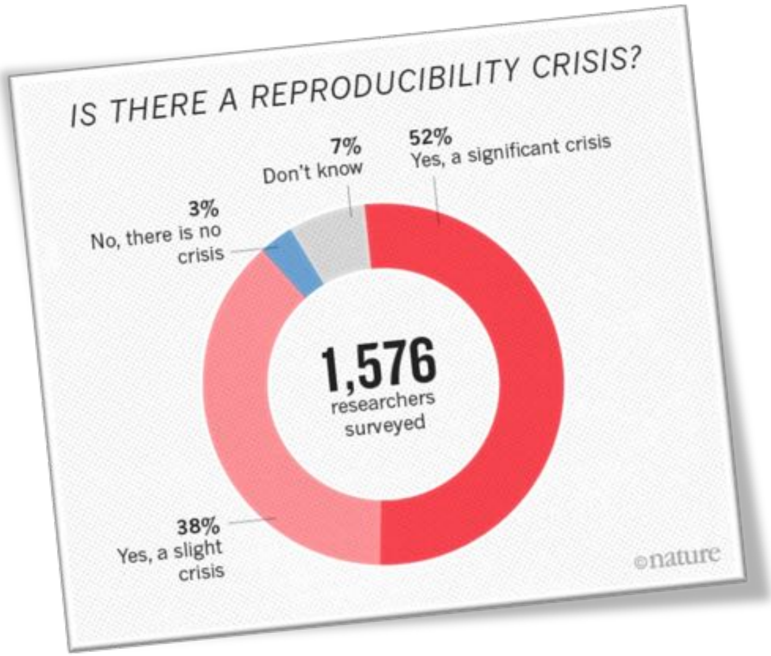
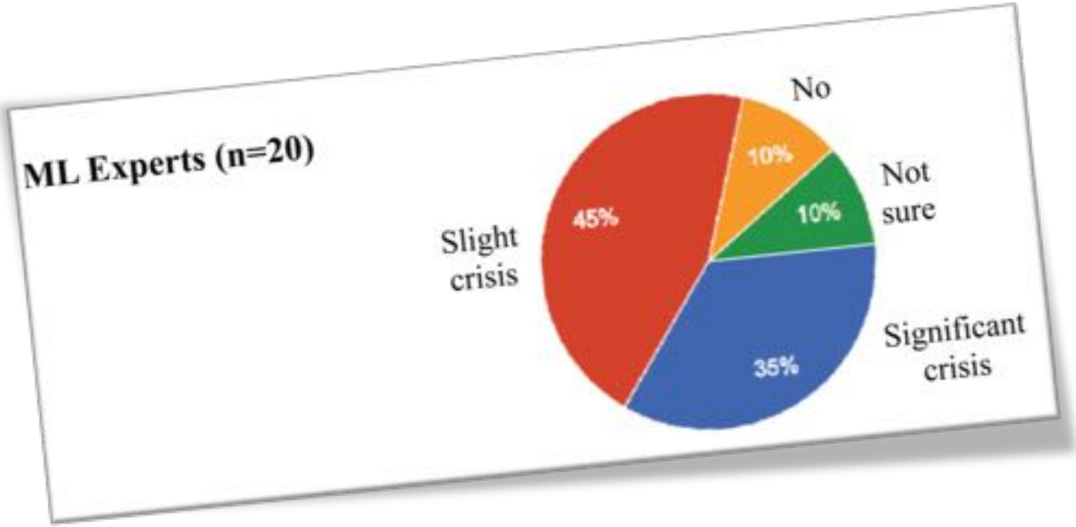
Towards Reproducibility in Machine Learning and AI





J. Pineau: „The ICLR 2018 Reproducibility Challenge". Talk at the MLTRAIN@RML ICML 2018 Workshop

M. Baker: „1,500 scientists lift the lid on reproducibility". Nature, 2016 May 26;533(7604):452-4. doi: 10.1038/533452
https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970?proof=true

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

Reproducibility Crisis in Science (2016)

M. Baker: „1,500 scientists lift the lid on reproducibility“. Nature, 2016 May 26;533(7604):452-4. doi: 10.1038/533452
https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970?proof=true

Do ML and AI make a difference?

