Safe Semi-Supervised Learning of Sum-Product Networks

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

[1]Signal Processing and Speech Communication Lab., Graz University of Technology, Austria

[2]Austrian Research Institute for Artificial Intelligence, Austria

[3]Computational and Biological Learning Lab., University of Cambridge, UK

Motivation

- In several domains, obtaining class labels is expensive
- Most semi-supervised approaches enforce restrictive assumptions on the data distribution
- **Goal:** Non-restrictive semi-supervised learning

Sum-Product Networks (SPNs)

- Deep probabilistic model capturing expressive variable interactions, while guaranteeing exact and efficient inference
- Generative parameter learning using efficient Expectation Maximisation (EM) [1, 2]
- Discriminative parameter learning using backpropagation for cond. llh maximisation [3]

Our Approach (MCP-SPN)

- First semi-supervised learning approach for SPNs
- Allows generative and discriminative learning of non-linear decision boundaries
- Guarantees that adding unlabelled data can increase, but not degrade, the training performance (safe)

$$L(\theta^{*}|\mathcal{X},\mathcal{U},\boldsymbol{q}) \geqslant L(\theta^{+}|\mathcal{X},\mathcal{U},\boldsymbol{q})$$

- Is computationally efficient (scales linearly)
- Does not enforce restrictive assumptions on the data distribution

Generative Training

- SPNs with soft-labels

$$\mathcal{S}[\boldsymbol{u},\boldsymbol{q}|\theta] = \sum_{k=1}^{K} q_k w_k S_k[\boldsymbol{u}|\boldsymbol{q},\theta]$$

- Optimisation given labelled and unlabelled data

$$\boxed{L(\theta|\mathcal{X},\mathcal{U},\boldsymbol{q}) = \sum_{n=1}^{N} \log \mathcal{S}[\boldsymbol{x}_n,\boldsymbol{y}_n|\theta] + \sum_{m=1}^{M} \log \mathcal{S}[\boldsymbol{u}_m,\boldsymbol{q}_m|\theta]}$$

- Pessimistic update of soft-labels

$$\nabla q_{mk} = \frac{\partial L(\theta^{*}|\mathcal{X},\mathcal{U},\boldsymbol{q})}{\partial q_{mk}} - \frac{\partial L(\theta^{+}|\mathcal{X},\mathcal{U},\boldsymbol{q})}{\partial q_{mk}}$$

Discriminative Training

- Optimisation of conditional log likelihood given labelled and unlabelled data using backpropagation

$$\boxed{CL(\theta|\mathcal{X},\mathcal{U},\boldsymbol{q}) = \sum_{n=1}^{N} \log \mathcal{S}[\boldsymbol{y}_n|\boldsymbol{x}_n,\theta] + \sum_{m=1}^{M} \log \mathcal{S}[\boldsymbol{q}_m|\boldsymbol{u}_m,\theta]}$$

- Pessimistic update of soft-labels as in the generative setting
- Initialisation of soft-labels using optimistic labelling for discriminative training and random draws from a Dirichlet distribution for generative training

