Deep Probabilistic Programming (for Healthcare)*

*Thanks for Sriraam Natarajan (UT Dallas) and many others for all the great collaborations

THE UNIVERSITY OF TEXAS AT DALLAS

Kristian Kersting

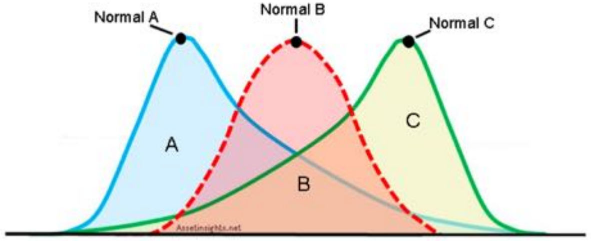
TECHNISCHE UNIVERSITÄT DARMSTADT

Centre for Cognitive Science

Fachbereich Informatik

ELLIS European Laboratory for Learning and Intelligent Systems

CLAIRE CONFEDERATION OF LABORATORIES FOR ARTIFICIAL INTELLIGENCE RESEARCH IN EUROPE



Getting deep systems that reason and know when they don’t know

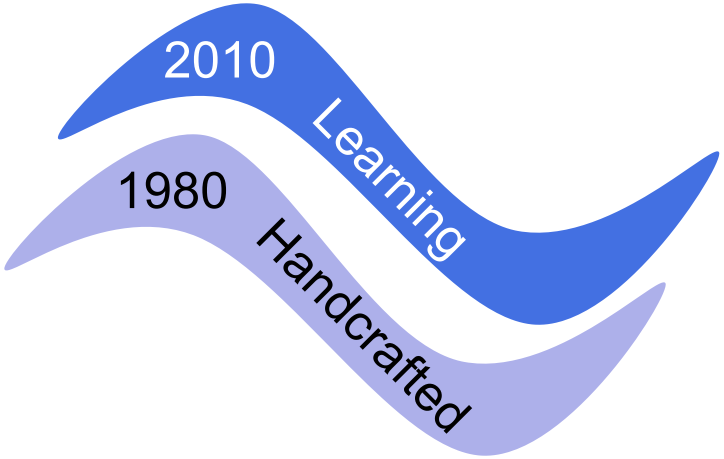
Responsible AI systems that explain their decisions and co-evolve with the humans

Open AI systems that are easy to realize and deal with complex data and knowledge

AI has impact

Data are now ubiquitous; there is great value from understanding this data, building models and making predictions

However, there are not enough data scientists, statisticians, machine learning and AI experts.



2010

Learning

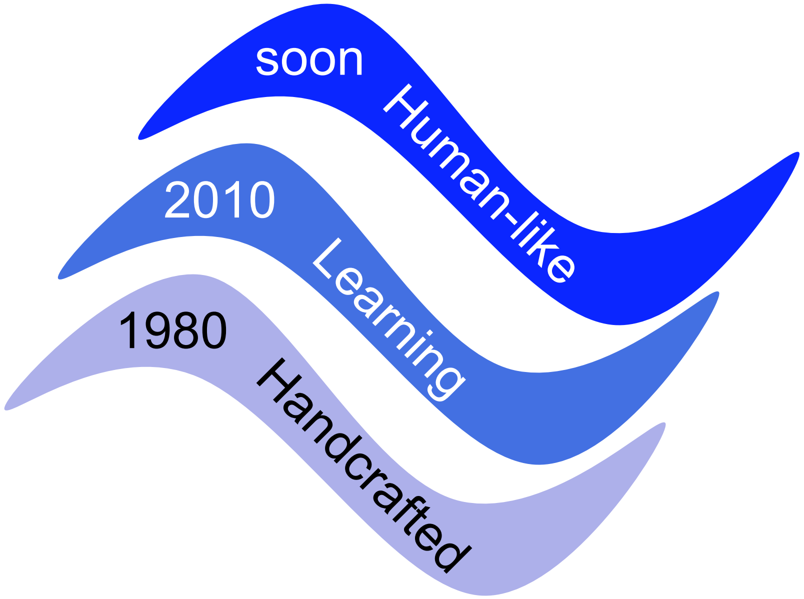
1980

Handcrafted

The third wave of AI

Data are now ubiquitous; there is great value from understanding this data, building models and making predictions

However, there are not enough data scientists, statisticians, machine learning and AI experts.



soon

Human-like

2010

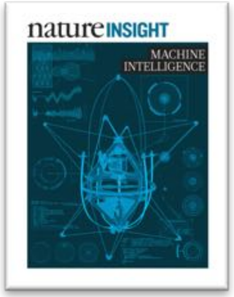
Learning

1980

Handcrafted

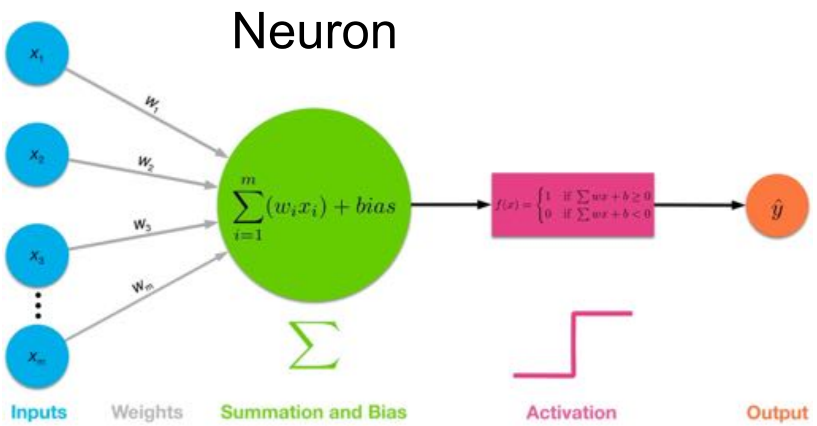
AI systems that can acquire human-like communication and reasoning capabilities, with the ability to recognise new situations and adapt to them.

Deep Neural Networks

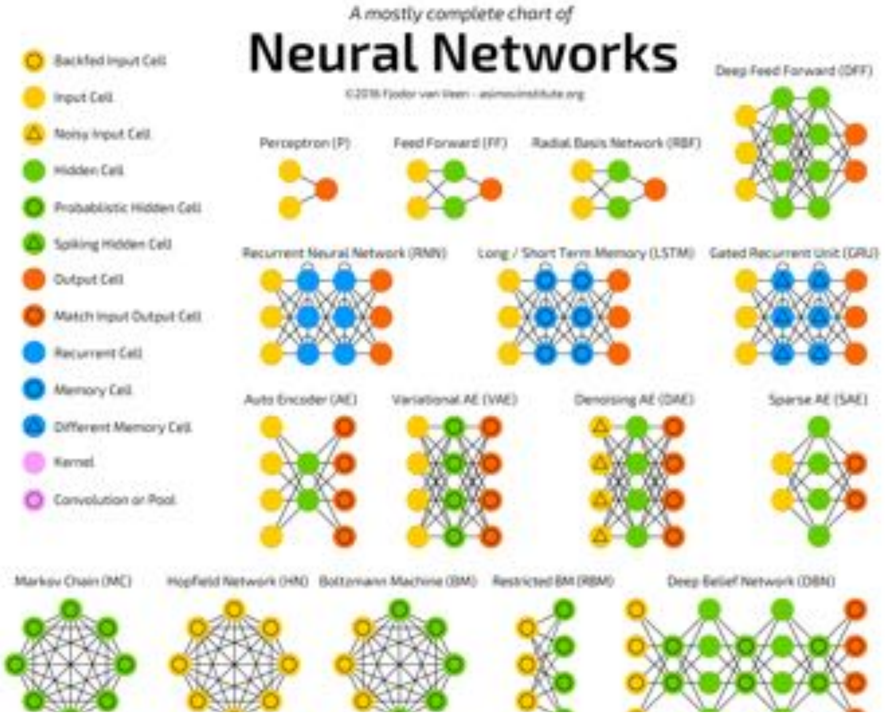


Potentially much more powerful than shallow architectures, represent computations

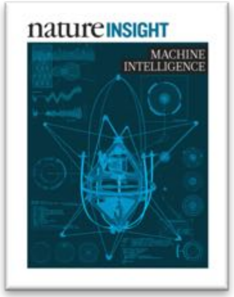
[LeCun, Bengio, Hinton Nature 521, 436–444, 2015]


Neuron
x_1
x_2
x_3
x_m
w_1
w_2
w_3
w_m
$\sum_{i=1}^{m}(w_i x_i) + bias$
$f(x) = \begin{cases} 1 & \text{if } \sum wx + b \geq 0 \\ 0 & \text{if } \sum wx + b < 0 \end{cases}$
$\hat{y}$
Σ
Inputs
Weights
Summation and Bias
Activation
Output

Differentiable Programming

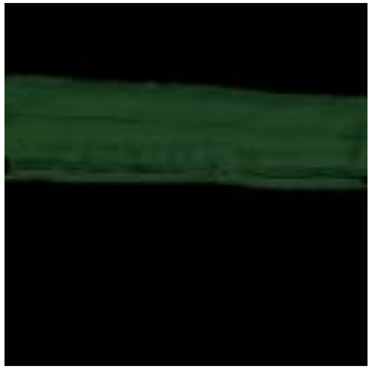


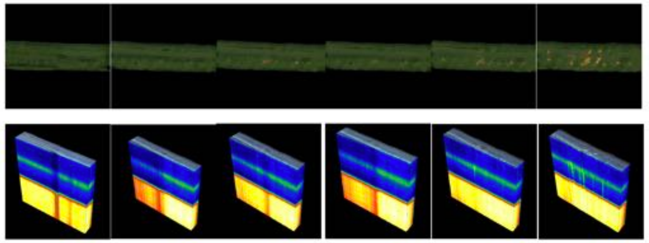
Deep Neural Networks

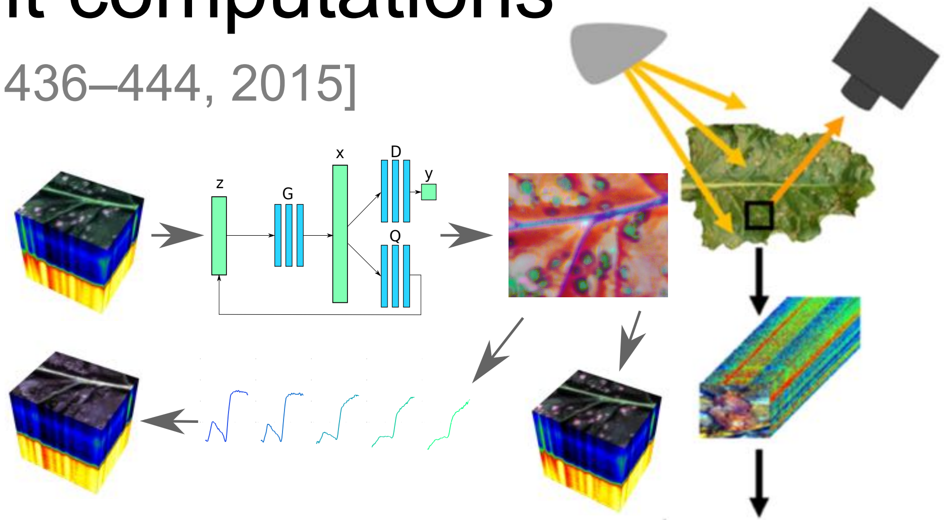


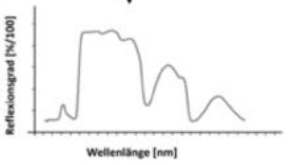
Potentially much more powerful than shallow architectures, represent computations

[LeCun, Bengio, Hinton Nature 521, 436–444, 2015]









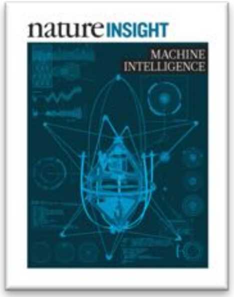
They “develop intuition” about complicated biological processes and generate scientific data

[Schramowski, Brugger, Mahlein, Kersting 2019]

DePhenSe

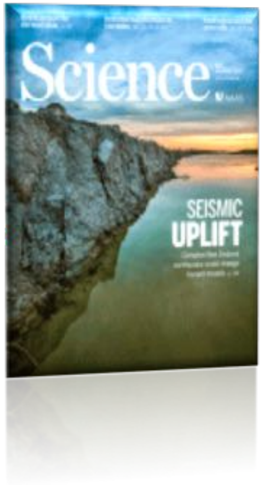
Bundesanstalt für Landwirtschaft und Ernährung

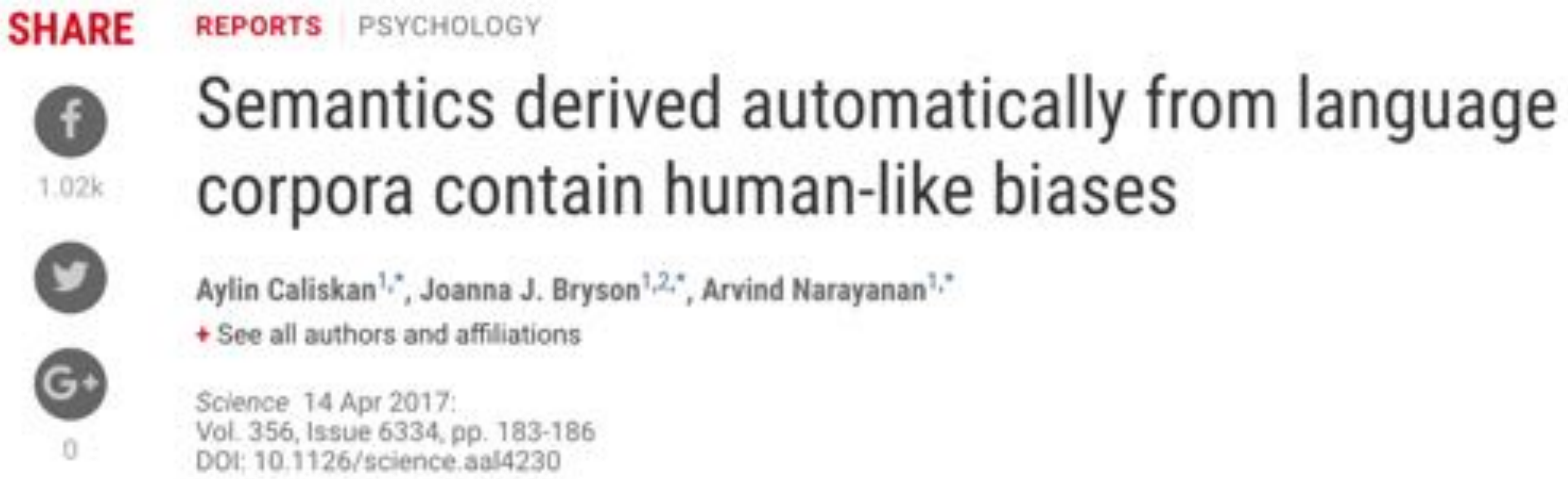
Deep Neural Networks



Potentially much more powerful than shallow architectures, represent computations

[LeCun, Bengio, Hinton Nature 521, 436–444, 2015]





REPORTS | PSYCHOLOGY

Semantics derived automatically from language corpora contain human-like biases

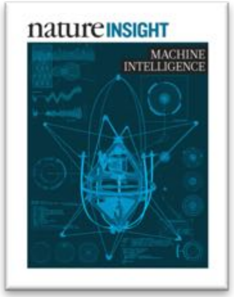
Aylin Caliskan[1,*], Joanna J. Bryson[1,2,*], Arvind Narayanan[1,*]

+ See all authors and affiliations

Science 14 Apr 2017:
Vol. 356, Issue 6334, pp. 183-186
DOI: 10.1126/science.aal4230

They “capture” stereotypes from human language

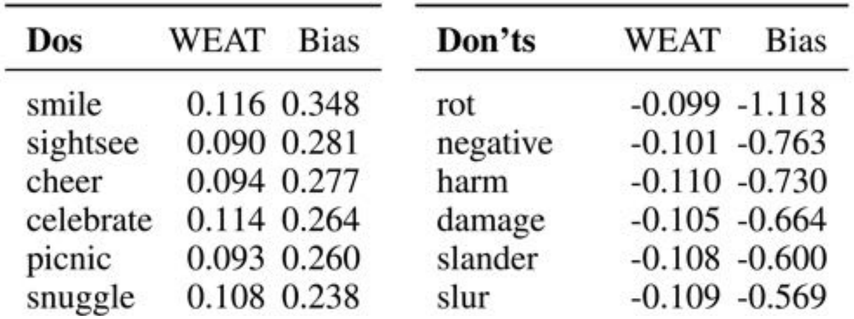
Deep Neural Networks



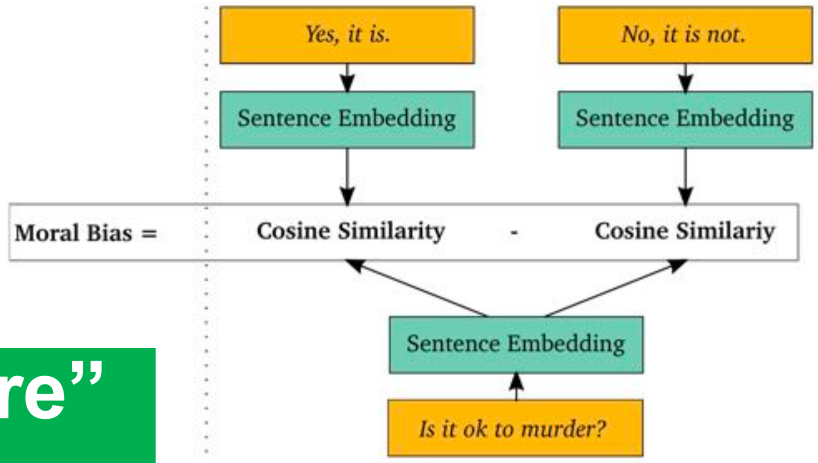
Potentially much more powerful than shallow architectures, represent computations

[LeCun, Bengio, Hinton Nature 521, 436–444, 2015]

The Moral Choice Machine



Dos	WEAT	Bias	Don’ts	WEAT	Bias
smile	0.116	0.348	rot	-0.099	-1.118
sightsee	0.090	0.281	negative	-0.101	-0.763
cheer	0.094	0.277	harm	-0.110	-0.730
celebrate	0.114	0.264	damage	-0.105	-0.664
picnic	0.093	0.260	slander	-0.108	-0.600
snuggle	0.108	0.238	slur	-0.109	-0.569



Yes, it is.