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

The New York Times

Opinion

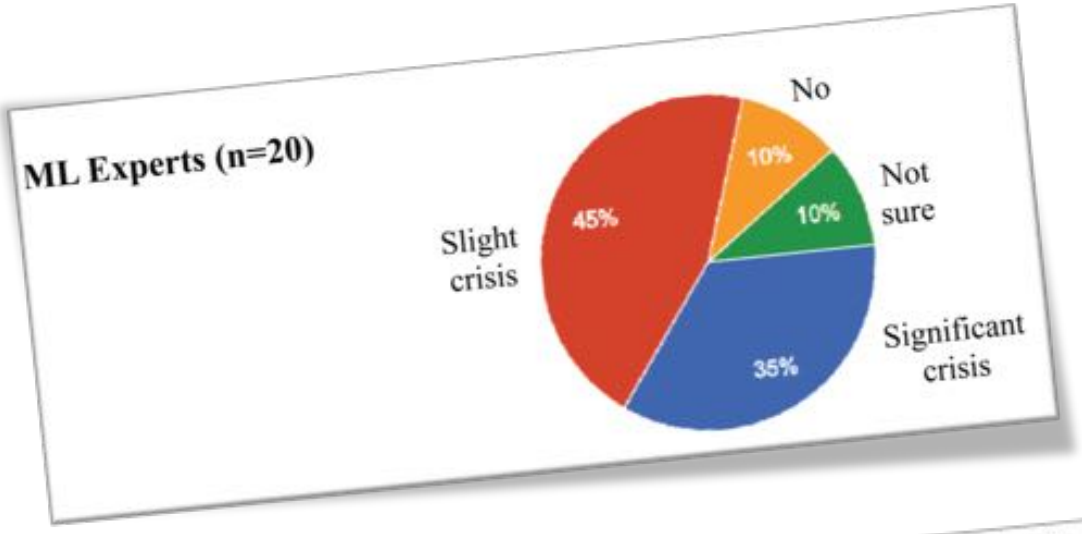
A.I. Is Harder Than You Think

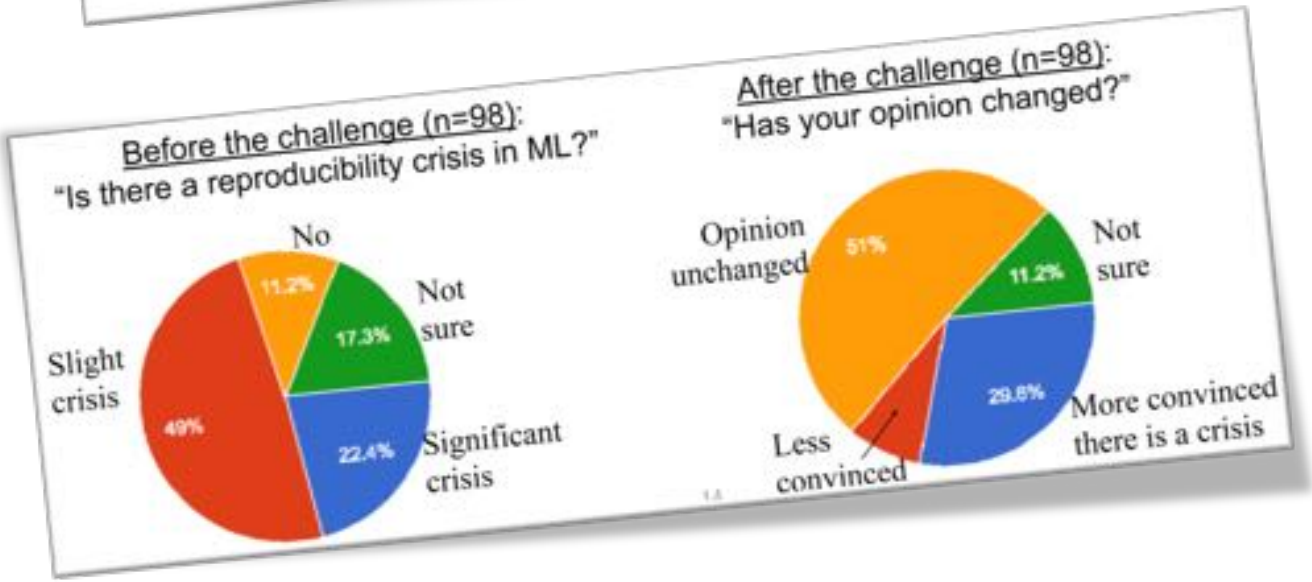
By Gary Marcus and Ernest Davis

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

May 18, 2018

Reproducibility Crisis in ML & AI (2018)





J. Pineau: „The ICLR 2018 Reproducibility Challenge".
Talk at the MLTRAIN@RML Workshop at ICML 2018

Joelle Pineau

McGill

Facebook AI Research (FAIR)

Survey participants:

- 54 challenge participants
- 30 authors of ICLR submissions targeted by reproducibility effort
- 14 others (random volunteers, other ICLR authors, ICLR area chair & reviewers, course instructors)

MLTRAIN

HOME EVENTS CORPORATE TRAINING ABOUT BLOG CONTACT

NIPS HIGHLIGHTS, LEARN HOW TO CODE A PAPER WITH STATE OF THE ART FRAMEWORKS

Dec 09 @ 08:50 AM - 06:05 PM NIPS, Los Angeles, California

Nikolaos Vasiloglou

ismion

ENABLING REPRODUCIBILITY IN MACHINE LEARNING MLTRAIN@RML (ICML 2018)

Jul 14 @ 08:30 AM - 06:00 PM Stockholmsmässan

Mila Yoshua Bengio

frontiers in Big Data | Machine Learning and Artificial Intelligence

First Machine Learning and Artificial Intelligence journal that explicitely welcomes replication studies and code review papers

Sriraam Natarajan

UT DALLAS
The University of Texas at Dallas

A lot of systems to support reproducible ML research

OpenML beta_2

Machine learning, better, together

Joaquin Vanschoren

Technische Universiteit Eindhoven University of Technology

20328 data sets — Find or add **data** to analyse

68724 tasks — Download or create scientific **tasks**

6994 flows — Find or add data analysis **flows**

9749541 runs — Upload and explore all **results** online.

Percy Lang

Stanford University

CodaLab

Accelerating reproducible computational research.

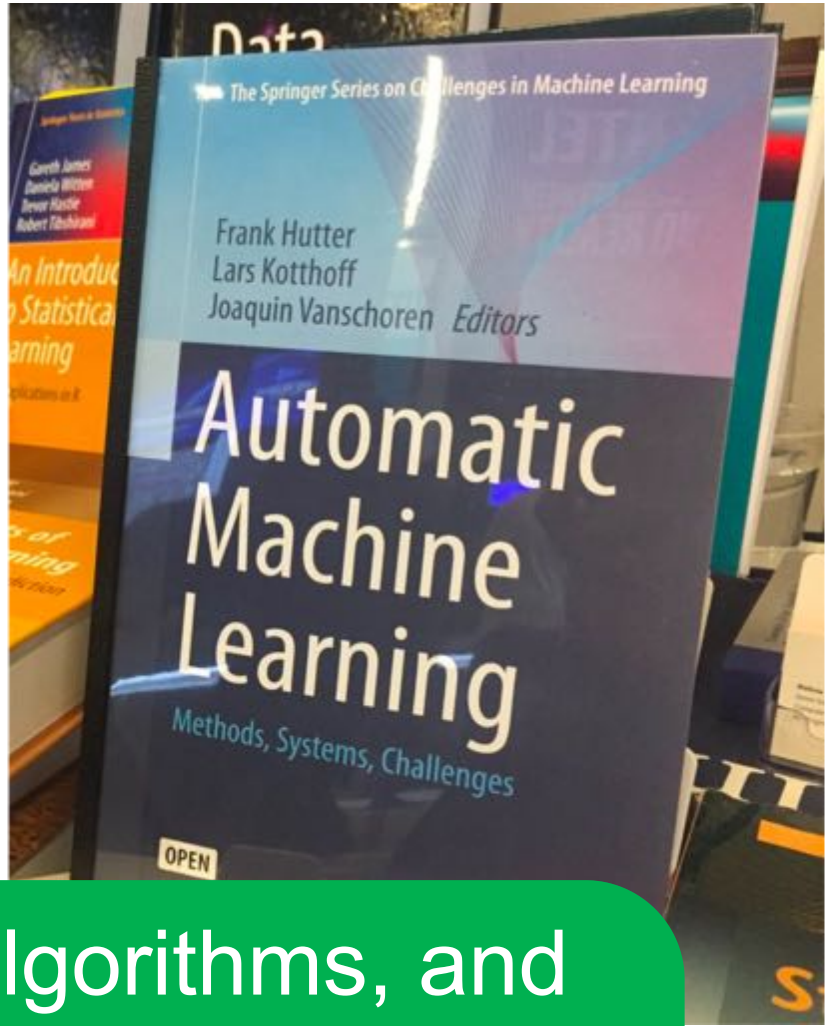
Worksheets

Run reproducible experiments and create executable papers using worksheets.

Competitions

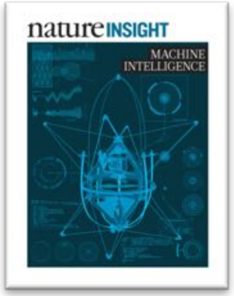
Enter an existing competition to solve challenging data problems, or host your own.

