6	Inference for Multicategorical Response Models	218
8.7	Multinomial Models with Hierarchically Structured Response	223
8.8	Discrete Choice Models	226
8.9	Nested Logit Model	231
8.10	Regularization for the Multinomial Model	233
8.11	Further Reading	238
8.12	Exercises	239
9	Ordinal Response Models	241
9.1	Cumulative Models	243
9.2	Sequential Models	252
9.3	Further Properties and Comparison of Models	255
9.4	Alternative Models	257
9.5	Inference for Ordinal Models	261

CONTENTS	vii
9.6 Further Reading	265
9.7 Exercises	265
10 Semi- and Non-Parametric Generalized Regression	269
10.1 Univariate Generalized Non-Parametric Regression	269
10.2 Non-Parametric Regression with Multiple Covariates	285
10.3 Structured Additive Regression	289
10.4 Functional Data and Signal Regression	307
10.5 Further Reading	313
10.6 Exercises	314
11 Tree-Based Methods	317
11.1 Regression and Classification Trees	317
11.2 Multivariate Adaptive Regression Splines	328
11.3 Further Reading	329
11.4 Exercises	329
12 The Analysis of Contingency Tables: Log-Linear and Graphical Models	331
12.1 Types of Contingency Tables	332
12.2 Log-Linear Models for Two-Way Tables	335
12.3 Log-Linear Models for Three-Way Tables	338
12.4 Specific Log-Linear Models	341
12.5 Log-Linear and Graphical Models for Higher Dimensions	345
12.6 Collapsibility	348
12.7 Log-Linear Models and the Logit Model	349
12.8 Inference for Log-Linear Models	350
12.9 Model Selection and Regularization	354
12.10 Mosaic Plots	357
12.11 Further Reading	358
12.12 Exercises	359
13 Multivariate Response Models	363
13.1 Conditional Modeling	365
13.2 Marginal Parametrization and Generalized Log-Linear Models	370
13.3 General Marginal Models: Association as Nuisance and GEEs	371
13.4 Marginal Homogeneity	385
13.5 Further Reading	392
13.6 Exercises	393
14 Random Effects Models and Finite Mixtures	395
14.1 Linear Random Effects Models for Gaussian Data	396
14.2 Generalized Linear Mixed Models	402
14.3 Estimation Methods for Generalized Mixed Models	407
14.4 Multicategorical Response Models	416
14.5 The Marginalized Random Effects Model	419
14.6 Latent Trait Models and Conditional ML	420
14.7 Semiparametric Mixed Models	420
14.8 Finite Mixture Models	422
14.9 Further Reading	426
14.10 Exercises	427

15 Prediction and Classification	429
15.1 Basic Concepts of Prediction	430
15.2 Methods for Optimal Classification	438
15.3 Basics of Estimated Classification Rules	445
15.4 Parametric Classification Methods	451
15.5 Non-Parametric Methods	457
15.6 Neural Networks	468
15.7 Examples	471
15.8 Variable Selection in Classification	473
15.9 Prediction of Ordinal Outcomes	474
15.10 Model-Based Prediction	480
15.11 Further Reading	481
15.12 Exercises	482
A Distributions	485
A.1 Discrete Distributions	485
A.2 Continuous Distributions	487
B Some Basic Tools	490
B.1 Linear Algebra	490
B.2 Taylor Approximation	491
B.3 Conditional Expectation, Distribution	493
B.4 EM Algorithm	494
C Constrained Estimation	496
C.1 Simplification of Penalties	496
C.2 Linear Constraints	498
C.3 Fisher Scoring with Penalty Term	499
D Kullback-Leibler Distance and Information-Based Criteria of Model Fit	500
D.1 Kullback-Leibler Distance	500
E Numerical Integration and Tools for Random Effects Modeling	504
E.1 Laplace Approximation	504
E.2 Gauss-Hermite Integration	505
E.3 Inversion of Pseudo-Fisher Matrix	507
List of Examples	509
Bibliography	513
Author Index	545
Subject Index	554