No, it is not.

Sentence Embedding

Sentence Embedding

Moral Bias = Cosine Similarity - Cosine Similariy

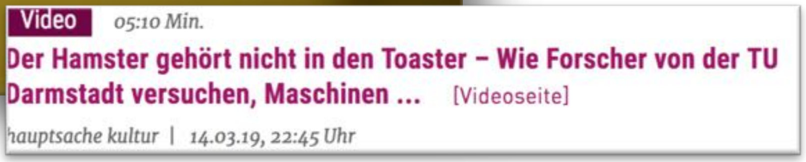
Sentence Embedding

Is it ok to murder?

But lucky they also “capture” our moral choices

[Jentzsch, Schramowski, Rothkopf, Kersting AIES 2019]

AAAI / ACM conference on
ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

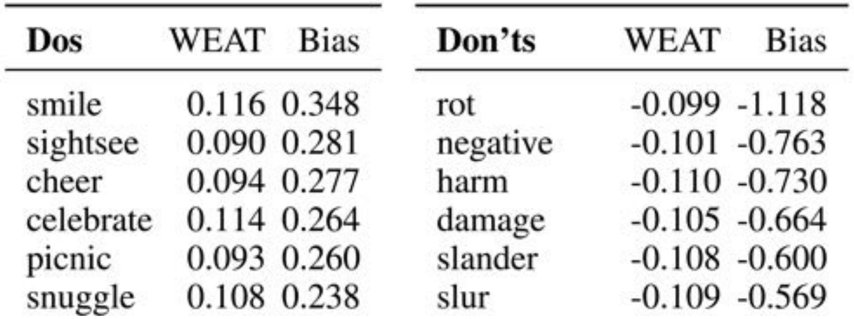


Video 05:10 Min.

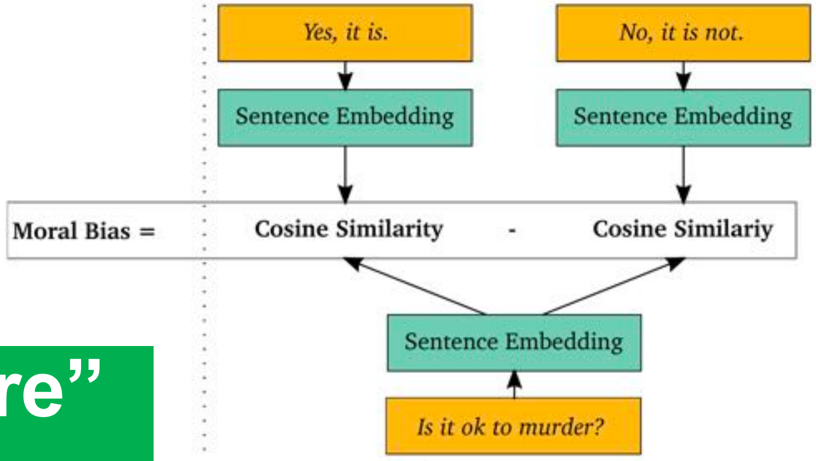
Der Hamster gehört nicht in den Toaster – Wie Forscher von der TU Darmstadt versuchen, Maschinen ... [Videoseite]

hauptsache kultur | 14.03.19, 22:45 Uhr

The Moral Choice Machine



Dos	WEAT	Bias	Don'ts	WEAT	Bias
smile	0.116	0.348	rot	-0.099	-1.118
sightsee	0.090	0.281	negative	-0.101	-0.763
cheer	0.094	0.277	harm	-0.110	-0.730
celebrate	0.114	0.264	damage	-0.105	-0.664
picnic	0.093	0.260	slander	-0.108	-0.600
snuggle	0.108	0.238	slur	-0.109	-0.569

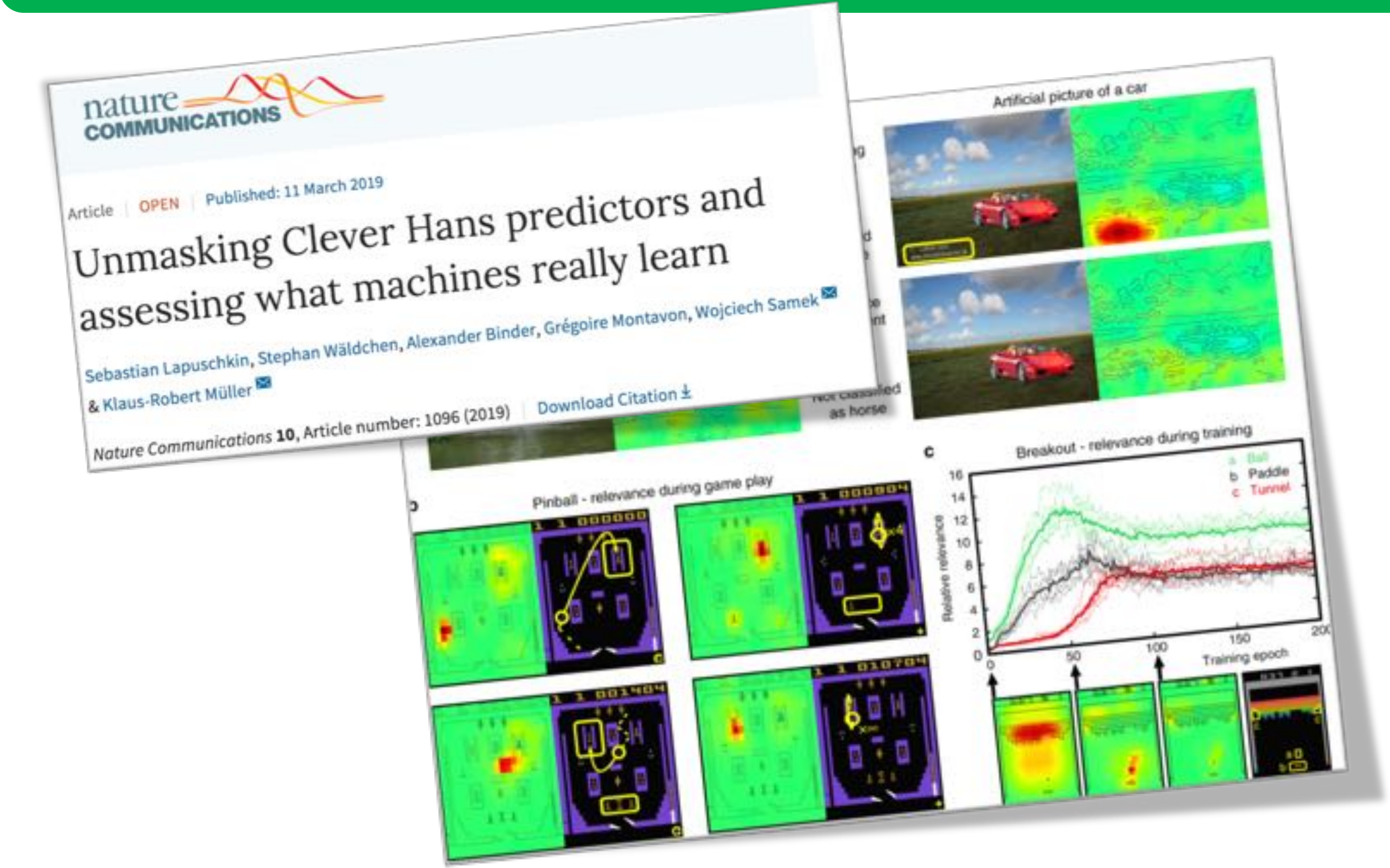


But lucky they also "capture" our moral choices

[Jentzsch, Schramowski, Rothkopf, Kersting AIES 2019]

AAAI / ACM conference on ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

Can we trust deep neural networks?



nature COMMUNICATIONS

Article | OPEN | Published: 11 March 2019

Unmasking Clever Hans predictors and assessing what machines really learn

Sebastian Lapuschkin, Stephan Wäldchen, Alexander Binder, Grégoire Montavon, Wojciech Samek & Klaus-Robert Müller

Nature Communications **10**, Article number: 1096 (2019) | Download Citation

DNNs do not quantify all of the uncertainty. They are not calibrated joint distributions.

$$P(Y|X) \neq P(Y,X)$$

MNIST

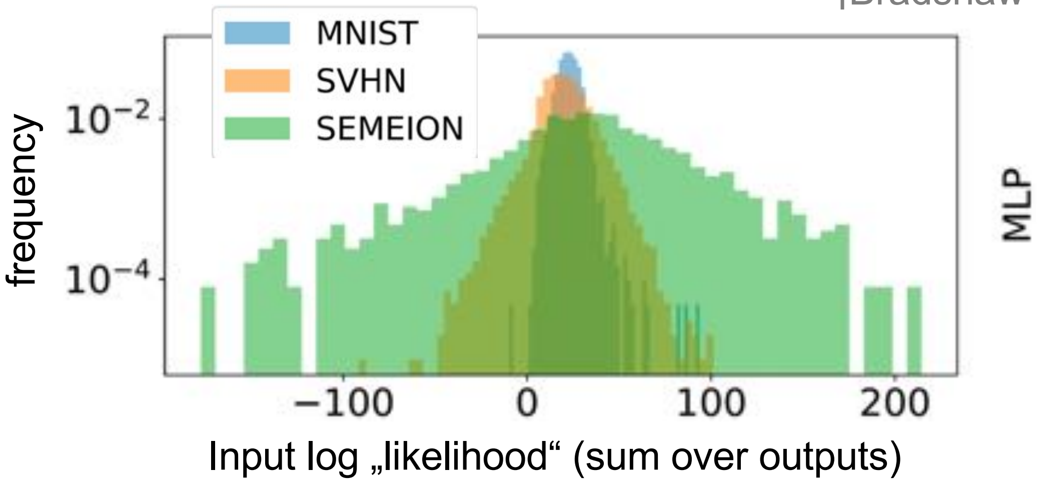
Train & Evaluate

SVHN

SEMEION

Transfer Testing

[Bradshaw et al. arXiv:1707.02476 2017]



DNNs cannot distinguish the datasets

[Peharz, Vergari, Molina, Stelzner, Trapp, Kersting, Ghahramani UDL@UAI 2018]

Getting deep systems that know when they don’t know.

Sum-Product Networks: A deep probabilistic learning framework

Adnan Darwiche UCLA

Pedro Domingos UW

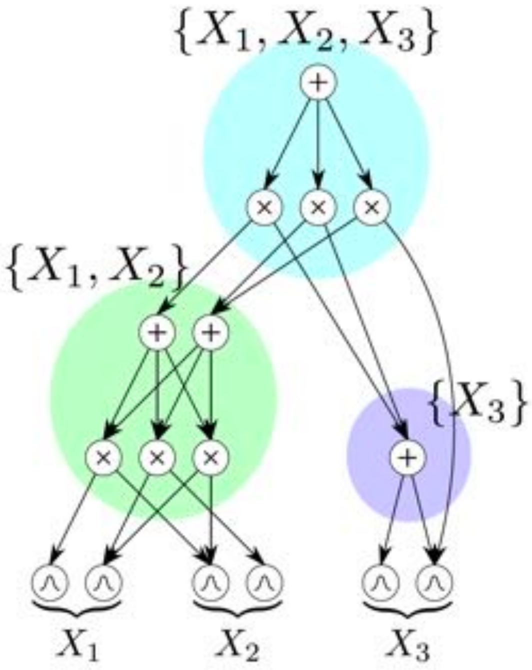
⊕ ...convex sum

⊗ ...product

(^) ...distribution

completeness
sum children: same scope

decomposability
product children:
non-overlapping scope

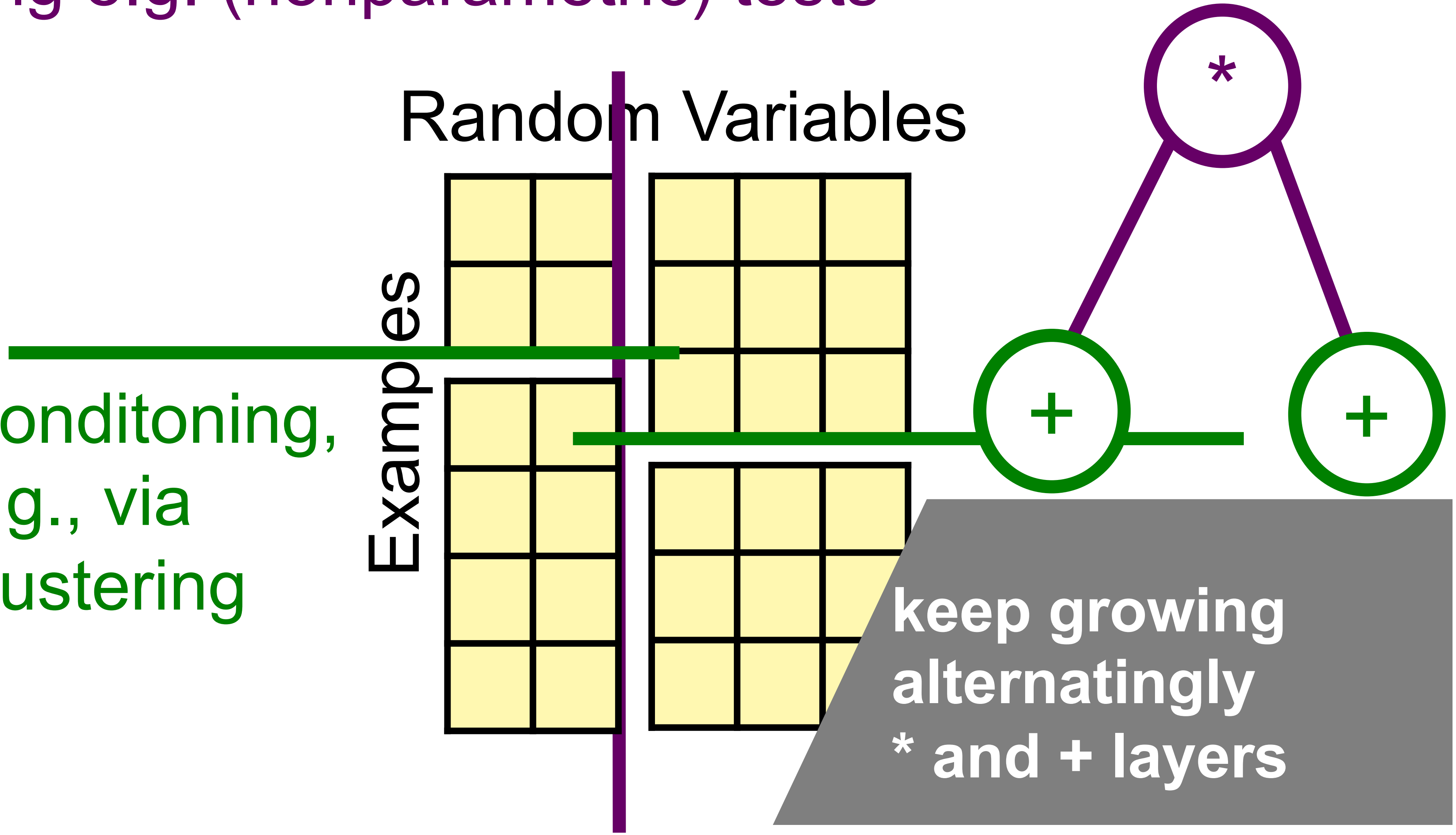


Computational graph (kind of TensorFlow graphs) that encodes how to compute probabilities

