On hardware-aware probabilistic frameworks for resource constrained embedded applications

Laura I. Galindez Olascoaga, Wannes Meert, Nimish Shah, Guy Van den Broeck and Marian Verhelst

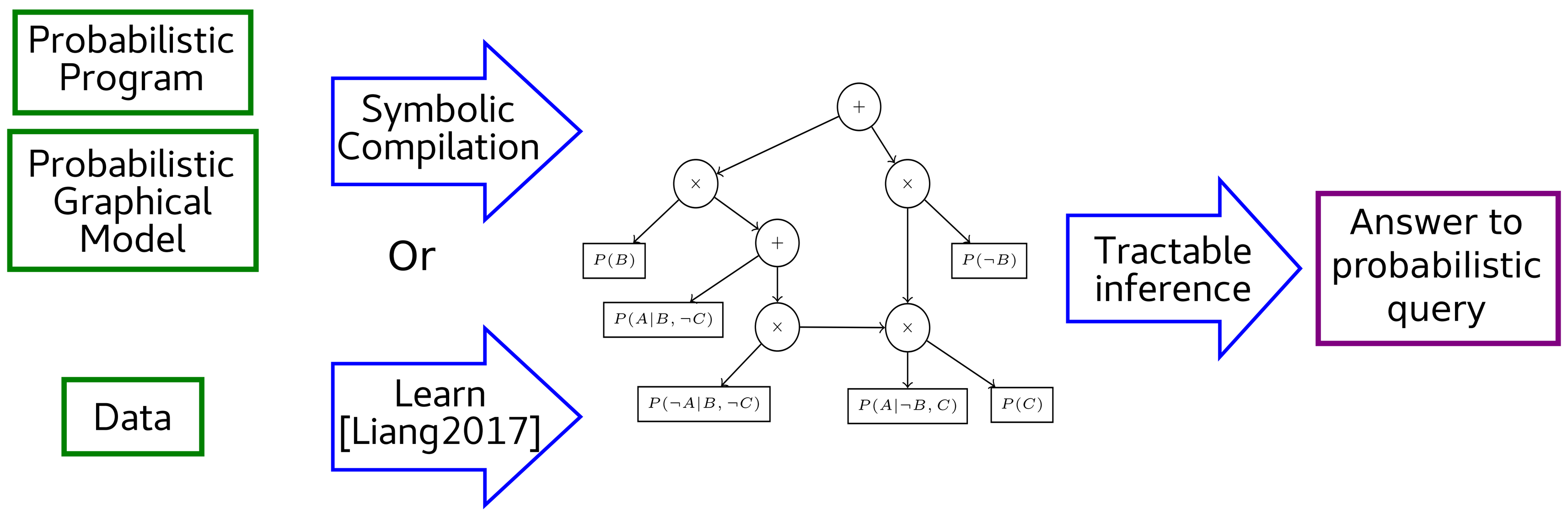
Background and motivation

Tractable probabilistic models

Tractable probabilistic models possess a set of traits that make them ideal for embedded sensory applications:

-**Robustness** to missing data allows them to cope with sensor failure.
-**Small data needs** allow them to adapt to different users quickly.
-Tractability enables **reliable and efficient inference** under constrained resources.

Arithmetic Circuits are one of those tractable representations [Darwiche2009]:

