Probabilistic Machine Learning

Tom Schierenbeck, 15.02.2024

Universität Bremen

Artificial Intelligence

About Me

Name: Tom Schierenbeck

E-Mail: tom_sch@uni-bremen.de

GitHub: tomsch420

Position: PhD Student at the Institute for Artificial Intelligence

Involved in the Projects:

- PyCRAM, Cognitive Architecture for Robots in Python
- Probabilistic Modelling
- Lecturer for Knowledge Acquisition and Knowledge Representation
- Project Leader of FAME, Learning plans from Videos

About the Institute for Artificial Intelligence

- AI-Powered & Cognition-enabled Robotics
- Perception, reasoning, learning, knowledge, decision-making, prospection, planning, action
- Cognitive robot architecture
- Robot agents
- Open education, research and innovation
- Successful robot applications

Goals for Today

After this presentation you will know...

- What is probabilistic Machine Learning
- Which classes exist inside Probabilistic Machine Learning
- How models have to be designed to match certain requirements
- What are Nyga Distributions
- What are Joint Probability Trees

Probabilistic Machine Learning

What is it?

- Subfield of Machine Learning and Artificial Intelligence
- Expresses explicit uncertainty over predictions instead of single point estimates
- Answers are typically entire distributions about all possible answers instead of single predictions

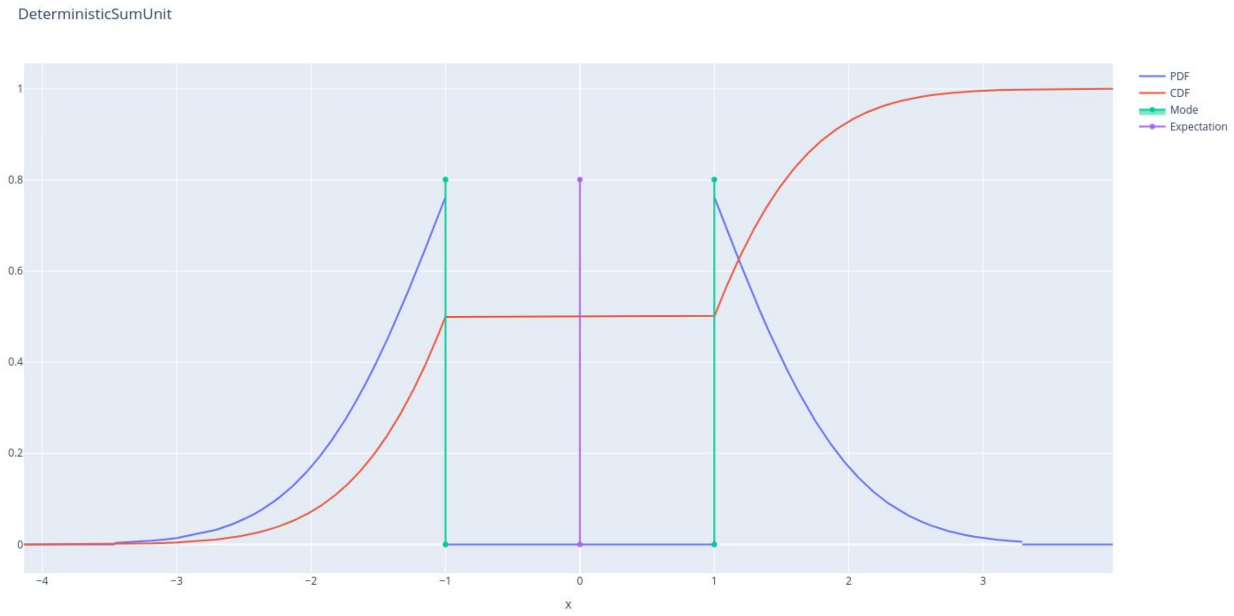
Why is it cool?