Inference is linear in size of network

And there is a principled approach to select SPNs from data

Testing independence of random variables using e.g. (nonparametric) tests



Conditoning, e.g., via clustering

keep growing alternatingly * and + layers

[Poon, Domingos UAI'11; Molina, Natarajan, Kersting AAAI'17; Vergari, Peharz, Di Mauro, Molina, Kersting, Esposito AAAI '18; Molina, Vergari, Di Mauro, Esposito, Natarajan, Kersting AAAI '18]

SPFlow: An Easy and Extensible Library for Sum-Product Networks

[Molina, Vergari, Stelzner, Peharz, Subramani, Poupart, Di Mauro, Kersting 2019]

TECHNISCHE UNIVERSITÄT DARMSTADT

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

UNIVERSITY OF WATERLOO

Max Planck Institute for Intelligent Systems

UNIVERSITY OF CAMBRIDGE

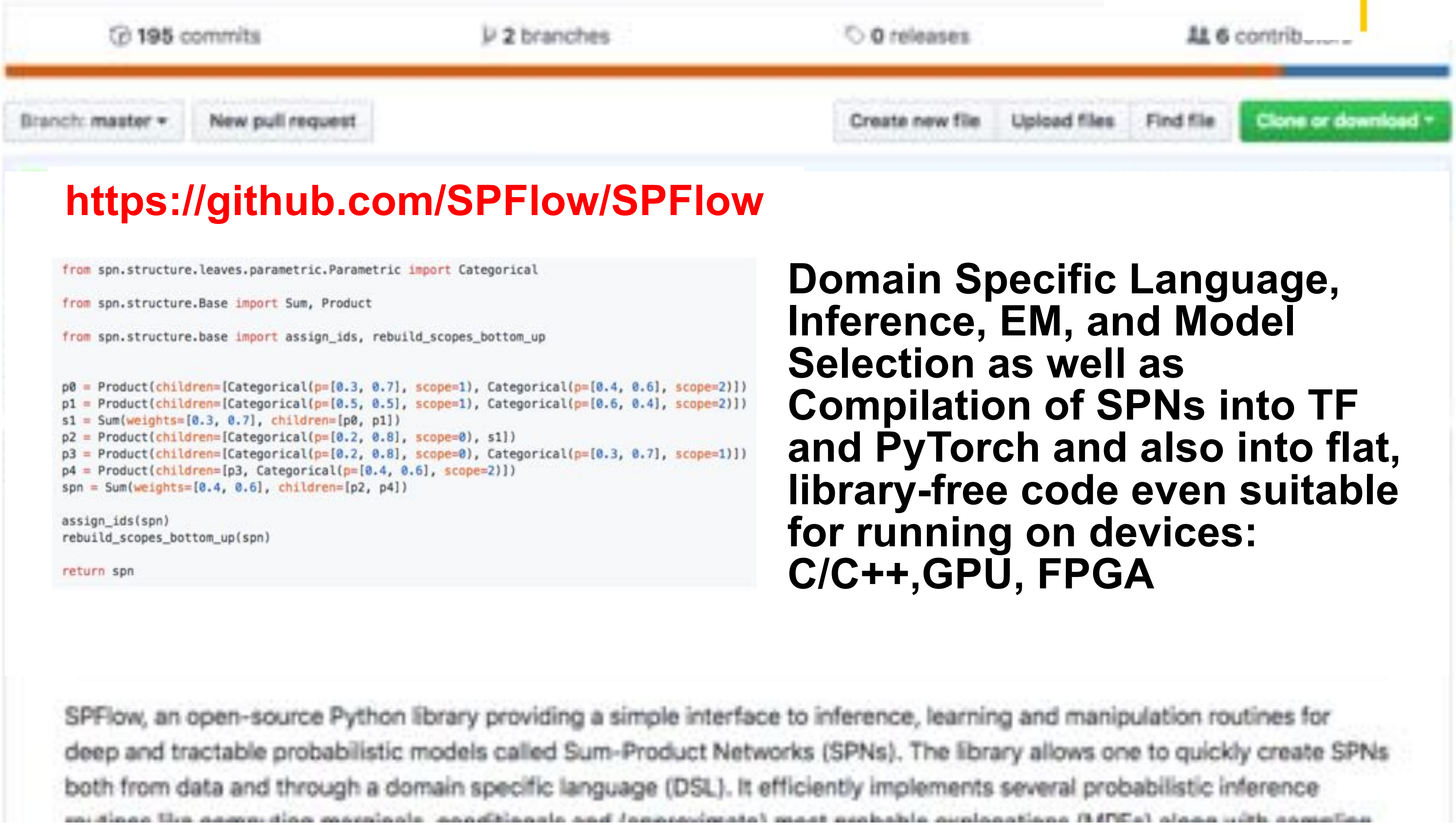
VECTOR INSTITUTE

CAML

MADESI

DFG

Federal Ministry of Education and Research



https://github.com/SPFlow/SPFlow

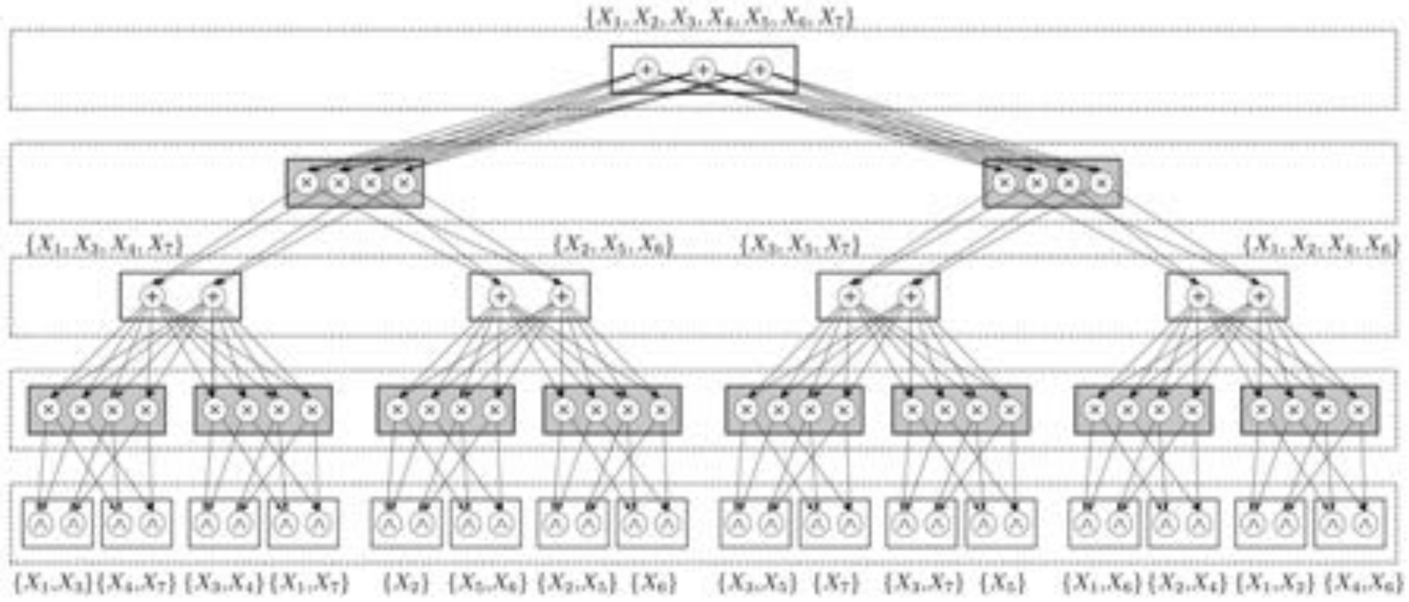
```
from spn.structure.leaves.parametric.Parametric import Categorical

from spn.structure.Base import Sum, Product

from spn.structure.base import assign_ids, rebuild_scopes_bottom_up


p0 = Product(children=[Categorical(p=[0.3, 0.7], scope=1), Categorical(p=[0.4, 0.6], scope=2)])
p1 = Product(children=[Categorical(p=[0.5, 0.5], scope=1), Categorical(p=[0.6, 0.4], scope=2)])
s1 = Sum(weights=[0.3, 0.7], children=[p0, p1])
p2 = Product(children=[Categorical(p=[0.2, 0.8], scope=0), s1])
p3 = Product(children=[Categorical(p=[0.2, 0.8], scope=0), Categorical(p=[0.3, 0.7], scope=1)])
p4 = Product(children=[p3, Categorical(p=[0.4, 0.6], scope=2)])
spn = Sum(weights=[0.4, 0.6], children=[p2, p4])

assign_ids(spn)
rebuild_scopes_bottom_up(spn)

return spn
```

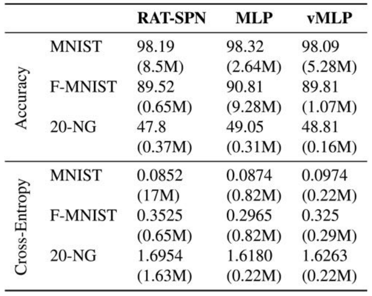
Domain Specific Language, Inference, EM, and Model Selection as well as Compilation of SPNs into TF and PyTorch and also into flat, library-free code even suitable for running on devices: C/C++,GPU, FPGA

SPFlow, an open-source Python library providing a simple interface to inference, learning and manipulation routines for deep and tractable probabilistic models called Sum-Product Networks (SPNs). The library allows one to quickly create SPNs both from data and through a domain specific language (DSL). It efficiently implements several probabilistic inference routines like computing marginals, conditionals and (approximate) most probable explanations (MPEs) along with sampling

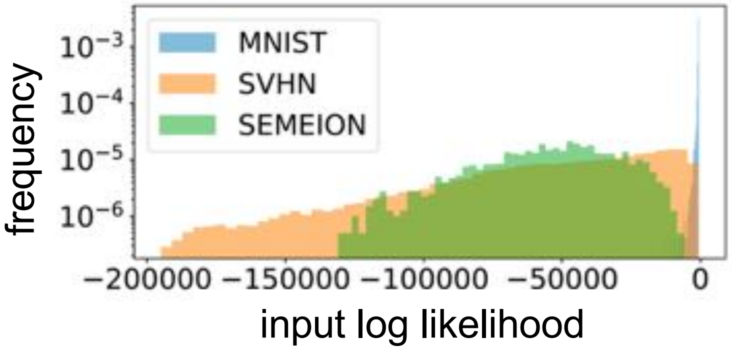
Random sum-product networks

[Peharz, Vergari, Molina, Stelzner, Trapp, Kersting, Ghahramani UDL@UAI 2018]


$\{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}$
$\{X_1, X_3, X_4, X_7\}$
$\{X_2, X_5, X_6\}$
$\{X_3, X_5, X_7\}$
$\{X_1, X_2, X_4, X_6\}$
$\{X_1, X_3\}$ $\{X_4, X_7\}$ $\{X_3, X_4\}$ $\{X_1, X_7\}$ $\{X_2\}$ $\{X_5, X_6\}$ $\{X_2, X_5\}$ $\{X_6\}$ $\{X_3, X_5\}$ $\{X_7\}$ $\{X_3, X_7\}$ $\{X_5\}$ $\{X_1, X_6\}$ $\{X_2, X_4\}$ $\{X_1, X_2\}$ $\{X_4, X_6\}$



		RAT-SPN	MLP	vMLP
Accuracy	MNIST	98.19 (8.5M)	98.32 (2.64M)	98.09 (5.28M)
	F-MNIST	89.52 (0.65M)	90.81 (9.28M)	89.81 (1.07M)
	20-NG	47.8 (0.37M)	49.05 (0.31M)	48.81 (0.16M)
Cross-Entropy	MNIST	0.0852 (17M)	0.0874 (0.82M)	0.0974 (0.22M)
	F-MNIST	0.3525 (0.65M)	0.2965 (0.82M)	0.325 (0.29M)
	20-NG	1.6954 (1.63M)	1.6180 (0.22M)	1.6263 (0.22M)



outliers

prototypes

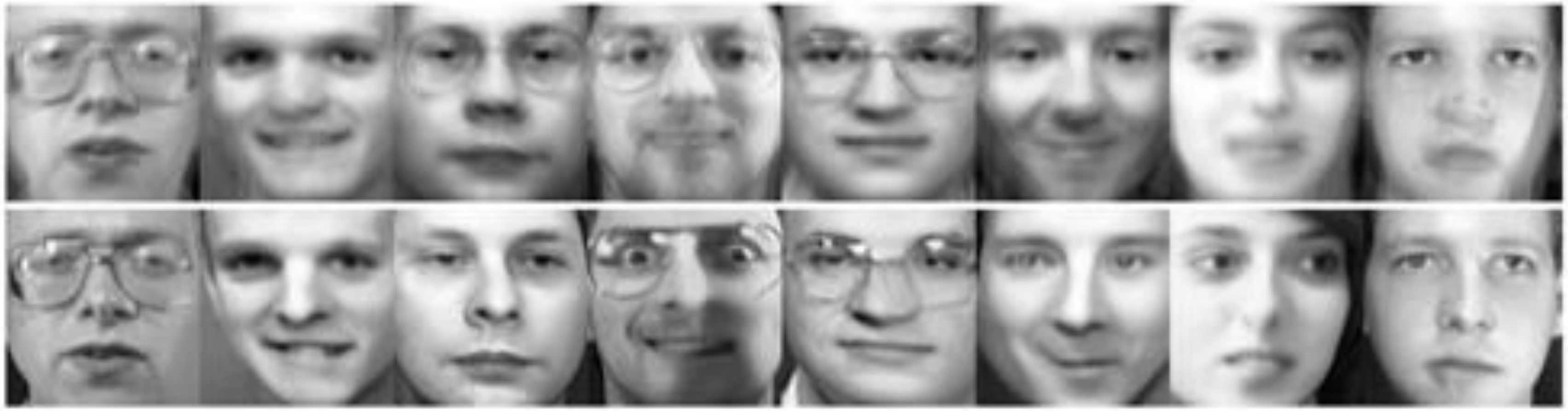
outliers

prototypes

Learning the Structure of Autoregressive Deep Models such as PixelCNNs [van den Oord et al. NIPS 2016]