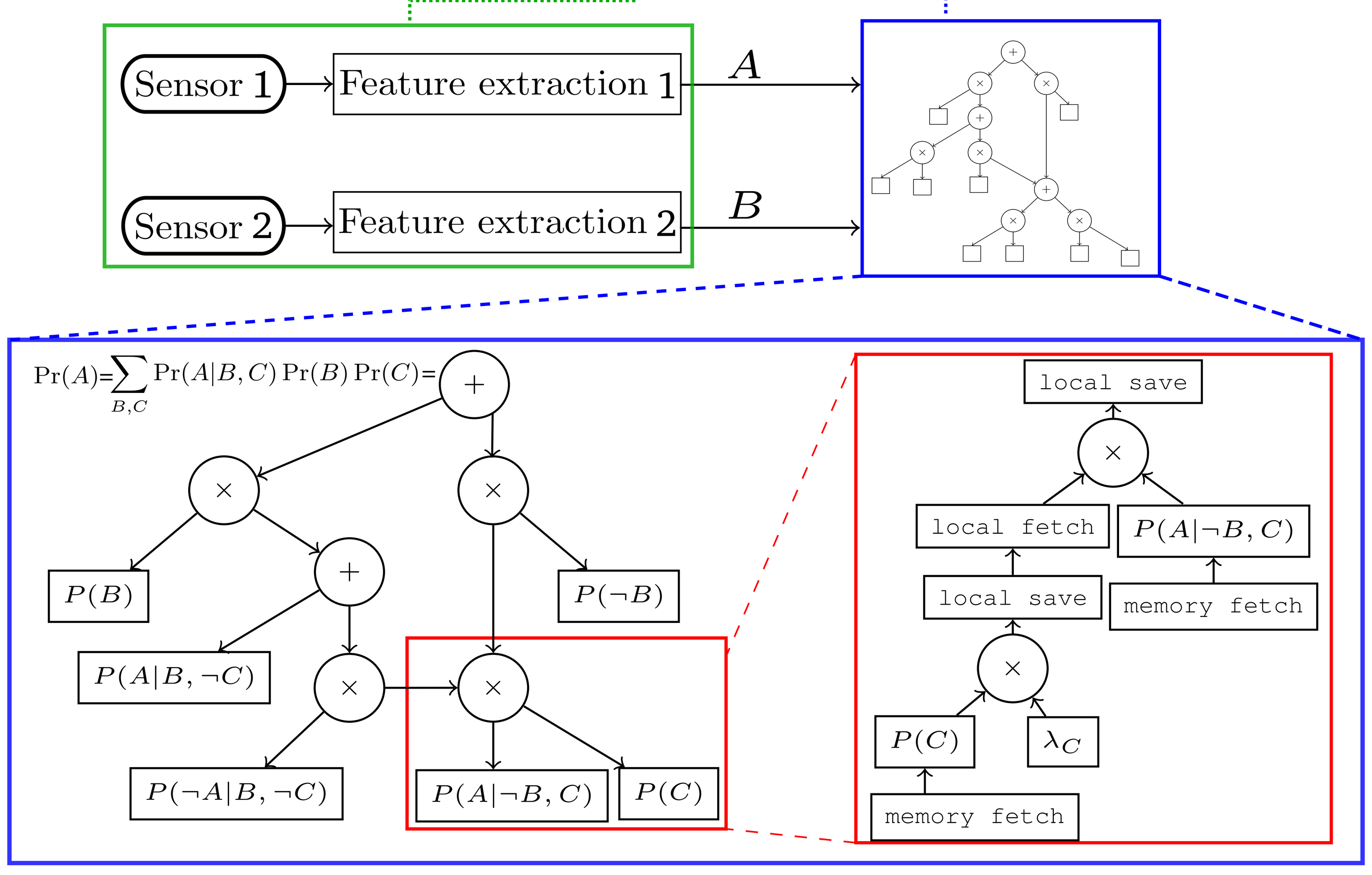


But current tractability notions disregard hardware-implementation details and are given in abstract tems such as time and space. Recently proposed **hardware-aware probabilistic frameworks** address these limitations by explicitely encoding hardware properties in the model.

Hardware-aware cost

For example, [Galindez2019] proposes the **Hardware-aware cost**, given in terms of relative energy consumption.

$$\mathrm{C}_{\mathrm{HA}}(\mathbf{S},\mathbf{F},\alpha,nb)=\sum_{S\in\mathbf{S}}\mathrm{C}_{\mathrm{SI}}(S,\mathbf{F}_S)+\mathrm{C}_{\mathrm{AC}}(\alpha,nb)$$



Hardware-aware probabilisic framework

Goal

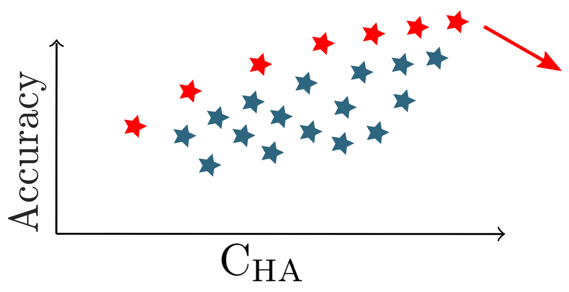
The notion of hardware-aware cost allows us to map the device's configuration, into a common resource vs. performance trade-off space.

