Contents

Preface v

List of Algorithms xiii

List of Figures xv

1 **Mathematical foundations** 1
- 1.1 Abstract algebras 1
  - Groups 1
  - Rings 3
- 1.2 Metrics 4
- 1.3 Vector spaces 5
  - Linear operators 7
  - Matrix algebra 7
  - Square and invertible matrices 8
  - Eigenvalues and eigenvectors 9
  - Special matrices 10
- 1.4 Probability and stochastic processes 12
  - Sample spaces, events, measures and distributions 12
  - Joint random variables: independence, conditionals, and marginals 14
  - Bayes' rule 16
  - Expectation, generating functions and characteristic functions 17
  - Empirical distribution function and sample expectations 19
  - Transforming random variables 20
  - Multivariate Gaussian and other limiting distributions 21
  - Stochastic processes 23
  - Markov chains 25
- 1.5 Data compression and information theory 28
  - The importance of the information map 31
  - Mutual information and Kullback-Leibler (K-L) divergence 32
- 1.6 Graphs 34
  - Special graphs 35
- 1.7 Convexity 36
- 1.8 Computational complexity 37
  - Complexity order classes and *big-O* notation 38

Tractable versus intractable problems: NP-completeness 38

2 Optimization **41**
2.1 Preliminaries 41
Continuous differentiable problems and critical points 41
Continuous optimization under equality constraints: Lagrange multipliers 42
Inequality constraints: duality and the Karush-Kuhn-Tucker conditions 44
Convergence and convergence rates for iterative methods 45
Non-differentiable continuous problems 46
Discrete (combinatorial) optimization problems 47
2.2 Analytical methods for continuous convex problems 48
L_2-norm objective functions 49
Mixed L_2-L_1 norm objective functions 50
2.3 Numerical methods for continuous convex problems 51
Iteratively reweighted least squares (IRLS) 51
Gradient descent 53
Adapting the step sizes: line search 54
Newton's method 56
Other gradient descent methods 58
2.4 Non-differentiable continuous convex problems 59
Linear programming 59
Quadratic programming 60
Subgradient methods 60
Primal-dual interior-point methods 62
Path-following methods 64
2.5 Continuous non-convex problems 65
2.6 Heuristics for discrete (combinatorial) optimization 66
Greedy search 67
(Simple) tabu search 67
Simulated annealing 68
Random restarting 69

3 Random sampling **71**
3.1 Generating (uniform) random numbers 71
3.2 Sampling from continuous distributions 72
Quantile function (inverse CDF) and inverse transform sampling 72
Random variable transformation methods 74
Rejection sampling 74
Adaptive rejection sampling (ARS) for log-concave densities 75
Special methods for particular distributions 78
3.3 Sampling from discrete distributions 79
Inverse transform sampling by sequential search 79

Rejection sampling for discrete variables 80
Binary search inversion for (large) finite sample spaces 81
3.4 Sampling from general multivariate distributions 81
Ancestral sampling 82
Gibbs sampling 83
Metropolis-Hastings 85
Other MCMC methods 88

4 Statistical modelling and inference 93
4.1 Statistical models 93
Parametric versus nonparametric models 93
Bayesian and non-Bayesian models 94
4.2 Optimal probability inferences 95
Maximum likelihood and minimum K-L divergence 95
Loss functions and empirical risk estimation 98
Maximum a-posteriori and regularization 99
Regularization, model complexity and data compression 101
Cross-validation and regularization 105
The bootstrap 107
4.3 Bayesian inference 108
4.4 Distributions associated with metrics and norms 110
Least squares 111
Least L_q-norms 111
Covariance, weighted norms and Mahalanobis distance 112
4.5 The exponential family (EF) 115
Maximum entropy distributions 115
Sufficient statistics and canonical EFs 116
Conjugate priors 118
Prior and posterior predictive EFs 122
Conjugate EF prior mixtures 123
4.6 Distributions defined through quantiles 124
4.7 Densities associated with piecewise linear loss functions 126
4.8 Nonparametric density estimation 129
4.9 Inference by sampling 130
MCMC inference 130
Assessing convergence in MCMC methods 130

5 Probabilistic graphical models 133
5.1 Statistical modelling with PGMs 133
5.2 Exploring conditional independence in PGMs 136
Hidden versus observed variables 136
Directed connection and separation 137
The Markov blanket of a node 138
5.3 Inference on PGMs 139

