Leon Tiefenböck

Learning Probabilistic Circuits through sliced score matching

Bachelors's Thesis

to achieve the university degree of

Bachelor of Science

Bachelor's degree programme: Computer Science

submitted to

Graz University of Technology

Supervisor

Ass.-Prof. Dipl.-Ing. Dr. techn. Robert Peharz

Graz, October 2024

Abstract

Probabilistic Circuits (PCs) have shown to be a promising approach to probabilistic modelling, due to their many effiecient and exact inference opportunities while still being complex enough to model arbitrary distibutions. However they still are difficult to train and lack in results compared to other state of the art approaches in probabilistic modelling, especially with large high dimensional datasets. (Researchers fear that during the gradient based Maximum Likelihood optimization the algorithm gets stuck in local minima.) In this thesis I try to experimentally see if by using other learning objectives, namely sliced score matching, which recently gained prominence in learning energy based models (EBMs), we can work around these issues and achieve better results with PCs. I do this by training PCs using different algorithms on 2D and also high dimensional image data and comparing results.

Contents

1 Introduction

2 Background

2.1 Probabilistic Modelling

In Machine Learning (ML) one of the core disciplines is modelling data in some manner to be able to extract information, make predictions and perform all sorts of tasks of inference. One way to achieve this is to use probability theory as a framework which provides a sound and consistent way to reason under uncertainty. We assume that the data was randomly drawn from some underlying unknown distribution *out there* and we try to learn the distribution back from the data. [1]

I will get to how this can be learned in the future sections.

2.2 Probabilistic Circuits

Recent progress in probabilistic modelling through .. have pushed the expressive capability (how close the approximated distribution is to the true one) ever forward. However what expressive gains these models achieve through their complexity they loose in tractability, the ability to perform inference tasks exactly and efficiently. More concretely tractability in one inference task means that a model can perform the task exactly, without approximation, and in polynomial time with respct to the model size. Except for the simplest taks, these models must rely on costly approximation.

On the other hand older approaches like .. which are less complex and therefore lack in expressiveness can perform all sort of inference tasks tractably.

Probabilistic Circuits are a framework for probabilistic models that try to be both expressive and tractable. In essence they try to achieve this by only introducing complexity in a controlled manner. To be valid a PC must adhere to certain structural constraints.

2.3 Sliced Score Matching

Why score matching instead of minimizing NLL? should i even write about unnormalized models? how much detail for sm and ssm?

As described before, we want to learn the PDF of the data. In the case of PCs we directly model a normalized (integrates to 1) density but most other approaches model a unnormalized density. This is because the normalization constant is very costly or impossible to compute in these scenarios.

In score matching [2]

Which leads to this score matching objective:

However this introduces another problem. With high dimensional data the trace of the hessian is very costly to compute.

Here sliced score matching [3] is introduced. Instead of doing the costly hessian calculation the multidimensional data is projected onto n random vectors (slices). The authors in the paper show that ..

Which leads to this final sliced score matching objective:

3 Methods

4 Experimental Results

4.1 Dataset

5 Discussion

5.1 Interpretation of Experimental Results

6 Conclusions and Future Work

Appendix

Bibliography

[1] YooJung Choi, Antonio Vergari, and Guy Van den Broeck. "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Models." In: (Oct. 2020). URL: http://starai.cs.ucla.edu/papers/ProbCirc20.pdf (cit. on p. 2).

[2] Aapo Hyvärinen and Peter Dayan. "Estimation of non-normalized statistical models by score matching." In: *Journal of Machine Learning Research* 6.4 (2005) (cit. on p. 3).

[3] Yang Song et al. "Sliced Score Matching: A Scalable Approach to Density and Score Estimation." In: (2019). arXiv: 1905.07088 [cs.LG]. URL: https://arxiv.org/abs/1905.07088 (cit. on p. 3).