CSPNs

PixelCNNs



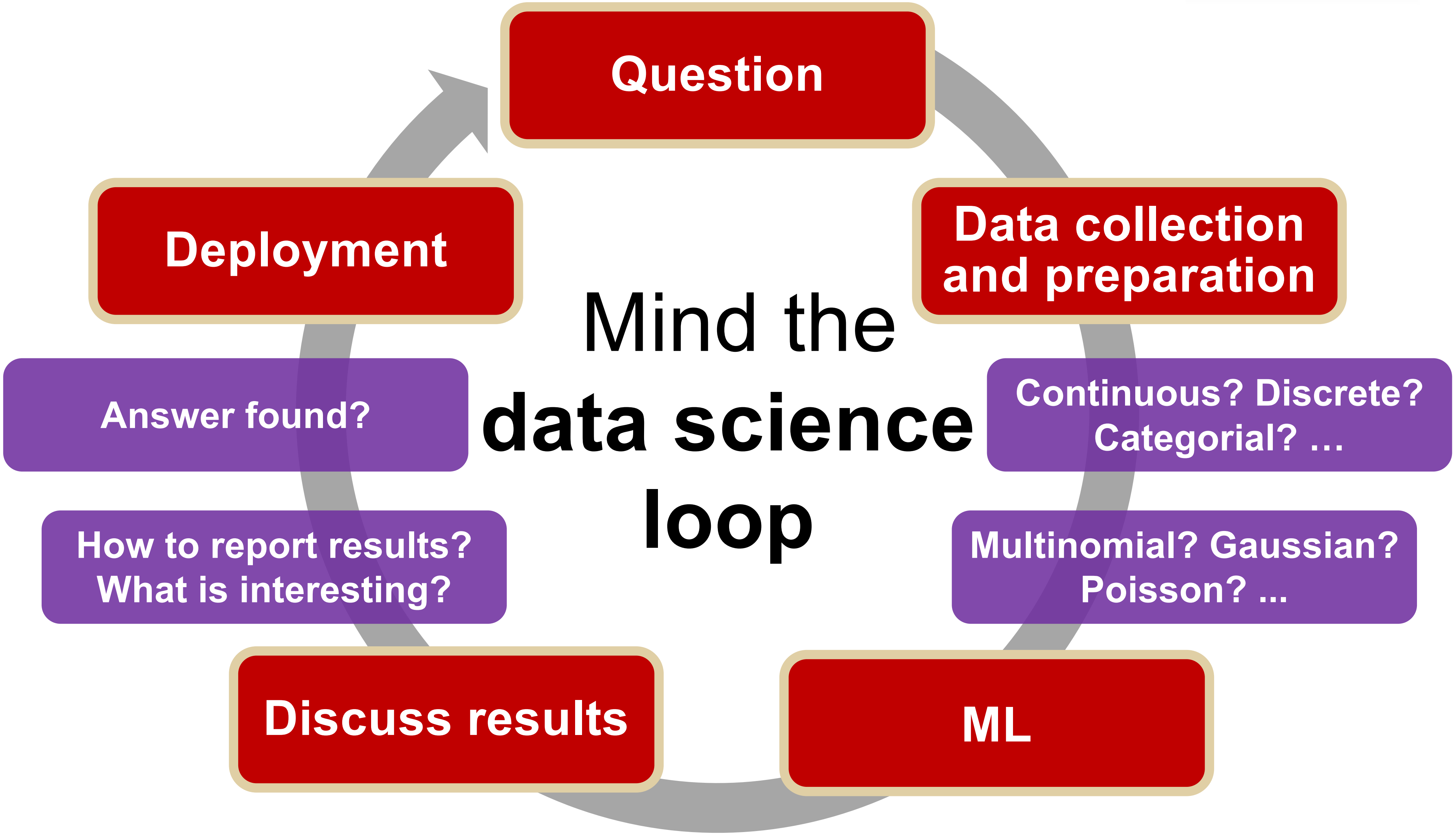
Learn Conditional SPN by testing conditional independence and using conditional clustering, using e.g. [Zhang et al. UAI 2011; Lee, Honovar UAI 2017; He et al. ICDM 2017; Zhang et al. AAAI 2018; Runge AISTATS 2018]

Conditional SPNs

[Shao, Molina, Vergari, Peharz, Kersting 2019]

CAML

DFG



Mind the **data science loop**

Question

Data collection and preparation

Continuous? Discrete? Categorial? …

Multinomial? Gaussian? Poisson? ...

ML

Discuss results

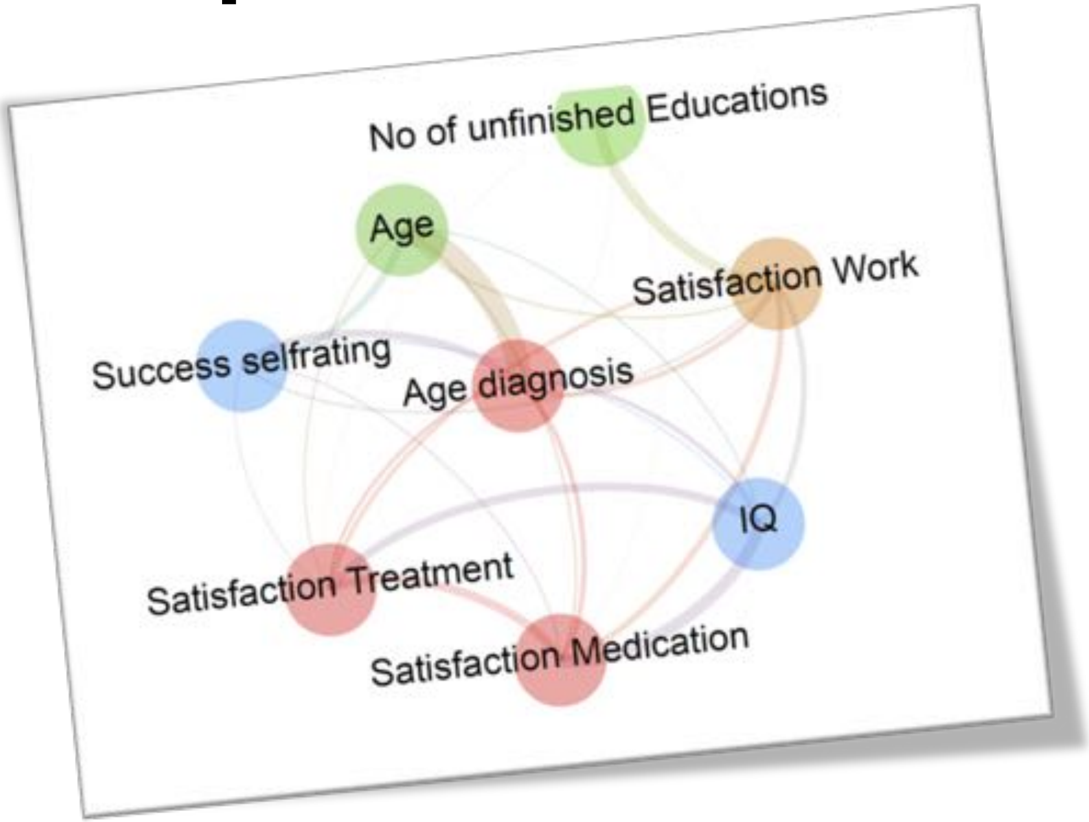
How to report results? What is interesting?

Answer found?

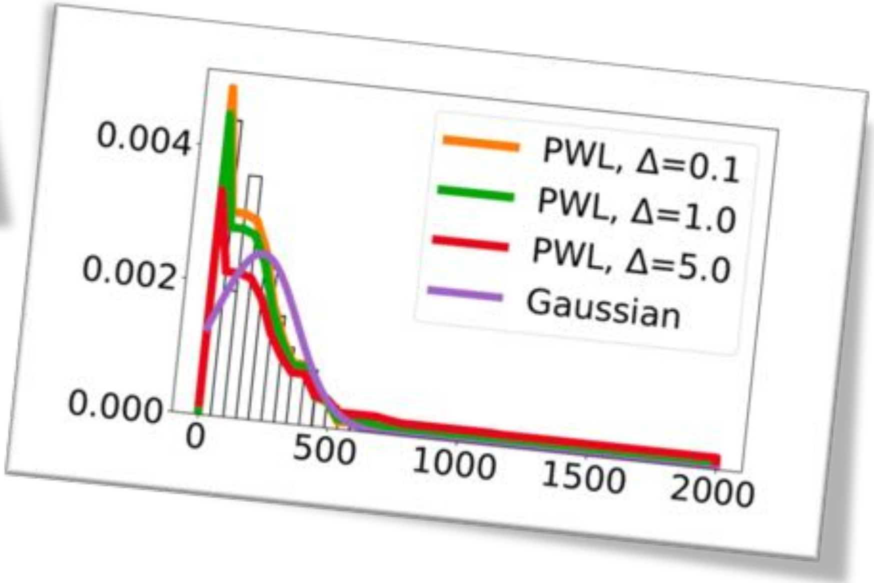
Deployment

[Molina, Natarajan, Vergari, Di Mauro, Esposito, Kersting AAAI 2018]

Distribution-agnostic Deep Probabilistic Learning

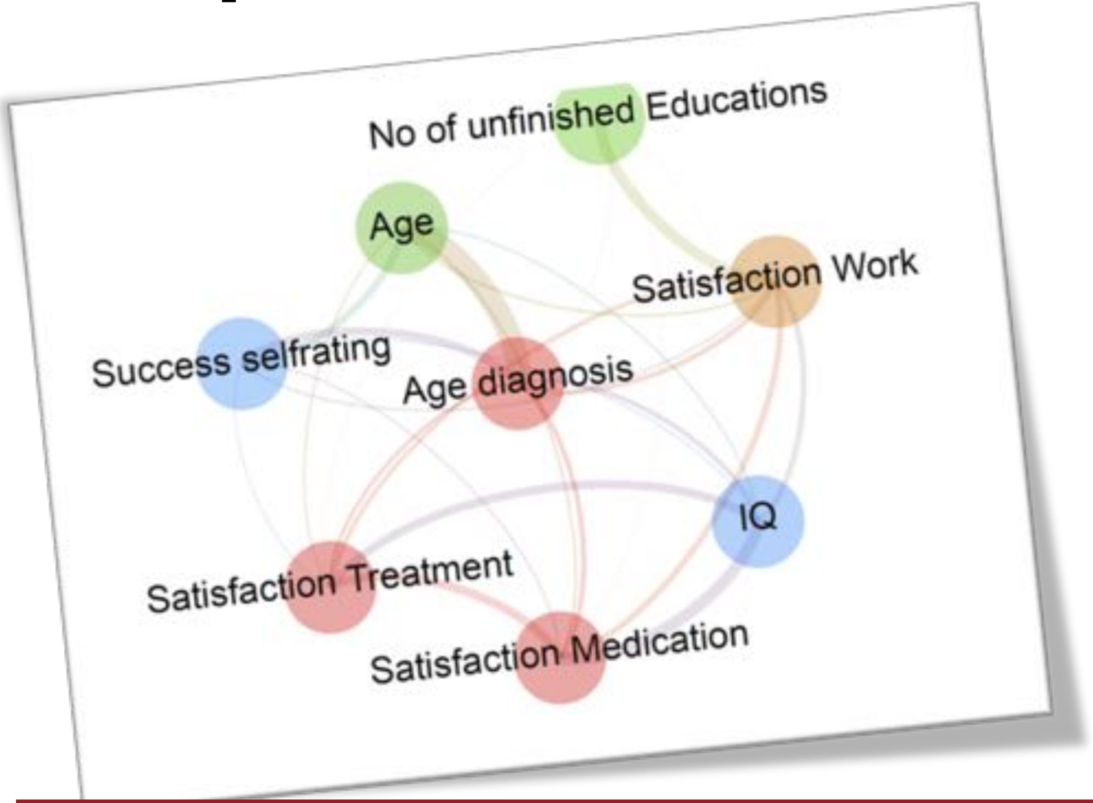


Use nonparametric independency tests and piece-wise linear approximations



[Molina, Natarajan, Vergari, Di Mauro, Esposito, Kersting AAAI 2018]

Distribution-agnostic Deep Probabilistic Learning



Use nonparametric independency tests and piece-wise linear approximations

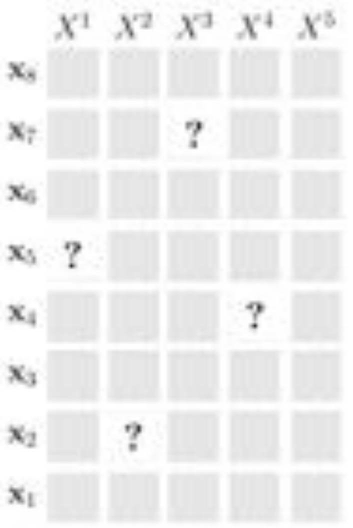
0.004
0.002
PWL, Δ=0.1
PWL, Δ=1.0
PWL, Δ=5.0

However, we have to provide the statistical types and do not gain insights into the parametric forms of the variables. **Are they Gaussians? Gammas? …**

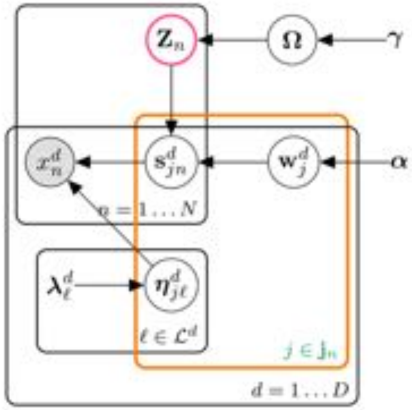
[Vergari, Molina, Peharz, Ghahramani, Kersting, Valera AAAI 2019]

The Explorative Automatic Statistician

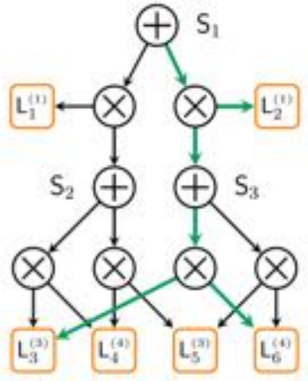
missing value



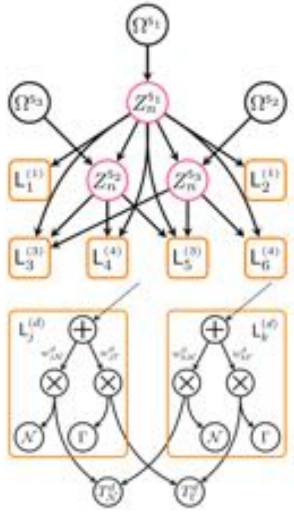
We can even automatically discovers the statistical types and parametric forms of the variables



Bayesian Type Discovery

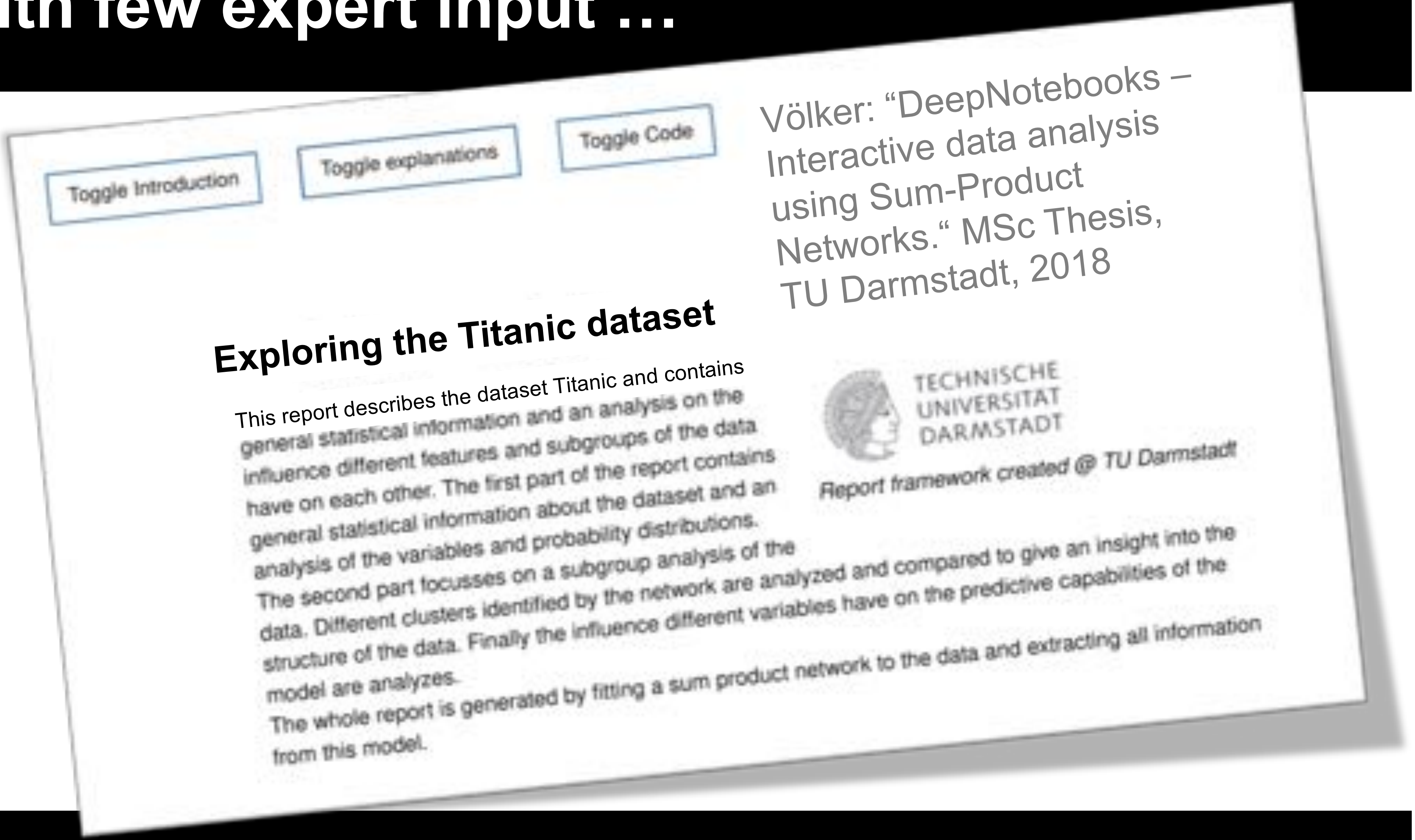


Mixed Sum-Product Network



Automatic Statistician

That is, the machine understands the data with few expert input …



Völker: “DeepNotebooks – Interactive data analysis using Sum-Product Networks.“ MSc Thesis, TU Darmstadt, 2018

