

Recent Advancements in Tractable Probabilistic Inference

Priyank Jaini^{*1}, Kristian Kersting^{*2}, Antonio Vergari^{*3}, and Max Welling^{*4}

1 Google – Toronto, CA. jaini.priyank1@gmail.com

2 TU Darmstadt, DE. kersting@cs.tu-darmstadt.de

3 University of Edinburgh, GB. aver@cs.ucla.edu

4 University of Amsterdam, NL. welling.max@gmail.com

Abstract

In several real-world scenarios, decision making involves advanced reasoning under uncertainty, i.e. the ability to answer probabilistic queries. Typically, it is necessary to compute these answers in a *limited amount of time*. Moreover, in many domains, such as healthcare and economical decision making, it is crucial that the result of these queries is reliable, i.e. either *exact* or comes with approximation guarantees. In all these scenarios, tractable probabilistic inference and learning are becoming increasingly important.

Research on representations and learning algorithms for tractable inference embraces very different fields, each one contributing its own perspective. These include automated reasoning, probabilistic modeling, statistical and Bayesian inference and deep learning.

Among the many recent emerging venues in these fields there are: tractable neural density estimators such as autoregressive models and normalizing flows; deep tractable probabilistic circuits such as sum-product networks and sentential decision diagrams; approximate inference routines with guarantees on the quality of the approximation.

Each of these model classes occupies a particular spot in the continuum between tractability and expressiveness. That is, different model classes might offer appealing advantages in terms of efficiency or representation capabilities while trading-off other of these aspects.

So far, clear connections and a deeper understanding of the key differences among them have been hindered by the different languages and perspectives adopted by the different “souls” that comprise the tractable probabilistic modeling community.

This Dagstuhl Seminar brought together experts from these sub-communities and provided the perfect venue to exchange perspectives, deeply discuss the recent advancements and build strong bridges that can greatly propel interdisciplinary research.

Seminar April 18–22, 2022 – <http://www.dagstuhl.de/22161>

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Computing methodologies → Machine learning

Keywords and phrases approximate inference with guarantees, deep generative models, probabilistic circuits, Tractable inference

Digital Object Identifier 10.4230/DagRep.12.4.13

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Recent Advancements in Tractable Probabilistic Inference, *Dagstuhl Reports*, Vol. 12, Issue 4, pp. 13–25

Editors: Priyank Jaini, Kristian Kersting, Antonio Vergari, and Max Welling



DAGSTUHL REPORTS

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Priyank Jaini (Google – Toronto, CA)

Kristian Kersting (TU Darmstadt, DE)

Antonio Vergari (University of Edinburgh, GB)

Max Welling (University of Amsterdam, NL)

License  Creative Commons BY 4.0 International license
© Priyank Jaini, Kristian Kersting, Antonio Vergari, and Max Welling

ML models and systems to enable and support decision making in real-world scenarios need to robustly and effectively *reason in the presence of uncertainty* over the configurations of the world that can be observed. *Probabilistic inference* provides a principled framework to carry on this reasoning process, and enables probabilistic *modeling*: a collection of principles to design and learn from data models that are capable of dealing with uncertainty. The main purpose for these models, once learned or built, is to answer *queries* – posed by humans or other autonomous systems – concerning some aspects of the represented world and quantifying some form of uncertainty over it. That is, that is computing some quantity of interest of the probability distribution that generated the observed data. For instance, the mean or the modes of such a distribution, the marginal or conditional probabilities of events, expected utilities of our policies, or decoding most likely assignments to variables (also known as MAP inference, cf. the Viterbi algorithm). Answering these queries reliably and efficiently is more important than ever: we need ML models and systems to perform inference based on well-calibrated uncertainty estimates throughout all reasoning steps, especially when informing and supporting humans in decision making processes in the real world.

For instance, consider a ML system learned from clinical data to support physicians and policy makers. Such a system would need to support *arbitrary queries* posed by physicians, that is, questions that are not known a priori. Moreover, these queries might involve complex probabilistic reasoning over possible states of the world, for instance involving maximization of some probabilities and the ability to marginalize over unseen or not available (missing) attributes like “At what age is a patient with this X-ray but no previous health record most likely to show *any* symptom of COVID-19?”, or counting and comparing sub-populations “What is the probability of there being more cases with fever given a BMI of 25 in this county than in the neighboring one?”. At the same time, it should guarantee that the uncertainty in its answers, modeled as probabilities, should be faithful to the real-world distribution as uncalibrated estimates might greatly mislead the decision maker.

