Positioning LANL for the 3rd wave of Machine Learning

Reid Porter, Kari Sentz CCS-3

9/14/2021

LA-UR-21-29828

The 3rd Wave of Machine Learning

Defense Advanced Research Projects Agency DARPA-BAA-16-53 (2016)

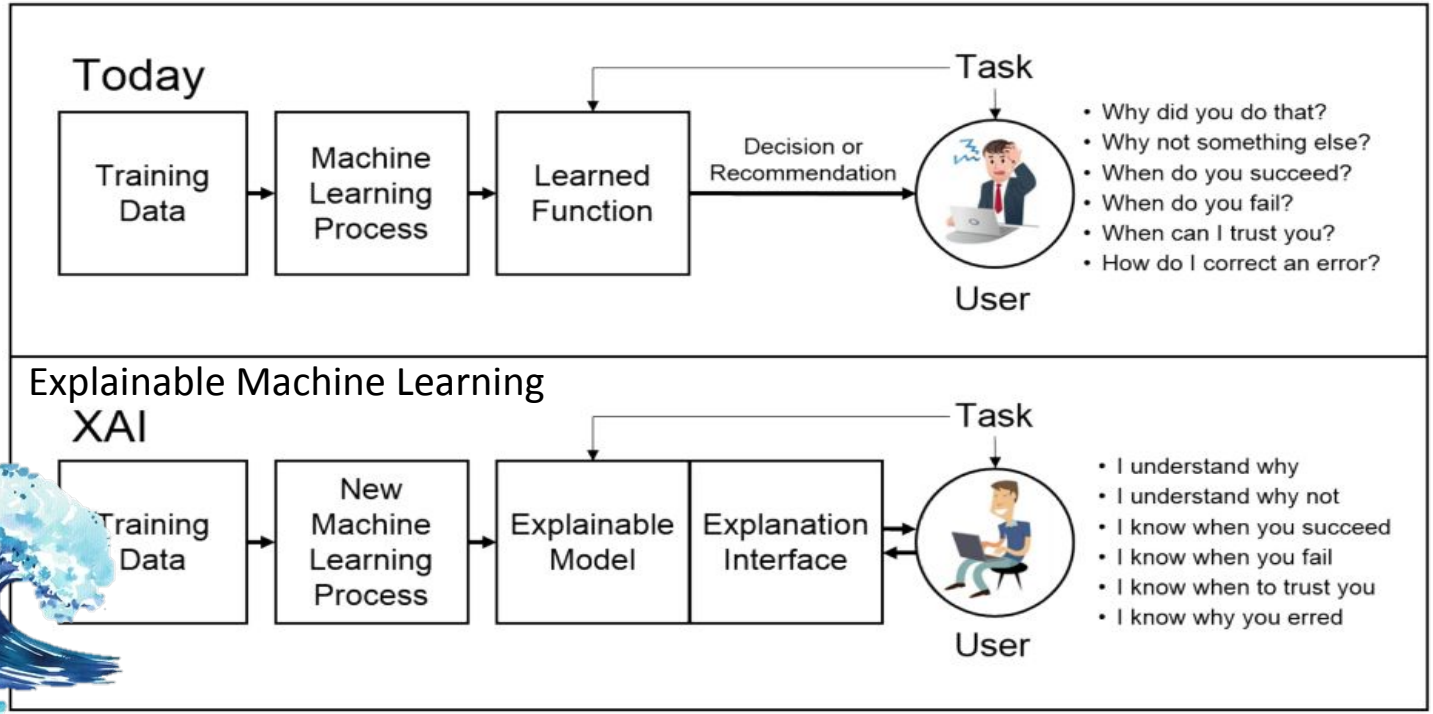


Figure 1: XAI Concept

Machine Learning Fallacies Feeding (Fueled by) XAI

- There is an accuracy / explainability tradeoff….

Machine Learning Fallacies Feeding (Fueled by) XAI

- There is an accuracy / explainability tradeoff…. a fictious plot!