- Key tool to automate tasks without specifying every detail
- Answers from machine learning systems are not taken for granted and come with attached probabilities
- Allows reasoning over every possible scenario

Conditional Expectation is not Enough

Most modern Machine Learning Algorithms calculate

$$\mathbb{E}(y \mid x)$$



Quantities of Interest

The most common queries to probability distributions are...

- Evaluation of the Likelihood $p(x)$
- Computing Integrals over Hyper Rectangles

$$p(E = e, Z \in \mathcal{I}) = \int_{\mathcal{I}} p(z, e) dZ$$

- Moments

$$\mathbb{M}_n(x) = \int (x - \mu)^n p(x) dx$$

- Finding the mode of the distribution

$$\hat{x} = \underset{x \in \mathcal{X}}{arg\,max}\; p(x)$$

The price to pay...

- The key operation in ordinary machine learning is optimization
- Integrals are the key operation to probabilistic machine learning
- Integration is computational very heavy and only doable under strict constraints
- We are interested in creating models that compute every interesting quantity in polynomial time
- Such constraints have been formalized in [1]

Probabilistic Circuits

Probabilistic Circuits are...

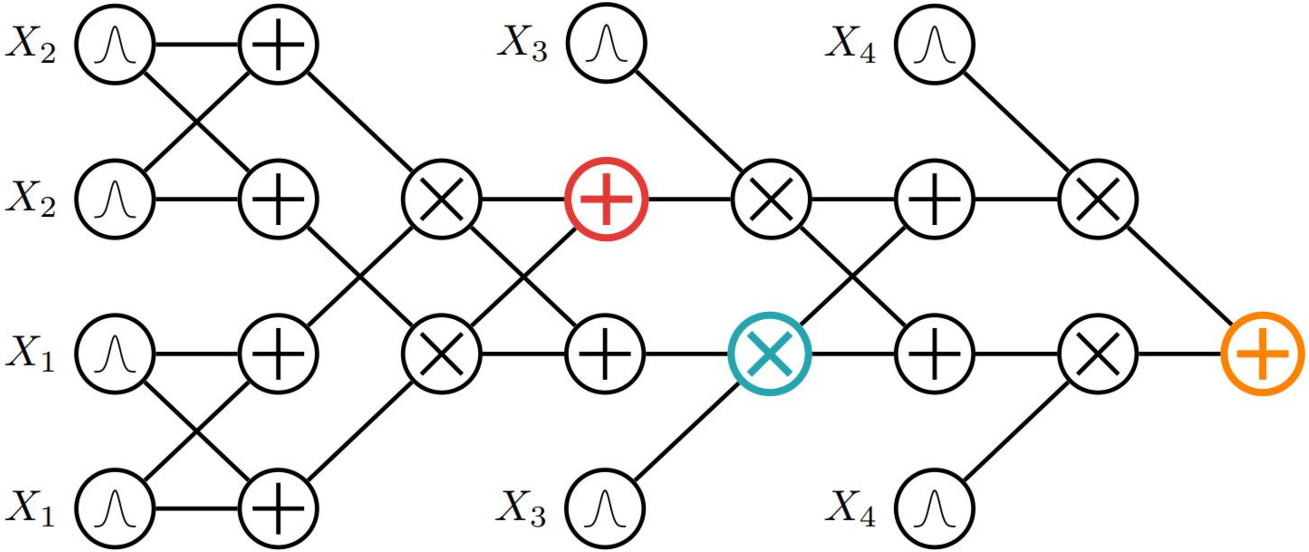
- The formalism that explains tractable inference
- Directed Acyclic Graphs that where the nodes are either…
  - a tractable distribution over encoded as a distribution unit $p(x)$
  - a product of PCs over subsets of

