

Probabilistic Graphical Models

Joschka Boedecker & Hao Zhu

Department of Computer Science
University of Freiburg

universität freiburg

1. Introduction

- Probabilistic graphical models
- Mathematical background
 - Basic concepts in probability
 - Random variables
 - Set conditional independence
 - Graphs
- Course information

Outline

- Probabilistic graphical models
- Mathematical background
 - Basic concepts in probability
 - Random variables
 - Set conditional independence
 - Graphs
- Course information

Probabilistic graphical models (PGMs)

example: medical diagnosis

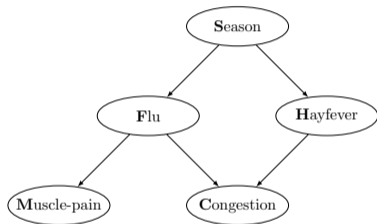
- 2 binary-valued diseases: **F**lu and **H**ayfever
- 4-valued variable: **S**ea^on
- 2 binary-valued symptoms: **C**ongestion and **M**uscle-pain

- probability space: $2 \times 2 \times 4 \times 2 \times 2 = 64$ values
- diagnosis query: how likely the patient is to have the flu given that it is fall, and that she has sinus congestion but no muscle pain

$$\mathbf{P}(F = \text{true} \mid S = \text{fall}, C = \text{true}, M = \text{false})$$

Probabilistic graphical models (PGMs)

graphical representation



independencies

$$\begin{aligned} & (F \perp\!\!\!\perp H \mid S) \\ & (C \perp\!\!\!\perp S \mid F, H) \\ & (M \perp\!\!\!\perp H, C \mid F) \\ & (M \perp\!\!\!\perp C \mid F) \end{aligned}$$

factorization

$$\begin{aligned} \mathbf{P}(S, F, H, C, M) &= \mathbf{P}(S) \mathbf{P}(F \mid S) \\ &\quad \mathbf{P}(H \mid S) \mathbf{P}(C \mid F, H) \mathbf{P}(M \mid F) \end{aligned}$$

- parameter space: $4 + 4 \times 2 + 4 \times 2 + 2 \times 2 \times 2 + 2 \times 2 = 32$ values
- diagnosis query: $\mathbf{P}(C \mid F, H, S) = \mathbf{P}(C \mid F, H)$

Probabilistic graphical models (PGMs)

- compact, tractable, transparent **representation** of variables and dependencies
 - human expert can understand and evaluate its semantics and properties
 - accurate reflection of our understanding of a domain
- effective **inference**
 - answering queries using the distribution as our model of the world
 - computing the posterior probability of some variables given evidence on others
- automatic **learning** of a model from data
 - provides a good approximation to our past experience
 - combining human expert knowledge and data information

Real-world applications

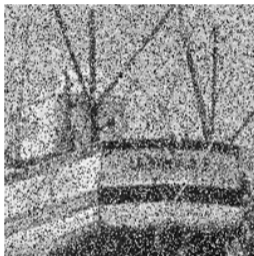
Where would you apply probabilistic graphical models?

Real-world applications

- **images**: generation, denoising
- **language**: generation, translation
- **audio**: super-resolution, speech synthesis, speech recognition
- **economics**: causal inference
- **science**: error-correcting codes, computational biology, ecology
- **health care and medicine**: diagnosis

Real-world applications

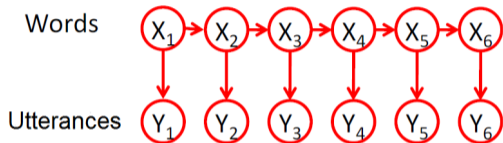
image denoising



- restore old photographs based on probabilistic graphical models that does a good job at modeling the posterior distribution $p(\text{original image} \mid \text{noisy image})$

Real-world applications

speech recognition

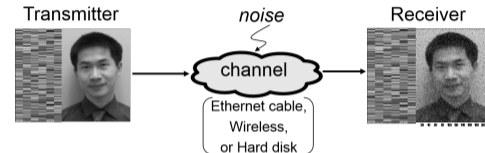


“He ate the cookies on the couch”

