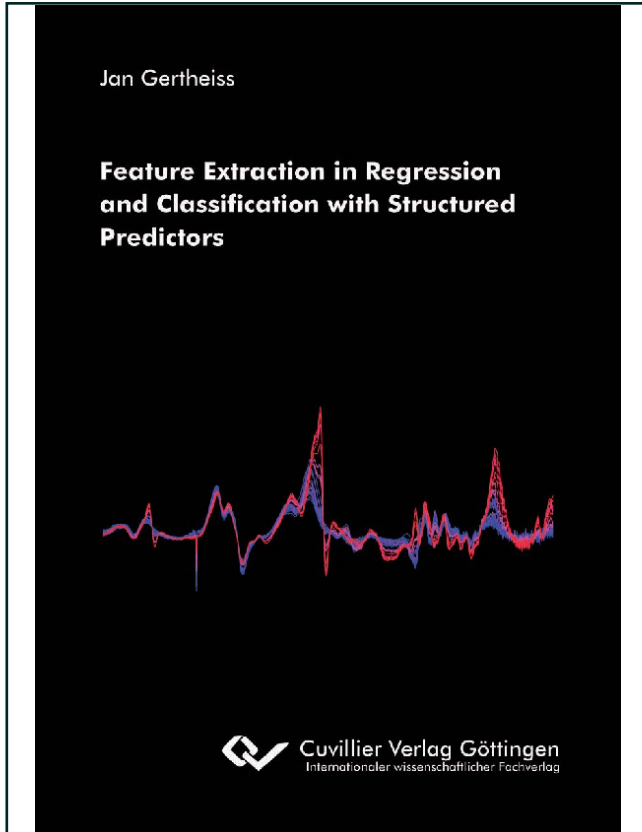




Jan Gertheiss (Autor)  
**Feature Extraction in Regression and Classification with  
Structured Predictors**



<https://cuvillier.de/de/shop/publications/452>

Copyright:

Cuvillier Verlag, Inhaberin Annette Jentsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen, Germany  
Telefon: +49 (0)551 54724-0, E-Mail: [info@cuvillier.de](mailto:info@cuvillier.de), Website: <https://cuvillier.de>

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Signal-like Predictors</b>	<b>7</b>
2.1	Introduction . . . . .	8
2.2	Feature Extraction in Signal Regression . . . . .	9
2.2.1	Introduction . . . . .	9
2.2.2	Feature Extraction and Blockwise Boosting . . . . .	12
2.2.3	Simulation Studies . . . . .	17
2.2.4	Evaluation by Real-World Data . . . . .	22
2.2.5	Summary and Discussion . . . . .	27
2.3	Supervised Feature Selection and Classification in Proteomics . . . . .	30
2.3.1	Introduction . . . . .	30
2.3.2	Methods for Binary Classification . . . . .	31
2.3.3	Application to Mass Spectrometry Data . . . . .	33
2.3.4	Generalizations to Other Non-Normal Outcomes . . . . .	36
2.3.5	Computational Issues . . . . .	38
2.3.6	Summary and Discussion . . . . .	40
<b>3</b>	<b>Categorical Predictors</b>	<b>41</b>
3.1	Introduction . . . . .	42
3.2	Quadratic Regularization for Ordinal Predictors . . . . .	44
3.2.1	Introduction . . . . .	44
3.2.2	Penalized Regression with Ordinal Predictors . . . . .	47
3.2.3	Simulation Studies . . . . .	51
3.2.4	Bias-Variance Calculations . . . . .	53
3.2.5	Applications to Real-World Data . . . . .	55
3.2.6	Handling Non-Normal Responses . . . . .	58
3.2.7	Summary and Discussion . . . . .	64
3.2.8	Appendix: Bias, Variances and an Approximate Hat Matrix . . . . .	65
3.3	Selection of Ordinally Scaled Independent Variables . . . . .	67
3.3.1	Introduction . . . . .	67
3.3.2	Ordinal Covariates . . . . .	69
3.3.3	Methods Proposed . . . . .	71
3.3.4	Application to the ICF Core Set for Chronic Widespread Pain . . . . .	76
3.3.5	Summary and Discussion . . . . .	81

3.3.6	Appendix: The ICF Core Sets for CWP . . . . .	83
3.4	Sparse Modeling of Categorical Explanatory Variables . . . . .	87
3.4.1	Introduction . . . . .	87
3.4.2	Lasso-type Regularization for Categorical Predictors . . . . .	90
3.4.3	Numerical Experiments . . . . .	96
3.4.4	Regularized Analysis of Munich Rent Standard Data . . . . .	103
3.4.5	Summary and Discussion . . . . .	109
3.4.6	Appendix: Propositions and Proofs . . . . .	110
3.5	Regularization and Model Selection with Categorical Effect Modifiers . . . . .	118
3.5.1	Introduction . . . . .	118
3.5.2	Penalized Estimation . . . . .	121
3.5.3	Large Sample Properties and Modifications . . . . .	124
3.5.4	Numerical Experiments . . . . .	126
3.5.5	Real World Data Evaluation . . . . .	130
3.5.6	Generalizations to Multiple Effect Modifiers . . . . .	134
3.5.7	Summary and Discussion . . . . .	136
3.5.8	Appendix: Derivations and Proofs . . . . .	138
3.6	Generalizations to Non-Normal Outcomes . . . . .	143
3.6.1	Introduction . . . . .	143
3.6.2	Penalized Likelihood Estimation . . . . .	144
3.6.3	Illustrative Examples . . . . .	144
3.6.4	Summary and Discussion . . . . .	149
<b>4</b>	<b>Nonparametric Feature Extraction</b>	<b>153</b>
4.1	Introduction . . . . .	154
4.2	Variable Scaling and Nearest Neighbor Methods . . . . .	156
4.2.1	Introduction . . . . .	156
4.2.2	Scaling by Pooled Variances . . . . .	156
4.2.3	Simulation Results . . . . .	157
4.2.4	Summary and Discussion . . . . .	159
4.3	Feature Selection and Weighting by Nearest Neighbor Ensembles . . . . .	160
4.3.1	Introduction . . . . .	160
4.3.2	Nearest Neighbor Ensembles . . . . .	160
4.3.3	Simulation Studies . . . . .	165
4.3.4	Evaluation of Real-World Data . . . . .	171
4.3.5	High-dimensional Settings . . . . .	175
4.3.6	Regression Problems . . . . .	180
4.3.7	Summary and Discussion . . . . .	180
4.3.8	Appendix: A Proposition about Class-Specific Weights . . . . .	181
<b>5</b>	<b>Conclusion and Outlook</b>	<b>183</b>
	<b>References</b>	<b>187</b>