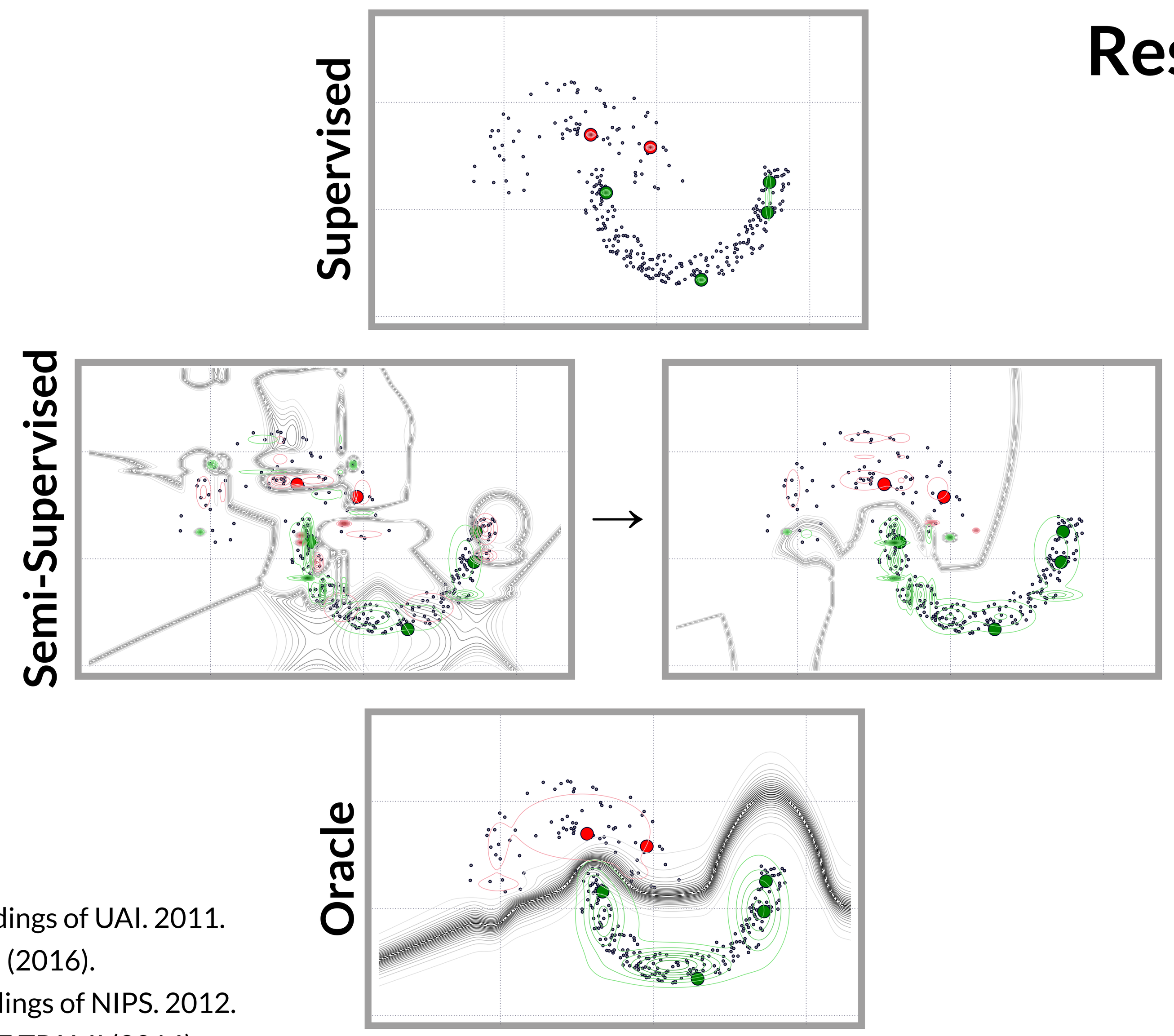
Contrastive Pessimistic Likelihood Estimation [4]

- Semi-supervised learning for generative linear models
- Maintains soft-labels for each unlabelled observation
- Alternates between optimistic parameter updates and pessimistic soft-label updates

$$\boxed{\theta^{*} = \operatorname*{argmax}_{\theta\in\Theta} \operatorname*{argmin}_{\boldsymbol{q}} L(\theta|\mathcal{X},\mathcal{U},\boldsymbol{q}) - L(\theta^{+}|\mathcal{X},\mathcal{U},\boldsymbol{q})}$$

- Results in a safe solution, since the supervised solution can allways be used in the worst case

Qualitative Results



Quantitative Results

Generative

Data Set	Supervised	SSL	Oracle	MCPLDA
BUPA	$-438.75\pm7\cdot10^{0}$	$\mathbf{-7.31}\pm6\cdot10^{-2}$	$-8.80\pm2\cdot10^{-1}$	$-9.07\pm3\cdot10^{-2}$
Fertility	$-3.31\pm3\cdot10^{-2}$	$\mathbf{-3.06}\pm7\cdot10^{-3}$	$-3.00\pm6\cdot10^{-3}$	$-12.68\pm5\cdot10^{-2}$
Haberman	$-138.63\pm4\cdot10^{0}$	$\mathbf{-5.05}\pm6\cdot10^{-2}$	$-5.14\pm6\cdot10^{-2}$	$-7.83\pm1\cdot10^{-1}$
ILPD	$-5.62\pm3\cdot10^{0}$	$\mathbf{-1.15}\pm2\cdot10^{-2}$	$-1.00\pm1\cdot10^{-2}$	$-37.54\pm1\cdot10^{-1}$
Ionosphere	$-2.83\pm5\cdot10^{-2}$	$\mathbf{-1.61}\pm1\cdot10^{-2}$	$-1.52\pm9\cdot10^{-3}$	$-46.12\pm5\cdot10^{-2}$
Iris	$-20.65\pm9\cdot10^{-1}$	$-3.78\pm3\cdot10^{-2}$	$-2.17\pm1\cdot10^{-2}$	$\mathbf{-2.65}\pm5\cdot10^{-2}$
Parkinsons	$\mathbf{-1.32}\pm4\cdot10^{-3}$	$-1.34\pm4\cdot10^{-3}$	$-1.30\pm2\cdot10^{-3}$	$-2.27\pm5\cdot10^{-2}$
WDBC	$\mathbf{-1.90}\pm1\cdot10^{-3}$	$-1.93\pm2\cdot10^{-3}$	$-1.88\pm3\cdot10^{-4}$	$-10.75\pm1\cdot10^{-2}$
Wine	$\mathbf{-2.47}\pm4\cdot10^{-3}$	$\mathbf{-2.47}\pm2\cdot10^{-3}$	$-2.44\pm9\cdot10^{-4}$	$-15.28\pm2\cdot10^{-2}$