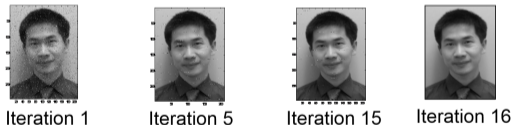
- infer spoken words from audio signals with variants of hidden Markov models

Real-world applications

error-correcting codes



Iterative message passing decoding



- detect and correct communication errors with graphical models

Deep learning and PGMs

Why do we still need PGMs in the age of deep learning?

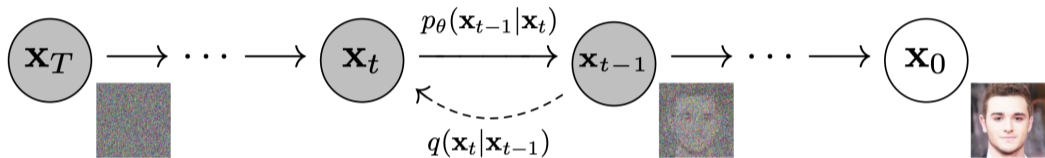
Deep learning and PGMs

combining deep learning architectures and probabilistic modeling techniques can be advantageous in

- **generative models:** variational autoencoders (VAEs), diffusion models
- **inference:** estimating the distribution of unobserved variables given observed data
- **interpretability and uncertainty:** interpretable representation of data and XAI
- **causality:** causal discovery and causal-based learning

Deep learning and PGMs

diffusion models



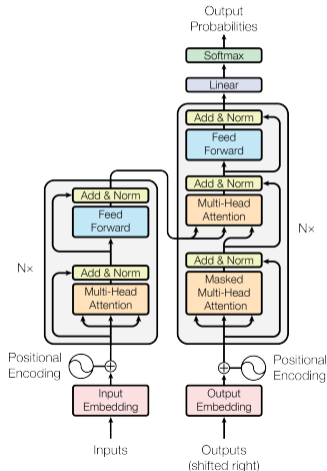
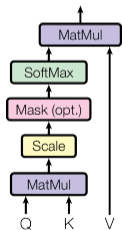
- directed Markov chain structure
- model a sequence of random variables $\{x_0, \dots, x_T\}$
- each x_t is an intermediate between a uniform random distribution and the data distribution

Deep learning and PGMs

transformer architecture

- fully connected graphical model
- scaled dot-product attention

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$



Outline

- Probabilistic graphical models
- Mathematical background
 - Basic concepts in probability
 - Random variables
 - Set conditional independence
 - Graphs
- Course information

Basic concepts in probability

axioms of probability

- $0 \leq \mathbf{P}(A) \leq 1$
- $\mathbf{P}(\text{sure proposition}) = 1$
- $\mathbf{P}(A \text{ or } B) = \mathbf{P}(A) + \mathbf{P}(B)$ if A and B are mutually exclusive

marginal probability

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A, B_i)$$

if $\{B_1, \dots, B_n\}$ is a set of exhaustive and mutually exclusive propositions

Basic concepts in probability

conditional probability $\mathbf{P}(A \mid B)$

- A and B are independent if $\mathbf{P}(A \mid B) = \mathbf{P}(A)$
- A and B are conditionally independent given C if $\mathbf{P}(A \mid B, C) = \mathbf{P}(A \mid C)$

product rule

$$\mathbf{P}(A, B) = \mathbf{P}(A \mid B) \mathbf{P}(B)$$

- chain rule

$$\mathbf{P}(E_1, \dots, E_n) = \mathbf{P}(E_1) \mathbf{P}(E_2 \mid E_1) \cdots \mathbf{P}(E_n \mid E_{n-1}, \dots, E_1)$$

