
Contents

1	Information Theory, Machine Learning, and Reproducing Kernel Hilbert Spaces	1
1.1	Introduction	1
1.2	Information Theory	7
1.3	Entropy	10
1.4	Mutual Information	14
1.5	Relative Entropy and Kullback–Leibler Divergence	16
1.6	Information Theory beyond Communications	19
1.7	Adaptive Model Building	24
1.8	Information-Theoretic Learning	29
1.9	ITL as a Unifying Learning Paradigm	30
1.10	Reproducing Kernel Hilbert Spaces	34
1.11	RKHS and ITL	38
1.12	Conclusions	44
2	Renyi’s Entropy, Divergence and Their Nonparametric Estimators	47
	<i>Chapter Coauthors: Dongxin Xu and Deniz Erdogmus</i>	
2.1	Introduction	47
2.2	Definition and Interpretation of Renyi’s Entropy	48
2.3	Quadratic Renyi’s Entropy Estimator	56
2.4	Properties of Renyi’s Nonparametric Entropy Estimators	60
2.5	Bias and Variance of the Information Potential Estimator	70
2.6	Physical Interpretation of Renyi’s Entropy Kernel Estimators	73
2.7	Extension to α -Information Potential with Arbitrary Kernels	79
2.8	Renyi’s Divergence and Mutual Information	80
2.9	Quadratic Divergences and Mutual Information	84
2.10	Information Potentials and Forces in the Joint Space	88
2.11	Fast Computation of IP and CIP	96
2.12	Conclusion	101

3	Adaptive Information Filtering with Error Entropy and Error Correntropy Criteria	103
	<i>Chapter Coauthors: Deniz Erdogmus and Weifeng Liu</i>	
3.1	Introduction	103
3.2	The Error Entropy Criterion (EEC) for Adaptation	104
3.3	Understanding the Error Entropy Criterion	105
3.4	Minimum Error Entropy Algorithm	111
3.5	Analysis of MEE Performance Surface	113
3.6	Error Entropy, Correntropy, and M Estimation	123
3.7	Correntropy Induced Metric and M-Estimation	128
3.8	Normalized Information Potential as a Pseudometric	131
3.9	Adaptation of the Kernel Size in Adaptive Filtering	135
3.10	Conclusions	139
4	Algorithms for Entropy and Correntropy Adaptation with Applications to Linear Systems	141
	<i>Chapter Coauthors: Deniz Erdogmus, Seungju Han, and Abhishek Singh</i>	
4.1	Introduction	141
4.2	Recursive Information Potential for MEE (MEE-RIP)	142
4.3	Stochastic Information Gradient for MEE (MEE-SIG)	146
4.4	Self-Adjusting Stepsize for MEE (MEE-SAS)	152
4.5	Normalized MEE (NMEE)	157
4.6	Fixed-Point MEE (MEE-FP)	162
4.7	Fast Gauss Transform in MEE Adaptation	167
4.8	Incomplete Cholesky Decomposition for MEE	170
4.9	Linear Filter Adaptation with MSE, MEE and MCC	171
4.10	Conclusion	177
5	Nonlinear Adaptive Filtering with MEE, MCC, and Applications	181
	<i>Chapter Coauthors: Deniz Erdogmus, Rodney Morejon, and Weifeng Liu</i>	
5.1	Introduction	181
5.2	Backpropagation of Information Forces in MLP Training	183
5.3	Advanced Search Methods for Nonlinear Systems	186
5.4	ITL Advanced Search Algorithms	190
5.5	Application: Prediction of the Mackey–Glass Chaotic Time Series	198
5.6	Application: Nonlinear Channel Equalization	204
5.7	Error Correntropy Criterion (ECC) in Regression	208
5.8	Adaptive Kernel Size in System Identification and Tracking	213
5.9	Conclusions	217

6 Classification with EEC, Divergence Measures, and Error Bounds 219
Chapter Coauthors: Deniz Erdogmus, Dongxin Xu, and Kenneth Hild II

