

Contents

1	Introduction	1
1.1	Two Fundamental Questions.	1
1.1.1	Why Should One Read the Book?	1
1.1.2	What Is the Book About?	2
1.2	The Structure of the Book	4
1.2.1	Part I: From Perception to Computation	4
1.2.2	Part II: Machine Learning.	5
1.2.3	Part III: Applications	6
1.2.4	Appendices.	7
1.3	How to Read This Book	8
1.3.1	Background and Learning Objectives.	8
1.3.2	Difficulty Level.	8
1.3.3	Problems	9
1.3.4	Software.	9
1.4	Reading Tracks.	9

Part I From Perception to Computation

2	Audio Acquisition, Representation and Storage	13
2.1	Introduction	13
2.2	Sound Physics, Production and Perception	15
2.2.1	Acoustic Waves Physics.	15
2.2.2	Speech Production.	18
2.2.3	Sound Perception	20
2.3	Audio Acquisition	22
2.3.1	Sampling and Aliasing.	23
2.3.2	The Sampling Theorem**	25
2.3.3	Linear Quantization	28
2.3.4	Nonuniform Scalar Quantization	30

2.4	Audio Encoding and Storage Formats	32
2.4.1	Linear PCM and Compact Discs	33
2.4.2	MPEG Digital Audio Coding	34
2.4.3	AAC Digital Audio Coding	35
2.4.4	Perceptual Coding	36
2.5	Time-Domain Audio Processing	38
2.5.1	Linear and Time-Invariant Systems	39
2.5.2	Short-Term Analysis	40
2.5.3	Time-Domain Measures	43
2.6	Linear Predictive Coding	47
2.6.1	Parameter Estimation	50
2.7	Conclusions	52
	Problems	52
	References	53
3	Image and Video Acquisition, Representation and Storage	57
3.1	Introduction	57
3.2	Human Eye Physiology	58
3.2.1	Structure of the Human Eye	58
3.3	Image Acquisition Devices	60
3.3.1	Digital Camera	60
3.4	Color Representation	63
3.4.1	Human Color Perception	63
3.4.2	Color Models	64
3.5	Image Formats	76
3.5.1	Image File Format Standards	76
3.5.2	JPEG Standard	77
3.6	Image Descriptors	81
3.6.1	Global Image Descriptors	81
3.6.2	SIFT Descriptors	85
3.7	Video Principles	88
3.8	MPEG Standard	89
3.8.1	Further MPEG Standards	90
3.9	Conclusions	93
	Problems	93
	References	95

Part II Machine Learning

4	Machine Learning	99
4.1	Introduction	99
4.2	Taxonomy of Machine Learning	100

4.2.1	Rote Learning	100
4.2.2	Learning from Instruction	101
4.2.3	Learning by Analogy	101
4.3	Learning from Examples	101
4.3.1	Supervised Learning	102
4.3.2	Reinforcement Learning	103
4.3.3	Unsupervised Learning	103
4.3.4	Semi-supervised Learning	104
4.4	Conclusions	105
	References.	105
5	Bayesian Theory of Decision	107
5.1	Introduction	107
5.2	Bayes Decision Rule	108
5.3	Bayes Classifier★	110
5.4	Loss Function	112
5.4.1	Binary Classification	114
5.5	Zero-One Loss Function.	115
5.6	Discriminant Functions	116
5.6.1	Binary Classification Case	117
5.7	Gaussian Density	118
5.7.1	Univariate Gaussian Density	118
5.7.2	Multivariate Gaussian Density	119
5.7.3	Whitening Transformation	120
5.8	Discriminant Functions for Gaussian Likelihood	122
5.8.1	Features Are Statistically Independent	122
5.8.2	Covariance Matrix Is the Same for All Classes	123
5.8.3	Covariance Matrix Is Not the Same for All Classes	125
5.9	Receiver Operating Curves	125
5.10	Conclusions	127
	Problems.	128
	References.	129
6	Clustering Methods.	131
6.1	Introduction	131
6.2	Expectation and Maximization Algorithm★	133
6.2.1	Basic EM★	134
6.3	Basic Notions and Terminology	136
6.3.1	Codebooks and Codevectors	136
6.3.2	Quantization Error Minimization	137
6.3.3	Entropy Maximization	138
6.3.4	Vector Quantization.	139

