Learning Deep Mixtures of Gaussian Process Experts Using Sum-Product Networks

Martin Trapp[1,2], Robert Peharz[3], Carl E. Rasmussen[3], and Franz Pernkopf[1]

[1] SPSC Lab, Graz University of Technology, Graz, Austria
[2] Austrian Research Institute for Artificial Intelligence, Vienna, Austria
[3] Computational and Bioligical Learning Lab, Universtiy of Cambridge, Cambridge, UK

Gaussian Processes (GPs)

Due to their non-parametric nature, Gaussian Processes (GPs) are a powerful and flexible principled way for non-linear probabilistic regression.

However, a limitation of GPs is that learning scales in $O(N^3)$ and has a memory consumption of $O(N(N+D))$.

The main computational burdens of GPs are the inversion of the covariance matrix and the computation of its determinant.

Sum-Product Networks (SPNs)

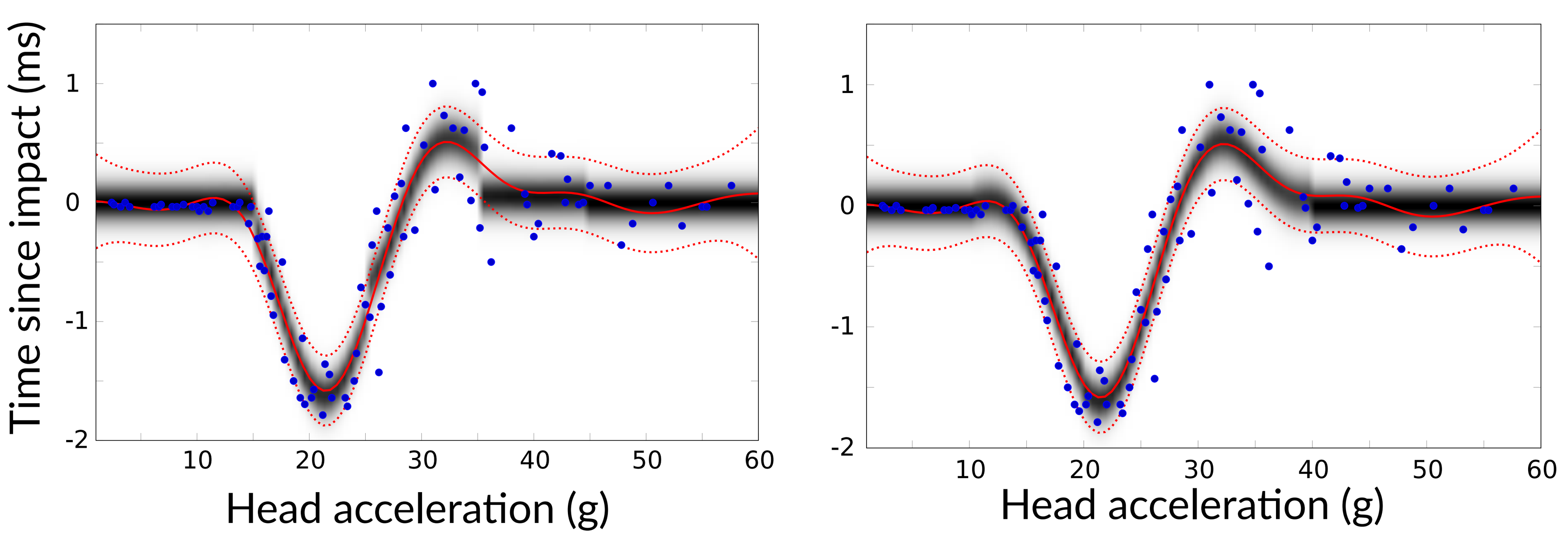
Sum-Product Networks (SPNs) are an expressive class of probabilistic models allowing exact and efficient inference.

They represent probability distributions by recursively utilizing factorization and mixtures according to an acyclic directed graph.