Recent successes in machine learning (ML) and particularly deep learning have delivered very expressive probabilistic models and learning algorithms. These have proven to be able to induce exceedingly richer models from larger datasets but, unfortunately, at an incredible cost: these models are vastly *intractable* for all but the most trivial of probabilistic reasoning tasks, and they have been demonstrated to provide unreliable uncertainty estimations. In summary, their applicability to real-world scenarios, like the one just described, is very limited.

Nevertheless all these required “ingredients” are within the grasp of several models which we group together under the umbrella name of *tractable probabilistic models*, the core interest of this seminar. *Tractability* here guarantees answering queries *efficiently* and *exactly*. Tractable probabilistic models (TPMs) have a long history rooted in several research fields such as classical probabilistic graphical models (low-treewidth and latent variable models), automated reasoning via knowledge compilation (logical and arithmetic circuits) and statistics (mixture models, Kalman filters). While these classical TPMs are known to

be limited in expressiveness, several recent advancements in deep tractable models (sum-product networks, probabilistic sentential decision diagrams, normalizing flows and neural autoregressive models) are inverting the trend and promising tractable probabilistic inference with little or no compromise when compared to the deep generative models discussed above. It becomes then more and more important to have a seminar on these recent successes of TPMs bringing together the different communities of tractable probabilistic modeling at the same table to propel collaborations by defining the goals and the agenda for future research.

These are the major topics around which the seminar brought up the aforementioned discussion:

- Advanced probabilistic query classes
- Deep tractable probabilistic modeling
- Robust and verifiable probabilistic inference
- Exploiting symmetries for probabilistic modelling and applications in science.

Advanced probabilistic query classes

Probabilistic inference can be reduced as computing probabilistic queries, i.e., functions whose output are certain properties of a probability distribution (e.g., its mass, density, mean, mode, etc.) as encoded by a probabilistic model. Probabilistic queries can be grouped into classes when they compute the same distributional properties and hence share the same computational effort to be answered. Among the most commonly used query classes there are complete evidence (EVI), marginals (MAR), conditionals (CON) and maximum a posteriori (MAP) inference. While these classes have been extensively investigated in theory and practice, they constitute a small portion of the probabilistic inference that might be required to support complex decision making in the real-world.

In fact, one might want to compute the probabilities of logical and arithmetic constraints, of structured objects such as rankings, comparing the likelihood and counts of groups of events or computing the expected predictions of discriminative model such as a classifier or regression w.r.t. some feature distribution. Tracing the exact boundaries of tractable probabilistic inference for these advanced probabilistic query classes and devising probabilistic models delivering efficient and reliable inference for them is an open challenge.

Deep tractable probabilistic modeling

A probabilistic model falls under the umbrella name of tractable probabilistic models (TPMs) if it guarantees exact and polytime inference for certain query classes. As different model classes can be tractable representations for different query classes, a *spectrum* of tractable inference emerges. Typically, this creates a tension with the extent of a model class supporting a larger set of tractable query classes, and its *expressive efficiency*, i.e., the set of functions it can represent compactly.

Recent deep generative models such as generative adversarial networks (GANs), regularized and variational autoencoders (VAEs) fall out of the TPM umbrella because they either have no explicit likelihood model or computing even the simplest class of queries, EVI, is hard in general. In fact, despite their successes, their inference capabilities are severely limited and one has to recur to approximations. However, the approximate inference routines available so far (such as the evidence lower bound and its variants) do not provide sufficiently strong guarantees on the quality of the approximation delivered to be safely deployed in real-world scenarios.

On the other hand, classical TPMs from the probabilistic graphical model community support larger classes of tractable queries comprising MAR, CON and MAP (to different extents based on the model class). Among these there are: i) low or bounded-treewidth probabilistic graphical models that exchange expressiveness for efficiency; ii) determinantal point processes which allow tractable inference for distributions over sets; iii) graphical models with high girth or weak potentials, that provide bounds on the performance of approximate inference methods; and iv) exchangeable probabilistic models that exploit symmetries to reduce inference complexity.

