

01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61

Short Table of Contents

Preface	xxiii
1 Introduction	1
1.1 What is machine learning?	1
1.2 Supervised learning	1
1.3 Unsupervised learning	7
1.4 Some basic concepts in machine learning	11
1.5 Exercises	28
2 Probability	29
2.1 Introduction	29
2.2 A brief review of probability theory	29
2.3 Some common discrete distributions	33
2.4 Some common continuous distributions	37
2.5 Joint probability distributions	41
2.6 Transformations of random variables	46
2.7 Monte Carlo approximation	49
2.8 Information theory	51
2.9 Exercises	53
3 Statistics	57
3.1 Introduction	57
3.2 Bayesian inference of a discrete parameter (the number game)	58
3.3 Bayesian inference of a continuous parameter (the healthy levels game)	63
3.4 Bayesian inference in everyday cognition *	65
3.5 The Beta-Binomial model	69
3.6 The Dirichlet-multinomial model	72
3.7 Summarizing posterior distributions	74
3.8 Bayesian model selection	78
3.9 Priors	83
3.10 Exercises	87
4 Gaussian models	93
4.1 Introduction	93
4.2 Inference in jointly Gaussian distributions	95
4.3 Linear Gaussian systems	100
4.4 Inferring the parameters of an MVN	105
4.5 Exercises	115
5 Generative models for classification	119
5.1 Introduction	119
5.2 Naive Bayes classifiers	119
5.3 Gaussian Discriminant analysis	126
5.4 Fisher's linear discriminant analysis (FLDA) *	132
5.5 Dealing with missing data	136
5.6 The exponential family *	136
5.7 Exercises	143

01		
02	6 Discriminative linear models	147
03	6.1 Introduction	147
04	6.2 Linear regression	148
05	6.3 Bayesian linear regression	154
06	6.4 Logistic regression	158
07	6.5 Bayesian logistic regression	164
08	6.6 Online learning and stochastic optimization	169
09	6.7 Generalized linear models (GLMs)	173
10	6.8 Probit regression	175
11	6.9 Exercises	177
12		
13	7 Graphical models	183
14	7.1 Introduction	183
15	7.2 Directed graphical models (Bayes nets)	185
16	7.3 Undirected graphical models (Markov random fields)	190
17	7.4 Factor graphs	195
18	7.5 Inference	196
19	7.6 Learning	197
20	7.7 CI properties of GMs	199
21	7.8 Gaussian graphical models *	205
22	7.9 Exercises	208
23		
24	8 Decision theory	213
25	8.1 Introduction	213
26	8.2 Bayesian decision theory	213
27	8.3 Frequentist decision theory	219
28	8.4 Frequentist heuristics for picking estimators *	227
29	8.5 Exercises	236
30		
31	9 Mixture models and the EM algorithm	239
32	9.1 Latent variable models	239
33	9.2 Mixture models	239
34	9.3 Parameter estimation for mixture models	244
35	9.4 The EM algorithm	247
36	9.5 Model selection for latent variable models	259
37	9.6 Fitting models with missing data	261
38	9.7 Exercises	263
39		
40	10 Latent linear models	267
41	10.1 Factor analysis	267
42	10.2 Principal components analysis (PCA)	271
43	10.3 Choosing the number of latent dimensions	279
44	10.4 PCA for categorical data	281
45	10.5 PCA for paired and multi-view data	283
46	10.6 Independent Component Analysis (ICA)	285
47	10.7 Exercises	292
48		
49	11 Hierarchical Bayes	295
50	11.1 Introduction	295
51	11.2 Empirical Bayes	296
52	11.3 Multi-task learning and mixed effect GLMs	300
53	11.4 Analysis of variance (ANOVA) *	302
54	11.5 Exercises	304
55		
56	12 Sparse linear models	305
57	12.1 Introduction	305
58	12.2 Bayesian variable selection	306
59	12.3 ℓ_1 regularization: basics	311
60	12.4 ℓ_1 regularization: algorithms	319
61	12.5 ℓ_1 regularization: extensions	324