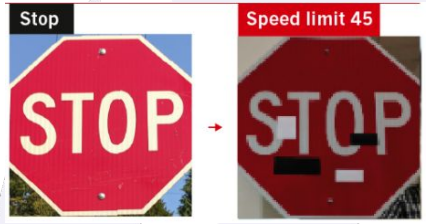
Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition, C. Rudin, Harvard Data Science Review. (2019)

Machine Learning Fallacies Feeding (Fueled by) XAI

- There is an accuracy / explainability tradeoff…. a fictious plot!

Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition, C. Rudin, Harvard Data Science Review. (2019)

- Its surprising that blackbox models are easily fooled...



Machine Learning Fallacies Feeding (Fueled by) XAI

- There is an accuracy / explainability tradeoff…. a fictious plot!

Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition, C. Rudin, Harvard Data Science Review. (2019)

- Its surprising that blackbox models are easily fooled…

…why wouldn't they be?

We don't need to see more vandalized road signs!

Machine Learning Fallacies Feeding (Fueled by) XAI

- There is an accuracy / explainability tradeoff…. a fictious plot!

Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition, C. Rudin, Harvard Data Science Review. (2019)

- Its surprising that blackbox models are easily fooled...

…why wouldn't they be?

We don't need to see more vandalized road signs!

- Post-hoc approximation and visualization of blackbox models helps interpretability.

Machine Learning Fallacies Feeding (Fueled by) XAI

- There is an accuracy / explainability tradeoff…. a fictious plot!

Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition, C. Rudin, Harvard Data Science Review. (2019)

- Its surprising that blackbox models are easily fooled...

…why wouldn't they be?

We don't need to see more vandalized road signs!

- Post-hoc approximation and visualization of blackbox models helps interpretability.

Models of blackbox models are easily fooled too!

Machine Learning Facts

1. Accurate interpretable models already exist in many domains.

   Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, C. Rudin, Nature Machine Intelligence **1**(5), 2019.

2. Accuracy and interpretability requirements are domain specific.

   We should be choosing the model appropriately.

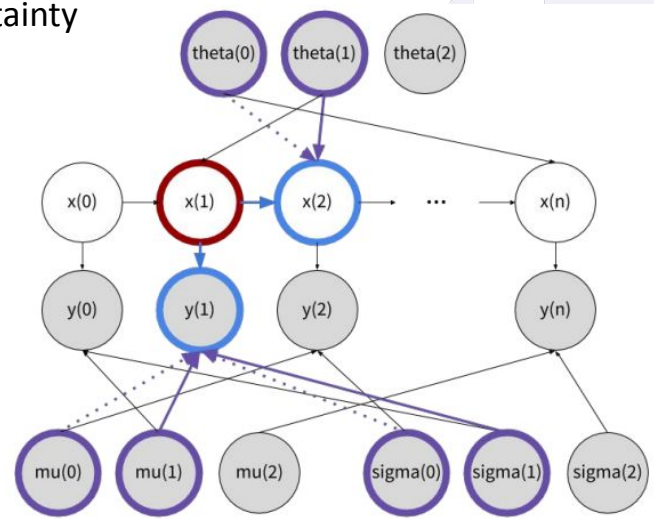
3. Accurate interpretable models can be expensive to construct.

   In terms of the computation and domain expertise required.

Probabilities to the Rescue!

Bayesian Networks

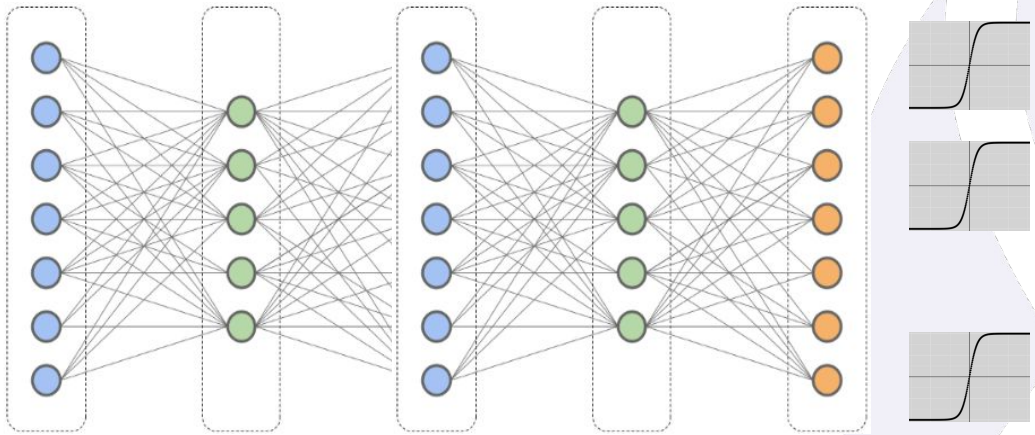
- A framework to handle uncertainty and produce robust models.
- Interpretable by design. (you may need a statistician)
- May not be computable. (may not be trustworthy)