A different perspective on tractability is brought by models compiling inference routines into efficient computational graphs such as arithmetic circuits, sum-product networks, cutset networks and probabilistic sentential decision diagrams have advanced the state-of-the-art inference performance by exploiting context-specific independence, determinism or by exploiting latent variables. These TPMs, as well as many classical tractable PGMs as listed above, can be cast under a unifying framework of probabilistic circuits (PCs), abstracting from the different graphical formalisms of each model. PCs with certain structural properties support tractable MAR, CON, MAP as well as some of the advanced query classes touched in the previous topic item. Guy Van den Broeck gave a long talk on the first day of the seminar to set the stage for participants for viewing tractable probabilistic models from the lens of probabilistic circuits.

More recently, the field of neural density estimators has gained momentum in the tractable probabilistic modeling community. This is due to model classes such as normalizing flows and autoregressive models. Autoregressive models and flows retain the expressiveness of GANs and VAEs, by leveraging powerful neural representations for probability factors or invertible transformations, while overcoming their limitations and delivering tractable EVI queries. As such, they position themselves in the spectrum of tractability in an antithetic position w.r.t. PCs: while the latter support more tractable query classes, the former are generally more expressive. On the first day of the seminar, Marcus Brubaker introduced these models to the seminar participants in a long talk. It is an interesting open challenge to combine TPM models from different regions of such a spectrum to leverage the “best of different worlds”, i.e., increase a model class expressive efficiency while retaining the largest set of supported tractable query classes as possible. The first day subsequently ended with a lively open discussion on the differences between TPMs and Neural Generative Models and what advantages and lessons they can provide the other models.

Robust and verifiable probabilistic inference

Along exactness and efficiency, one generally requires inference routines to be robust to adversarial conditions (noise, malicious attacks, etc.) and to be allow exactness and efficiency to be formally provable. This is crucial to deploy reliable probabilistic models in real-world scenarios (cf. other topic). Recent advancements in learning tractable and intractable probabilistic models from data have raised the question if the learned models are just exploiting spurious correlations in input space, thus ultimately delivering an unfaithful image of the probability distribution they try to encode. This raises several issues, as in tasks like anomaly detection and model comparison, which rely on correctly calibrated probabilities, one can be highly misled by such unfaithful probabilistic models. Furthermore, one might want to verify a priori or ex-post (e.g., in presence of adversarial interventions) if one probabilistic inference algorithm truly guarantees exact inference. Questions like this have just very recently been tackled in a formal verification setting, where proofs of the correctness of inference can be verified with less resources than it takes to execute inference.

Over the course of the seminar, through informal discussions and formal talks by the participants discussed the above mentioned issues in tractable probabilistic inference through topics such as Bayesian Deep Learning, Incorporating symmetries in probabilistic modelling using equivariance with applications in sciences, explainable AI etc.

Overall, the seminar produced numerous insights into how efficient, expressive, flexible, and robust tractable probabilistic models can be built. Specially, the discussions and talks at the seminar spurred a renewed interest in the community to:

- develop techniques and approaches that bring together key ideas from several different fields that include deep generative models, probabilistic circuits, knowledge compilation, and approximate inference.
- create bridges between researchers in these different fields and identify ways in which enhanced interaction between the communities can continue.
- generate a set of goals, research directions, and challenges for researchers in these field to develop robust and principled probabilistic models.
- provide a unified view of the current undertakings in these different fields towards probabilistic modelling and identifying ways to incorporate ideas from several fields together.
- develop a new systematic and unified set of development tools encompassing these different areas of probabilistic modelling.

2 Table of Contents

Executive Summary

Priyank Jaini, Kristian Kersting, Antonio Vergari, and Max Welling 14

Overview of Talks

Causality and Tractable Probabilistic Models
Alessandro Antonucci 19

A tutorial on Normalizing Flows
Marcus A. Brubaker 19

Solving Marginal MAP Exactly by Probabilistic Circuit Transformations
YooJung Choi 19

Towards Robust Classification with Deep Generative Forests
Cassio de Campos 20

Exploiting Symmetries for Probabilistic Generative Modelling
Priyank Jaini 20

Equivariant Probabilistic Models for Physics
Danilo Jimenez Rezende 21

Predictive Complexity Priors
Eric Nalisnick 21

Extracting context specific independencies from sum product networks
Sriyaam Natarajan 21