6.4	K-Means	141
6.4.1	Batch K-Means	142
6.4.2	Online K-Means	143
6.4.3	K-Means Software Packages	146
6.5	Self-Organizing Maps	146
6.5.1	SOM Software Packages	148
6.5.2	SOM Drawbacks	148
6.6	Neural Gas and Topology Representing Network	149
6.6.1	Neural Gas	149
6.6.2	Topology Representing Network	150
6.6.3	Neural Gas and TRN Software Package	151
6.6.4	Neural Gas and TRN Drawbacks	151
6.7	General Topographic Mapping★	151
6.7.1	Latent Variables★	152
6.7.2	Optimization by EM Algorithm★	153
6.7.3	GTM Versus SOM★	154
6.7.4	GTM Software Package	155
6.8	Fuzzy Clustering Algorithms	155
6.8.1	FCM	156
6.9	Hierarchical Clustering	157
6.10	Mixtures of Gaussians	159
6.10.1	The E-Step	160
6.10.2	The M-Step	161
6.11	Conclusion	163
	Problems	164
	References	165
7	Foundations of Statistical Learning and Model Selection	169
7.1	Introduction	169
7.2	Bias-Variance Dilemma	170
7.2.1	Bias-Variance Dilemma for Regression	170
7.2.2	Bias-Variance Decomposition for Classification★	171
7.3	Model Complexity	173
7.4	VC Dimension and Structural Risk Minimization	176
7.5	Statistical Learning Theory★	179
7.5.1	Vapnik-Chervonenkis Theory	180
7.6	AIC and BIC Criteria	182
7.6.1	Akaike Information Criterion	182
7.6.2	Bayesian Information Criterion	183
7.7	Minimum Description Length Approach	184

7.8	Crossvalidation	186
7.8.1	Generalized Crossvalidation	186
7.9	Conclusion	188
	Problems	188
	References	189
8	Supervised Neural Networks and Ensemble Methods	191
8.1	Introduction	191
8.2	Artificial Neural Networks and Neural Computation	192
8.3	Artificial Neurons	193
8.4	Connections and Network Architectures	196
8.5	Single-Layer Networks	198
8.5.1	Linear Discriminant Functions and Single-Layer Networks	199
8.5.2	Linear Discriminants and the Logistic Sigmoid	200
8.5.3	Generalized Linear Discriminants and the Perceptron	201
8.6	Multilayer Networks	203
8.6.1	The Multilayer Perceptron	204
8.7	Multilayer Networks Training	205
8.7.1	Error Back-Propagation for Feed-Forwards Networks*	206
8.7.2	Parameter Update: The Error Surface	208
8.7.3	Parameters Update: The Gradient Descent*	210
8.7.4	The Torch Package	212
8.8	Learning Vector Quantization	212
8.8.1	The <i>LVQ_PAK</i> Software Package	214
8.9	Nearest Neighbour Classification	215
8.9.1	Probabilistic Interpretation	217
8.10	Ensemble Methods	217
8.10.1	Classifier Diversity and Ensemble Performance*	218
8.10.2	Creating Ensemble of Diverse Classifiers	220
8.11	Conclusions	224
	Problems	224
	References	225
9	Kernel Methods	229
9.1	Introduction	229
9.2	Lagrange Method and Kuhn Tucker Theorem	231
9.2.1	Lagrange Multipliers Method	231
9.2.2	Kuhn Tucker Theorem	233
9.3	Support Vector Machines for Classification	235
9.3.1	Optimal Hyperplane Algorithm	236
9.3.2	Support Vector Machine Construction	238

