Structure Inference in Sum-Product Networks Using Infinite Sum-Product Trees

Martin Trapp[1,2], Robert Peharz[3], Marcin Skowron[1], Tamas Madl[1], Franz Pernkopf[2] and Robert Trappl[1]

[1]Austrian Research Institute for Artificial Intelligence
[2]Graz University of Technology
[3]Medical University of Graz

Motivation

Sum-Product Networks (SPNs) are a **highly efficient** type of a **deep probabilistic model** that allows **exact inference** in time linear in the size of the network. Previous work on Bayesian sum-product networks neglected induced trees in their posterior construction and could not report quantitative results and comparisons to existing approaches. We introduce the first **Bayesian nonparametric extension** of sum-product networks with a posterior distribution based on induced trees. We show that our infinite Sum-Product Trees (SPTs) allow to discover structures with **high modelling performance** while maintaining a good generalisation behaviour.

Sum-Product Networks

Sum-Product Networks (SPNs) can be defined recursively, as weighted sums and products of smaller SPNs, with univariate or multivariate probability distributions as leaf nodes. In a complete and decomposable SPN, all children of a sum node have the same variable scope as the sum, whereas the children of each product partition the product's scope into non-empty disjoint sub-scopes. Complete and decomposable SPNs can be represented as a sum of induced trees. The generative process of normalized SPNs can be described by:

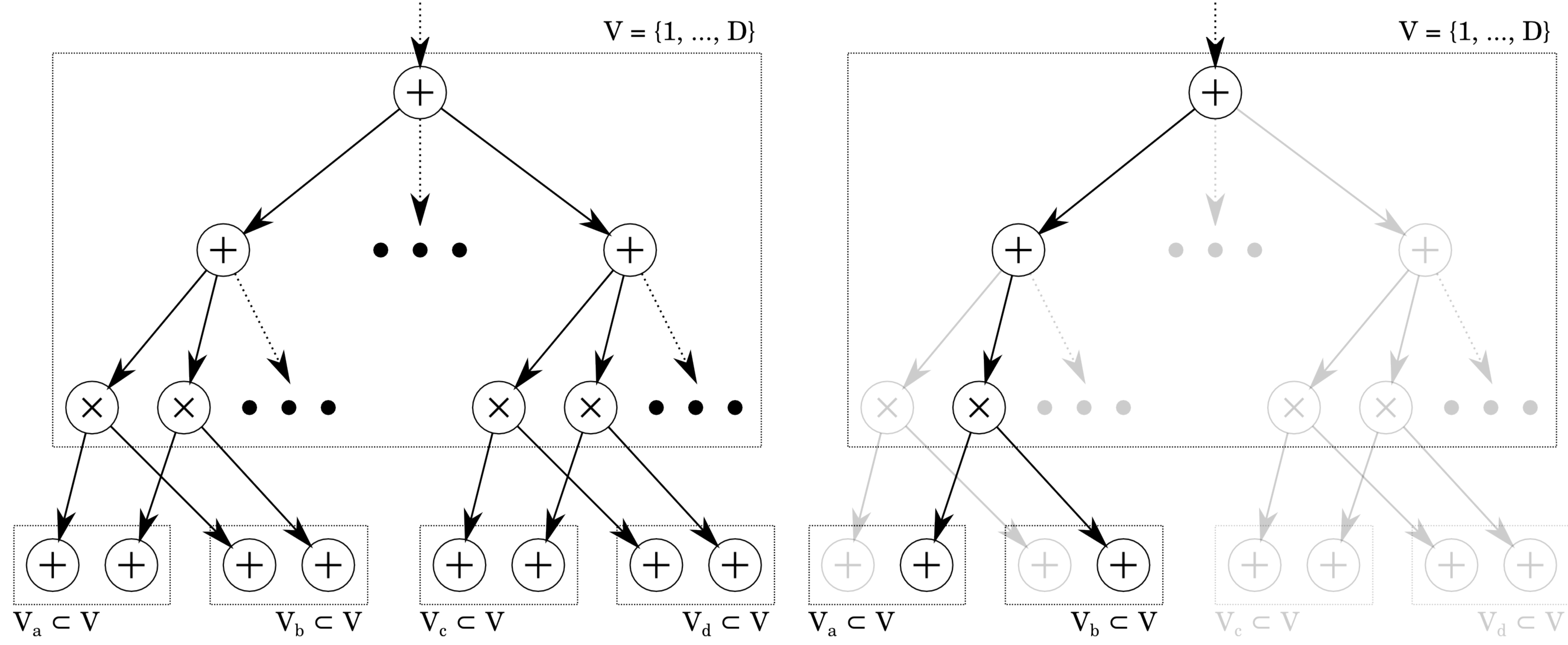
1. Selecting an induced tree with probability proportional to its weights: $P(\mathcal{T}) \propto \prod_{w \in \mathcal{T}} w$
2. Sampling the observation from the leaf nodes of $\mathcal{T}$.