Implicit MLE: Backpropagating Through Discrete Exponential Family Distributions
Mathias Niepert 22

Rapid Adaptation in Robot Learning
Deepak Pathak 22

Exact and Efficient Adversarial Robustness with Decomposable Neural Networks
Robert Peharz 23

Probabilistic Circuits: Representations, Inference, Learning and Applications
Guy Van den Broeck 23

Conditional Generative Models and Where to Apply Them
Max Welling 24

Bayesian Deep Learning and a Probabilistic Perspective of Model Construction
Andrew G. Wilson 24

Participants 25

Remote Participants 25

3 Overview of Talks

3.1 Causality and Tractable Probabilistic Models

Alessandro Antonucci (IDSIA – Manno, CH)

License © Creative Commons BY 4.0 International license
© Alessandro Antonucci

Probabilistic sentential decision diagrams (PSDDs) are a popular class of probabilistic circuits intended to implement generative models consistent with a propositional knowledge base. We discuss a number of results related to these models. This includes: the sensitivity analysis of the inferences with respect to perturbations in the local probabilistic parameters of the circuit; a structural learning algorithm for these models based on a relaxation of the closed-world assumption for the training data; and a discussion on the benefits and the challenges related to the embedding of knowledge bases in ML tasks.

3.2 A tutorial on Normalizing Flows

Marcus A. Brubaker (York University – Toronto, CA)

License © Creative Commons BY 4.0 International license
© Marcus A. Brubaker

Normalizing flows (NFs) offer an answer to a long-standing question in computer vision: How can one define faithful probabilistic models for complex high-dimensional data like natural images? NFs solve this problem by means of non-linear bijective mappings from simple distributions (e.g. multivariate normal) to the desired target distributions. These mappings are implemented with invertible neural networks and thus have high expressive power and can be trained by gradient descent in the usual way. Thanks to bijectivity, NFs can work forward and backward, serving as both discriminative and generative models alike, and are especially suitable for inverse problems. This tutorial will explain the theoretical underpinnings of NFs, show various practical implementation options, clarify their relationships with GANs, VAEs, and non-linear ICA. Particular emphasis will be given to successful applications in the field of computer vision.

3.3 Solving Marginal MAP Exactly by Probabilistic Circuit Transformations

YooJung Choi (UCLA, US)

License © Creative Commons BY 4.0 International license
© YooJung Choi
Joint work of YooJung Choi, Antonio Vergari, Guy Van den Broeck

Probabilistic circuits (PCs) are a class of tractable probabilistic models that allow efficient, often linear-time, inference of queries such as marginals and most probable explanations (MPE). However, marginal MAP, which is central to many decision-making problems, remains a hard query for PCs unless they satisfy highly restrictive structural constraints. In this paper, we develop a pruning algorithm that removes parts of the PC that are irrelevant to a marginal MAP query, shrinking the PC while maintaining the correct solution. This pruning

technique is so effective that we are able to build a marginal MAP solver based solely on iteratively transforming the circuit—no search is required. We empirically demonstrate the efficacy of our approach on real-world datasets.

3.4 Towards Robust Classification with Deep Generative Forests

Cassio de Campos (TU Eindhoven, NL)

License  Creative Commons BY 4.0 International license
© Cassio de Campos

Joint work of Alvaro H. C. Correia, Robert Peharz, Cassio de Campos

Main reference Alvaro H. C. Correia, Robert Peharz, Cassio P. de Campos: “Towards Robust Classification with Deep Generative Forests”, CoRR, Vol. abs/2007.05721, 2020.

URL <https://arxiv.org/abs/2007.05721>

Decision Trees (DTs) and Random Forests (RFs) are powerful discriminative learners and tools of central importance to the everyday machine learning practitioner and data scientist. Due to their discriminative nature, however, they lack principled methods to process inputs with missing features or to detect outliers, which requires pairing them with imputation techniques or a separate generative model. In this paper, we demonstrate that DTs and RFs can naturally be interpreted as generative models, by drawing a connection to Probabilistic Circuits, a prominent class of tractable probabilistic models. This reinterpretation equips them with a full joint distribution over the feature space and leads to Generative Decision Trees (GeDTs) and Generative Forests (GeFs), a family of novel hybrid generative-discriminative models. This family of models retains the overall characteristics of DTs and RFs while additionally being able to handle missing features by means of marginalisation. Under certain assumptions, frequently made for Bayes consistency results, we show that consistency in GeDTs and GeFs extend to any pattern of missing input features, if missing at random. Empirically, we show that our models often outperform common routines to treat missing data, such as K-nearest neighbour imputation, and moreover, that our models can naturally detect outliers by monitoring the marginal probability of input features.