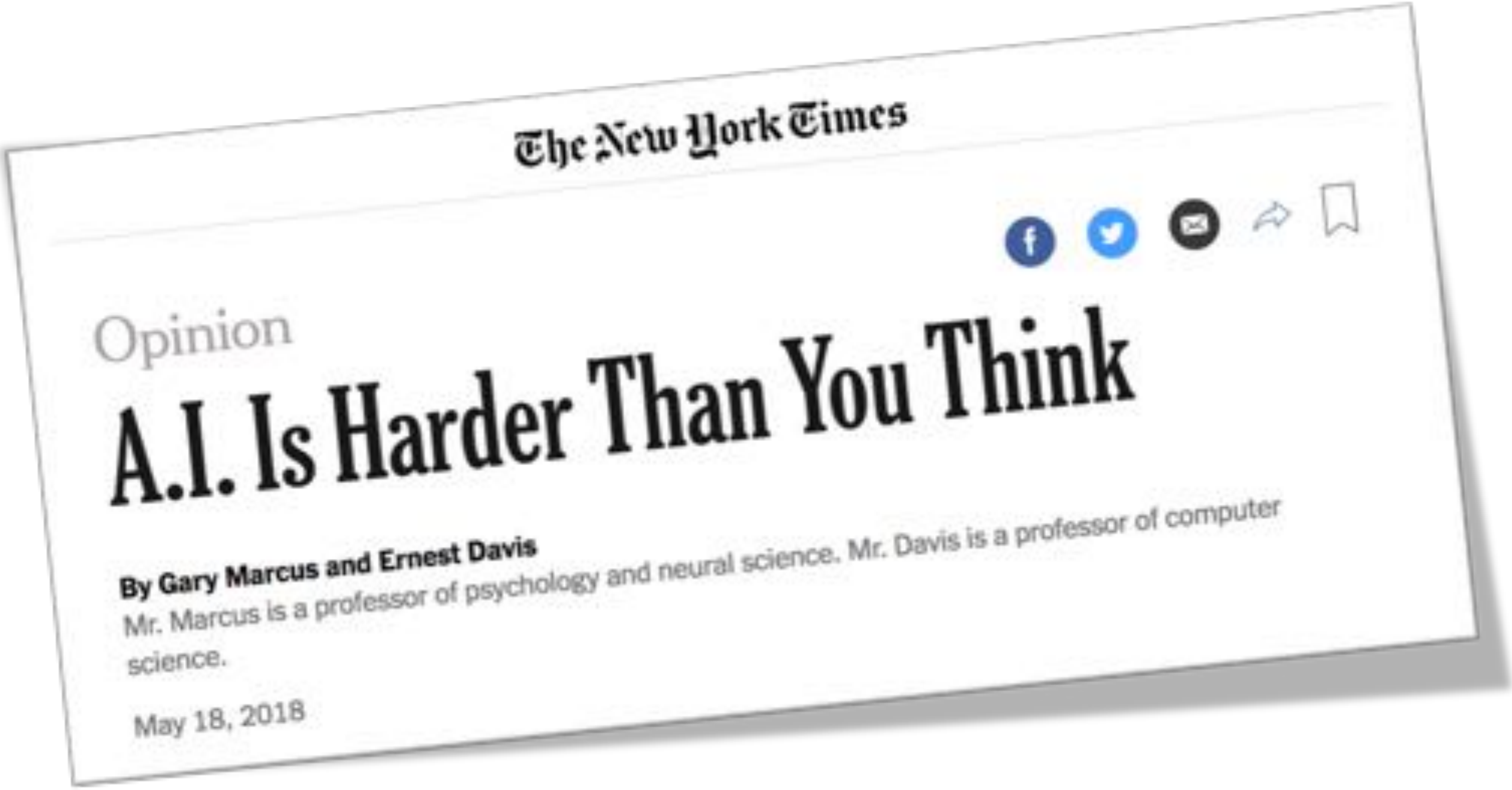
Exploring the Titanic dataset

This report describes the dataset Titanic and contains general statistical information and an analysis on the influence different features and subgroups of the data have on each other. The first part of the report contains general statistical information about the dataset and an analysis of the variables and probability distributions. The second part focusses on a subgroup analysis of the data. Different clusters identified by the network are analyzed and compared to give an insight into the structure of the data. Finally the influence different variables have on the predictive capabilities of the model are analyzes. The whole report is generated by fitting a sum product network to the data and extracting all information from this model.

Report framework created @ TU Darmstadt

…and can compile data reports automatically

P(heart attack | ▦)?



The New York Times

Opinion

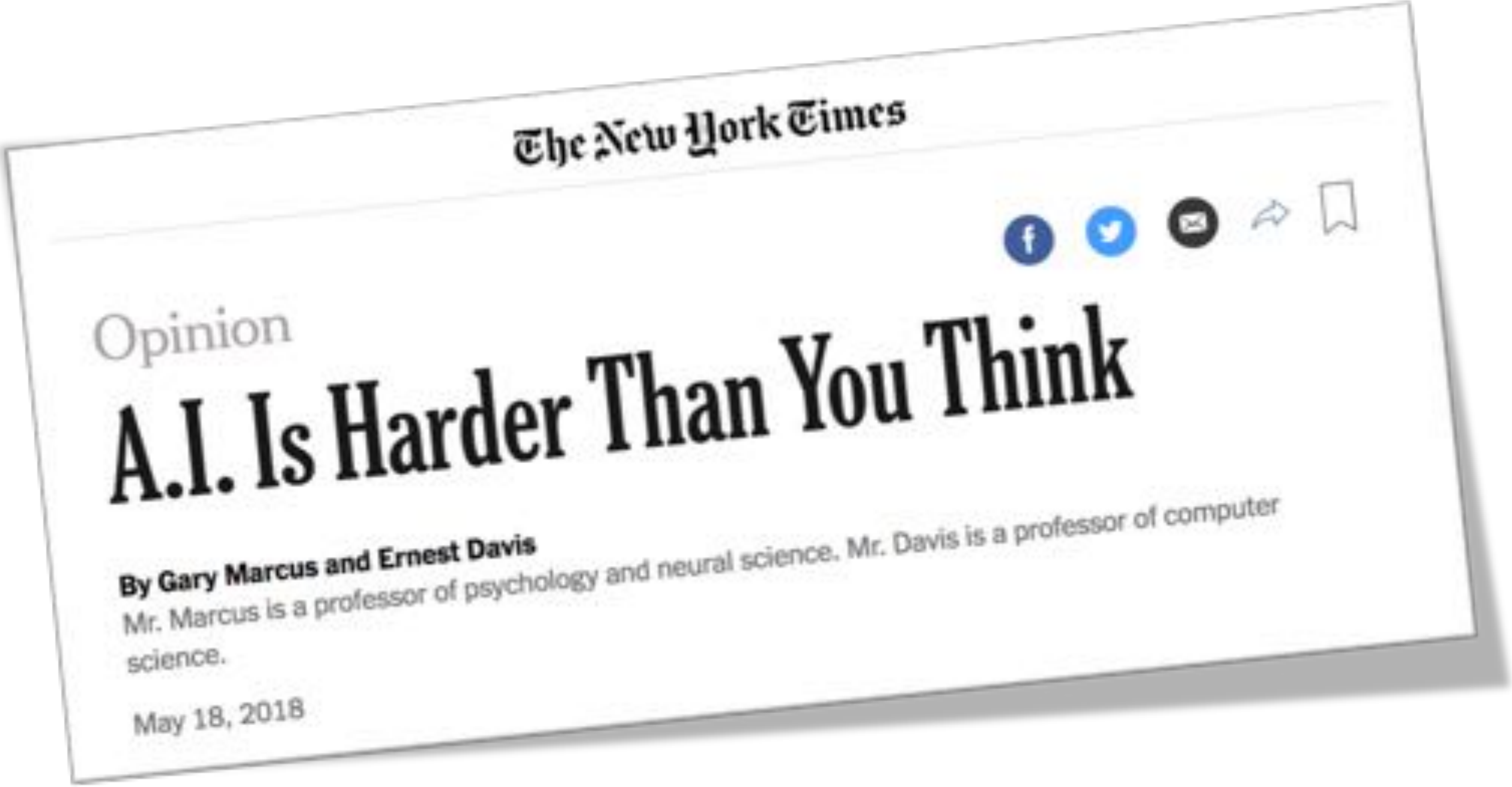
A.I. Is Harder Than You Think

By Gary Marcus and Ernest Davis

Mr. Marcus is a professor of psychology and neural science. Mr. Davis is a professor of computer science.

May 18, 2018

P(heart attack |)?



The New York Times

Opinion

A.I. Is Harder Than You Think

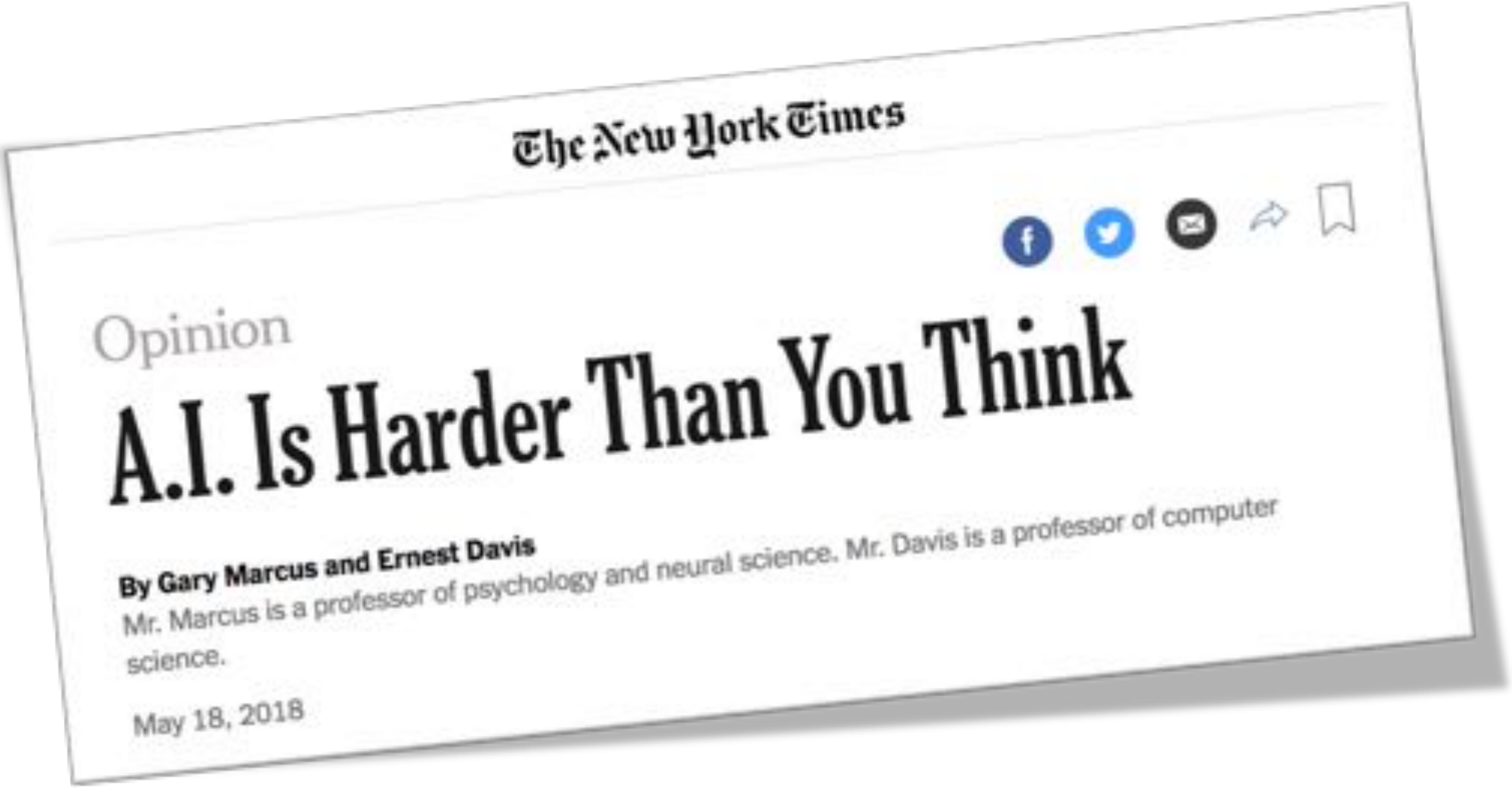
By Gary Marcus and Ernest Davis

Mr. Marcus is a professor of psychology and neural science. Mr. Davis is a professor of computer science.

May 18, 2018

P(heart attack |

)?



The New York Times

Opinion

A.I. Is Harder Than You Think

By Gary Marcus and Ernest Davis

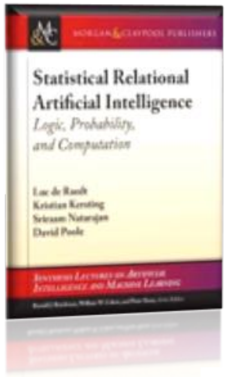
Mr. Marcus is a professor of psychology and neural science. Mr. Davis is a professor of computer science.

May 18, 2018

P(heart attack |)?

Crossover of ML and DS with data & programming abstractions

De Raedt, Kersting, Natarajan, Poole: Statistical Relational Artificial Intelligence: Logic, Probability, and Computation. Morgan and Claypool Publishers, ISBN: 9781627058414, 2016.