9.3.3	Algorithmic Approaches to Solve Quadratic Programming	241
9.3.4	Sequential Minimal Optimization.	242
9.3.5	Other Optimization Algorithms	244
9.3.6	SVM and Regularization Methods [★]	244
9.4	Multiclass Support Vector Machines	247
9.4.1	One-Versus-Rest Method	247
9.4.2	One-Versus-One Method	247
9.4.3	Other Methods	248
9.5	Support Vector Machines for Regression	248
9.5.1	Regression with Quadratic ϵ -Insensitive Loss	249
9.5.2	Kernel Ridge Regression	252
9.5.3	Regression with Linear ϵ -Insensitive Loss.	254
9.5.4	Other Approaches to Support Vector Regression	256
9.6	Gaussian Processes	256
9.6.1	Regression with Gaussian Processes.	257
9.7	Kernel Fisher Discriminant.	258
9.7.1	Fisher's Linear Discriminant.	258
9.7.2	Fisher Discriminant in Feature Space	260
9.8	Kernel PCA	262
9.8.1	Centering in Feature Space	262
9.9	One-Class SVM	264
9.9.1	One-Class SVM Optimization	267
9.10	Kernel Clustering Methods	269
9.10.1	Kernel K-Means	270
9.10.2	Kernel SOM	272
9.10.3	Kernel Neural Gas.	272
9.10.4	One-Class SVM Extensions	273
9.10.5	Kernel Fuzzy Clustering Methods	274
9.11	Spectral Clustering	278
9.11.1	Shi and Malik Algorithm	280
9.11.2	Ng-Jordan-Weiss' Algorithm.	281
9.11.3	Other Methods	282
9.11.4	Connection Between Spectral and Kernel Clustering Methods	283
9.12	Software Packages.	287
9.13	Conclusion	287
	Problems.	288
	References.	289
10	Markovian Models for Sequential Data	295
10.1	Introduction	295
10.2	Hidden Markov Models	296
10.2.1	Emission Probability Functions	300

10.3	The Three Problems	300
10.4	The Likelihood Problem and the Trellis**	301
10.5	The Decoding Problem**	304
10.6	The Learning Problem**	308
10.6.1	Parameter Initialization	309
10.6.2	Estimation of the Initial State Probabilities	310
10.6.3	Estimation of the Transition Probabilities	311
10.6.4	Emission Probability Function Parameters Estimation	312
10.7	HMM Variants	315
10.8	Linear-Chain Conditional Random Fields	317
10.8.1	From HMMs to Linear-Chain CRFs	319
10.8.2	General CRFs	321
10.8.3	The Three Problems.	322
10.9	The Inference Problem for Linear Chain CRFs	323
10.10	The Training Problem for Linear Chain CRFs.	323
10.11	N -gram Models and Statistical Language Modeling	325
10.11.1	N -gram Models	325
10.11.2	The Perplexity	326
10.11.3	N -grams Parameter Estimation	327
10.11.4	The Sparseness Problem and the Language Case	328
10.12	Discounting and Smoothing Methods for N -gram Models**	330
10.12.1	The Leaving-One-Out Method	331
10.12.2	The Turing Good Estimates	333
10.12.3	Katz's Discounting Model	334
10.13	Building a Language Model with N -grams	336
	Problems	337
	References	338
11	Feature Extraction Methods and Manifold Learning Methods . . .	341
11.1	Introduction	341
11.2	*The Curse of Dimensionality	343
11.3	Data Dimensionality	344
11.3.1	Local Methods	345
11.3.2	Global Methods	347
11.3.3	Mixed Methods	355
11.4	Principal Component Analysis	357
11.4.1	PCA as ID Estimator	359
11.4.2	Nonlinear Principal Component Analysis	361
11.5	Independent Component Analysis	362
11.5.1	Statistical Independence	363
11.5.2	ICA Estimation	364

11.5.3	ICA by Mutual Information Minimization	367
11.5.4	FastICA Algorithm	369
11.6	Multidimensional Scaling Methods	370
11.6.1	Sammon's Mapping.	371
11.7	Manifold Learning.	372
11.7.1	The Manifold Learning Problem	372
11.7.2	Isomap.	374
11.7.3	Locally Linear Embedding	375
11.7.4	Laplacian Eigenmaps	378
11.8	Conclusion.	379
	Problems.	379
	References.	381

