Contents

Preface ix

1 Graphical models and probabilistic reasoning 1
- 1.1 Introduction 1
- 1.2 Axioms of probability and basic notations 4
- 1.3 The Bayes update of probability 9
- 1.4 Inductive learning 11
  - 1.4.1 Bayes' rule 12
  - 1.4.2 Jeffrey's rule 13
  - 1.4.3 Pearl's method of virtual evidence 13
- 1.5 Interpretations of probability and Bayesian networks 14
- 1.6 Learning as inference about parameters 15
- 1.7 Bayesian statistical inference 17
- 1.8 Tossing a thumb-tack 20
- 1.9 Multinomial sampling and the Dirichlet integral 24

Notes 28

Exercises: Probabilistic theories of causality, Bayes' rule, multinomial sampling and the Dirichlet density 31

2 Conditional independence, graphs and *d*-separation 37
- 2.1 Joint probabilities 37
- 2.2 Conditional independence 38
- 2.3 Directed acyclic graphs and *d*-separation 41
  - 2.3.1 Graphs 41
  - 2.3.2 Directed acyclic graphs and probability distributions 45
- 2.4 The Bayes ball 50
  - 2.4.1 Illustrations 51
- 2.5 Potentials 53
- 2.6 Bayesian networks 58
- 2.7 Object oriented Bayesian networks 63
- 2.8 *d*-Separation and conditional independence 66

Koski, Timo
Bayesian networks
2009

digitalisiert durch:
IDS Basel Bern

2.9 Markov models and Bayesian networks 67
2.10 I-maps and Markov equivalence 69
2.10.1 The trek and a distribution without a faithful graph 72
Notes 73
Exercises: Conditional independence and d-separation 75

3 Evidence, sufficiency and Monte Carlo methods 81
3.1 Hard evidence 82
3.2 Soft evidence and virtual evidence 85
3.2.1 Jeffrey's rule 86
3.2.2 Pearl's method of virtual evidence 87
3.3 Queries in probabilistic inference 88
3.3.1 The chest clinic problem 89
3.4 Bucket elimination 89
3.5 Bayesian sufficient statistics and prediction sufficiency 92
3.5.1 Bayesian sufficient statistics 92
3.5.2 Prediction sufficiency 95
3.5.3 Prediction sufficiency for a Bayesian network 97
3.6 Time variables 98
3.7 A brief introduction to Markov chain Monte Carlo methods 100
3.7.1 Simulating a Markov chain 103
3.7.2 Irreducibility, aperiodicity and time reversibility 104
3.7.3 The Metropolis-Hastings algorithm 108
3.7.4 The one-dimensional discrete Metropolis algorithm 111
Notes 112
Exercises: Evidence, sufficiency and Monte Carlo methods 113

4 Decomposable graphs and chain graphs 123
4.1 Definitions and notations 124
4.2 Decomposable graphs and triangulation of graphs 127
4.3 Junction trees 131
4.4 Markov equivalence 133
4.5 Markov equivalence, the essential graph and chain graphs 138
Notes 144
Exercises: Decomposable graphs and chain graphs 145

5 Learning the conditional probability potentials 149
5.1 Initial illustration: maximum likelihood estimate for a fork connection 149
5.2 The maximum likelihood estimator for multinomial sampling 151
5.3 MLE for the parameters in a DAG: the general setting 155
5.4 Updating, missing data, fractional updating 160
Notes 161
Exercises: Learning the conditional probability potentials 162

6 Learning the graph structure 167
6.1 Assigning a probability distribution to the graph structure 168

6.2 Markov equivalence and consistency 171
6.2.1 Establishing the DAG isomorphic property 173
6.3 Reducing the size of the search 176
6.3.1 The Chow-Liu tree 177
6.3.2 The Chow-Liu tree: A predictive approach 179
6.3.3 The K2 structural learning algorithm 183
6.3.4 The MMHC algorithm 184
6.4 Monte Carlo methods for locating the graph structure 186
6.5 Women in mathematics 189
Notes 191
Exercises: Learning the graph structure 192

7 Parameters and sensitivity 197
7.1 Changing parameters in a network 198
7.2 Measures of divergence between probability distributions 201
7.3 The Chan-Darwiche distance measure 202
7.3.1 Comparison with the Kullback-Leibler divergence and euclidean distance 209
7.3.2 Global bounds for queries 210
7.3.3 Applications to updating 212
7.4 Parameter changes to satisfy query constraints 216
7.4.1 Binary variables 218
7.5 The sensitivity of queries to parameter changes 220
Notes 224
Exercises: Parameters and sensitivity 225

8 Graphical models and exponential families 229
8.1 Introduction to exponential families 229
8.2 Standard examples of exponential families 231
8.3 Graphical models and exponential families 233
8.4 Noisy 'or' as an exponential family 234
8.5 Properties of the log partition function 237
8.6 Fenchel Legendre conjugate 239
8.7 Kullback-Leibler divergence 241
8.8 Mean field theory 243
8.9 Conditional Gaussian distributions 246
8.9.1 CG potentials 249
8.9.2 Some results on marginalization 249
8.9.3 CG regression 250
Notes 251
Exercises: Graphical models and exponential families 252

9 Causality and intervention calculus 255
9.1 Introduction 255
9.2 Conditioning by observation and by intervention 257
9.3 The intervention calculus for a Bayesian network 258

9.3.1 Establishing the model via a controlled experiment 262
9.4 Properties of intervention calculus 262
9.5 Transformations of probability 265
9.6 A note on the order of 'see' and 'do' conditioning 267
9.7 The 'Sure Thing' principle 268
9.8 Back door criterion, confounding and identifiability 270
Notes 273
Exercises: Causality and intervention calculus 275

10 The junction tree and probability updating 279
10.1 Probability updating using a junction tree 279
10.2 Potentials and the distributive law 280
10.2.1 Marginalization and the distributive law 283
10.3 Elimination and domain graphs 284
10.4 Factorization along an undirected graph 288
10.5 Factorizing along a junction tree 290
10.5.1 Flow of messages initial illustration 292
10.6 Local computation on junction trees 294
10.7 Schedules 296
10.8 Local and global consistency 302
10.9 Message passing for conditional Gaussian distributions 305
10.10 Using a junction tree with virtual evidence and soft evidence 311
Notes 313
Exercises: The junction tree and probability updating 314

11 Factor graphs and the sum product algorithm 319
11.1 Factorization and local potentials 319
11.1.1 Examples of factor graphs 320
11.2 The sum product algorithm 323
11.3 Detailed illustration of the algorithm 329
Notes 332
Exercise: Factor graphs and the sum product algorithm 333

References 335

Index 343