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



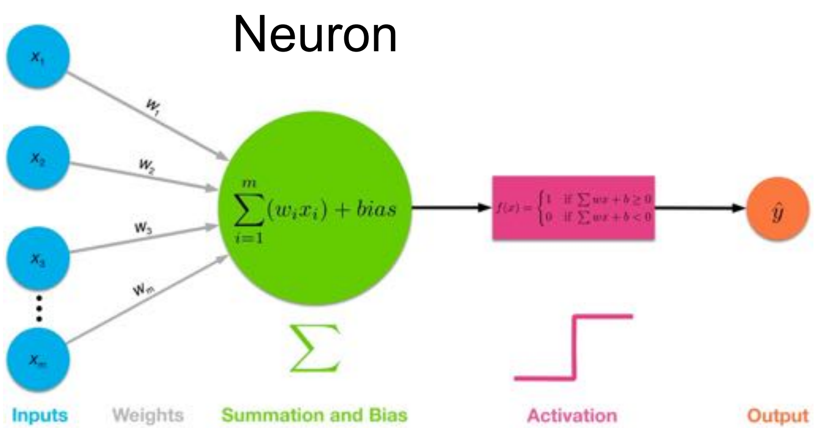
Provide the foundations, algorithms, and tools to develop systems that ease and support building ML/AI models as much as possible and in turn help reproducing and even justifying our results

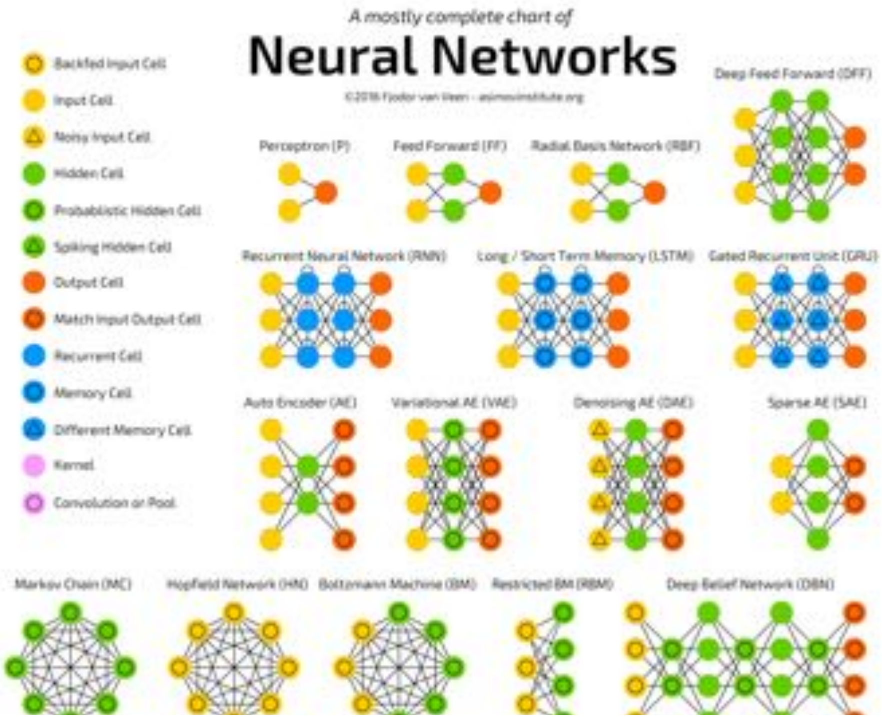
Deep Neural Networks



Potentially much more powerful than shallow architectures, represent computations

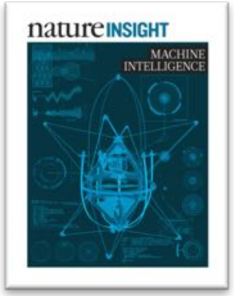
[LeCun, Bengio, Hinton Nature 521, 436–444, 2015]





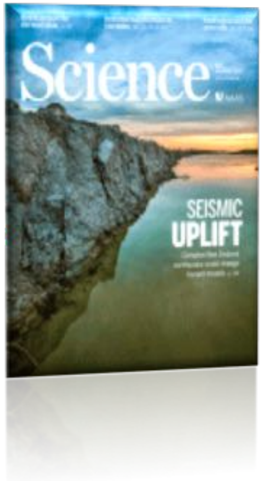
Differentiable Programming

Deep Neural Networks



Potentially much more powerful than shallow architectures, represent computations

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



SHARE

1.02k

0

REPORTS | PSYCHOLOGY

Semantics derived automatically from language corpora contain human-like biases

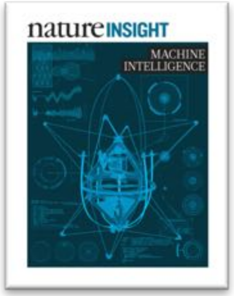
Aylin Caliskan[1,*], Joanna J. Bryson[1,2,*], Arvind Narayanan[1,*]

+ 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

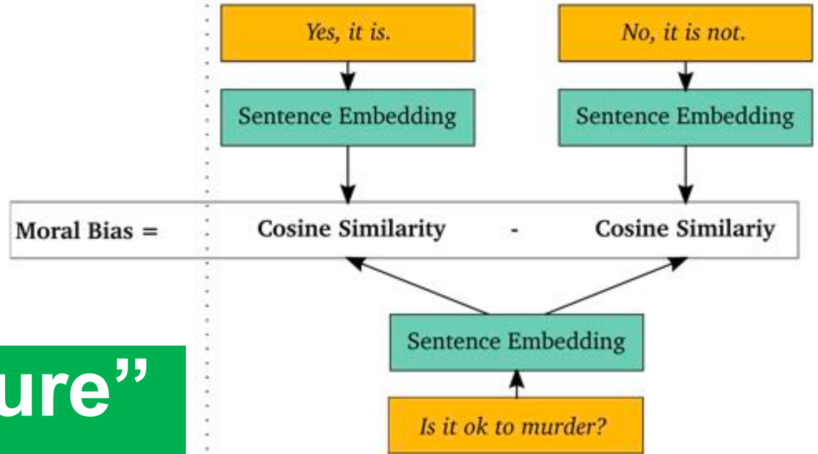


Potentially much more powerful than shallow architectures, represent computations

[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



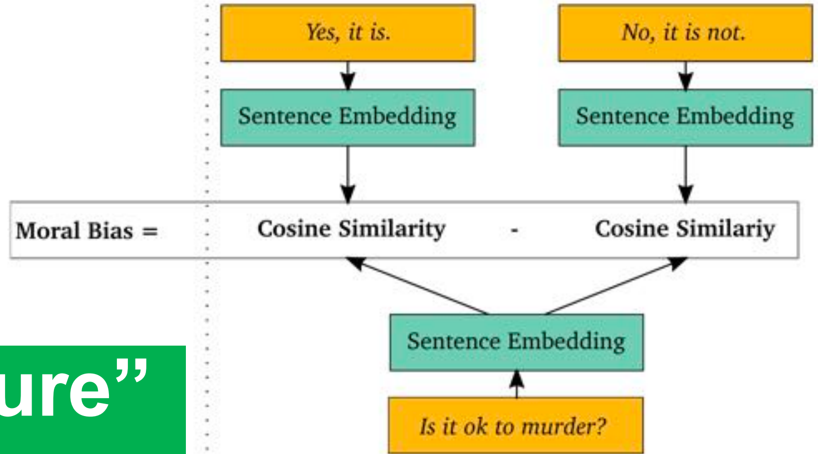
But luckily they also "capture" our moral choices

[Jentzsch, Schramowski, Rothkopf, Kersting AIES 2019]

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

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 luckily they also "capture" our moral choices

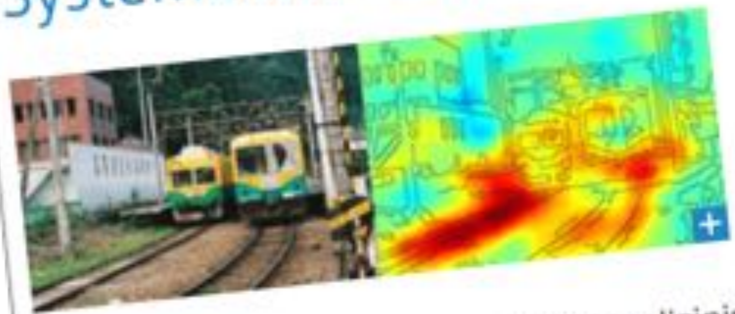
[Jentzsch, Schramowski, Rothkopf, Kersting AIES 2019]

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

Can we trust neural networks?