Basic concepts in probability

Bayes' theorem

$$\mathbf{P}(H | e) = \frac{\mathbf{P}(e | H) \mathbf{P}(H)}{\mathbf{P}(e)}$$

- $\mathbf{P}(H)$: prior probability
- $\mathbf{P}(e | H)$: likelihood
- $\mathbf{P}(H | e)$: posterior probability

Random variables

given a random variable X with $\text{dom } X \subseteq \mathbf{R}$

- expected value of X :

$$\mathbf{E} X = \sum_x x \mathbf{P}(x)$$

- expectation of function g of X :

$$\mathbf{E} g(X) = \sum_x g(x) \mathbf{P}(x).$$

- variance of X :

$$\mathbf{var} X = \mathbf{E} (X - \mathbf{E} X)^2$$

– standard deviation of X : $\sigma(X) = \sqrt{\mathbf{var} X}$

Random variables

given two random variables X, Y with $\mathbf{dom} X, \mathbf{dom} Y \subseteq \mathbf{R}$

- covariance of X and Y :

$$\mathbf{cov}(X, Y) = \mathbf{E}(X - \mathbf{E} X)(Y - \mathbf{E} Y)$$

- correlation coefficient of X and Y :

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

- regression coefficient of X on Y :

$$r(X, Y) = \rho(X, Y) \frac{\sigma(X)}{\sigma(Y)} = \frac{\mathbf{cov}(X, Y)}{\mathbf{var} Y}$$

Random variables

continuous random variables

$$\mathbf{P}(a \leq X \leq b) = \int_a^b p(x) dx$$

- $p(x)$: density function
- translation between discrete and continuous random variables:

$$\int_{-\infty}^{\infty} p(x) dx \iff \sum_x \mathbf{P}(x)$$

- expected value of continuous random variable X :

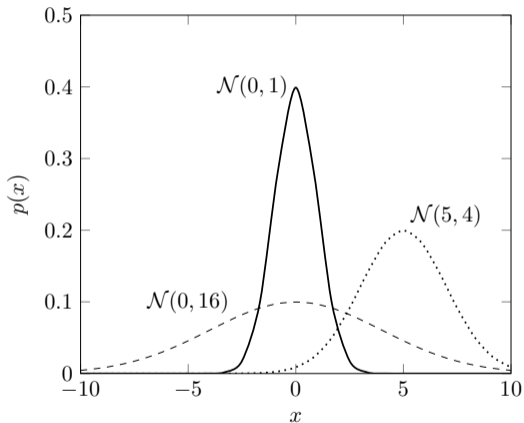
$$\mathbf{E} X = \int_{-\infty}^{\infty} xp(x) dx$$

Random variables

example: Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

- μ : mean of random variable X
- σ^2 : variance of random variable X
- standard Gaussian distribution: $\mathcal{N}(0, 1)$



Set conditional independence

let $V = \{V_1, V_2, \dots\}$, and $X, Y, Z \subseteq V$

conditionally independent set X and Y given Z :

$$(X \perp\!\!\!\perp Y \mid Z) \implies \mathbf{P}(x \mid y, z) = \mathbf{P}(x \mid z) \text{ for all } \{y, z \mid \mathbf{P}(y, z) > 0\}$$

- marginal independence: $(X \perp\!\!\!\perp Y \mid \emptyset)$
- $(X \perp\!\!\!\perp Y \mid Z) \implies \mathbf{P}(V_i \mid V_j, Z) = \mathbf{P}(V_i \mid Z)$, for all $V_i \in X, V_j \in Y$
 - the converse is not necessarily true