$$\mu_k \sim \text{Normal}(\alpha, \beta)$$
$$\sigma_k \sim \text{Gamma}(\nu, \rho)$$
$$\theta_k \sim \text{Dirichlet}(\kappa)$$
$$x_i \sim \begin{cases} \text{Categorical}(init) & \text{if } i = 0 \\ \text{Categorical}(\theta_{x_{i-1}}) & \text{if } i > 0 \end{cases}$$
$$y_i \sim \text{Normal}(\mu_{x_i}, \sigma_{x_i})$$

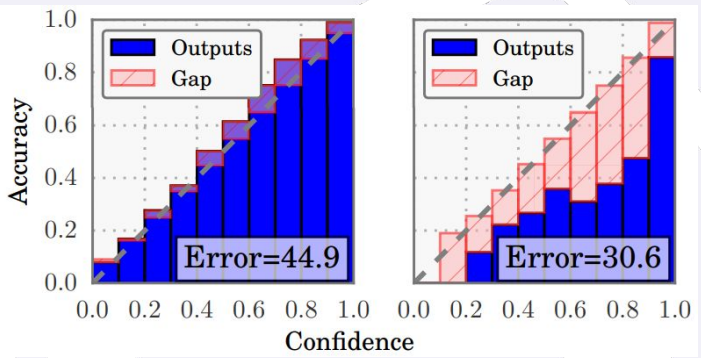
Probabilities and Neural Networks (output)

- Prediction confidence helps interpretability: $p(y|x)$



LeNet (1998)
5 layers

ResNet (2016)
101 layers

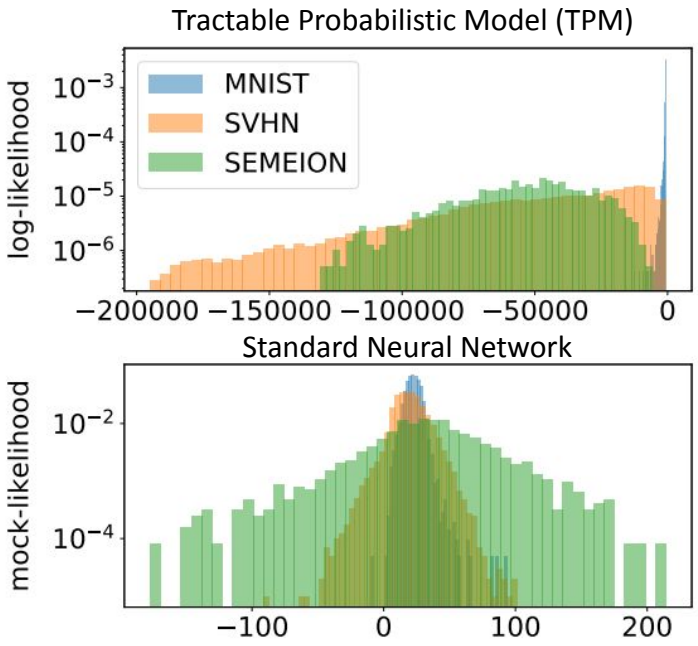


- Data goes in => probabilities come out (at least is explicit!).
- Probabilities may not be trustworthy.