01			
02	12.6	Non-convex regularizers	330
03	12.7	Automatic relevance determination (ARD)/ sparse Bayesian learning (SBL)	333
04	12.8	Sparse coding *	337
05	12.9	Exercises	340
06			
07	13	Kernels	345
08	13.1	Introduction	345
09	13.2	Kernel functions	345
10	13.3	Using kernels inside GLMs	350
11	13.4	The kernel trick	351
12	13.5	Support vector machines (SVMs)	358
13	13.6	Comparison of discriminative kernel methods	364
14	13.7	Kernels for building generative models	365
15	13.8	Exercises	370
16			
17	14	Gaussian processes	371
18	14.1	Introduction	371
19	14.2	GPs for regression	372
20	14.3	GPs meet GLMs	379
21	14.4	Connection with other methods	384
22	14.5	GP latent variable model	390
23	14.6	Approximation methods for large datasets *	392
24	14.7	Exercises	397
25			
26	15	Adaptive basis function models	399
27	15.1	Introduction	399
28	15.2	Classification and regression trees (CART)	400
29	15.3	Generalized additive models	405
30	15.4	Boosting	407
31	15.5	Feedforward neural networks (multilayer perceptrons)	413
32	15.6	Experimental comparison	424
33	15.7	Interpreting black-box models	426
34	15.8	Exercises	427
35			
36	16	Markov and hidden Markov Models	429
37	16.1	Introduction	429
38	16.2	Markov models	429
39	16.3	Hidden Markov models	439
40	16.4	Inference in HMMs	441
41	16.5	Learning for HMMs	448
42	16.6	Generalizations of HMMs	453
43	16.7	Exercises	460
44			
45	17	State space models	461
46	17.1	Introduction	461
47	17.2	Applications of SSMs	462
48	17.3	Inference in LG-SSM	469
49	17.4	Learning for LG-SSM	472
50	17.5	Approximate online inference for non-linear, non-Gaussian SSMs	475
51	17.6	Hybrid discrete/ continuous SSMs	482
52	17.7	Exercises	486
53			
54	18	Conditional random fields	487
55	18.1	Introduction	487
56	18.2	Applications of CRFs	490
57	18.3	Inference	495
58	18.4	Learning	495
59	18.5	Structural SVMs and max margin Markov networks *	502
60	18.6	Exercises	502
61			

01		
02	19 Exact inference algorithms for graphical models	503
03	19.1 Introduction	503
04	19.2 Belief propagation for trees	503
05	19.3 The variable elimination algorithm	509
06	19.4 The junction tree algorithm *	514
07	19.5 Computational intractability of exact inference in the worst case	523
08	19.6 Exercises	524
09		
10	20 Mean field inference algorithms	527
11	20.1 Introduction	527
12	20.2 Variational inference	527
13	20.3 The mean field method	530
14	20.4 Structured mean field	533
15	20.5 Variational Bayes	534
16	20.6 Variational Bayes EM	540
17	20.7 Variational message passing and VIBES	544
18	20.8 Exercises	544
19		
20		
21	21 Other variational inference algorithms	547
22	21.1 Introduction	547
23	21.2 Local variational bounds	547
24	21.3 Variational inference for discrete graphical models : a view from the marginal polytope *	552
25	21.4 Loopy belief propagation	555
26	21.5 Convex belief propagation *	563
27	21.6 MAP state estimation in discrete graphical models	565
28	21.7 Expectation propagation	569
29	21.8 Exercises	571
30		
31		
32	22 Monte Carlo inference algorithms	573
33	22.1 Introduction	573
34	22.2 Sampling from standard distributions	573
35	22.3 Rejection sampling	574
36	22.4 Importance sampling	576
37	22.5 Particle filtering	578
38	22.6 Rao-Blackwellised particle filtering (RBPF)	584
39	22.7 Approximating the marginal likelihood	588
40	22.8 Exercises	589
41		
42		
43	23 MCMC inference algorithms	591
44	23.1 Introduction	591
45	23.2 Gibbs sampling	591
46	23.3 Metropolis Hastings algorithm	599
47	23.4 Simulated annealing	607
48	23.5 Auxiliary variable MCMC *	607
49	23.6 Speed and accuracy of MCMC	611
50	23.7 Exercises	617
51		
52		
53	24 Clustering	621
54	24.1 Introduction	621
55	24.2 Dirichlet process mixture models	623
56	24.3 Affinity propagation	630
57	24.4 Spectral clustering	631
58	24.5 Hierarchical clustering	633
59	24.6 Clustering datapoints and features	639
60	24.7 Exercises	641
61		

01		
02	25 Graphical model structure learning	643
03	25.1 Introduction	643
04	25.2 Quick and dirty ways to learn graph structure	643
05	25.3 Learning tree structures	646
06	25.4 Learning DAG structures	649
07	25.5 Learning DAG structure with latent variables	655
08	25.6 Learning causal DAGs	661
09	25.7 Learning undirected Gaussian graphical models	666
10	25.8 Learning undirected discrete graphical models	669
11	25.9 Exercises	670
12		
13	26 Two-layer latent variable models	673
14	26.1 Introduction	673
15	26.2 LVMs for modeling vectors and bags of counts and tokens	673
16	26.3 Latent Dirichlet allocation (LDA)	676
17	26.4 Extensions of LDA	684
18	26.5 LVMs for graph-structured data	691
19	26.6 LVMs for relational data	696
20	26.7 Restricted Boltzmann machines (RBMs)	702
21	26.8 Exercises	709
22		
23	27 Deep learning	711
24	27.1 Introduction	711
25	27.2 Deep generative models	712
26	27.3 Training deep networks	714
27	27.4 Applications of deep networks	716
28	27.5 RBMs with inputs	720
29	27.6 Other kinds of deep models	724
30	27.7 Bridging the signal-to-symbol divide	727
31	27.8 Learning structural forms	728
32	27.9 Exercises	736
33		
34	I Index	737
35	Index to code	762
36	Index to keywords	764
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		