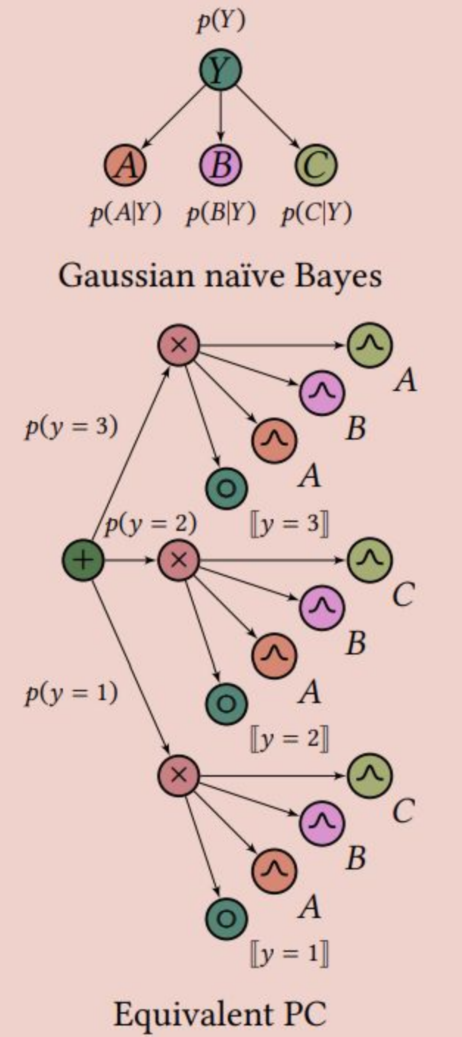
$$X : C(x) = \prod_i C_i(x)$$

  - a convex weighted sum of PCs over subsets of

$$X : C(x) = \sum_i w_i C_i(x)$$

Circuit Examples





Tractable Inference

A product node is decomposable if the scopes of its input units do not share variables.

A PC is decomposable if all of its product units are decomposable.

$$\phi(c_i) \cap \phi(c_j) = \emptyset, \forall c_i, c_j \in in(n), i \neq j$$

A sum node is deterministic if, for any fully-instantiated input, the output of at most one of its children is nonzero. Their input units do not share support.

A circuit is deterministic if all of its sum nodes are deterministic.