Statistical Relational Artificial Intelligence
Logic, Probability, and Computation

building general-purpose data science and ML machines

make the ML/DS expert more effective and employing domain knowledge

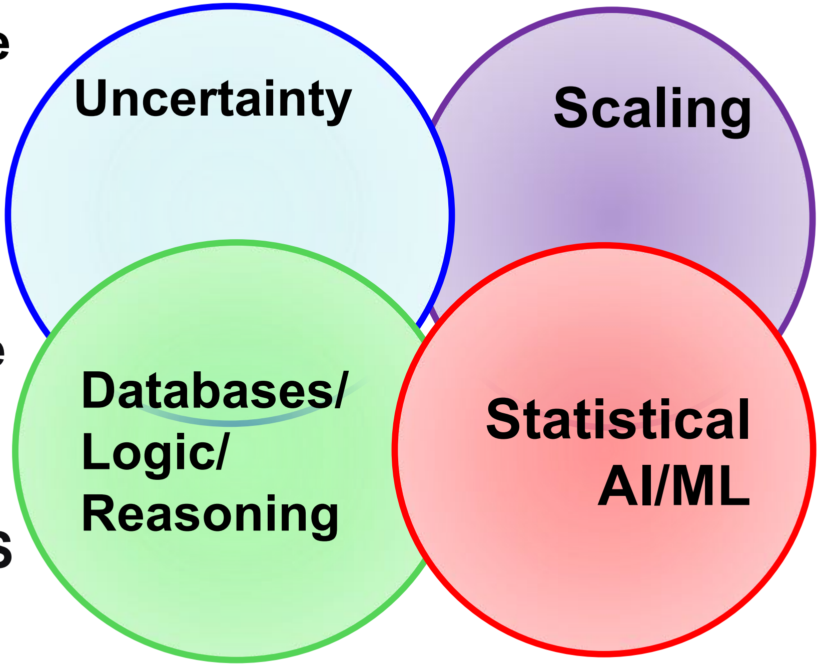
increases the number of people who can successfully build ML/DS applications

KATHOLIEKE UNIVERSITEIT LEUVEN

THE UNIVERSITY OF TEXAS AT DALLAS

TECHNISCHE UNIVERSITÄT DARMSTADT

UBC THE UNIVERSITY OF BRITISH COLUMBIA



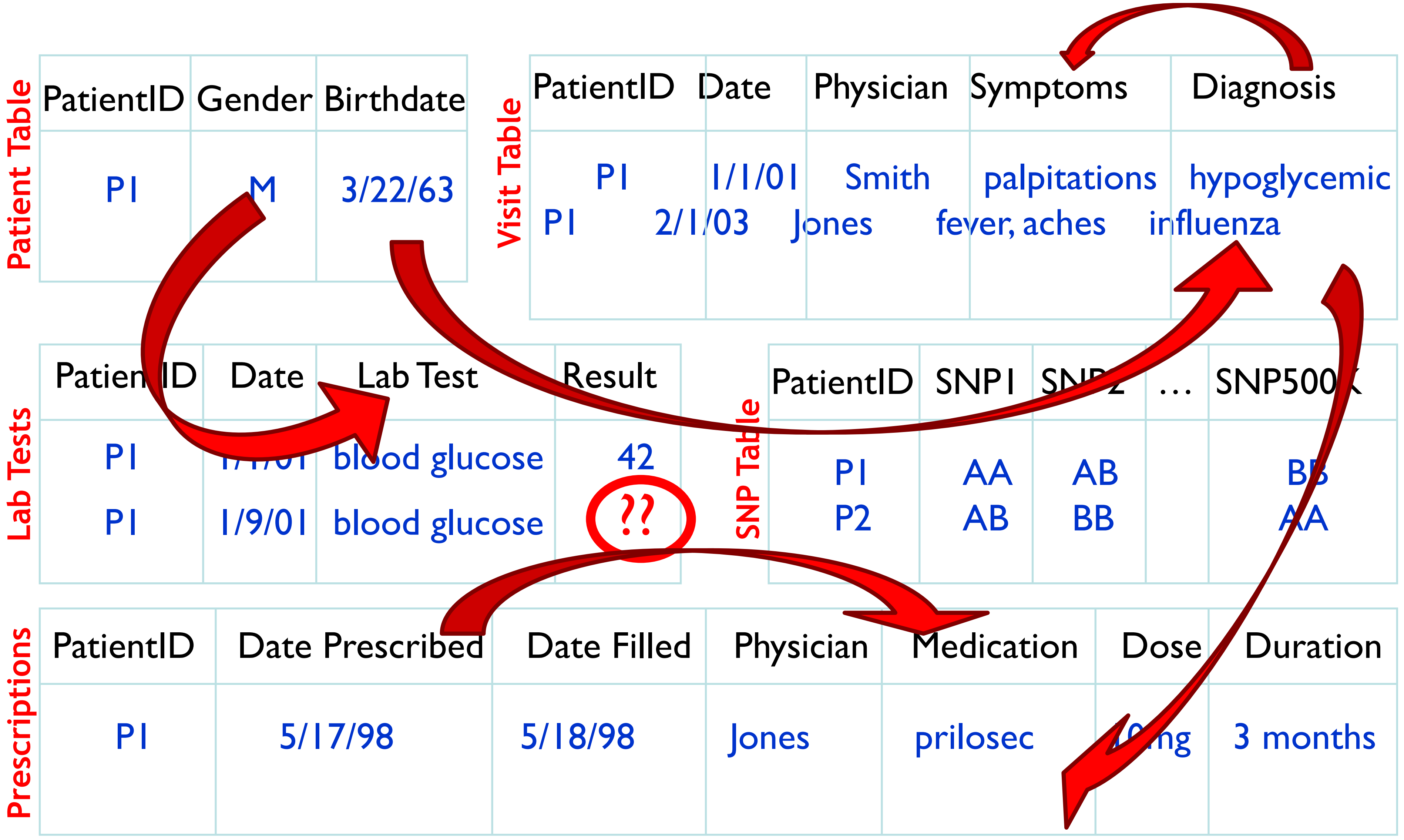
Nat Rev Genet. 2012 May 2;13(6):395-405

Heart diseases and strokes – cardiovascular disease – are expensive for the world

According to the World Heart Federation, cardiovascular disease cost the European Union EURO169 billion in 2003 and the USA about EURO310.23 billion in direct and indirect annual costs. By comparison, the estimated cost of all cancers is EURO146.19 billion and HIV infections, EURO22.24 billion

Electronic Health Records
A new opportunity for AI to save our Lifes

EHRs are dirty and interconnected



Patient Table

PatientID	Gender	Birthdate
P1	M	3/22/63

Visit Table

PatientID	Date	Physician	Symptoms	Diagnosis
P1	1/1/01	Smith	palpitations	hypoglycemic
P1	2/1/03	Jones	fever, aches	influenza

Lab Tests

PatientID	Date	Lab Test	Result
P1	1/1/01	blood glucose	42
P1	1/9/01	blood glucose	??

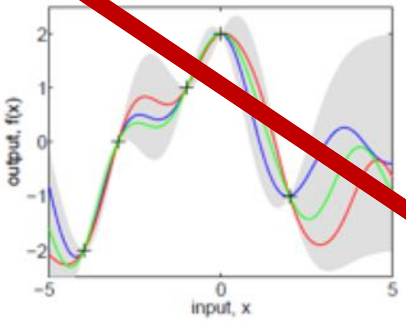
SNP Table

PatientID	SNP1	SNP2	…	SNP500K
P1	AA	AB		BB
P2	AB	BB		AA

Prescriptions

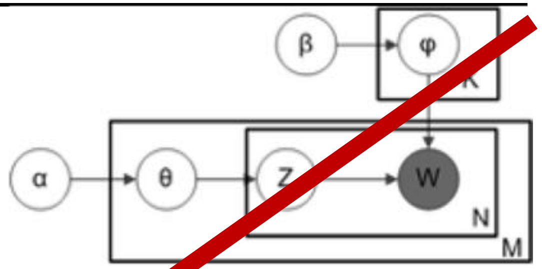
PatientID	Date Prescribed	Date Filled	Physician	Medication	Dose	Duration
P1	5/17/98	5/18/98	Jones	prilosec	10mg	3 months

Standard machine learning

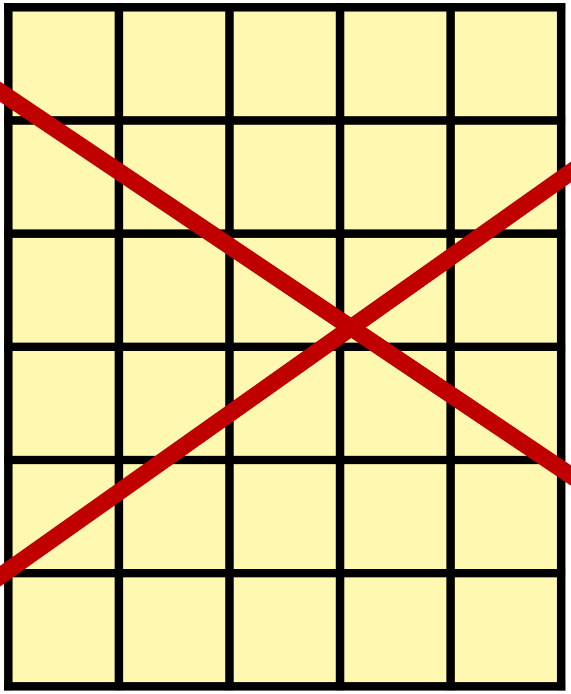


Gaussian Processes

Graphical models

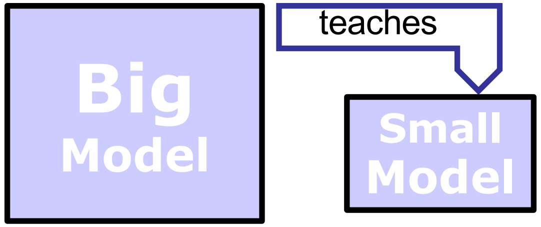


Features



Objects

Graph Mining



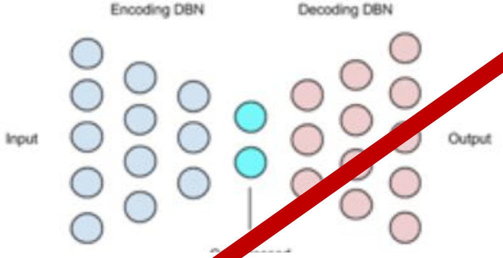
Big Model

teaches

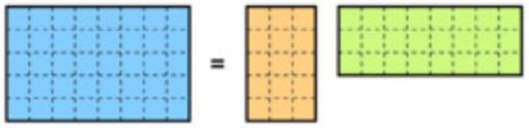
Small Model

Distillation/LUPI

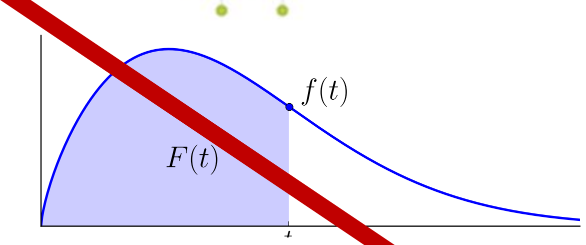
Boosting



Autoencoder,
Deep Learning



Big Data Matrix Factorization

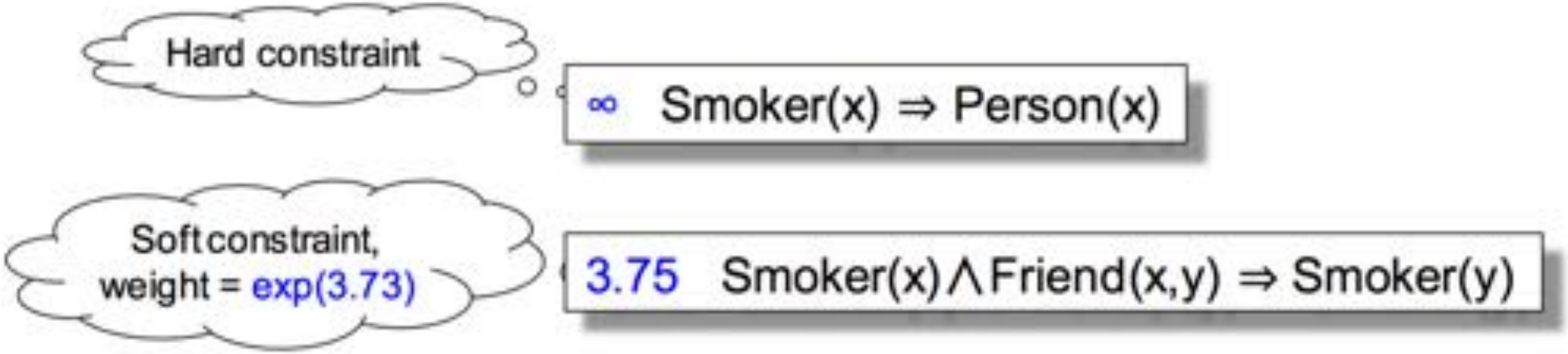


Diffusion Models

and many more …

Statistical Relational Models

Weighted logical formulas / uncertain databases



Hard constraint

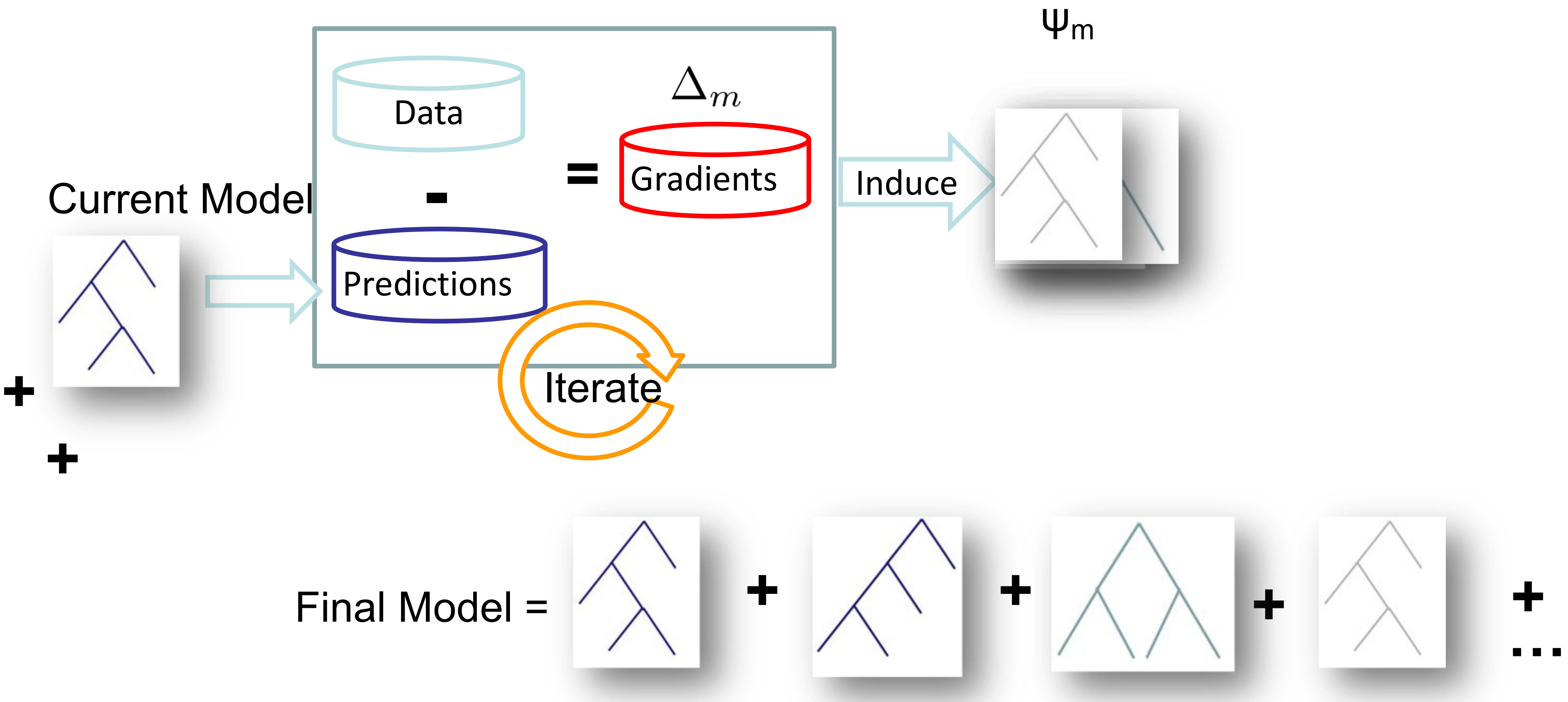
∞ Smoker(x) $\Rightarrow$ Person(x)

Soft constraint, weight = exp(3.73)