Part III Applications

12	Speech and Handwriting Recognition	389
12.1	Introduction	389
12.2	The General Approach.	390
12.3	The Front End	392
12.3.1	The Handwriting Front End	393
12.3.2	The Speech Front End	394
12.4	HMM Training	397
12.4.1	Lexicon and Training Set	397
12.4.2	Hidden Markov Models Training.	398
12.5	Recognition and Performance Measures	400
12.5.1	Recognition	400
12.5.2	Performance Measurement	401
12.6	Recognition Experiments	403
12.6.1	Lexicon Selection	404
12.6.2	N -gram Model Performance	405
12.6.3	Cambridge Database Results	407
12.6.4	IAM Database Results	408
12.7	Speech Recognition Results	409
12.8	Applications	411
12.8.1	Applications of Handwriting Recognition	411
12.8.2	Applications of Speech Recognition.	413
	References.	415
13	Automatic Face Recognition	421
13.1	Introduction	421
13.2	Face Recognition: General Approach	423
13.3	Face Detection and Localization	424

13.3.1	Face Segmentation and Normalization with <i>TorchVision</i>	426
13.4	Lighting Normalization	428
13.4.1	Center/Surround Retinex	428
13.4.2	Gross and Brajovic's Algorithm	429
13.4.3	Normalization with <i>TorchVision</i>	429
13.5	Feature Extraction	430
13.5.1	Holistic Approaches	430
13.5.2	Local Approaches	434
13.5.3	Feature Extraction with <i>TorchVision</i>	434
13.6	Classification	437
13.7	Performance Assessment	439
13.7.1	The FERET Database	440
13.7.2	The FRVT Database	441
13.8	Experiments	442
13.8.1	Data and Experimental Protocol	443
13.8.2	Euclidean Distance-Based Classifier	443
13.8.3	SVM-Based Classification	445
	References	445
14	Video Segmentation and Keyframe Extraction	449
14.1	Introduction	449
14.2	Applications of Video Segmentation	451
14.3	Shot Boundary Detection	452
14.3.1	Pixel-Based Approaches	453
14.3.2	Block-Based Approaches	455
14.3.3	Histogram-Based Approaches	455
14.3.4	Clustering-Based Approaches	456
14.3.5	Performance Measures	457
14.4	Shot Boundary Detection with <i>Torchvision</i>	458
14.5	Keyframe Extraction	460
14.6	Keyframe Extraction with <i>Torchvision</i> and <i>Torch</i>	462
	References	463
15	Real-Time Hand Pose Recognition	467
15.1	Introduction	467
15.2	Hand Pose Recognition Methods	468
15.3	Hand Pose Recognition by a Data Glove	471
15.4	Hand Pose Color-Based Recognition	475
15.4.1	Segmentation Module	476
15.4.2	Feature Extraction	478
15.4.3	The Classifier	479
15.4.4	Experimental Results	480
	References	483

16	Automatic Personality Perception	485
16.1	Introduction	485
16.2	Previous Work	486
16.2.1	Nonverbal Behaviour.	487
16.2.2	Social Media.	488
16.3	Personality and Its Measurement.	488
16.4	Speech-Based Automatic Personality Perception	490
16.4.1	The SSPNet Speaker Personality Corpus	491
16.4.2	The Approach.	492
16.4.3	Extraction of Short-Term Features.	492
16.4.4	Extraction of Statisticals.	493
16.4.5	Prediction.	493
16.5	Experiments and Results	494
16.6	Conclusions	496
	References.	497
 Part IV Appendices		
	Appendix A: Statistics	501
	Appendix B: Signal Processing.	513
	Appendix C: Matrix Algebra	525
	Appendix D: Mathematical Foundations of Kernel Methods	531
	Index	551

Machine Learning for Audio, Image and Video Analysis

Theory and Applications

Camastra, F.; Vinciarelli, A.

2015, XVI, 561 p. 119 illus., Hardcover

ISBN: 978-1-4471-6734-1