$$supp(c_i) \cap supp(c_j) = \emptyset, \forall c_i, c_j \in in(n), i \neq j$$

Nyga Distribution Tutorial

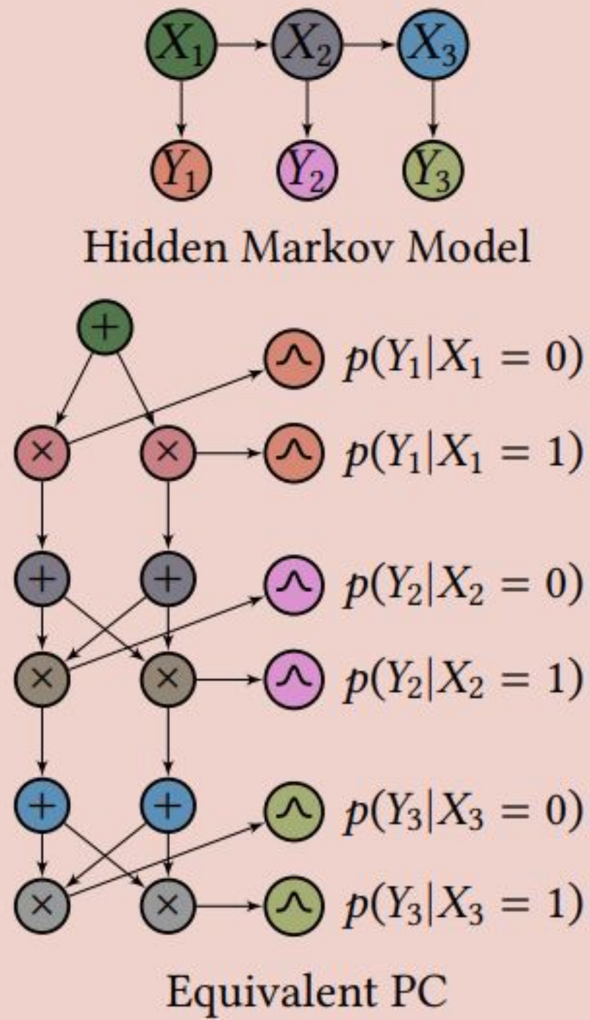
Joint Probability Trees Tutorial

Dynamic Circuits

- Consider that we now want to reason about dynamic worlds, such as time or relations
- Hidden Markov Models are a prominent example of doing so
- Since complexity of inference in PGMs is exponential heavy in their bounded treewidth, they also expand to relations pretty well
- Due to their shallow form JPTs offer a perfect interface for template modelling
- JPTs can be force to be marginal deterministic and hence provide the necessary amount of parameters to interact well with other (dynamic) concepts

Dynamic Models as Circuit

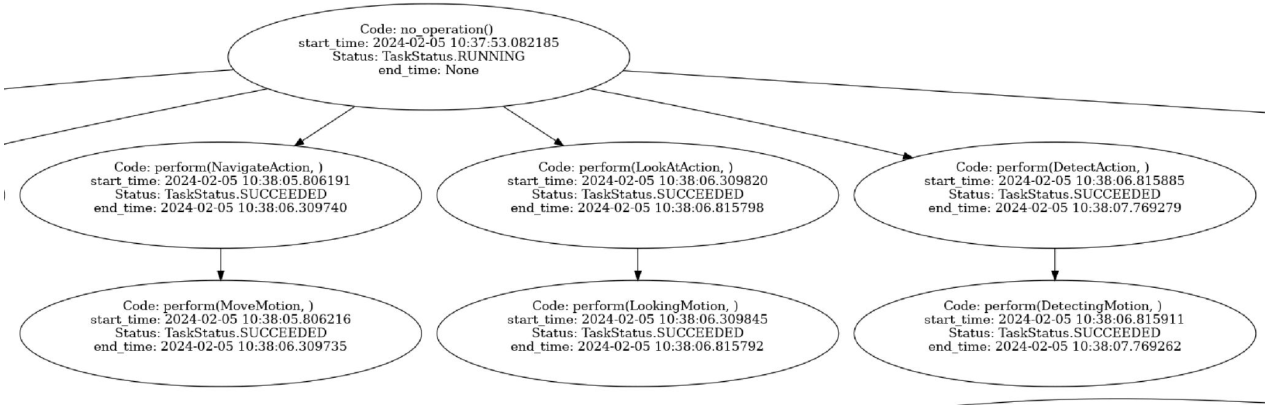
$$\sum_i \theta_i p_i(x) = P(i)P(x|i)$$



What is coming next?

- [Probabilistic Circuits in template models]
- SQL as query language for probabilistic circuits
- Large scale evaluation on template JPTs with PyCRAM
- Transformations in Circuits
- Convolutions of Circuits
- Metrics for Circuits

...



Research Implementation

- Flexible
- More general circuits
- Unified interface
- Higher Usability
- GitHub

Production Implementation

- Cython Backend
- JPTs only
- Basic Interface
- super fast
- GitHub

References

1. Choi, Y., Vergari, A., & Van den Broeck, G. (2020). Probabilistic circuits: A unifying framework for tractable probabilistic models. UCLA. URL: http://starai. cs. ucla. edu/papers/ProbCirc20. pdf.
2. Geh, Renato Lui Scalable Learning of Probabilistic Circuits / Renato Lui Geh; orientador, Denis Deratani Mauá. - São Paulo, 2022. 144 p.: il. Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Computação / Instituto de Matemática e Estatística / Universidade de São Paulo. Bibliografia Versão corrigida 1. Circuitos probabilísticos. 2. Aprendizado de máquina. 3. Modelos probabilísticos. 4. Inteligência Artificial. I. Mauá, Denis Deratani. II. Título.
3. Nyga, D., Picklum, M., Schierenbeck, T., & Beetz, M. (2023). Joint Probability Trees. arXiv preprint arXiv:2302.07167.

Thanks for your attention!

JOINT PROBABILITY
TREES

Questions?

Universität
Bremen

Artificial
Intelligence