

Generalized Additive Models

T.J. HASTIE

*AT&T Bell Laboratories
New Jersey*

and

R.J. TIBSHIRANI

*Department of Preventive Medicine and Biostatistics
and Department of Statistics
University of Toronto*

C

CHAPMAN & HALUCRC

Boca Raton London New York Washington, D.C.

Contents

Preface	xiii
1 Introduction	1
1.1 What's in this book	1
1.2 A word of caution	6
1.3 How to read this book	6
1.4 Data sets used in this book	6
1.5 Notation	7
2 Smoothing	9
2.1 What is a smoother?	9
2.2 Scatterplot smoothing: definition.	13
2.3 Parametric regression	14
2.4 Bin smoothers	14
2.5 Running-mean and running-line smoothers	15
2.6 Kernel smoothers $p <$	18
2.7 Running medians-and enhancements	20
2.8 Equivalent kernels •	20
2.9 Regression splines	22
2.10 Cubic smoothing splines •	27
2.11 Locally-weighted running-line smoothers	29
2.12 Smoothers for multiple predictors	32
2.13 Bibliographic notes	34
2.14 Further results and exercises ,2.	35
3 Smoothing in detail	39
3.1 Introduction '	39
3.2 A formal model for scatterplot smoothing	39
3.3 The bias-variance trade-off '	40

3.4	Automatic selection of smoothing parameters	42
3.4.1	Cross-validation	42
3.4.2	The bias-variance trade-off for linear smoothers	44
3.4.3	Cross-validation for linear smoothers	46
3.4.4	The C_p statistic	48
3.4.5	Variations on cross-validation and a comparison with C_p	49
3.4.6	Discussion	52
3.5	Degrees of freedom of a smoother	52
3.6	A Bayesian model for smoothing	55
3.7	Eigenanalysis of a smoother and spectral smoothing	57
3.8	Variance of a smooth and confidence bands	60
3.8.1	Pointwise standard-error bands	60
3.8.2	Global confidence sets	61
3.9	Approximate F-tests	65
3.10	Asymptotic behaviour of smoothers	68
3.11	Special topics	70
3.11.1	Nonlinear smoothers	70
3.11.2	Kriging	71
3.11.3	Smoothing and penalized least-squares	72
3.11.4	Weighted smoothing	72
3.11.5	Tied predictor values	74
3.11.6	Resistant smoothing	74
3.12	Bibliographical notes	76
3.13	Further results and exercises	79
	Additive models	82
4.1	Introduction	82
4.2	Multiple regression and linear models	82
4.3	Additive models	86
4.4	Fitting additive models	89
4.5	Generalized additive models: logistic regression	95
4.6	Bibliographic notes	102
4.7	Further results and exercises	103
	Some theory for additive models	105
5.1	Introduction	105
5.2	Estimating equations for additive models	106
5.2.1	L_2 function spaces	107
5.2.2	Penalized least-squares	110

5.2.3	Reproducing-kernel Hilbert-spaces	112
5.3	Solutions to the estimating equations	114
5.3.1	Introduction	114
5.3.2	Projection smoothers	115
5.3.3	Semi-parametric models	118
5.3.4	Backfitting with two smoothers	118
5.3.5	Existence and uniqueness of p-smoothers	121
5.3.6	Convergence of backfitting: p-smoothers	122
5.3.7	Summary of the main results of the section	123
5.4	Special topics	124
5.4.1	Weighted additive models	124
5.4.2	A modified backfitting algorithm	124
5.4.3	Explicit solutions to the estimating equations	126
5.4.4	Standard errors	127
5.4.5	Degrees of freedom	128
5.4.6	A Bayesian version of additive models	129
5.5	Bibliographic notes	130
5.6	Further results and exercises	132
6	Generalized additive models	136
6.1	Introduction	136
6.2	Fisher scoring for generalized linear models	137
6.3	Local scoring for generalized additive models	140
6.4	Illustrations	141
6.4.1	Clotting times of blood	141
6.4.2	Warm cardioplegia	143
6.5	Derivation of the local-scoring procedure	148
6.5.1	L_2 function spaces	148
6.5.2	Penalized likelihood	149
6.6	Convergence of the local-scoring algorithm	151
6.7	Semi-parametric generalized linear models	152
6.8	Inference	155
6.8.1	Analysis of deviance	155
6.8.2	Standard error bands	156
6.8.3	Degrees of freedom	157
6.9	Smoothing parameter selection	159
6.10	Overinterpreting additive fits	161
6.11	Missing predictor values	166
6.12	Estimation of the link function	166

