

Contents

1	Introduction	1
1.1	Data Classification	1
1.2	Data-Based Learning	4
1.3	Outline of the Book	9
2	Continuous Risk Functionals	13
2.1	Classic Risk Functionals	14
2.1.1	The Mean-Square-Error Risk	14
2.1.2	The Cross-Entropy Risk	19
2.2	Risk Functionals Reappraised	21
2.2.1	The Error Distribution	21
2.2.2	MSE and CE Risks in Terms of the Error	22
2.3	The Error-Entropy Risk	26
2.3.1	EE Risks	26
2.3.2	EE Risks and Information Theoretic Learning	28
2.3.3	MEE Is Harder for Classification than for Regression	29
2.3.4	The Quest for Minimum Entropy	31
2.3.5	Risk Functionals and Probability of Error	34
3	MEE with Continuous Errors	41
3.1	Theoretical and Empirical MEE	42
3.1.1	Computational Issues	42
3.1.2	Empirical EE Gradient Descent	44
3.1.3	Fat Estimation of Error PDF	45
3.2	The Linear Discriminant	50
3.2.1	Gaussian Inputs	50
3.2.2	Consistency and Generalization	55
3.3	The Continuous-Output Perceptron	59
3.3.1	Motivational Examples	60
3.3.2	Theoretical and Empirical MEE Behaviors	67
3.3.3	The Arctangent Perceptron	71

3.4	The Hypersphere Neuron	80
3.4.1	Motivational Examples	80
3.4.2	Theoretical and Empirical MEE in Realistic Datasets	84
3.5	The Data Splitter	86
3.6	Kernel Smoothing Revisited	88
4	MEE with Discrete Errors	93
4.1	The Data Splitter Setting	94
4.1.1	Motivational Examples	96
4.1.2	SEE Critical Points	98
4.1.3	Empirical SEE Splits	104
4.1.4	MEE Splits and Classic Splits	110
4.2	The Discrete-Output Perceptron	116
5	EE-Inspired Risks	121
5.1	Zero-Error Density Risk	122
5.1.1	Motivational Examples	122
5.1.2	Empirical ZED	124
5.2	A Parameterized Risk	133
5.2.1	The Exponential Risk Functional	133
6	Applications	139
6.1	MLPs with Error Entropy Risks	139
6.1.1	Back-Propagation of Error Entropy	140
6.1.2	The Batch-Sequential Algorithm	149
6.1.3	Experiments with Real-World Datasets	152
6.2	Recurrent Neural Networks	156
6.2.1	Real Time Recurrent Learning	157
6.2.2	Long Short-Term Memory	161
6.3	Complex-Valued Neural Network	170
6.3.1	Introduction	170
6.3.2	Single Layer Complex-Valued NN	171
6.3.3	MMSE Batch Algorithm	172
6.3.4	MEE Algorithm	173
6.3.5	Experiments	174
6.4	An Entropic Clustering Algorithm	176
6.4.1	Introduction	176
6.4.2	Overview of Clustering Algorithms	177
6.4.3	Dissimilarity Matrix	178
6.4.4	The LEGClust Algorithm	179
6.4.5	Experiments	192
6.5	Task Decomposition and Modular Neural Networks	199
6.5.1	Modular Neural Networks	199
6.5.2	Experiments	201
6.6	Decision Trees	204

6.6.1	The MEE Tree Algorithm	205
6.6.2	Application to Real-World Datasets	208
A	Maximum Likelihood and Kullback-Leibler Divergence	215
A.1	Maximum Likelihood	215
A.2	Kullback-Leibler Divergence	216
A.3	Equivalence of ML and KL Empirical Estimates	216
B	Properties of Differential Entropy	219
B.1	Shannon's Entropy	219
B.2	Rényi's Entropy	220
C	Entropy and Variance of Partitioned PDFs	221
D	Entropy Dependence on the Variance	223
D.1	Saturating Functions	223
D.2	PDF Families with Up-Saturating $H(V)$	225
D.2.1	The Pearson System	225
D.2.2	List of PDFs with USF $H(V)$	229
E	Optimal Parzen Window Estimation	235
E.1	Parzen Window Estimation	235
E.2	Optimal Bandwidth and IMSE	238
F	Entropy Estimation	241
F.1	Integral and Plug-in Estimates	241
F.2	Integral Estimate of Rényi's Quadratic Entropy	241
F.3	Plug-in Estimate of Shannon's Entropy	243
F.4	Plug-in Estimate of Rényi's Entropy	244
	References	245
	Index	257