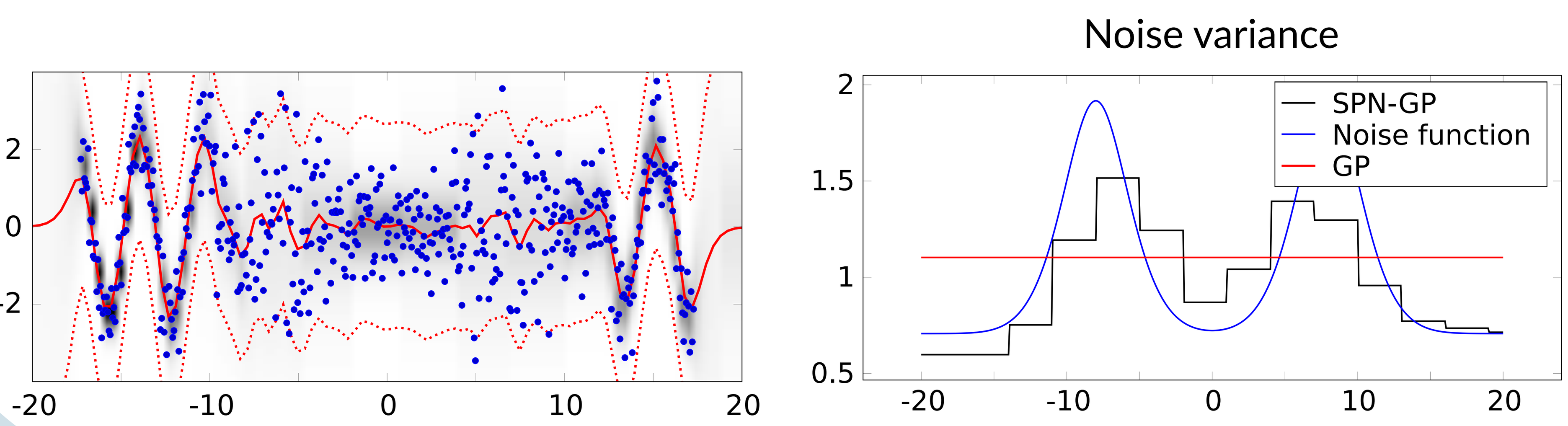
The crucial insight exploited in this paper is that SPNs are a sound language for any probabilistic model used as leaves.

Qualitative Results

Analysis of the challenges and capacities of the SPN-GP model. Right figure illustrates the density region of SPN-GPs with reduced discontinuities.



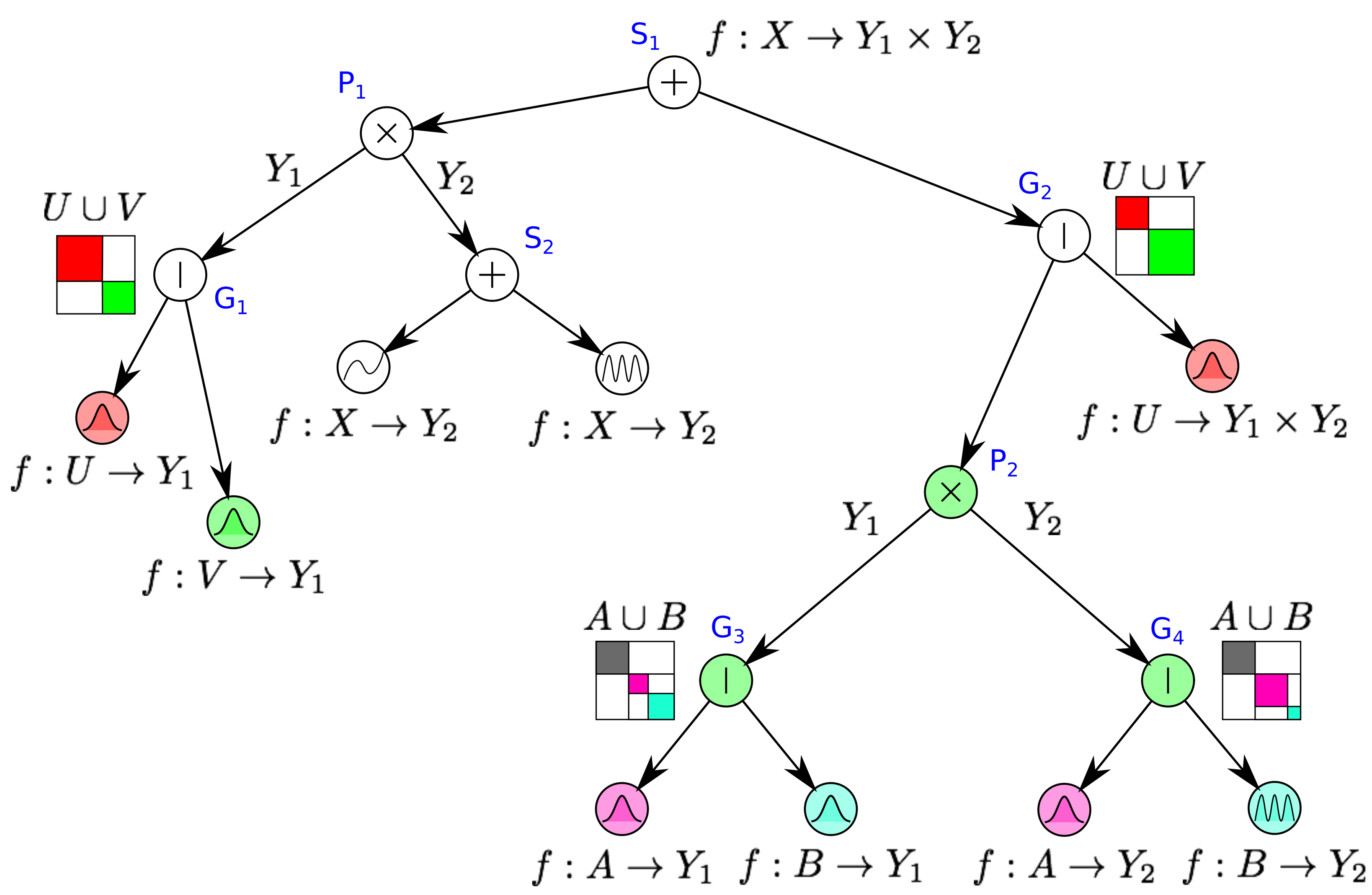
Qualitative experiment using non-stationary data with conditional heteroscedastic noise. SPN-GPs naturally account for input-dependent hyper-paramters.