12. März 2019

Paper bei Nature Commmunications erschienen: Wissenschaftler stellen KI-Systeme auf den Prüfstand



Algorithmen der Künstlichen Intelligenz (KI) und des Maschinellen Lernens wie beispielsweise Deep Learning erobern immer mehr Bereiche unseres Lebens: Sie ermöglichen digitale Sprachassistenten oder Übersetzungsdienste, verbessern die medizinische Diagnostik und sind unverzichtbarer Bestandteil von Zukunftstechnologien wie dem autonomen Fahren. Gestützt durch eine stetig wachsende Anzahl verfügbarer Daten und leistungsfähiger Rechnerarchitekturen, scheinen Lernalgorithmen der menschlichen Leistungsfähigkeit gleichgestellt oder sogar überlegen. Das Problem: Bislang bleibt es den Wissenschaftlern und Wissenschaftlerinnen meistens verborgen, wie die KI-Systeme zu ihren Entscheidungen kommen. Damit bleibt oft auch unklar, ob es sich wirklich um intelligente Entscheidungen oder statistisch erfolgreiche Verfahren handelt.

Deep neural networks do not quantify their uncertainty
They are not calibrated probabilistic models

MNIST

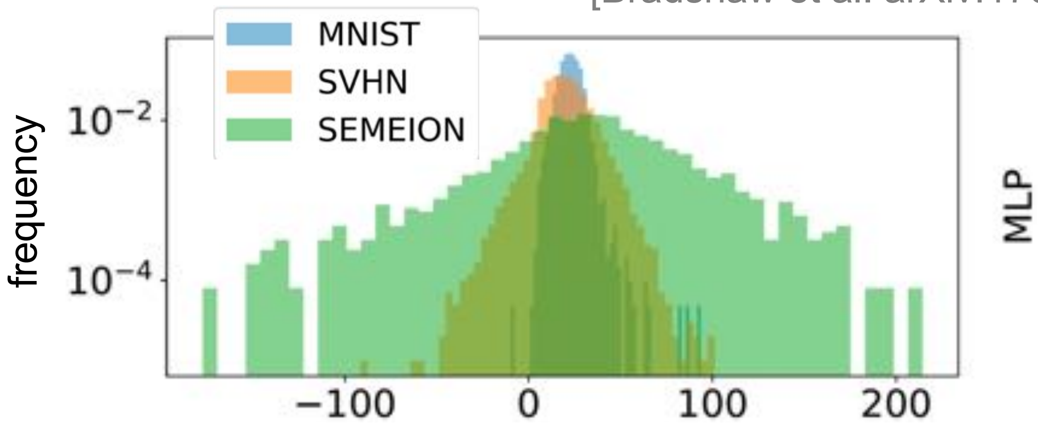
Train & Evaluate

SVHN

SEMEION

Transfer Testing

[Bradshaw et al. arXiv:1707.02476 2017]

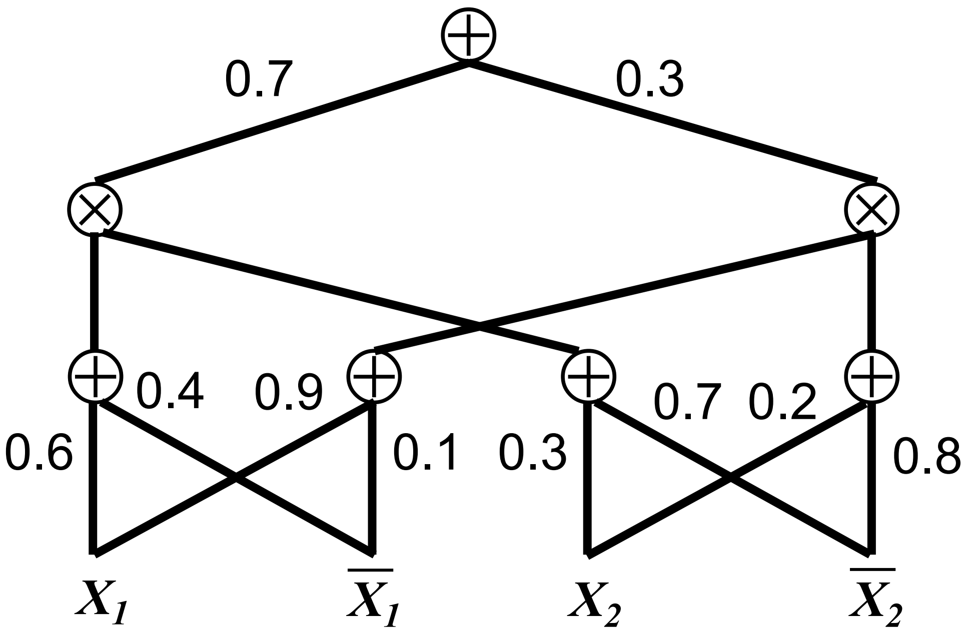


Input log „likelihood“ (sum over outputs)

[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



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

Inference is linear in size of network

[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

195 commits | 2 branches | 0 releases | 6 contrib...

Branch: master ▾ | New pull request | Create new file | Upload files | Find file | Clone or download ▾