Discriminative

Data Set	Supervised	SSL	Oracle	TSVM	ICLSC	MER
BUPA	$0.41\pm1\cdot10^{-2}$	$0.40\pm1\cdot10^{-2}$	$0.48\pm5\cdot10^{-3}$	$0.36\pm2\cdot10^{-2}$	$\mathbf{0.47}\pm7\cdot10^{-3}$	$0.42\pm1\cdot10^{-2}$
Fertility	$0.07\pm2\cdot10^{-2}$	$0.03\pm1\cdot10^{-2}$	$0.06\pm2\cdot10^{-2}$	$0.07\pm2\cdot10^{-2}$	$0.07\pm2\cdot10^{-2}$	$\mathbf{0.12}\pm2\cdot10^{-2}$
Haber.	$0.23\pm2\cdot10^{-2}$	$0.28\pm2\cdot10^{-2}$	$0.25\pm0.$	$0.20\pm2\cdot10^{-2}$	$\mathbf{0.33}\pm1\cdot10^{-2}$	$0.27\pm1\cdot10^{-2}$
ILPD	$0.17\pm2\cdot10^{-2}$	$0.20\pm1\cdot10^{-2}$	$0.24\pm4\cdot10^{-3}$	$0.23\pm2\cdot10^{-2}$	$0.29\pm1\cdot10^{-2}$	$\mathbf{0.33}\pm2\cdot10^{-2}$
Ionos.	$0.79\pm4\cdot10^{-3}$	$\mathbf{0.82}\pm4\cdot10^{-3}$	$0.87\pm0.$	$0.66\pm9\cdot10^{-3}$	$0.61\pm9\cdot10^{-3}$	$0.70\pm7\cdot10^{-3}$
Iris	$0.73\pm1\cdot10^{-2}$	$\mathbf{0.88}\pm1\cdot10^{-2}$	$0.93\pm0.$	$0.72\pm1\cdot10^{-2}$	$0.74\pm2\cdot10^{-2}$	$0.80\pm6\cdot10^{-3}$
Parkins.	$0.72\pm1\cdot10^{-2}$	$\mathbf{0.77}\pm4\cdot10^{-3}$	$0.82\pm4\cdot10^{-3}$	$0.74\pm1\cdot10^{-2}$	$0.66\pm2\cdot10^{-2}$	$0.68\pm1\cdot10^{-2}$
PID	$0.38\pm1\cdot10^{-2}$	$0.45\pm1\cdot10^{-2}$	$0.64\pm8\cdot10^{-4}$	$0.45\pm1\cdot10^{-2}$	$0.54\pm7\cdot10^{-3}$	$\mathbf{0.57}\pm9\cdot10^{-3}$
WDBC	$0.85\pm3\cdot10^{-3}$	$0.90\pm2\cdot10^{-3}$	$0.92\pm3\cdot10^{-4}$	$0.91\pm4\cdot10^{-3}$	$0.88\pm4\cdot10^{-3}$	$\mathbf{0.92}\pm3\cdot10^{-3}$
Wine	$0.82\pm7\cdot10^{-3}$	$\mathbf{0.97}\pm2\cdot10^{-3}$	$0.97\pm4\cdot10^{-3}$	$0.96\pm2\cdot10^{-3}$	$0.95\pm7\cdot10^{-3}$	$0.95\pm9\cdot10^{-3}$

[1] Poon, Hoifung and Domingos, Pedro. "Sum-product networks: a new deep architecture." In proceedings of UAI. 2011.
[2] Peharz, Robert, et al. "On the latent variable interpretation in sum-product networks." IEEE TPAMI (2016).
[3] Gens, Robert, and Pedro Domingos. "Discriminative learning of sum-product networks." In proceedings of NIPS. 2012.
[4] Loog, Marco. "Contrastive pessimistic likelihood estimation for semi-supervised classification." IEEE TPAMI (2016).

Conference on Uncertainty in Artificial Intelligence
Sydney, Australia
August 11-15, 2017
uai2017

This work was partially funded by the Austrian Science Fund grant no. P 27530 and grant no. P 27803-N15.
We further want to acknowledge the support through the UAI Scholarship and the Rudolf Chaudoire-Stifftung.