SPNs with GP Leaves (SPN-GP)

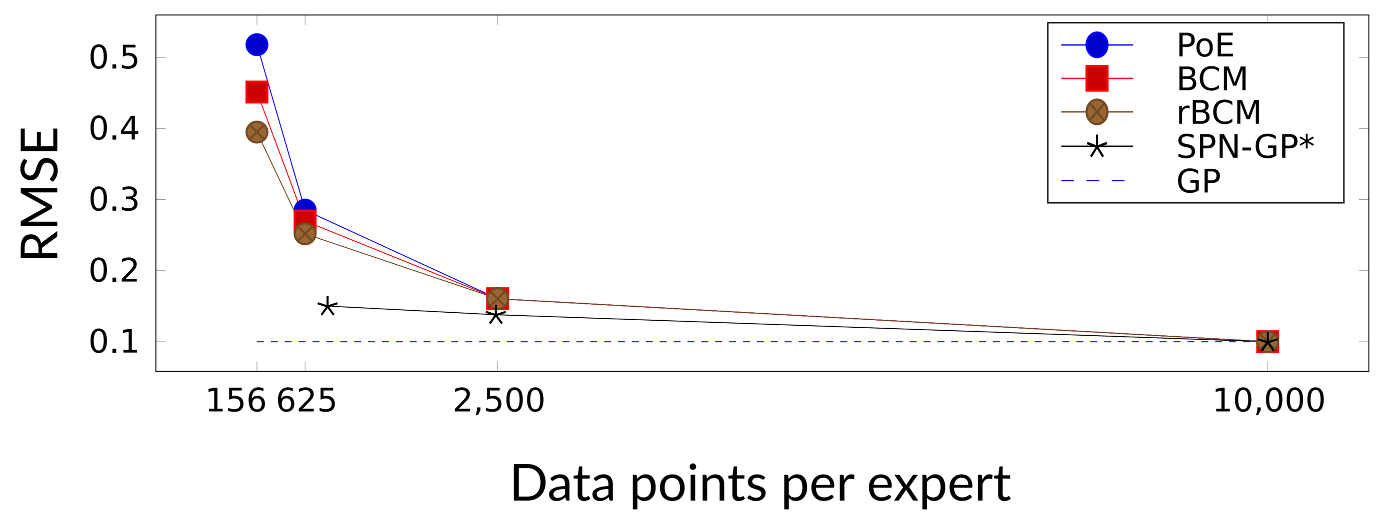
An SPN-GP is a hierarchically structured mixture model which recursively combines mixtures over sub-SPN-GPs or GPs with sub-divisions of the input space.

The SPN-GP model is a very flexible model which allows to: (1) mix over different block-diagonal covariance re-presentations (2) mix over GPs with different kernel functions, (3) hierarchically sub-divides the input space and (4) partitions the dependent variables subject to the covariates.



Quantitative Results

Assessment of the approximation error on the Kin40K data set. The results suggest that SPN-GPs do not suffer from vanishing variances or the effect of weak experts.



Assessment of the performance of SPN-GPs against different baseline approaches: mean prediction, least squares (LLS), ridge regression and a full GP.

Method	Energy	Concrete	CCPP
Mean	9.83 ± 0.09	16.45 ± 0.10	17.00 ± 0.05
LLS	3.08 ± 0.02	10.33 ± 0.25	4.63 ± 0.04
Ridge	3.08 ± 0.02	10.33 ± 0.25	4.63 ± 0.04
GP	2.44 ± 0.17	$\mathbf{6.25} \pm 0.14$	-
SPN-GP	2.23 ± 0.11	6.27 ± 0.20	$\mathbf{4.10} \pm 0.05$
SPN-GP*	$\mathbf{2.07} \pm 0.04$	$\mathbf{6.25} \pm 0.14$	4.11 ± 0.04

ICML 2018 Workshop on Tractable Probabilistic Models
July 15th, 2018
Stockholm, Sweden

UNIVERSITY OF CAMBRIDGE
TU Graz
OFAI