3.5 Exploiting Symmetries for Probabilistic Generative Modelling

Priyank Jaini (Google – Toronto, CA)

License  Creative Commons BY 4.0 International license
© Priyank Jaini

Joint work of Priyank Jaini, Lars Holdijk, Max Welling

Main reference Priyank Jaini, Lars Holdijk, Max Welling: “Learning Equivariant Energy Based Models with Equivariant Stein Variational Gradient Descent”, CoRR, Vol. abs/2106.07832, 2021.

URL <https://arxiv.org/abs/2106.07832>

Symmetries play a crucial role in Physics and Mathematics. In this talk, I will explore generative models for efficient sampling and inference by incorporating inductive biases in the form of symmetries. I will begin by introducing Equivariant Stein Variational Gradient Descent (SVGD) algorithm — an equivariant sampling method based on Stein’s identity for sampling from symmetric distributions. Subsequently, I will discuss training equivariant energy based models using Equivariant-SVGD to model invariant probability distributions with applications in many-body particle systems and molecular structure generation.

3.6 Equivariant Probabilistic Models for Physics

Danilo Jimenez Rezende (Google DeepMind – London, GB)

License  Creative Commons BY 4.0 International license
© Danilo Jimenez Rezende

The study of symmetries in physics has revolutionized our understanding of the world. Inspired by this, the development of methods to incorporate internal (Gauge) and external (space-time) symmetries into machine learning models is a very active field of research. We will present our work on invariant generative models and its applications to lattice-QCD and molecular dynamics simulations. In the molecular dynamics front, we'll talk about how we constructed permutation and translation-invariant normalizing flows on a torus for free-energy estimation. In lattice-QCD, we'll present our work that introduced the first $U(N)$ and $SU(N)$ Gauge-equivariant normalizing flows for pure Gauge simulations and its extensions to incorporate fermions.

3.7 Predictive Complexity Priors

Eric Nalisnick (University of Amsterdam, NL)

License  Creative Commons BY 4.0 International license
© Eric Nalisnick

Specifying a Bayesian prior is notoriously difficult for complex models such as neural networks. Reasoning about parameters is made challenging by the high-dimensionality and over-parameterization of the space. Priors that seem benign and uninformative can have unintuitive and detrimental effects on a model's predictions. To help cope with these problems, I will describe our work on predictive complexity priors: a prior that is defined by comparing the model's predictions to those of a reference model.

3.8 Extracting context specific independencies from sum product networks

Sriraam Natarajan (University of Texas – Dallas, US)

License  Creative Commons BY 4.0 International license
© Sriraam Natarajan

Joint work of Sriraam Natarajan, Athresh Karanam, Saurabh Sanjay Mathur, Predrag Radivojac, Kristian Kersting

I present the problem of explaining a class of tractable deep probabilistic model, the Sum-Product Networks (SPNs). First, I motivate how knowledge as qualitative constraints could be extracted from SPNs and then present an algorithm EXSPN to generate explanations. To this effect, I define the notion of a context-specific independence tree(CSI-tree) and present an iterative algorithm that converts an SPN to a CSI-tree. The resulting CSI-tree is both interpretable and explainable to the domain expert. We achieve this by extracting the conditional independencies encoded by the SPN and approximating the local context specified by the structure of the SPN. Our extensive empirical evaluations on synthetic, standard, and real-world clinical data sets demonstrate that the resulting models exhibit superior explainability.

3.9 Implicit MLE: Backpropagating Through Discrete Exponential Family Distributions

Mathias Niepert (Universität Stuttgart, DE)

License  Creative Commons BY 4.0 International license
 Mathias Niepert

Joint work of Mathias Niepert, Pasquale Minervini, Luca Franceschi

Main reference Mathias Niepert, Pasquale Minervini, Luca Franceschi: “Implicit MLE: Backpropagating Through Discrete Exponential Family Distributions”, CoRR, Vol. abs/2106.01798, 2021.