https://github.com/SPFlow/SPFlow

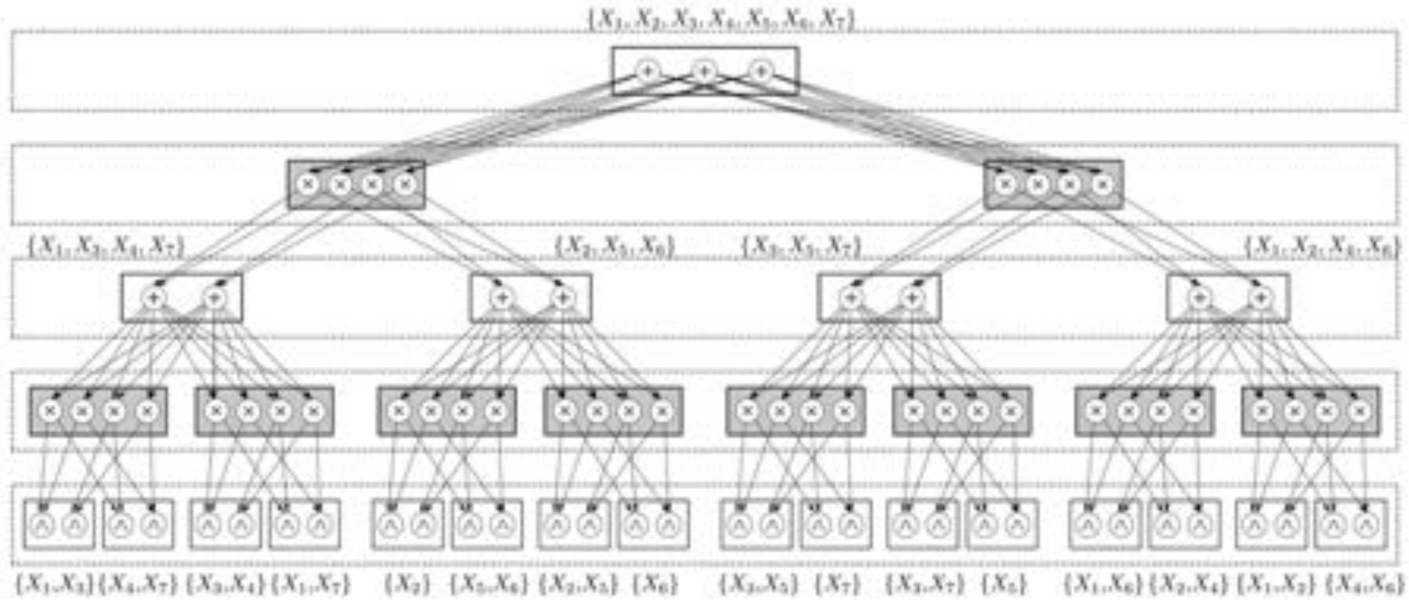
```
from spn.structure.leaves.parametric.Parametric import Categorical

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

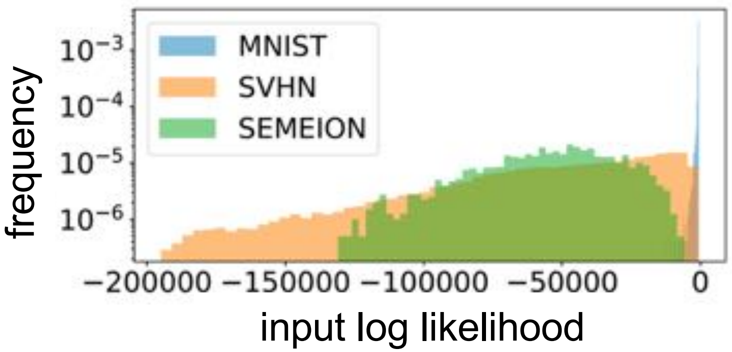
SPFlow, an open-source Python library providing a simple interface to inference, learning and manipulation routines for deep and tractable probabilistic models called Sum-Product Networks (SPNs). The library allows one to quickly create SPNs both from data and through a domain specific language (DSL).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

Random sum-product networks

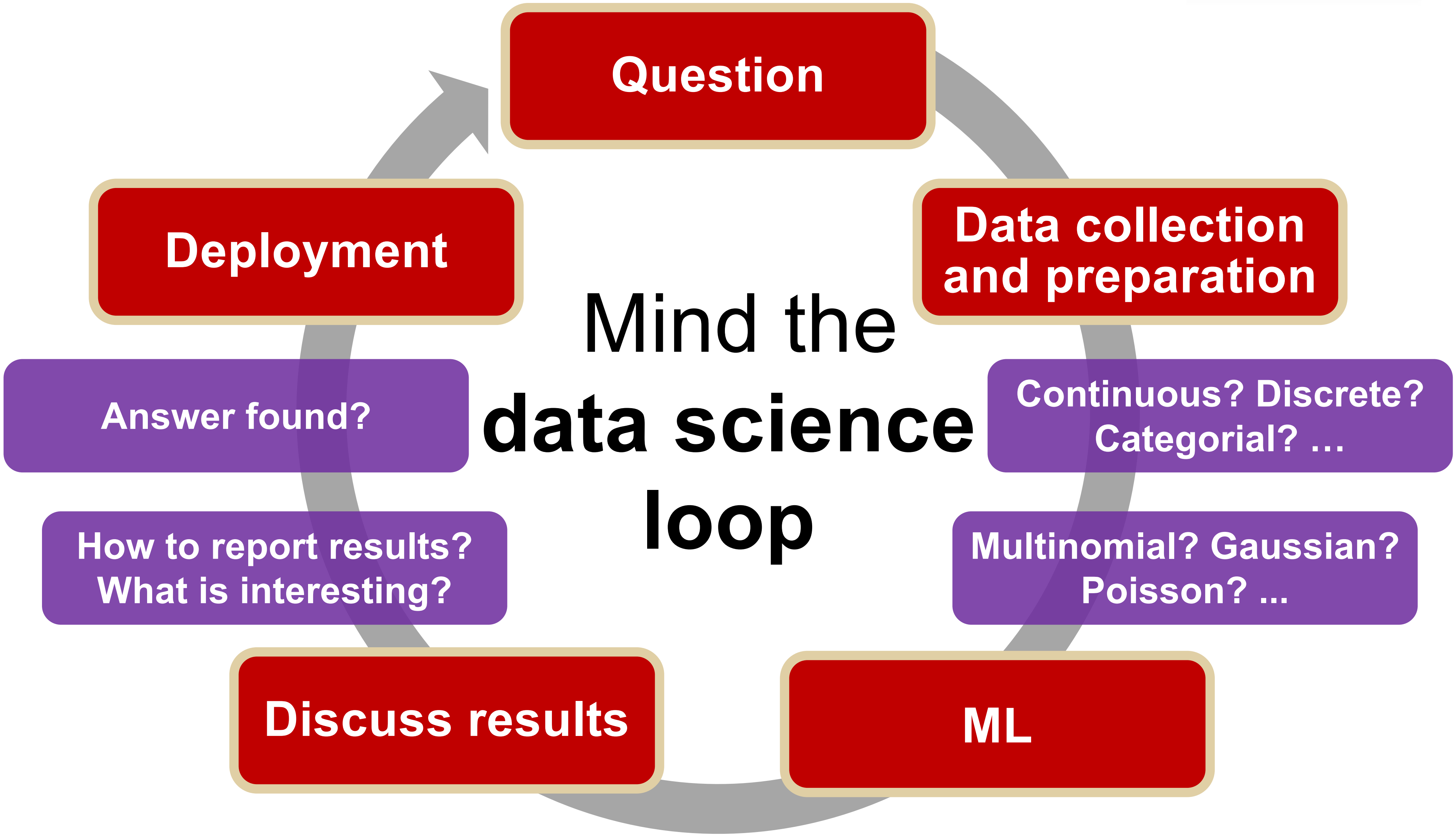
[Peharz, Vergari, Molina, Stelzner, Trapp, Kersting, Ghahramani UDL@UAI 2018]



		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


Question
Data collection and preparation
Continuous? Discrete? Categorial? …
Multinomial? Gaussian? Poisson? ...
ML
Discuss results
How to report results? What is interesting?
Answer found?
Deployment
Mind the data science loop

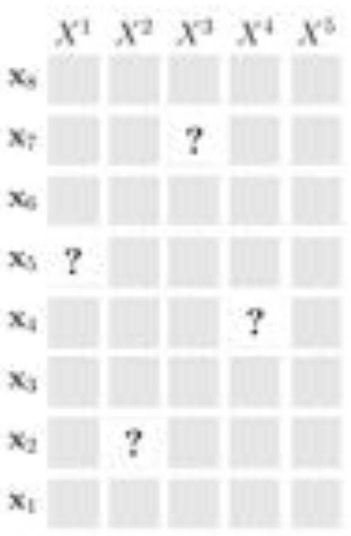
[Vergari, Molina, Peharz, Ghahramani, Kersting, Valera AAAI 2019]

CAML DFG MADESI Federal Ministry of Education and Research

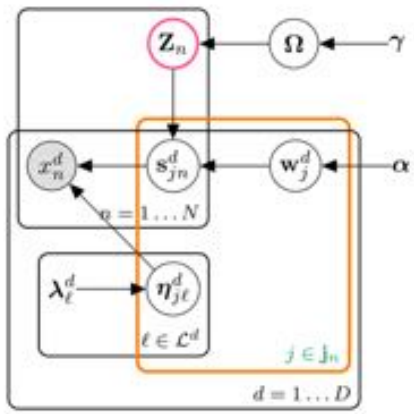
The Explorative Automatic Statistician

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

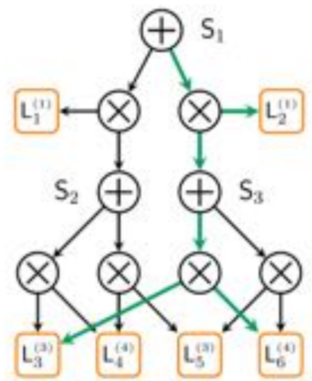
missing value



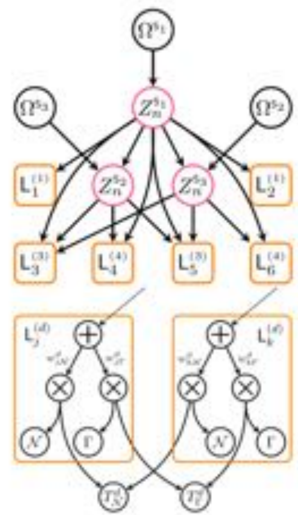
We can even automatically discovers the statistical types and parametric forms of the variables



Bayesian Type Discovery

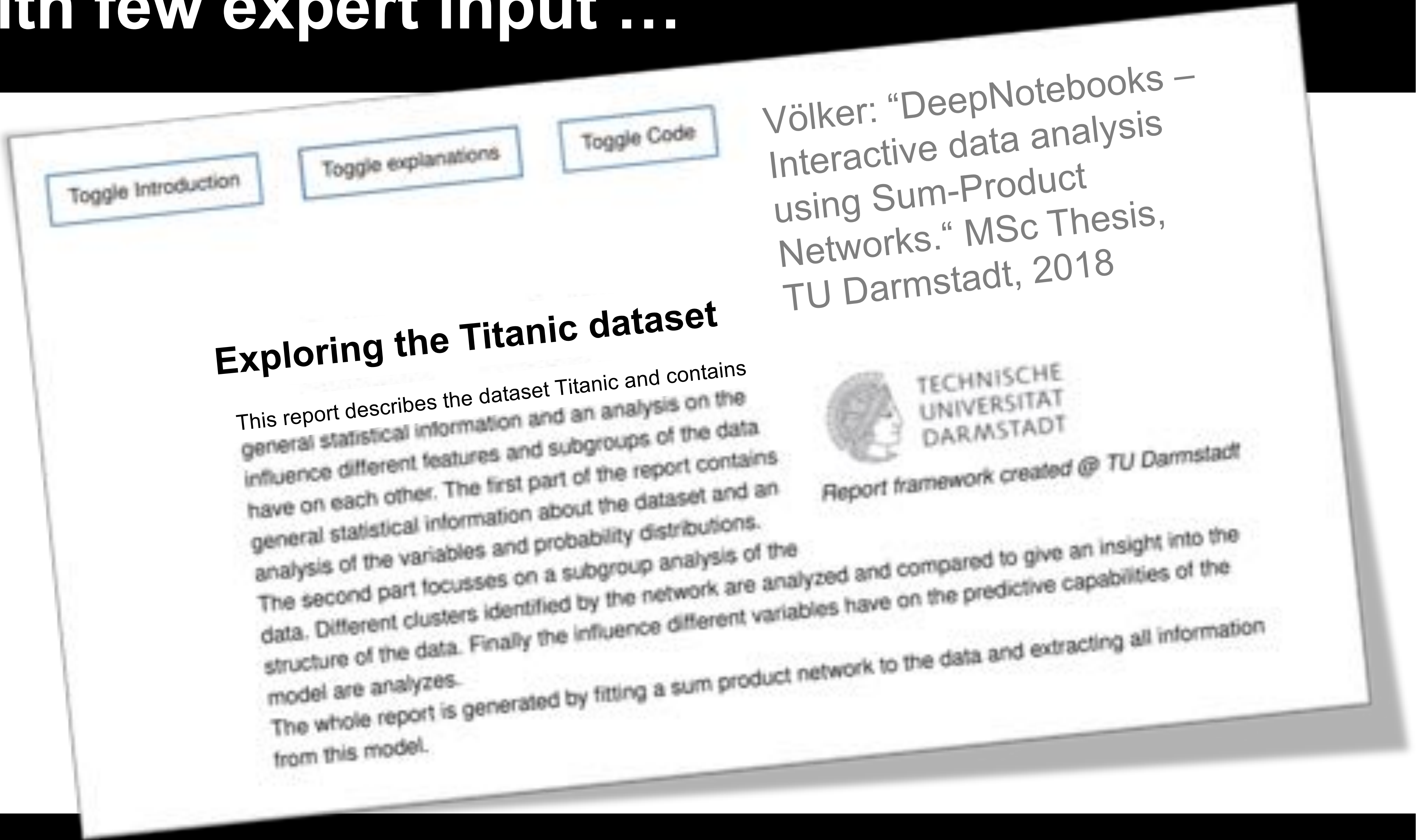


Mixed Sum-Product Network



Automatic Statistician

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