3.75 Smoker(x) $\wedge$ Friend(x,y) $\Rightarrow$ Smoker(y)

Learning statistical models over databases: Functional Gradient Boosting

Learn multiple weak is easier than a single complex model



Friedman et al 2001, Dietterich et al. 2004, Natarajan et al. MLJ 2012

Functional Gradients for SRL Models

Pseudo probability of an example

$$P(x_i = true|\mathbf{Pa}(x_i)) \quad = \quad \frac{e^{\psi(x_i;\mathbf{Pa}(x_i))}}{e^{\psi(x_i;\mathbf{Pa}(x_i))} + 1}$$

x	Δ
target(x1)	0.7
target(x2)	-0.2
target(x3)	-0.9

Functional gradient

Maximize e.g. Pseudo Log Likelihood

$$LL(\mathbf{X} = \mathbf{x}) \quad = \quad \sum_{x_i \in \mathbf{x}} \log P(x_i|\mathbf{Pa}(x_i))$$

Gradient of pseudo log-likelihood w.r.t ψ for learning gradient models

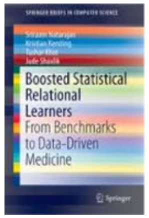
$$\Delta(x_i) = \frac{\partial \log P(\mathbf{X} = \mathbf{x})}{\partial \psi(x_i; \mathbf{Pa}(x_i))} = I(x_i = true; \mathbf{Pa}(x_i)) - P(x_i = true; \mathbf{Pa}(x_i))$$

Sum all gradient models to get final ψ

$$\psi_m = \psi_0 + \Delta_1 + \ldots + \Delta_m$$

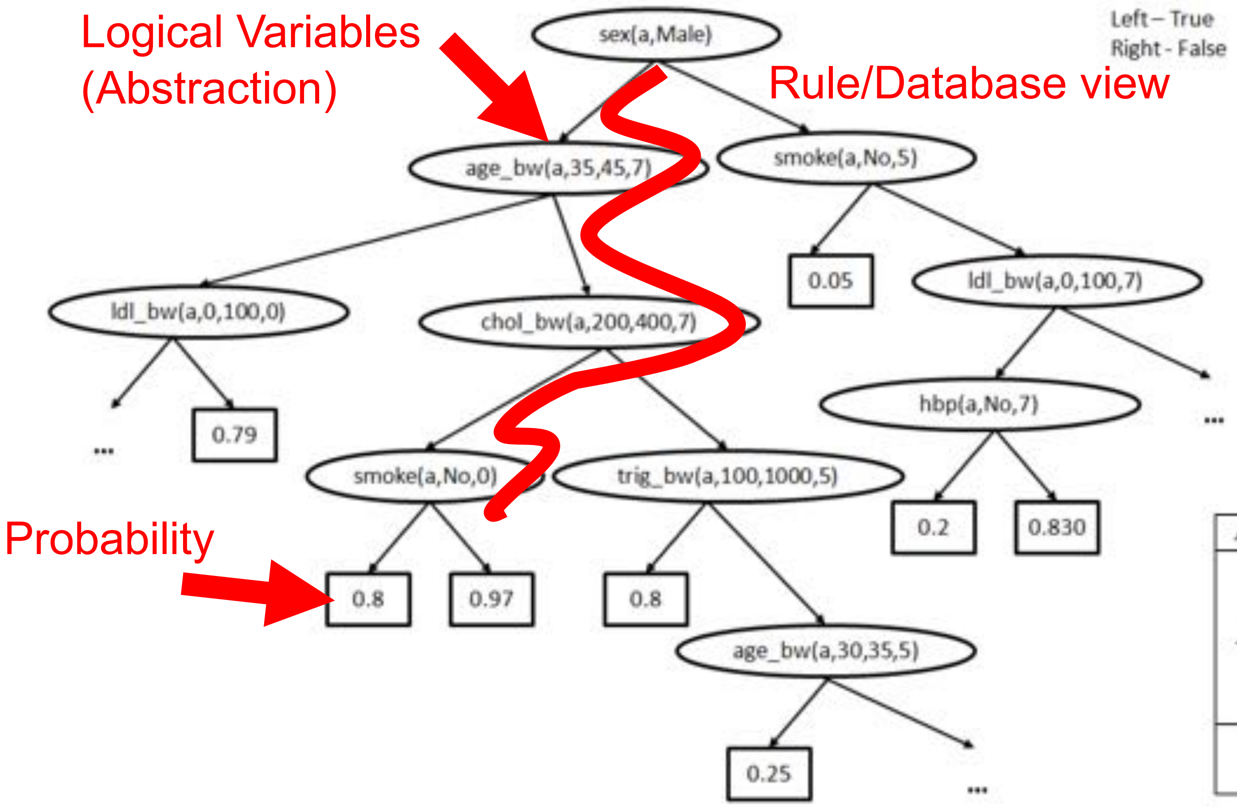
Extended to multiple SRL models & in presence of hidden data

Natarajan, Khot, Kersting, Shavlik. Boosted Statistical Relational Learners. Springer Brief 2015


Boosted Statistical Relational Learners
From Benchmarks to Data-Driven Medicine

Understanding Electronic Health Records

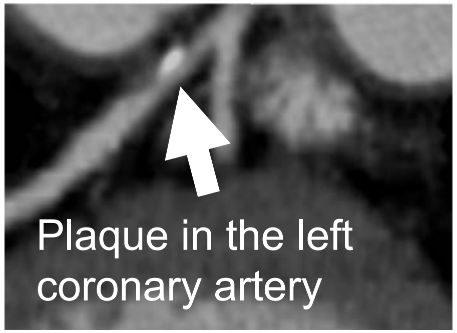
Atherosclerosis is the cause of the majority of Acute Myocardial Infarctions (heart attacks)



Logical Variables (Abstraction)

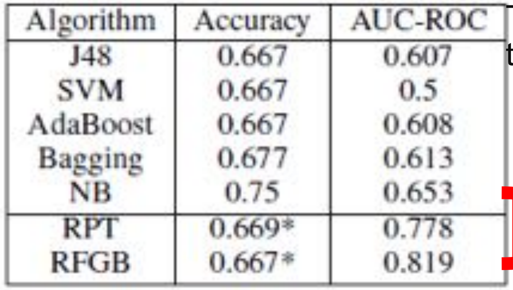
Rule/Database view

Probability



Plaque in the left coronary artery

[Circulation; 92(8), 2157-62, 1995; JACC; 43, 842-7, 2004]



Algorithm	Accuracy	AUC-ROC
J48	0.667	0.607
SVM	0.667	0.5
AdaBoost	0.667	0.608
Bagging	0.677	0.613
NB	0.75	0.653
RPT	0.669*	0.778
RFGB	0.667*	0.819

The higher, the better

25%

Algorithm for Mining Markov Logic Networks	Likelihood The higher, the better	AUC-ROC The higher, the better	AUC-PR The higher, the better	Time The lower, the better
Boosting	0.81	0.96	0.93	**9s**
LSM	0.73	0.54	0.62	**93 hrs**

11%

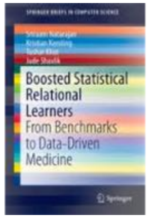
78%

50%

37200x faster

state-of-the-art

[Kersting, Driessens ICML´08; Karwath, Kersting, Landwehr ICDM´08; Natarajan, Joshi, Tadepelli, Kersting, Shavlik. IJCAI´11; Natarajan, Kersting, Ip, Jacobs, Carr IAAI `13; Yang, Kersting, Terry, Carr, Natarajan AIME ´15; Khot, Natarajan, Kersting, Shavlik ICDM´13, MLJ´12, MLJ´15, Yang, Kersting, Natarajan BIBM`17]

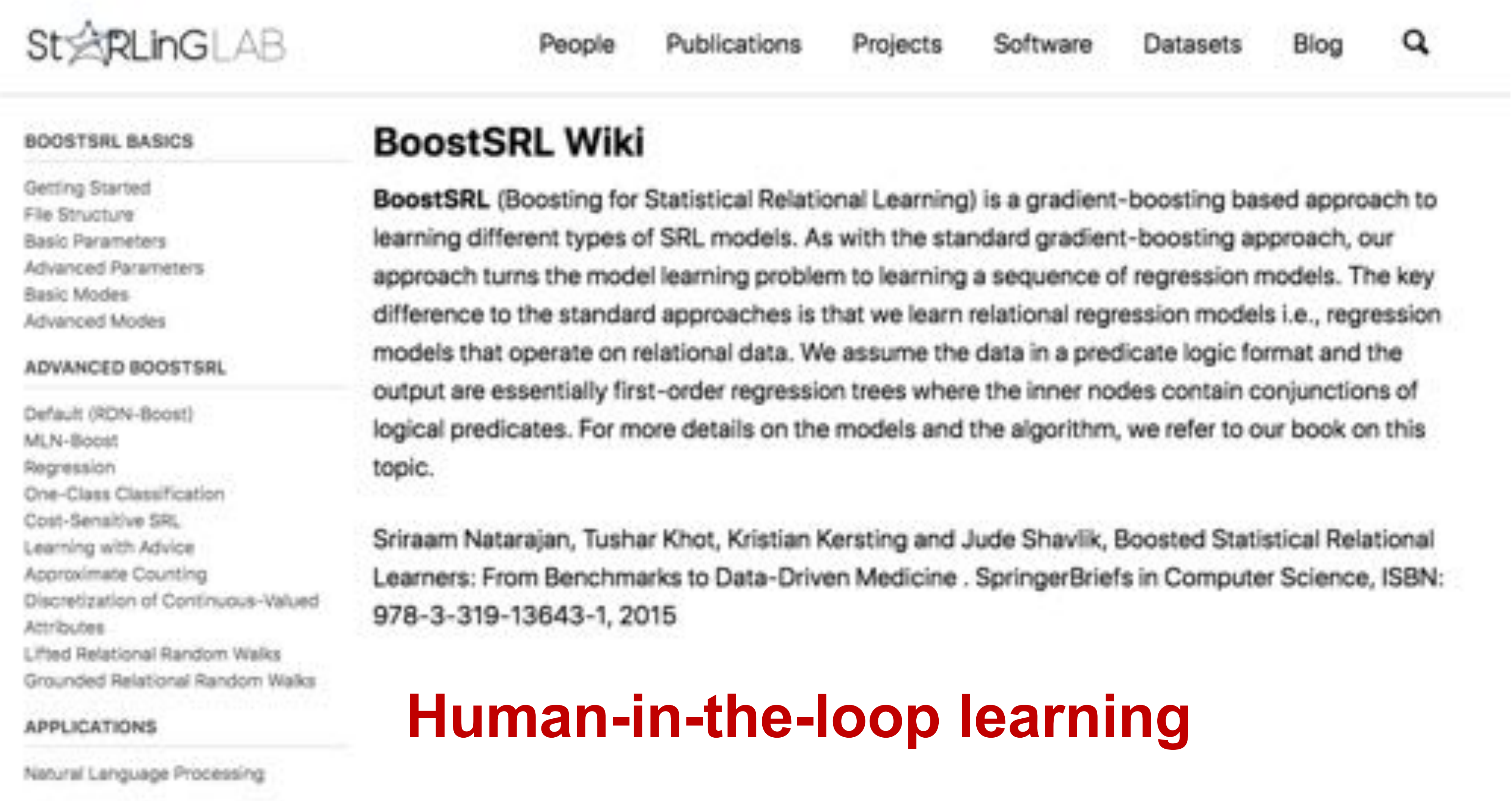

Boosted Statistical Relational Learners
From Benchmarks to Data-Driven Medicine

Natarajan, Khot, Kersting, Shavlik. Boosted Statistical Relational Learners. Springer Brief 2015

TECHNISCHE UNIVERSITÄT DARMSTADT

UTD THE UNIVERSITY OF TEXAS AT DALLAS

https://starling.utdallas.edu/software/boostsrl/wiki/



BOOSTSRL BASICS

- Getting Started
- File Structure
- Basic Parameters
- Advanced Parameters
- Basic Modes
- Advanced Modes

ADVANCED BOOSTSRL

- Default (RDN-Boost)
- MLN-Boost
- Regression
- One-Class Classification
- Cost-Sensitive SRL
- Learning with Advice
- Approximate Counting
- Discretization of Continuous-Valued Attributes
- Lifted Relational Random Walks
- Grounded Relational Random Walks

APPLICATIONS

- Natural Language Processing

BoostSRL Wiki

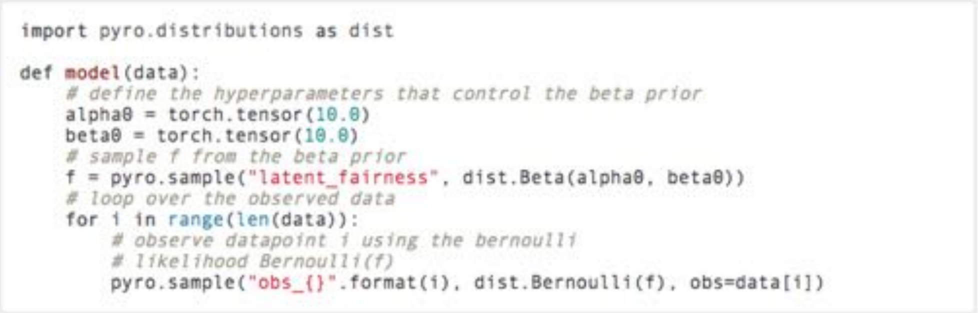
BoostSRL (Boosting for Statistical Relational Learning) is a gradient-boosting based approach to learning different types of SRL models. As with the standard gradient-boosting approach, our approach turns the model learning problem to learning a sequence of regression models. The key difference to the standard approaches is that we learn relational regression models i.e., regression models that operate on relational data. We assume the data in a predicate logic format and the output are essentially first-order regression trees where the inner nodes contain conjunctions of logical predicates. For more details on the models and the algorithm, we refer to our book on this topic.

Sriraam Natarajan, Tushar Khot, Kristian Kersting and Jude Shavlik, Boosted Statistical Relational Learners: From Benchmarks to Data-Driven Medicine . SpringerBriefs in Computer Science, ISBN: 978-3-319-13643-1, 2015

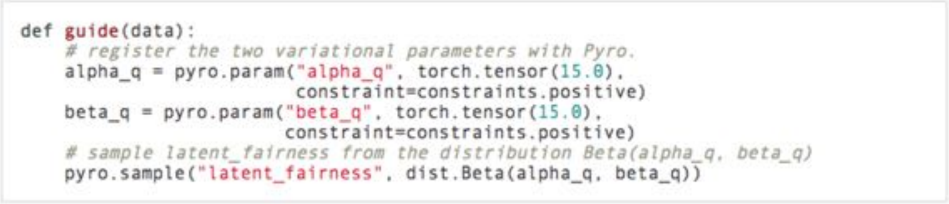
Human-in-the-loop learning

New field: Deep Probabilistic Programming

In general, computing the exact posterior is intractable, i.e., inverting the generative process to determine the state of latent variables corresponding to an input is time-consuming and error-prone.