URL <https://arxiv.org/abs/2106.01798>

Combining discrete probability distributions and combinatorial optimization problems with neural network components has numerous applications in learning and reasoning but poses several challenges. We propose Implicit Maximum Likelihood Estimation (I-MLE), a framework for end-to-end learning of models combining discrete exponential family distributions and differentiable neural components. I-MLE is widely applicable as it only requires the ability to compute the most probable states and does not rely on smooth relaxations. The framework encompasses several approaches such as perturbation-based implicit differentiation and recent methods to differentiate through black-box combinatorial solvers. We introduce a novel class of noise distributions for approximating marginals via perturb-and-MAP. Moreover, we show that I-MLE simplifies to maximum likelihood estimation when used in some recently studied learning settings that involve combinatorial solvers. Experiments on several datasets suggest that I-MLE is competitive with and often outperforms existing approaches which rely on problem-specific relaxations. Lastly we discuss potential connections with more sophisticated reasoning scenarios with tractable models.

3.10 Rapid Adaptation in Robot Learning

Deepak Pathak (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 4.0 International license
 Deepak Pathak

Generalization, i.e., the ability to adapt to novel scenarios, is the hallmark of human intelligence. While we have systems that excel at cleaning floors, playing complex games, and occasionally beating humans, they are incredibly specific in that they only perform the tasks they are trained for and are miserable at generalization. One of the fundamental reasons is that, unlike humans, most of these artificial agents start tabula-rasa without any prior knowledge and learn only towards a fixed goal. Could actually optimizing towards fixed external goals be hindering the generalization instead of aiding it? In this talk, I will present our initial efforts toward endowing artificial agents with an ability to generalize in diverse scenarios. The main insight is to first allow the agent to learn general-purpose skills in a completely self-directed manner, without optimizing for any external goal. These skills are then later repurposed to perform complex tasks. I will discuss how this framework can be instantiated to develop curiosity-driven agents (virtual as well as real) that can learn to play games, learn to walk, and learn to perform real-world object manipulation without any rewards or supervision. These curious robotic agents, after exploring the environment, can generalize to find their way in office environments, tie knots using rope and rearrange object configuration.

3.11 Exact and Efficient Adversarial Robustness with Decomposable Neural Networks

Robert Peharz (TU Graz, AT)

License © Creative Commons BY 4.0 International license
© Robert Peharz

Joint work of Robert Peharz, Pranav Shankar Subramani, Antonio Vergari, Gautam Kamath

Main reference Pranav Shankar Subramani, Antonio Vergari, Gautam Kamath, Robert Peharz: “Exact and Efficient Adversarial Robustness with Decomposable Neural Networks”, in Proc. of the The 4th Workshop on Tractable Probabilistic Modeling, 2021.

URL <https://openreview.net/forum?id=5E7V1tCwLq>

As deep neural networks are notoriously vulnerable to adversarial attacks, there has been significant interest in defenses with provable guarantees. Recent solutions advocate for a randomized smoothing approach to provide probabilistic guarantees, by estimating the expectation of a network’s output when the input is randomly perturbed. As the convergence of the estimated expectations depends on the number of Monte Carlo samples, and hence network evaluations, these techniques come at the price of considerable additional computation at inference time. We take a different route and introduce a novel class of deep models – decomposable neural networks (DecoNets) – which are hierarchical multi-linear functions over non-linear input features. DecoNets can compute the expectation over the outputs in closed form in a *single network evaluation*, thus providing *exact* smoothing guarantees. Our empirical analysis shows the promising nature of DecoNets: they achieve the same or better certified accuracy in comparison to models of equivalent size on benchmark datasets, while providing exact guarantees one or two orders of magnitude faster.

3.12 Probabilistic Circuits: Representations, Inference, Learning and Applications

Guy Van den Broeck (UCLA, US)

License © Creative Commons BY 4.0 International license
© Guy Van den Broeck

Joint work of Antonio Vergari, Guy Van den Broeck

URL <https://web.cs.ucla.edu/~guyvdb/talks/IJCAI20-tutorial/>

Exact and efficient probabilistic inference and learning are becoming more and more mandatory when we want to quickly take complex decisions in presence of uncertainty in real-world scenarios where approximations are not a viable option. In this tutorial, we will introduce probabilistic circuits (PCs) as a unified computational framework to represent and learn deep probabilistic models guaranteeing tractable inference. Differently from other deep neural estimators such as variational autoencoders and normalizing flows, PCs enable large classes of tractable inference with little or no compromise in terms of model expressiveness. Moreover, after showing a unified view to learn PCs from data and several real-world applications, we will cast many popular tractable models in the framework of PCs while leveraging it to theoretically trace the boundaries of tractable probabilistic inference.