Toggle Introduction | Toggle explanations | Toggle Code

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.

TECHNISCHE UNIVERSITAT DARMSTADT

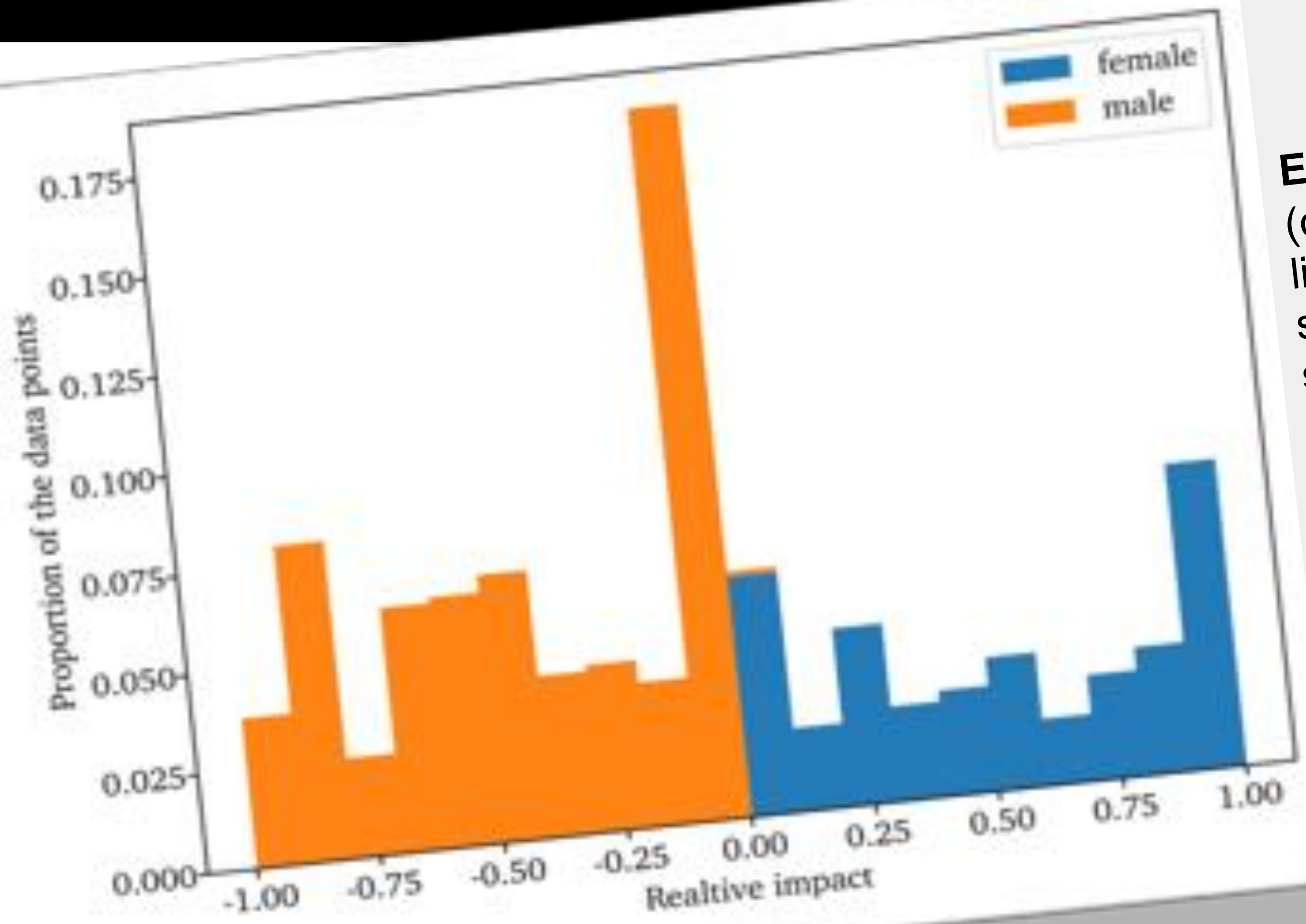
Report framework created @ TU Darmstadt

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

…and can compile data reports automatically

*Similar to [Lapuschkin et al., Nature Communication 10:1096, 2019] nature COMMUNICATIONS

The machine understands the data with no expert input …



Explanations* (computable in linear time in the size of the SPN) **showing the impact of "gender" on the chances of survival for the Titanic dataset**

…and can compile data reports automatically

Programming languages for Systems AI,

the computational and mathematical modeling of complex AI systems.

[Laue et al. NeurIPS 2018; Kordjamshidi, Roth, Kersting: “Systems AI: A Declarative Learning Based Programming Perspective.“ IJCAI-ECAI 2018]

Th ... y

…bu ... e,
d

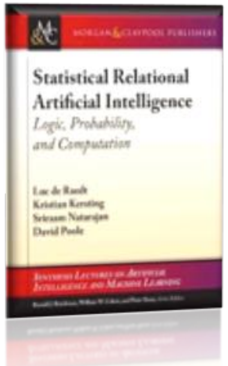
Eric Schmidt, Executive Chairman, Alphabet Inc.: Just Say "Yes”, Stanford Graduate School of Business, May 2, 2017.https://www.youtube.com/watch?v=vbb-AjiXyh0.

Since science is more than a single table !

P(heart attack | )?

Crossover of ML and AI with data & programming abstractions

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



KATHOLIEKE UNIVERSITEIT LEUVEN

TECHNISCHE UNIVERSITÄT DARMSTADT

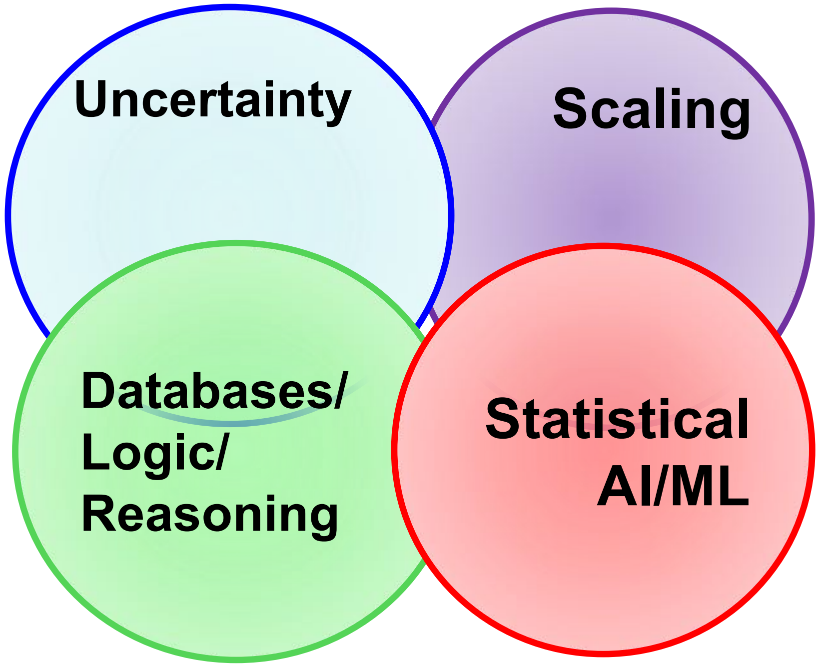
THE UNIVERSITY OF TEXAS AT DALLAS

UBC THE UNIVERSITY OF BRITISH COLUMBIA

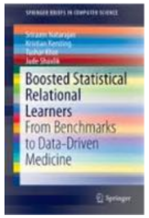
building general-purpose AI and ML machines

make the ML/AI expert more effective

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