Infinite Sum-Product Trees

Starting at the root node with scope V, for each observation $n = 1, \dots, N$

1. If the scope $V_S \subseteq V$ for the current node S is multivariate:
   - Draw weights $\boldsymbol{w}_S \sim \mathrm{Dir}(\alpha_S)$ to the group nodes directly below S.
   - Draw the latent assignments $c_{S,n} \sim \mathrm{Multi}(\boldsymbol{w}_S)$ to the group nodes and draw the partition of each group node $u_{c_{S,n}} \sim \mathrm{U}(1, \left\{ {|V_S| \atop 2} \right\})$ without replacement.
   - For each selected group node, draw latent assignments $z_{c_{S,n}} \sim \mathrm{CRP}(\beta_{c_{S,n}})$ for the observations to product nodes.
   - For each selected product, partition the scope into non-empty disjoint sub-scopes and for each child of the product, apply the infinite SPT process recursively.
2. Else for node S with univariate scope $d \in V$:
   - Draw latent leaf assignments $c_{S,n} \sim \mathrm{CRP}(\gamma_S)$ and draw distribution parameters $\theta_{c_{S,n}}$ from an appropriate prior.
   - Generate the value of the d^{th} dimension for the n^{th} observation from an appropriate leaf node distribution parametrized with $\theta_{c_{S,n}}$.

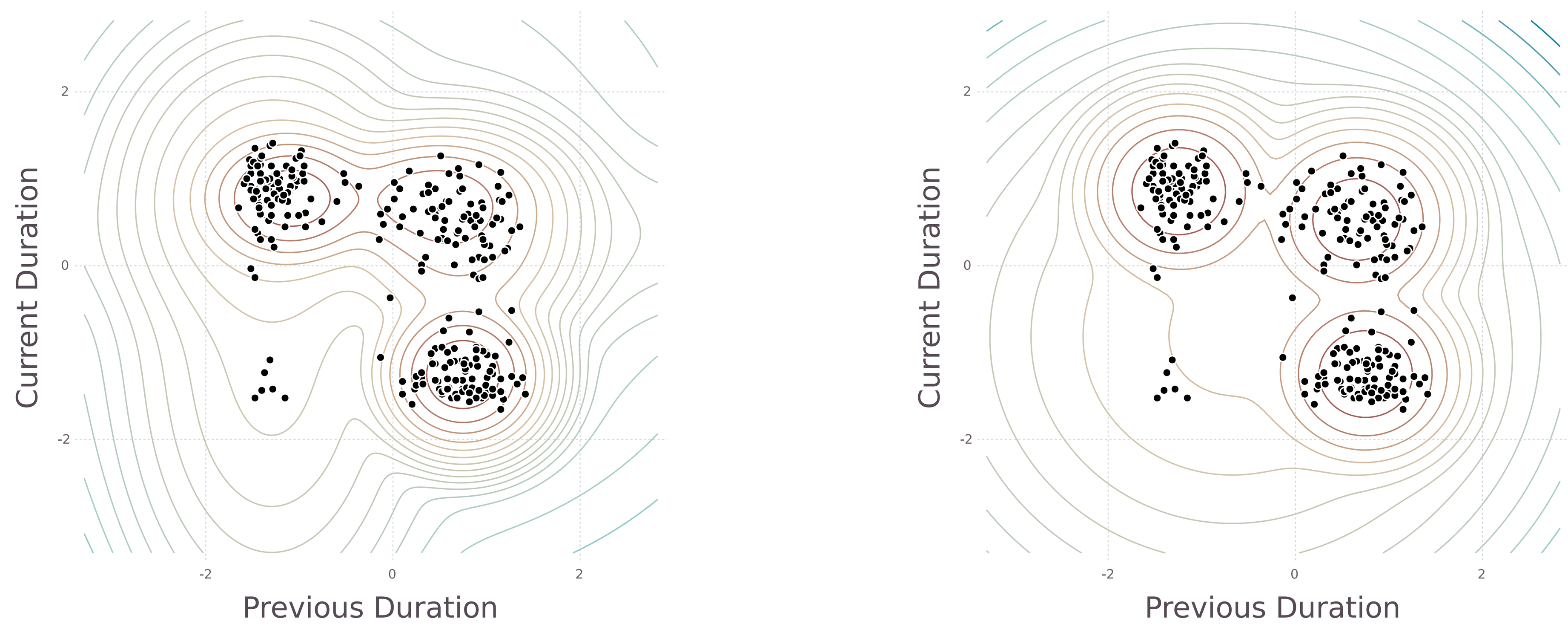
Figure: Illustration of infinite sum-product trees. Note that $V_i \neq V_j \; \forall (i,j) \in \{a,b,c,d\}$ and all V_i with $i \in \{a,b,c,d\}$ are non-empty sub-scopes of V. Moreover, $V_a \cup V_b = V$, $V_a \cap V_b = \emptyset$ and $V_c \cup V_d = V$, $V_c \cap V_d = \emptyset$.



