Probabilistic Graphical Models

Zusammenfassung

Fabian Damken

17. Februar 2022

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Inhaltsverzeichnis

1 Introduction **8**
- 1.1 Examples . . . 8
- 1.2 Fundamental Questions . . . 8

2 Foundations **9**
- 2.1 Probability Theory . . . 9
  - 2.1.1 (Conditional) Independence . . . 9
  - 2.1.2 Inference . . . 9
  - 2.1.3 Potentials . . . 9
- 2.2 Machine Learning . . . 9
  - 2.2.1 (Document) Classification . . . 9

3 Bayesian Networks **10**
- 3.1 The Naive Bayes Model . . . 11
  - 3.1.1 Classification . . . 11
  - 3.1.2 Maximum Likelihood Parameter Estimation . . . 11
  - 3.1.3 Application . . . 11
- 3.2 Definition and Independence Assumptions . . . 11
  - 3.2.1 Local Markov Assumption . . . 11
  - 3.2.2 „Explaining Away“ / Berkson’s Paradox . . . 11
  - 3.2.3 Representation Theorem . . . 11
  - 3.2.4 Building a Bayesian Network . . . 11
- 3.3 Encoded Independencies . . . 11
  - 3.3.1 Dependency Structures . . . 11
  - 3.3.2 d-Separation . . . 11
  - 3.3.3 Faithful Distributions . . . 11
  - 3.3.4 Context-Specific Independence (CSI) . . . 11
  - 3.3.5 The Bayes’ Ball Algorithm . . . 11

4 Inference **12**
- 4.1 Chain Models . . . 12
- 4.2 Variable Elimination . . . 12
  - 4.2.1 Evidence . . . 12
  - 4.2.2 Complexity . . . 12
  - 4.2.3 VE for Potentials . . . 12
- 4.3 Abductive Inference . . . 12
  - 4.3.1 Consistency . . . 12
  - 4.3.2 Finding Most Probable Explanations (MPEs) . . . 12
- 4.4 Complexity of Conditional Queries . . . 12
- 4.5 Moralizing . . . 12

4.6 Variable Elimination in Moral Graphs . . . 12
4.6.1 Perfect Elimination Sequences . . . 12
4.6.2 Complexity . . . 12
4.6.3 Induced Graph . . . 12
4.6.4 Induced Treewidth . . . 12
4.6.5 Elimination on Trees . . . 12
4.6.6 General Networks . . . 12

5 Markov Random Fields 13
5.1 Bayesian Networks as MRFs . . . 13
5.2 Triangulated Graphs . . . 13
5.3 Join Trees . . . 13
5.4 Junction Trees . . . 13
5.4.1 Collecting Evidence . . . 13
5.4.2 Distributing Evidence . . . 13
5.5 Non-Triangulated Graphs . . . 13

6 Learning 14
6.1 Complete and Incomplete Data Sets . . . 14
6.1.1 Hidden Variables . . . 14
6.2 Parameter Estimation . . . 14
6.2.1 Known Structure, Complete Data . . . 14
6.2.2 Known Structure, Incomplete Data (Expectation-Maximization) . . . 14
6.2.3 Gradient Ascent . . . 14
6.2.4 Bayesian Parameter Estimation . . . 14
6.2.5 Summary . . . 15
6.3 Structure Learning / Model Selection . . . 15
6.3.1 Minimal I-Maps . . . 15
6.3.2 Perfect Maps (P-Maps) . . . 15
6.3.3 I-Equivalence . . . 15
6.3.4 Obtaining a P-Map . . . 15
6.3.5 Accurate Structures . . . 15
6.3.6 Learning . . . 15
6.3.7 Structure Search as Optimization . . . 15
6.3.8 Structural EM . . . 15
6.3.9 Summary . . . 15

7 Dynamic Bayesian Networks 16
7.1 Hidden Markov Models . . . 17
7.2 Inference . . . 17
7.2.1 Decoding . . . 17
7.2.2 Best State Sequence . . . 17
7.2.3 Parameter Estimation . . . 17
7.3 State Estimation (Kalman Filter) . . . 17
7.3.1 Recursive Bayesian Updating . . . 17
7.3.2 (Modeling) Actions . . . 17
7.3.3 Bayes Filter . . . 17
7.3.4 Discrete-Time Kalman Filter . . . 17

7.4 General Dynamic Bayesian Networks . . . 17
7.4.1 Exact Inference . . . 17
7.4.2 Tractable, Approximate Inference . . . 17

8 Approximate Inference 18
8.1 Message Passing . . . 18
8.1.1 Sum-Product Belief Propagation . . . 18
8.1.2 (Acyclic) Belief Propagation as Dynamic Programming . . . 18
8.1.3 Loopy Belief Propagation . . . 18
8.2 Sampling . . . 18
8.2.1 Forward Sampling (Without Evidence) . . . 18
8.2.2 Forward Sampling (With Evidence) . . . 18
8.2.3 Gibbs Sampling . . . 18
8.2.4 Likelihood Weighting . . . 18

9 Tractable Probabilistic Models 19
9.1 Deep Learning . . . 19
9.2 Probabilistic Circuits . . . 19
9.3 Sum-Product Networks . . . 19
9.3.1 Inference . . . 19
9.3.2 Learning . . . 19
9.3.3 Inference on Devices . . . 19

10 Deep Generative Models 20
10.1 Likelihood-Based . . . 20
10.1.1 Autoregressive Generatie Models . . . 20
10.1.2 Variational Auto-Encoders . . . 20
10.1.3 Normalizing Flows . . . 20
10.2 Likelihood-Free . . . 20
10.2.1 Generative Adversarial Networks . . . 20
10.3 Applications in Scientific Discovery . . . 20