Graphs

A graph G consists of

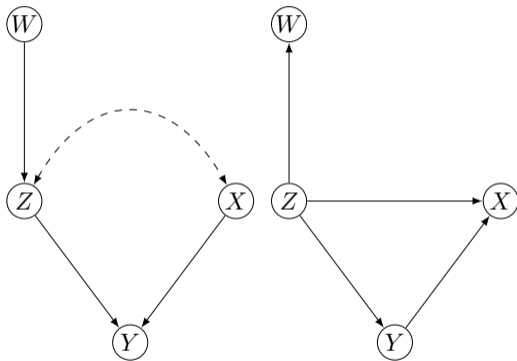
- a set V of vertices (or nodes)
- a set E of edges (or links)
 - directed or undirected

path: $((W, Z), (Z, Y), (Y, X), (X, Z))$

- directed: $((W, Z), (Z, Y))$
- undirected: $((W, Z), (Z, Y), (Y, X))$

directed acyclic graph (DAG)

- a directed graph without directed cycles



Outline

- Probabilistic graphical models
- Mathematical background
 - Basic concepts in probability
 - Random variables
 - Set conditional independence
 - Graphs
- **Course information**

Contents

1. Introduction
2. Bayesian classifiers
3. Markov models
4. Bayesian networks
5. Inference with Monte Carlo methods
6. Markov decision problems
7. Control as probabilistic inference
8. Graphical causal models
9. Deep learning and graphical models

Organization

- 6 ETCS
- course number: 11E13MO-**1228**
- every Tuesday 16:15 – 17:45
- course material and forum on ILIAS
- exam:
 - open-book
 - date: tbd (most likely some day between Sep. 8th and Sep. 12th)

Exercises

- **not** mandatory
- upload your solutions with possible questions to ILIAS
- submissions before the release of the sample solution will be corrected
- sample solutions will be available on ILIAS, 1 weeks after exercise release
- answer questions of others and ask your own questions in the ILIAS forum
- time management options:

7 – 9 hours per ex.

5 – 6 hours per ex.

0 hours per ex.

little exam prep.

more exam prep.

∞ exam prep.

RECOMMENDED

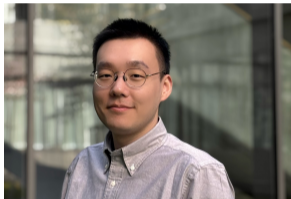
MINIMUM

IMPOSSIBLE

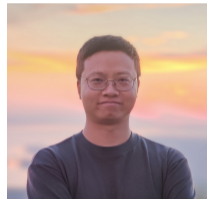
Team



Joschka Boedecker

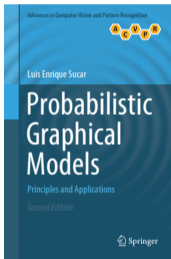


Hao Zhu



Shengchao Yan

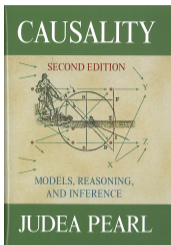
Some useful materials



Probabilistic Graphical Models: Principles and Applications (2nd edition)
Luis Enrique Sucar (2021)

<https://link.springer.com/book/10.1007/978-3-030-61943-5>

Some useful materials

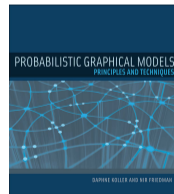


Causality (2nd edition)
Judea Pearl (2009)

<http://bayes.cs.ucla.edu/BOOK-2K>

Probabilistic Graphical Models: Principles and Techniques
Daphne Koller and Nir Friedman (2009)

<https://mitpress.mit.edu/9780262013192/probabilistic-graphical-models>



Some useful materials

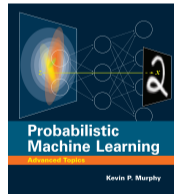


Probabilistic Machine Learning: An introduction
Kevin P. Murphy (2022)

<https://probml.github.io/pml-book/book1.html>

Probabilistic Machine Learning: Advanced Topics
Kevin P. Murphy (2023)

<https://probml.github.io/pml-book/book2.html>



Some useful materials

- Probabilistic Graphical Models Specialization (Coursera)
 - Probabilistic Graphical Models 1: Representation
 - Probabilistic Graphical Models 2: Inference
 - Probabilistic Graphical Models 3: Learning
- notes from CS228 - Probabilistic Graphical Models (Stanford University)