Density Estimation Results

Infinite SPTs favour deep structures allowing the model to fit complex distributions more easily than shallow architectures.

Figure: Log density modelled on Old Faithful data by an infinite SPT and an infinite Gaussian mixture.



Infinite SPT   Infinite Gaussian mixture

Table: Average 10-fold cross-validation log predictive densities and Mann-Whitney U p-values.

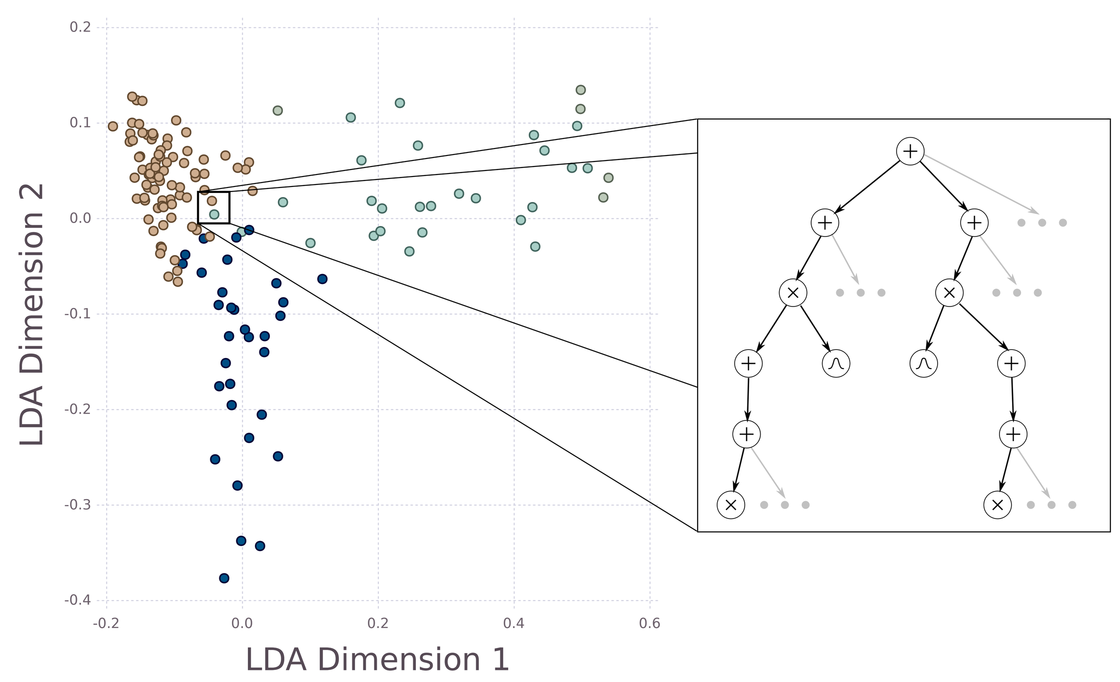
Dataset	infinite Gaussian mixture	infinite SPT	p-values
Old Faithful	-1.737	-1.700	< 0.01
Chemical Diabetes	-3.022	-2.879	< 0.01
Iris	-3.943	-3.744	< 0.01

[1] A. Darwiche. A differential approach to inference in Bayesian networks. *Journal of the ACM (JACM)*, 50(3):280–305, 2003.
[2] H. Poon and P. Domingos. Sum-product networks: A new deep architecture. In *UAI*, pages 337–346, 2011.
[3] S.W. Lee, C.J. Watkins, and B.T. Zhang. Non-parametric bayesian sum-product networks. In *Workshop on LTPM*, 2014.
[4] H. Zhao, T. Adel, G. Gordon, and B. Amos. Collapsed variational inference for sum-product networks. In *ICML*, pages 1310–1318, 2016.

Variable Discovery

In contrast to infinite Gaussian mixtures, observations are generated from induced trees. We therefore implicitly obtain an assignment hierarchy which can be analysed using a dendrogram constructed on the induced tree assignments.

Figure: Group assignments estimated by an infinite SPT on the Chemical Diabetes dataset. Colouring of the assignments encodes the induced trees mapped to a one dimensional embedding. The enlarged section illustrates the induced trees of two observations.



Acknowledgments

Funded by the Austrian Science Fund, grant no. P 27530.