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

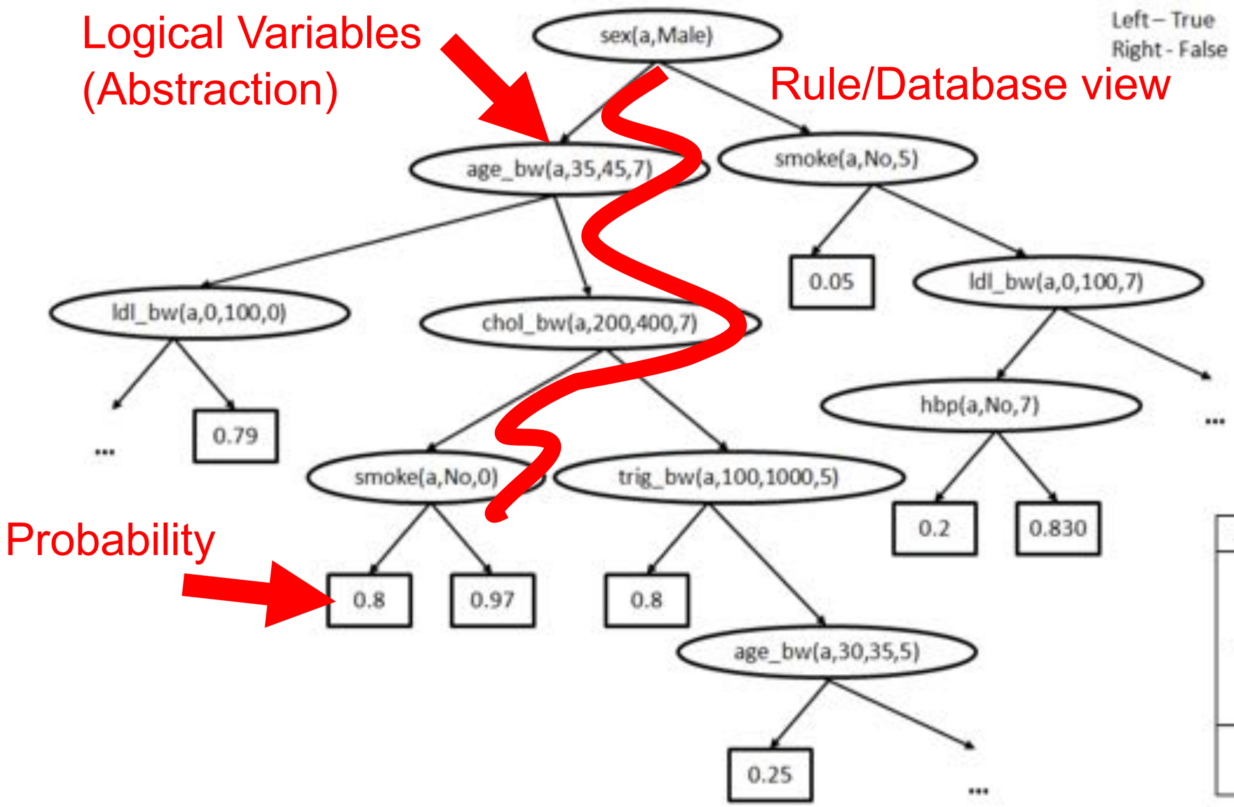


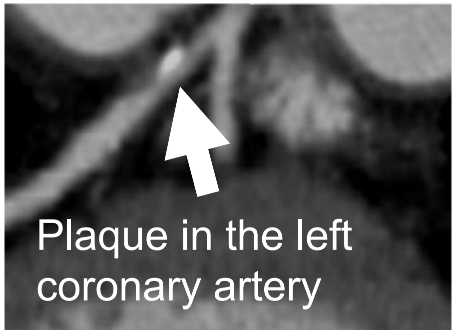
Understanding Electronic Health Records

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

TECHNISCHE UNIVERSITÄT DARMSTADT

THE UNIVERSITY OF TEXAS AT DALLAS





[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%

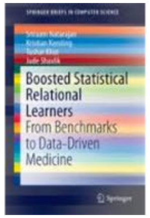
Algorithm for Mining Markov Logic Networks	Likelihood The higher, the better	AUC-ROC The higher, the better	AUC-PR The higher, the better	Time The lower, the better
Boosting	0.81	0.96	0.93	**9s**
LSM	0.73	0.54	0.62	**93 hrs**

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]

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



TECHNISCHE UNIVERSITÄT DARMSTADT

THE UNIVERSITY OF TEXAS AT DALLAS

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

StARLinG LAB

People Publications Projects Software Datasets Blog

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

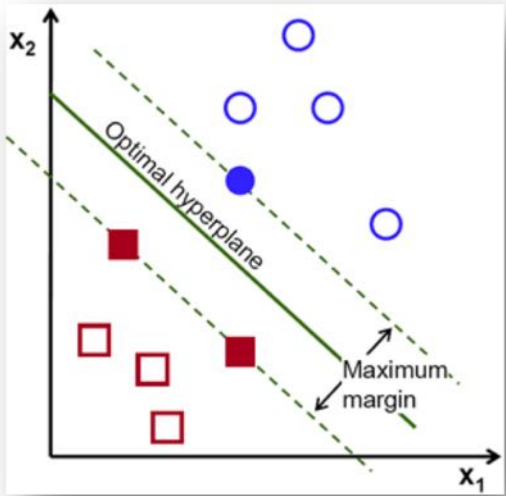
Human-in-the-loop learning

Not every scientist likes to turn math into code

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \mathcal{P}(\mathbf{w},b,\boldsymbol{\xi}) = \frac{1}{2}\mathbf{w}^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to} \begin{cases} \forall i & y_i(\mathbf{w}^\top \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \\ \forall i & \xi_i \geq 0 \end{cases}$$

Support Vector Machines
Cortes, Vapnik MLJ 20(3):273-297, 1995



High-level Languages for Mathematical Programs

TECHNISCHE UNIVERSITÄT DARMSTADT

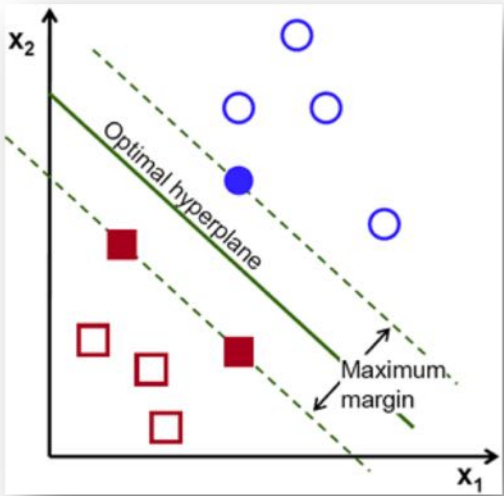
Write down SVM in „paper form.“ The machine compiles it into solver form.