Scalable system properies:

(1) Model complexity α
(2) Feature set $\mathbf{F}$
(3) Sensor set S
(4) Precision nb

System configuration
$\sigma=\{\alpha,\mathbf{F},S,nb\}$

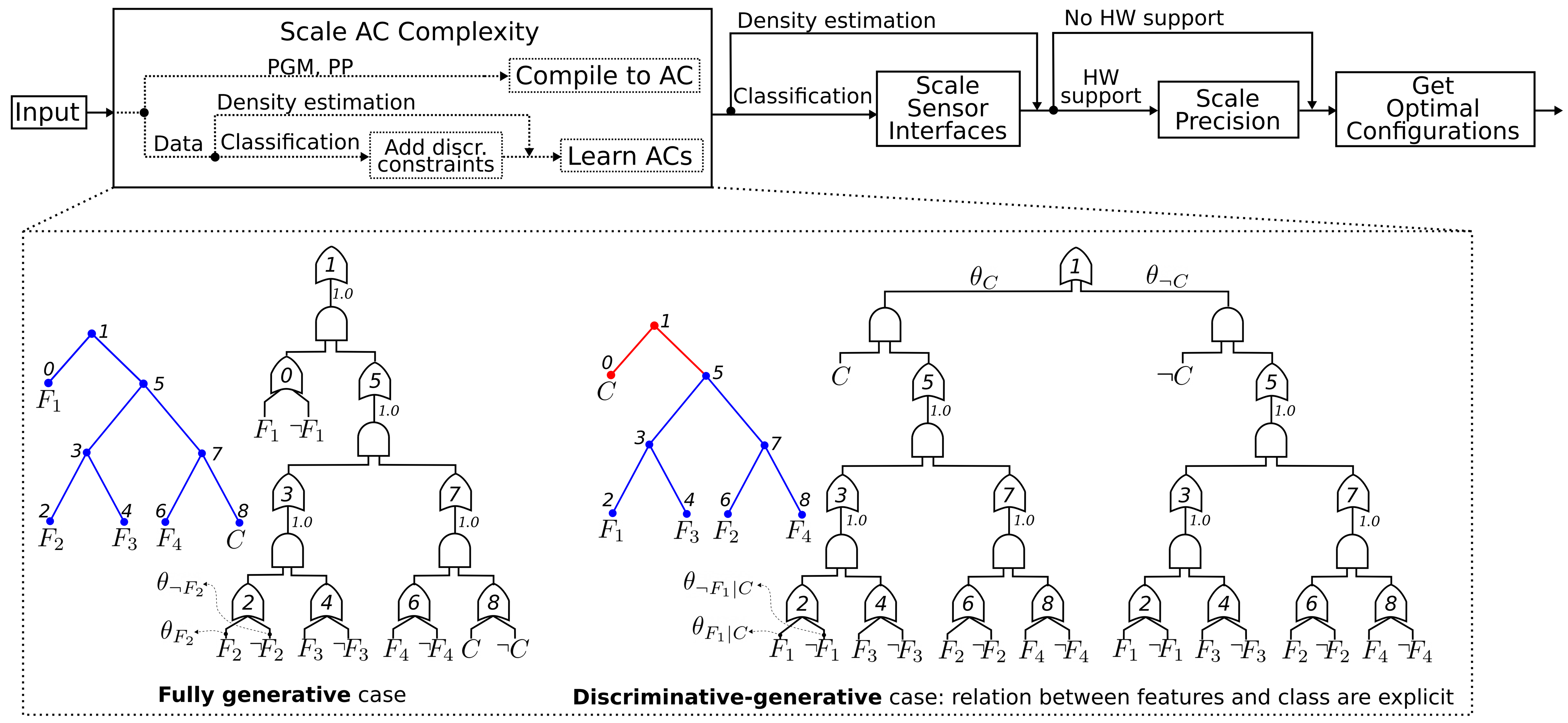
Map to trade-off space



Find the optimal configuration for a given accuracy and cost σ^*

Framework

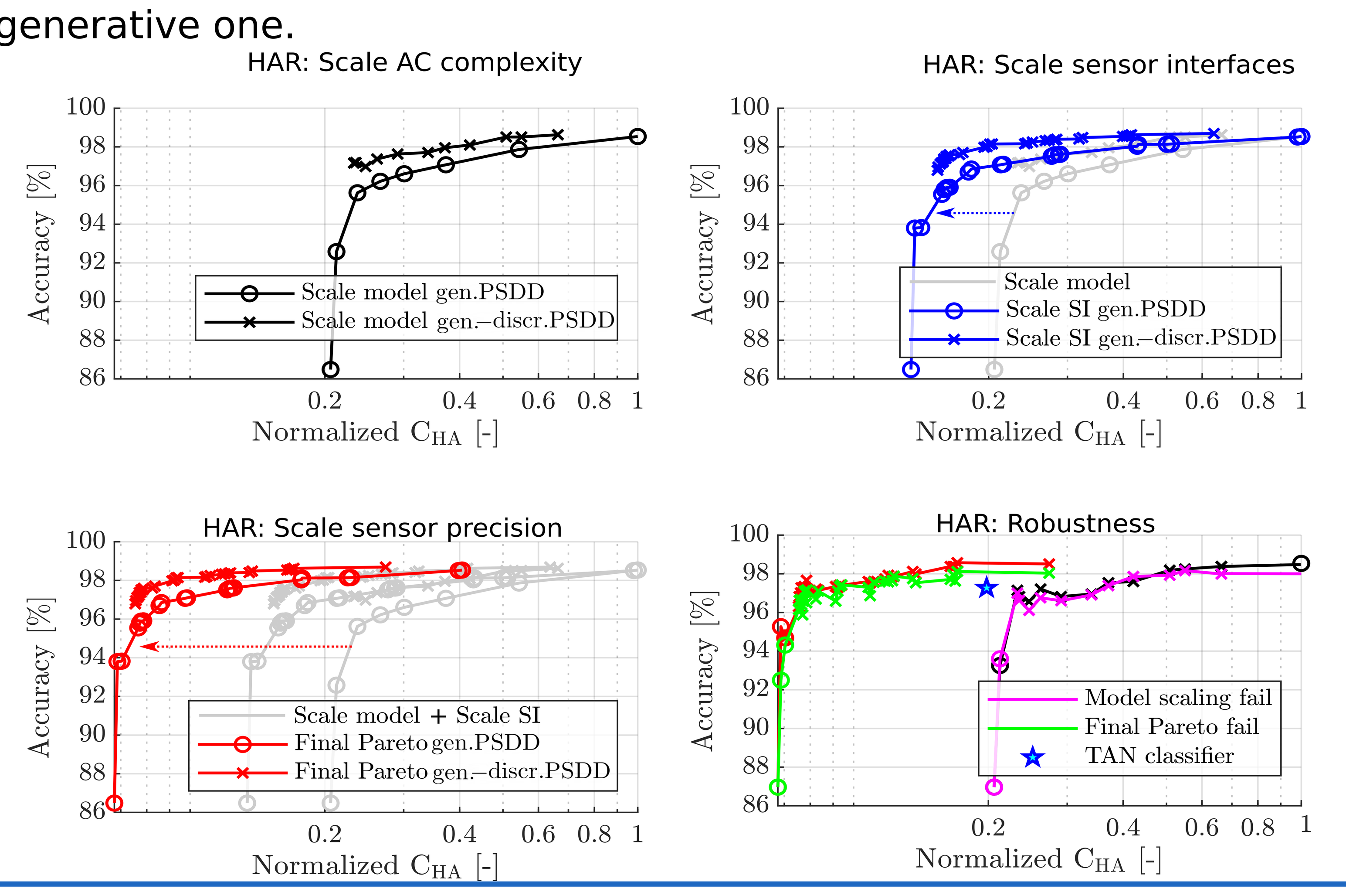
The framework relies on a **sequence of greedy searches** that scale each system property individually. It also accomodates density estimation tasks and classification tasks, as well of different AC learning strategies. Moreover, ACs allow to encode logical constraints, which we exploit to add a discriminative bias to the otherwise generative model learning procedure [Liang2017].