Abbildungsverzeichnis

Tabellenverzeichnis

Liste der Algorithmen

1 Introduction

1.1 Examples

1.2 Fundamental Questions

2 Foundations

2.1 Probability Theory

2.1.1 (Conditional) Independence

Monty Hall Problem

2.1.2 Inference

Information Theory

2.1.3 Potentials

2.2 Machine Learning

2.2.1 (Document) Classification

3 Bayesian Networks

3.1 The Naive Bayes Model

3.1.1 Classification

3.1.2 Maximum Likelihood Parameter Estimation

3.1.3 Application

3.2 Definition and Independence Assumptions

3.2.1 Local Markov Assumption

3.2.2 „Explaining Away“ / Berkson’s Paradox

3.2.3 Representation Theorem

3.2.4 Building a Bayesian Network

3.3 Encoded Independencies

3.3.1 Dependency Structures

3.3.2 d-Separation

(Active) Trails

Independencies

Soundness

Completeness

3.3.3 Faithful Distributions

3.3.4 Context-Specific Independence (CSI)

Tree CPD

Determinism

3.3.5 The Bayes’ Ball Algorithm

4 Inference

4.1 Chain Models

4.2 Variable Elimination

4.2.1 Evidence

4.2.2 Complexity

4.2.3 VE for Potentials

4.3 Abductive Inference

4.3.1 Consistency

4.3.2 Finding Most Probable Explanations (MPEs)

4.4 Complexity of Conditional Queries

4.5 Moralizing

4.6 Variable Elimination in Moral Graphs

4.6.1 Perfect Elimination Sequences

4.6.2 Complexity

4.6.3 Induced Graph

4.6.4 Induced Treewidth

4.6.5 Elimination on Trees

Polytrees

4.6.6 General Networks

5 Markov Random Fields

5.1 Bayesian Networks as MRFs

5.2 Triangulated Graphs

5.3 Join Trees

5.4 Junction Trees

5.4.1 Collecting Evidence

5.4.2 Distributing Evidence

5.5 Non-Triangulated Graphs

6 Learning

6.1 Complete and Incomplete Data Sets

6.1.1 Hidden Variables

6.2 Parameter Estimation

6.2.1 Known Structure, Complete Data

Maximum Likelihood

Decomposability of the Likelihood

Likelihood for (Conditional) Bi- and Multinomials

6.2.2 Known Structure, Incomplete Data (Expectation-Maximization)

EM Idea

Complete-Data Likelihood

EM for (Conditional) Multinomials

Monotonicity

6.2.3 Gradient Ascent

6.2.4 Bayesian Parameter Estimation

Laplace Estimation

Bayesian Prediction

Conjugate Priors

Binomial Prior

Dirichlet Prior

Bayesian Networks and Bayesian Prediction

6.2.5 Summary

6.3 Structure Learning / Model Selection

6.3.1 Minimal I-Maps

6.3.2 Perfect Maps (P-Maps)

6.3.3 I-Equivalence

Skeleton and Immoralities

6.3.4 Obtaining a P-Map

Identifying the Skeleton

Identifying Immoralities

From Immoralities to Structures

6.3.5 Accurate Structures

6.3.6 Learning

Constrained-Based

Score-Based

Likelihood Score

Bayesian Score and Bayesian Information Criterion

6.3.7 Structure Search as Optimization

Learning Trees (Complete Data)

Heuristic (Local) Search

6.3.8 Structural EM

6.3.9 Summary

7 Dynamic Bayesian Networks

7.1 Hidden Markov Models

7.2 Inference

7.2.1 Decoding

Forward Pass

Backward Pass

7.2.2 Best State Sequence

Viterbi Algorithm

7.2.3 Parameter Estimation

7.3 State Estimation (Kalman Filter)

7.3.1 Recursive Bayesian Updating

7.3.2 (Modeling) Actions

7.3.3 Bayes Filter

7.3.4 Discrete-Time Kalman Filter

Dynamics and Observations

Belief Update: Prediction

Belief Update: Correction

7.4 General Dynamic Bayesian Networks

7.4.1 Exact Inference

7.4.2 Tractable, Approximate Inference

Assumed Density Filtering

8 Approximate Inference

8.1 Message Passing

8.1.1 Sum-Product Belief Propagation

8.1.2 (Acyclic) Belief Propagation as Dynamic Programming

8.1.3 Loopy Belief Propagation

8.2 Sampling

8.2.1 Forward Sampling (Without Evidence)

8.2.2 Forward Sampling (With Evidence)

8.2.3 Gibbs Sampling

Burn-In

Irreducibility, Aperiodicity, and Ergodicity

Convergence

Performance

Speeding Convergence

Skipping Samples

Randomized Variable Order

Blocking

Rao-Blackwellization

Multiple Chains

8.2.4 Likelihood Weighting

9 Tractable Probabilistic Models

9.1 Deep Learning

9.2 Probabilistic Circuits

9.3 Sum-Product Networks

9.3.1 Inference

9.3.2 Learning

Directly Learning SPNs

9.3.3 Inference on Devices

10 Deep Generative Models

10.1 Likelihood-Based

10.1.1 Autoregressive Generatie Models

Learning and Inference

Parametrization

10.1.2 Variational Auto-Encoders

Inference as Optimization

Variational Bayes

Learning and Inference

Open Questions

10.1.3 Normalizing Flows

Learning and Inference

10.2 Likelihood-Free

10.2.1 Generative Adversarial Networks

Inference

10.3 Applications in Scientific Discovery