
Contents

Deutsche Zusammenfassung	v
Abstract	vii
Acknowledgments	ix
Foreword	1
1 Introduction	5
1.1 Classification and Regression	5
1.1.1 Classification	7
1.1.2 Regression	8
1.2 Empirical Risk Minimization	10
1.2.1 The Curses of Dimensionality and Dataset Sparsity	10
1.3 Bayesian Inference	13
1.3.1 Example: Bayesian Neural Networks	17
2 Automatic Relevance Determination	25
2.1 The ARD Framework	26
2.1.1 Bayesian Neural Networks	27
2.1.2 Choice of Network Priors	29
2.1.3 Automatic Relevance Determination Models	30
2.2 Bayesian Linear Regression in High Dimensional Spaces	32
2.2.1 Prediction Model	32
2.2.2 Hierarchical Prior Distribution	34
2.2.3 Maximum A-Posteriori Hypothesis	38
2.2.4 Simulation Results	40

3	Predicting Cancer Survival Time	43
3.1	Gene Expression and Cancer	44
3.2	Models of Survival	45
3.2.1	The Cox Regression Model	46
3.3	Handling of Censored Observations	48
3.3.1	Estimating the Survivor Function $F(t)$ from Data	49
3.3.2	Estimating Cox' θ under Censoring	51
3.4	Bayesian Regression with Censored Data	52
3.5	Results	54
3.5.1	Datasets used	55
3.5.2	Results – Simulated Data	62
3.5.3	Results – Diffuse Large B-Cell Lymphoma	67
3.5.4	Results – Adenocarcinoma of the Lung	80
4	Markov Chain Monte Carlo Integration	91
4.1	Introduction	91
4.2	Basic Markov Chain Theory	92
4.3	Gibbs Sampling	96
4.4	Metropolis-Hastings	97
4.5	The Hybrid Monte Carlo Algorithm	99
5	Feature Relevance and Label Density	105
5.1	Feature Relevance	106
5.1.1	Linear Model	108
5.1.2	Survival Time Regression Model	109
5.2	The Label Density $p(y x, \mathcal{D})$	110
5.2.1	Linear Model	112
5.2.2	Survival Time Regression Model	113
5.3	Results	114
5.3.1	Linear Model	115
5.3.2	Survival Time Regression Model	121
6	Further Extensions and Discussion	139
6.1	Related Problems and Further Work	139
6.1.1	Adaptations for Classification	139
6.1.2	Learning of Gene Regulatory Networks	141
6.1.3	Other Extensions and Applications	144
6.2	Discussion	146
6.2.1	A Subjective View on Model Advantages	147
6.2.2	An Equally Subjective View on Limitations	150
6.2.3	Lessons Learned	152
6.2.4	Open Questions	156

A Mathematical Appendix	159
A.1 Numerical Methods	159
A.1.1 Conjugate Gradient Descent	159
A.1.2 Downhill Simplex Method	160
A.1.3 Laplace Approximation	160
A.1.4 Gaussian Quadrature	161
A.2 Gaussian Integrals	163
A.2.1 $x \in \mathbf{R}$	163
A.2.2 $x \in \mathbf{R}^n$	164
A.2.3 Inclusion of a Linear Term	165
A.2.4 Individual Integrals	165
A.3 Time Dependent ROC Curves	167
A.4 Comparing Survival Curves	168
B DLBCL Cancer Genes	171
B.1 MAP Analysis - Highly ranked Genes	171
B.2 Repetitive Analysis – Frequent Genes	174
B.3 Repetitive Analysis – High Genes	177
C Adenocarcinoma of the Lung Genes	181
C.1 MAP Hypothesis – Genes	181
C.2 Discussion of Genes	183
Bibliography	189
Index	201
List of Symbols	205
Abbreviations	209
Erklärung	211