Experiments and references

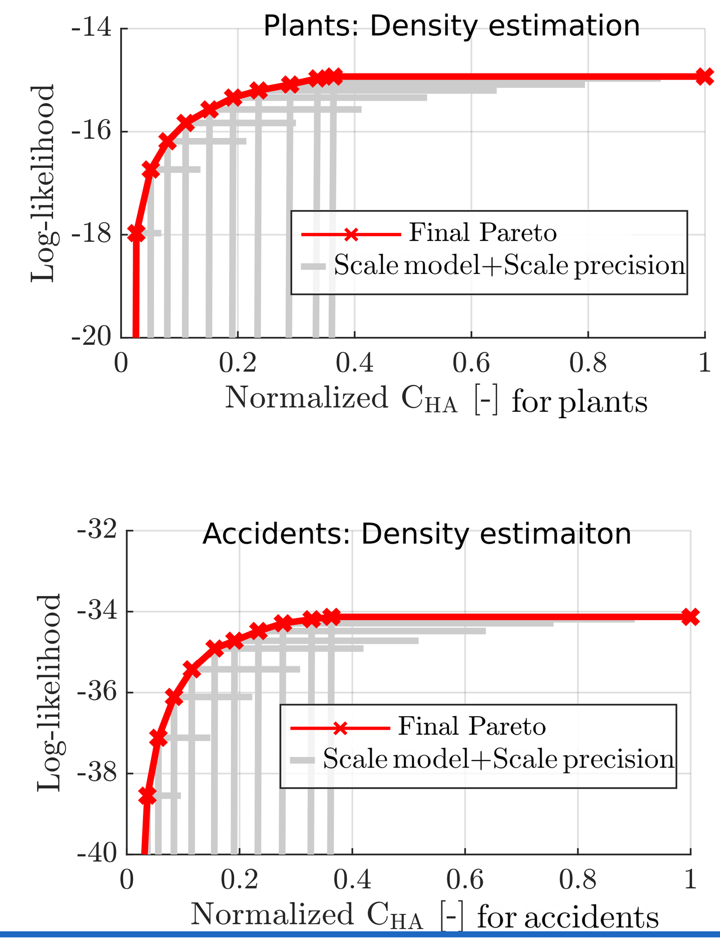
Classification

Experiments on a Human Activity Recognition (HAR) benchmark show significant cost savings with minimal accuracy loss. Moreover, the discriminative approach attains higher accuracy than the purely generative one.



Density estimation

Experiments on standard density estimation benchmarks show that the framework in [NeurIPS2019] can benefit tasks other than classification.



References

[Darwiche2009]:A. Darwiche. Modeling and Reasoning with Bayesian Networks. Cambridge University Press, 2009.

[Liang2017]: Y. Liang, J. Bekker and G. Van den Broeck. Learning the Structure of Probabilistic Sentential Decision Diagrams. UAI 2017.

[Galindez 2019]: L. Galindez Olascoaga, W. Meert, N. Shah, M. Verhelst and G. Van den Broeck. Towards Hardware-Aware Tractable Learning of Probabilistic Models. NeurIPS 2019

Contact: Laura I. Galindez Olascoaga (laura.galindez@esat.kuleuven.be)

EMC^2: 5th Edition Co-located with NeurIPS'19