6.13	Local-likelihood estimation	167
6.14	Bibliographic notes	169
6.15	Further results and exercises 6	171
	Response transformation models	174
7.1	Introduction	174
7.2	The ACE algorithm	175
7.2.1	Introduction	175
7.2.2	ACE in L_2 function spaces	179
7.2.3	ACE and penalized least-squares	181
7.2.4	Convergence of ACE with linear smoothers	182
7.2.5	A close ancestor to ACE: canonical correlation	183
7.2.6	Some anomalies of ACE	184
7.3	Response transformations for regression	187
7.3.1	Introduction	187
7.3.2	Generalizations of the Box-Cox procedure	187
7.3.3	Comparison of generalized Box-Cox and ACE	189
7.4	Additivity and variance stabilization	190
7.4.1	Some properties of AVAS	193
7.5	Further topics	194
7.5.1	Prediction from a transformation model	194
7.5.2	Methods for inference	195
7.6	Bibliographical notes	196
7.7	Further results and exercises 7	197
	Extensions to other settings	201
8.1	Introduction	201
8.2	Matched case-control data	202
8.2.1	Background	202
8.2.2	Estimation	203
8.2.3	Maximizing the conditional, likelihood for the linear model	204
8.2.4	Spline estimation for a single function	205
8.2.5	Algorithm for the additive model	207
8.2.6	A simulated example	209
8.3	The proportional-hazards model	211
8.3.1	Background	211
8.3.2	Estimation	211
8.3.3	Further details of the computations	213
8.3.4	An example	214

8.4	The proportional-odds model	219
8.4.1	Background	219
8.4.2	Fitting the additive model	220
8.4.3	Illustration	223
8.5	Seasonal decomposition of time series	224
8.5.1	The STL procedure	224
8.5.2	Eigen-analysis and bandwidth selection	230
8.6	Bibliographic notes	231
8.7	Further results and exercises 8	232
9	Further topics	235
9.1	Introduction	235
9.2	Resistant fitting	236
9.2.1	Resistant fitting of additive models	236
9.2.2	Resistant fitting of generalized additive models	240
9.2.3	Illustration	242
9.2.4	Influence and resistance	244
9.3	Parametric additive models	245
9.3.1	Regression splines	246
9.3.2	Simple knot-selection schemes for regression splines	247
9.3.3	A simulated example	247
9.3.4	Adaptive knot-selection strategies	249
9.3.5	Discussion of regression splines	251
9.3.6	Generalized ridge regression — pseudo additive models,	254
9.3.7	Illustration: diagnostics for additive models	256
9.4	Model selection techniques	259
9.4.1	Backward and forward stepwise selection techniques	260
9.4.2	Adaptive regression splines: TURBO'	261
9.4.3	Adaptive backfitting: BRUTO	262
9.5	Modelling interactions	264
9.5.1	Simple interactions	264
9.5.2	Interactions resulting in separate curves	265
9.5.3	Hierarchical models	266
9.5.4	Examining residuals for interactions	268
9.5.5	Illustration	268
9.5.6	Regression trees	271

9.5.7 Multivariate adaptive regression splines: MARS	, 275
9.6 Bibliographic notes	277
9.7 Further results and exercises	278
10 Case studies	281
10.1 Introduction	281
10.2 Kyphosis in laminectomy patients	282
10.3 Atmospheric ozone concentration	294
Appendices	301
A Data	301
B Degrees of freedom	305
C Software	307
References	311
Author index	325
Subject index	329