Exact inference 140
Approximate inference 143

6 Statistical machine learning 149
6.1 Feature and kernel functions 149
6.2 Mixture modelling 150
Gibbs sampling for the mixture model 150
E-M for mixture models 152
6.3 Classification 154
Quadratic and linear discriminant analysis (QDA and LDA) 155
Logistic regression 156
Support vector machines (SVM) 158
Classification loss functions and misclassification count 161
Which classifier to choose? 161
6.4 Regression 162
Linear regression 162
Bayesian and regularized linear regression 163
Linear-in parameters regression 164
Generalized linear models (GLMs) 165
Nonparametric, nonlinear regression 167
Variable selection 169
6.5 Clustering 171
K-means and variants 171
Soft K-means, mean shift and variants 174
Semi-supervised clustering and classification 176
Choosing the number of clusters 177
Other clustering methods 178
6.6 Dimensionality reduction 178
Principal components analysis (PCA) 179
Probabilistic PCA (PPCA) 182
Nonlinear dimensionality reduction 184

7 Linear-Gaussian systems and signal processing 187
7.1 Preliminaries 187
Delta signals and related functions 187
Complex numbers, the unit root and complex exponentials 189
Marginals and conditionals of linear-Gaussian models 190
7.2 Linear, time-invariant (LTI) systems 191
Convolution and impulse response 191
The discrete-time Fourier transform (DTFT) 192
Finite-length, periodic signals: the discrete Fourier transform (DFT) 198
Continuous-time LTI systems 201
Heisenberg uncertainty 203
Gibb's phenomena 205
Transfer function analysis of discrete-time LTI systems 206

	Fast Fourier transforms (FFT)	208
7.3	LTI signal processing	212
	Rational filter design: FIR, IIR filtering	212
	Digital filter recipes	220
	Fourier filtering of very long signals	222
	Kernel regression as discrete convolution	224
7.4	Exploiting statistical stability for linear-Gaussian DSP	226
	Discrete-time Gaussian processes (GPs) and DSP	226
	Nonparametric power spectral density (PSD) estimation	231
	Parametric PSD estimation	236
	Subspace analysis: using PCA in DSP	238
7.5	The Kalman filter (KF)	242
	Junction tree algorithm (JT) for KF computations	243
	Forward filtering	244
	Backward smoothing	246
	Incomplete data likelihood	247
	Viterbi decoding	247
	Baum-Welch parameter estimation	249
	Kalman filtering as signal subspace analysis	251
7.6	Time-varying linear systems	252
	Short-time Fourier transform (STFT) and perfect reconstruction	253
	Continuous-time wavelet transforms (CWT)	255
	Discretization and the discrete wavelet transform (DWT)	257
	Wavelet design	261
	Applications of the DWT	262
8	**Discrete signals: sampling, quantization and coding**	**265**
8.1	Discrete-time sampling	266
	Bandlimited sampling	267
	Uniform bandlimited sampling: Shannon-Whittaker interpolation	267
	Generalized uniform sampling	270
8.2	Quantization	273
	Rate-distortion theory	275
	Lloyd-Max and entropy-constrained quantizer design	278
	Statistical quantization and dithering	282
	Vector quantization	286
8.3	Lossy signal compression	288
	Audio companding	288
	Linear predictive coding (LPC)	289
	Transform coding	291
8.4	Compressive sensing (CS)	293
	Sparsity and incoherence	294
	Exact reconstruction by convex optimization	295
	Compressive sensing in practice	296

9 Nonlinear and non-Gaussian signal processing **299**
9.1 Running window filters 299
Maximum likelihood filters 300
Change point detection 301
9.2 Recursive filtering 302
9.3 Global nonlinear filtering 302
9.4 Hidden Markov models (HMMs) 304
Junction tree (JT) for efficient HMM computations 305
Viterbi decoding 306
Baum-Welch parameter estimation 306
Model evaluation and structured data classification 309
Viterbi parameter estimation 309
Avoiding numerical underflow in message passing 310
9.5 Homomorphic signal processing 311

10 Nonparametric Bayesian machine learning and signal processing **313**
10.1 Preliminaries 313
Exchangeability and de Finetti's theorem 314
Representations of stochastic processes 316
Partitions and equivalence classes 317
10.2 Gaussian processes (GP) 318
From basis regression to kernel regression 318
Distributions over function spaces: GPs 319
Bayesian GP kernel regression 321
GP regression and Wiener filtering 325
Other GP-related topics 326
10.3 Dirichlet processes (DP) 327
The Dirichlet distribution: canonical prior for the categorical distribution 328
Defining the Dirichlet and related processes 331
Infinite mixture models (DPMMs) 334
Can DP-based models actually infer the number of components? 343

Bibliography **345**

Index **353**