```
#QUADRATIC OBJECTIVE
minimize: sum{J in feature(I,J)} weight(J)**2 + c1 * slack + c2 * coslack;

#labeled examples should be on the correct side
subject to forall {I in labeled(I)}: labeled(I)*predict(I) >= 1 - slack(I);

#slacks are positive
subject to forall {I in labeled(I)}: slack(I) >= 0;
```

Embedded within Python s.t. loops and rules can be used

reloop

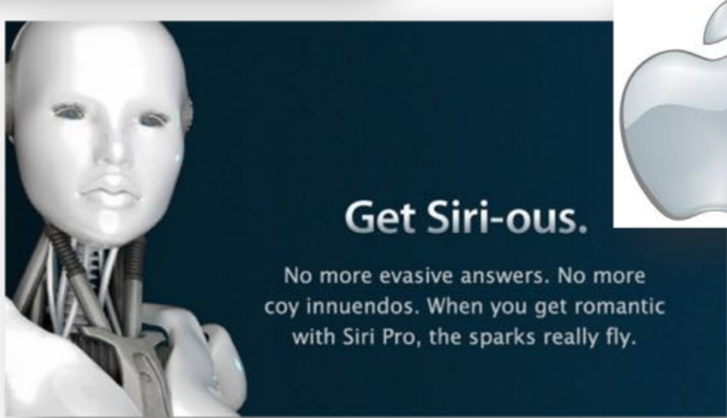
RELOOP: A Toolkit for Relational Convex Optimization



Support Vector Machines
Cortes, Vapnik MLJ 20(3):273-297, 1995

There are strong invests into high-level programming languages for AI/ML

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



Overall, AI/ML/DS indeed refine “formal” science, but …

- **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 ML/AI and high-level programming languages for ML/AI help to capture this complexity and makes using ML/AI simpler
- **Science is more than just Machine Learners and Statisticians**
- **Together, machines and scientists can help making science more reproducible**

A lot left to be done