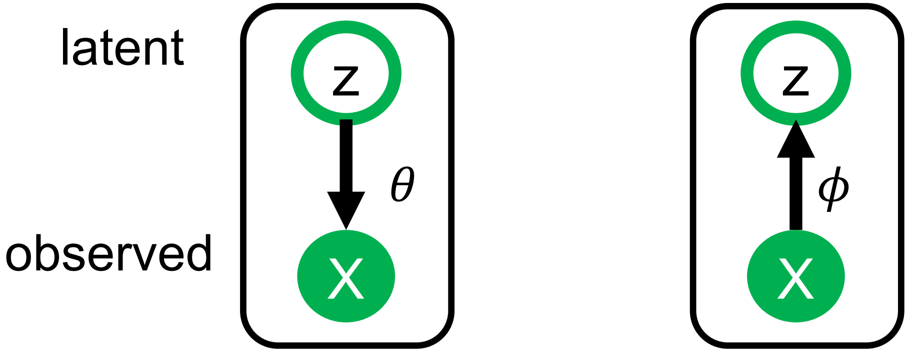


```python
import pyro.distributions as dist

def model(data):
    # define the hyperparameters that control the beta prior
    alpha0 = torch.tensor(10.0)
    beta0 = torch.tensor(10.0)
    # sample f from the beta prior
    f = pyro.sample("latent_fairness", dist.Beta(alpha0, beta0))
    # loop over the observed data
    for i in range(len(data)):
        # observe datapoint i using the bernoulli
        # likelihood Bernoulli(f)
        pyro.sample("obs_{}".format(i), dist.Bernoulli(f), obs=data[i])
```



```python
def guide(data):
    # register the two variational parameters with Pyro.
    alpha_q = pyro.param("alpha_q", torch.tensor(15.0),
                         constraint=constraints.positive)
    beta_q = pyro.param("beta_q", torch.tensor(15.0),
                        constraint=constraints.positive)
    # sample latent_fairness from the distribution Beta(alpha_q, beta_q)
    pyro.sample("latent_fairness", dist.Beta(alpha_q, beta_q))
```

(2) Ease the implementation by some high-level, probabilistic programming language



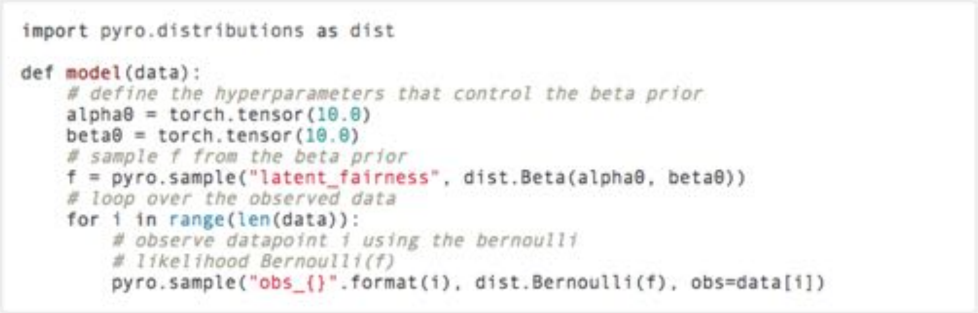
Deep Neural Network

(1) Instead of optimizating variational parameters for every new data point, use a deep network to predict the posterior given X [Kingma, Welling 2013, Rezende et al. 2014]

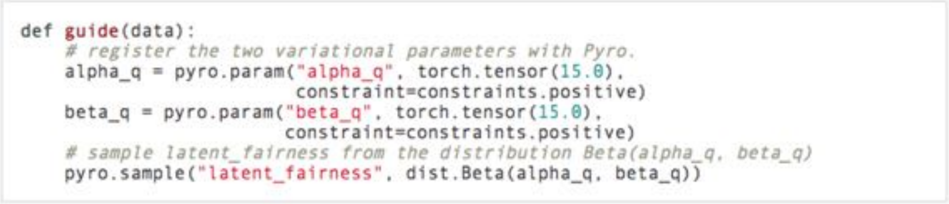
UBER AI Labs UNIVERSITY OF CAMBRIDGE Max Planck Institute for Intelligent Systems UNI GRAZ TECHNISCHE UNIVERSITÄT DARMSTADT

[Stelzner, Molina, Peharz, Vergari, Trapp, Valera, Ghahramani, Kersting ProgProb 2018]

Sum-Product Probabilistic Programming

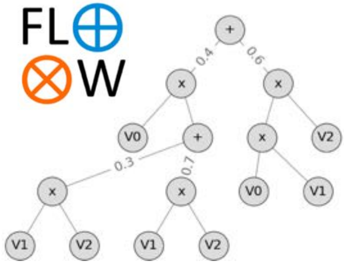


```
import pyro.distributions as dist

def model(data):
    # define the hyperparameters that control the beta prior
    alpha0 = torch.tensor(10.0)
    beta0 = torch.tensor(10.0)
    # sample f from the beta prior
    f = pyro.sample("latent_fairness", dist.Beta(alpha0, beta0))
    # loop over the observed data
    for i in range(len(data)):
        # observe datapoint i using the bernoulli
        # likelihood Bernoulli(f)
        pyro.sample("obs_{}".format(i), dist.Bernoulli(f), obs=data[i])
```

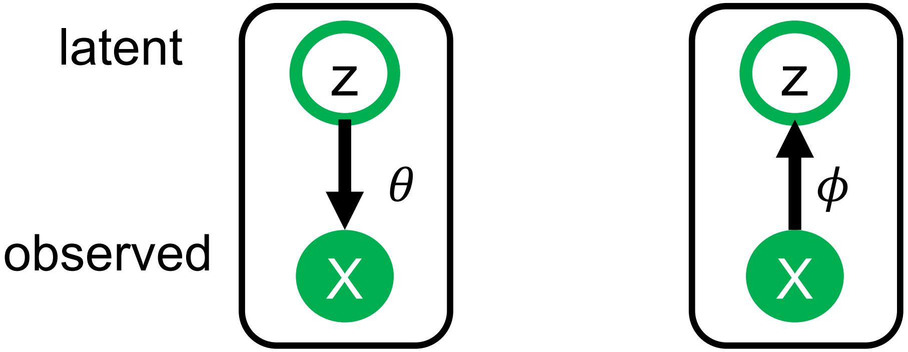


```
def guide(data):
    # register the two variational parameters with Pyro.
    alpha_q = pyro.param("alpha_q", torch.tensor(15.0),
                         constraint=constraints.positive)
    beta_q = pyro.param("beta_q", torch.tensor(15.0),
                        constraint=constraints.positive)
    # sample latent_fairness from the distribution Beta(alpha_q, beta_q)
    pyro.sample("latent_fairness", dist.Beta(alpha_q, beta_q))
```

Sum-Product Network



(2) Ease the implementation by some high-level, probabilistic programming language

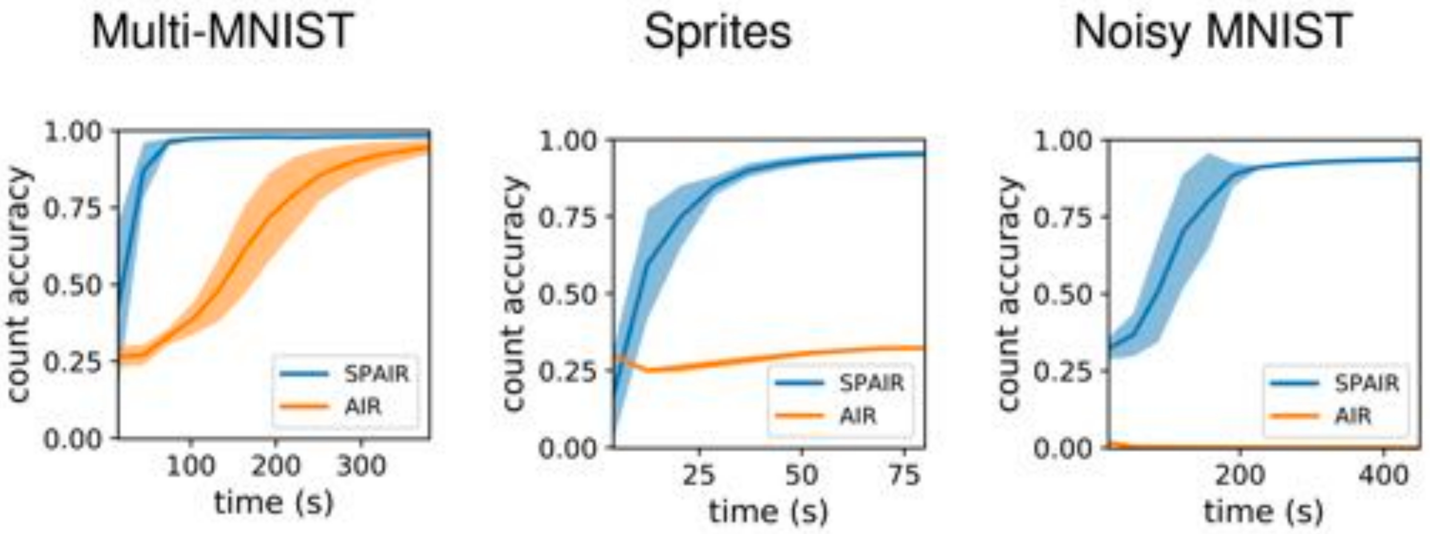


Deep Neural Network

(1) Instead of optimizating variational parameters for every new data point, use a deep network to predict the posterior given X [Kingma, Welling 2013, Rezende et al. 2014]

Sum-Product Attent-Infer Repeat

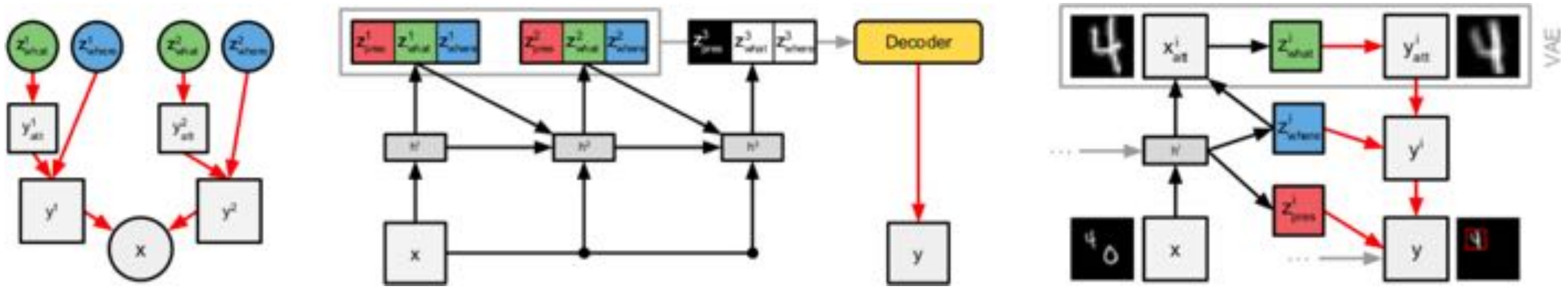
Replace VAE by SPN



[Stelzner, Peharz, Kersting ICML 2019]

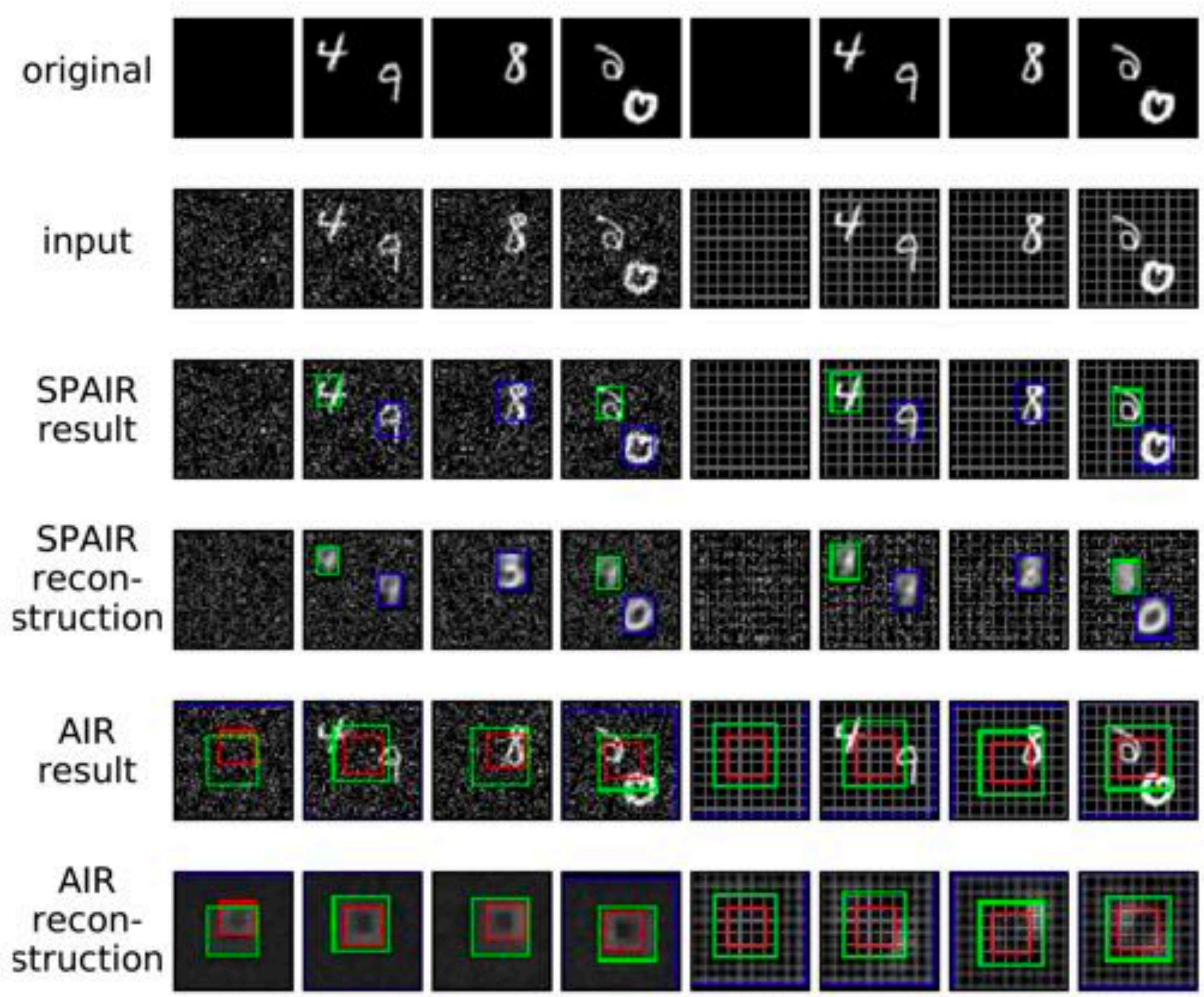
UNIVERSITY OF CAMBRIDGE

TECHNISCHE UNIVERSITÄT DARMSTADT



A graphical model implemented in neural fashion using an VAE as object represenation [Eslami, Heess, Weber, Tassa, Szepesvari, Kavukcuoglu, Hinton NIPS 2016]

Sum-Product Attent-Infer Repeat



[Stelzner, Peharz, Kersting ICML 2019]

There are strong invests into (deep) probabilistic programming

UBER AI Labs

RelationalAI, Apple, Microsoft and Uber are investing hundreds of millions of US dollars

Get Siri-ous.

No more evasive answers. No more coy innuendos. When you get romantic with Siri Pro, the sparks really fly.

Microsoft® Research

relationalAI

AI for the enterprise

UNIVERSITY OF CAMBRIDGE

UBER AI Labs

Max Planck Institute for Intelligent Systems

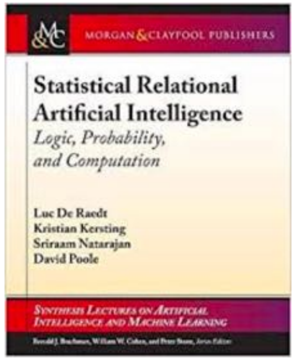
UNI GRAZ

FLOW

TECHNISCHE UNIVERSITÄT DARMSTADT

Microsoft Research

UBER AI Labs


MORGAN & CLAYPOOL PUBLISHERS

Statistical Relational Artificial Intelligence
Logic, Probability, and Computation

Luc De Raedt
Kristian Kersting
Sriraam Natarajan
David Poole

Synthesis Lectures on Artificial Intelligence and Machine Learning

Ronald J. Brachman, William W. Cohen, and Peter Stone, *Series Editors*

relationalAI
AI for the enterprise

Getting deep systems that reason and know when they don’t know

Responsible AI systems that explain their decisions and co-evolve with the humans

Open AI systems that are easy to realize and understandable for the domain experts

„Tell the AI when it is right for the wrong reasons and it adapts its behavior“

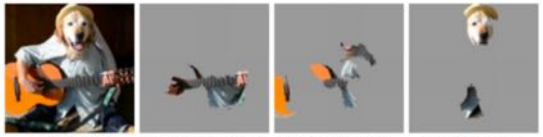

(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Figure 4: Explaining an image classification prediction made by Google’s Inception network, highlighting positive pixels. The top 3 classes predicted are “Electric Guitar” ($p = 0.32$), “Acoustic guitar” ($p = 0.24$) and “Labrador” ($p = 0.21$)

Teso, Kersting AIES 2019

AAAI / ACM conference on
ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

The third wave of AI

- **AI is more than deep neural networks.** Probabilistic and causal models are whiteboxes that provide insights into applications
- **AI is more than a single table.** Loops, graphs, different data types, relational DBs, … are central to data science, and high-level programming languages for DS help to capture this complexity
- **AI is more than just Machine Learners and Statisticians**

Healthcare calls for AI systems that can acquire human-like communication and reasoning capabilities, with the ability to recognise new situations and adapt to them