6.1 Introduction 219

6.2 Brief Review of Classification 220

6.3 Error Entropy Criterion in Classification 222

6.4 Nonparametric Classifiers 228

6.5 Classification with Information Divergences 231

6.6 ITL Algorithms for Divergence and Mutual Information 233

6.6.1 Case Study: Automatic Target Recognition (ATR) with ITL 237

6.7 The Role of ITL Feature Extraction in Classification 246

6.8 Error Bounds for Classification 253

6.9 Conclusions 260

7 Clustering with ITL Principles 263
Chapter Coauthors: Robert Jenssen and Sudhir Rao

7.1 Introduction 263

7.2 Information-Theoretic Clustering 264

7.3 Differential Clustering Using Renyi’s Entropy 266

7.4 The Clustering Evaluation Function 267

7.5 A Gradient Algorithm for Clustering with D_{CS} 273

7.6 Mean Shift Algorithms and Renyi’s Entropy 282

7.7 Graph-Theoretic Clustering with ITL 286

7.8 Information Cut for Clustering 292

7.9 Conclusion 297

8 Self-Organizing ITL Principles for Unsupervised Learning . . 299
Chapter Coauthors: Sudhir Rao, Deniz Erdogmus, Dongxin Xu, and Kenneth Hild II

8.1 Introduction 299

8.2 Entropy and Cross-Entropy Optimization 301

8.3 The Information Maximization Principle 304

8.4 Exploiting Spatial Structure for Self-Organization 307

8.5 Principle of Redundancy Reduction 308

8.6 Independent Component Analysis (ICA) 310

8.7 The Information Bottleneck (IB) Method 313

8.8 The Principle of Relevant Information (PRI) 315

8.9 Self-Organizing Principles with ITL Estimators 329

8.10 Conclusions 349

9	A Reproducing Kernel Hilbert Space Framework for ITL . . .	351
	<i>Chapter Coauthors: Jianwu Xu, Robert Jenssen, Antonio Paiva, and Il Park</i>	
9.1	Introduction	351
9.2	A RKHS Framework for ITL	353
9.3	ITL Cost Functions in the RKHS Framework	358
9.4	ITL Estimators in RKHS	360
9.5	Connection Between ITL and Kernel Methods via RKHS H_{ν} .	364
9.6	An ITL Perspective of MAP and SVM Classifiers	368
9.7	Case Study: RKHS for Computation with Spike Train	376
9.8	Conclusion	383
10	Correntropy for Random Variables: Properties and Applications in Statistical Inference	385
	<i>Chapter Coauthors: Weifeng Liu, Puskal Pokharel, Jianwu Xu, and Sohan Seth</i>	
10.1	Introduction	385
10.2	Cross-Correntropy: Definitions and Properties	386
10.3	Centered Cross-Correntropy and Correntropy Coefficient	396
10.4	Parametric Cross-Correntropy and Measures of Dependence . .	399
10.5	Application: Matched Filtering	402
10.6	Application: Nonlinear Coupling Tests	409
10.7	Application: Statistical Dependence Tests	410
10.8	Conclusions	412
11	Correntropy for Random Processes: Properties and Applications in Signal Processing	415
	<i>Chapter Coauthors: Puskal Pokharel, Ignacio Santamaria, Jianwu Xu, Kyu-hwa Jeong, and Weifeng Liu</i>	
11.1	Introduction	415
11.2	Autocorrentropy Function: Definition and Properties	417
11.3	Cross-Correntropy Function: Definition and Properties	424
11.4	Optimal Linear Filters in H_{ν}	425
11.5	Correntropy MACE (CMACE) Filter in H_{ν}	427
11.6	Application: Autocorrentropy Function as a Similarity Measure over Lags	431
11.7	Application: Karhunen-Loeve Transform in H_{ν}	438
11.8	Application: Blind Source Separation	442
11.9	Application: CMACE for Automatic Target Recognition	448
11.10	Conclusion	454
A	PDF Estimation Methods and Experimental Evaluation of ITL Descriptors	457
	<i>Chapter Coauthors: Deniz Erdogmus and Rodney Morejon</i>	
A.1	Introduction	457
A.2	Probability Density Function Estimation	457

A.3	Nonparametric Entropy Estimation	472
A.4	Estimation of Information–Theoretic Descriptors	474
A.5	Convolution Smoothing	494
A.6	Conclusions	497
Bibliography		499
Index		517



<http://www.springer.com/978-1-4419-1569-6>

Information Theoretic Learning
Renyi's Entropy and Kernel Perspectives
Principe, J.C.
2010, XIV, 448 p., Hardcover
ISBN: 978-1-4419-1569-6