3.13 Conditional Generative Models and Where to Apply Them

Max Welling (University of Amsterdam, NL)

License  Creative Commons BY 4.0 International license
© Max Welling

I talked about how we can use flow and diffusion models to generate data from the equilibrium distribution, but that it seems much harder to generate from conditional generative models of the form $F : (z, x) \rightarrow y$ with $z \sim p(z)$ and x some conditioning statement. These models are important for searching through chemical space, for proposing moves in a MCMC algorithm, for modeling domain shifts, etc. This talk will be mostly asking questions: why is this problem hard (harder than sampling from the unconditional distribution $F : z \rightarrow y$)?

3.14 Bayesian Deep Learning and a Probabilistic Perspective of Model Construction

Andrew G. Wilson (New York University, US)

License  Creative Commons BY 4.0 International license
© Andrew G. Wilson

Main reference Andrew Gordon Wilson, Pavel Izmailov: “Bayesian Deep Learning and a Probabilistic Perspective of Generalization”, CoRR, Vol. abs/2002.08791, 2020.

URL <https://arxiv.org/abs/2002.08791>

The key distinguishing property of a Bayesian approach is marginalization, rather than using a single setting of weights. Bayesian marginalization can particularly improve the accuracy and calibration of modern deep neural networks, which are typically underspecified by the data, and can represent many compelling but different solutions. We show that deep ensembles provide an effective mechanism for approximate Bayesian marginalization, and propose a related approach that further improves the predictive distribution by marginalizing within basins of attraction, without significant overhead. We also investigate the prior over functions implied by a vague distribution over neural network weights, explaining the generalization properties of such models from a probabilistic perspective. From this perspective, we explain results that have been presented as mysterious and distinct to neural network generalization, such as the ability to fit images with random labels, and show that these results can be reproduced with Gaussian processes. We also show that Bayesian model averaging alleviates double descent, resulting in monotonic performance improvements with increased flexibility. Finally, we provide a Bayesian perspective on tempering for calibrating predictive distributions.

Participants

- Alessandro Antonucci
IDSIA – Manno, CH
- Michael Chertkov
University of Arizona –
Tucson, US
- YooJung Choi
UCLA, US
- Alvaro Correia
TU Eindhoven, NL
- Priyank Jaini
Google – Toronto, CA
- Kristian Kersting
TU Darmstadt, DE
- Stefan Mengel
University of Artois/CNRS –
Lens, FR
- Eric Nalisnick
University of Amsterdam, NL
- Sriraam Natarajan
University of Texas – Dallas, US
- Mathias Niepert
Universität Stuttgart, DE
- Robert Peharz
TU Graz, AT
- Xiaoting Shao
TU Darmstadt, DE
- Guy Van den Broeck
UCLA, US
- Antonio Vergari
University of Edinburgh, GB
- Andrew G. Wilson
New York University, US



Remote Participants

- Marcus A. Brubaker
York University – Toronto, CA
- Cassio de Campos
TU Eindhoven, NL
- Nicola Di Mauro
University of Bari, IT
- Laurent Dinh
Montreal, CA
- Danilo Jimenez Rezende
Google DeepMind – London, GB
- Mikko Koivisto
University of Helsinki, FI
- Sara Magliacane
University of Amsterdam, NL
- Lilith Francesca Mattei
IDSIA – Lugano, CH
- Denis D. Mauá
University of Sao Paulo, BR
- Karthika Mohan
Oregon State University, US
- David Montalvan Hernandez
TU Eindhoven, NL
- Deepak Pathak
Carnegie Mellon University –
Pittsburgh, US
- Tahrira Rahman
University of Texas – Dallas, US
- Jakub Tomczak
VU University Amsterdam, NL
- Aki Vehtari
Aalto University, FI
- Max Welling
University of Amsterdam, NL
- Yaoliang Yu
University of Waterloo, CA
- Han Zhao
University of Illinois –
Urbana-Champaign, US