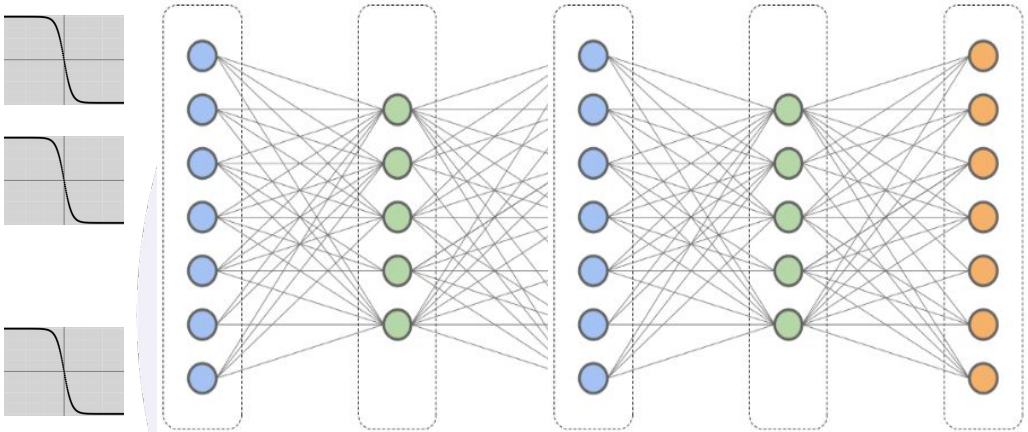
On calibration of modern neural networks, Guo, C., et al. International Conference on Machine Learning, PMLR, 2017.

Probabilities and Neural Networks (input)

- Input probabilities



$p(x)$ tells us what the model was trained on

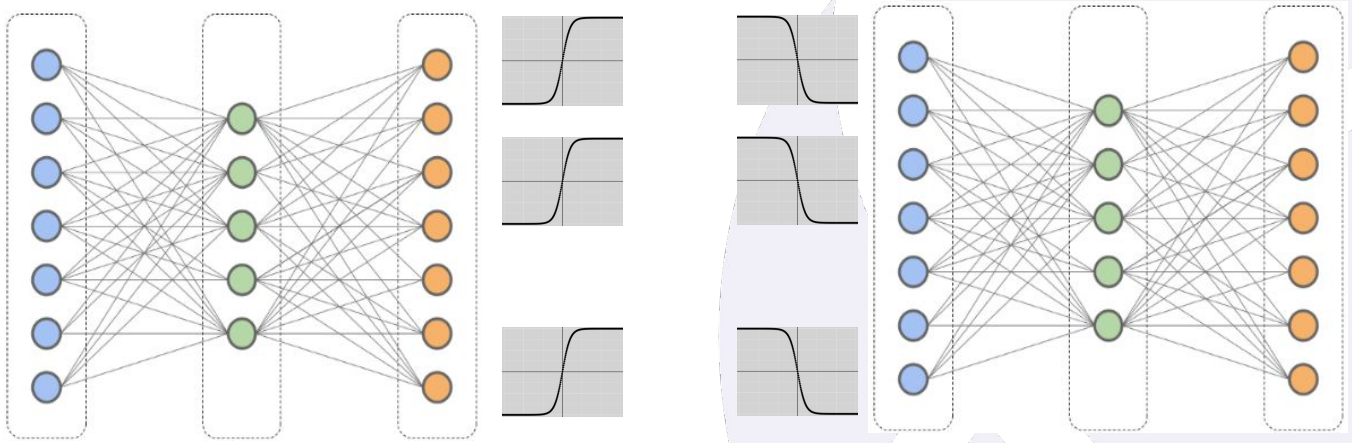


- Networks are trained on MNIST (blue)
- Input probabilities for unseen data SVHN (orange), SEMEION (green)

Peharz, R., et al. (2020). Random sum-product networks: A simple and effective approach to probabilistic deep learning. Uncertainty in Artificial Intelligence, PMLR.

Probabilities and Neural Networks (middle)

- Probabilities in mid-layers $p(z|x)$ $p(z|y)$ represent latent variables.



- Latent variables are comparable and their impact can be quantified.
- Impose desired constraints on latent variables.

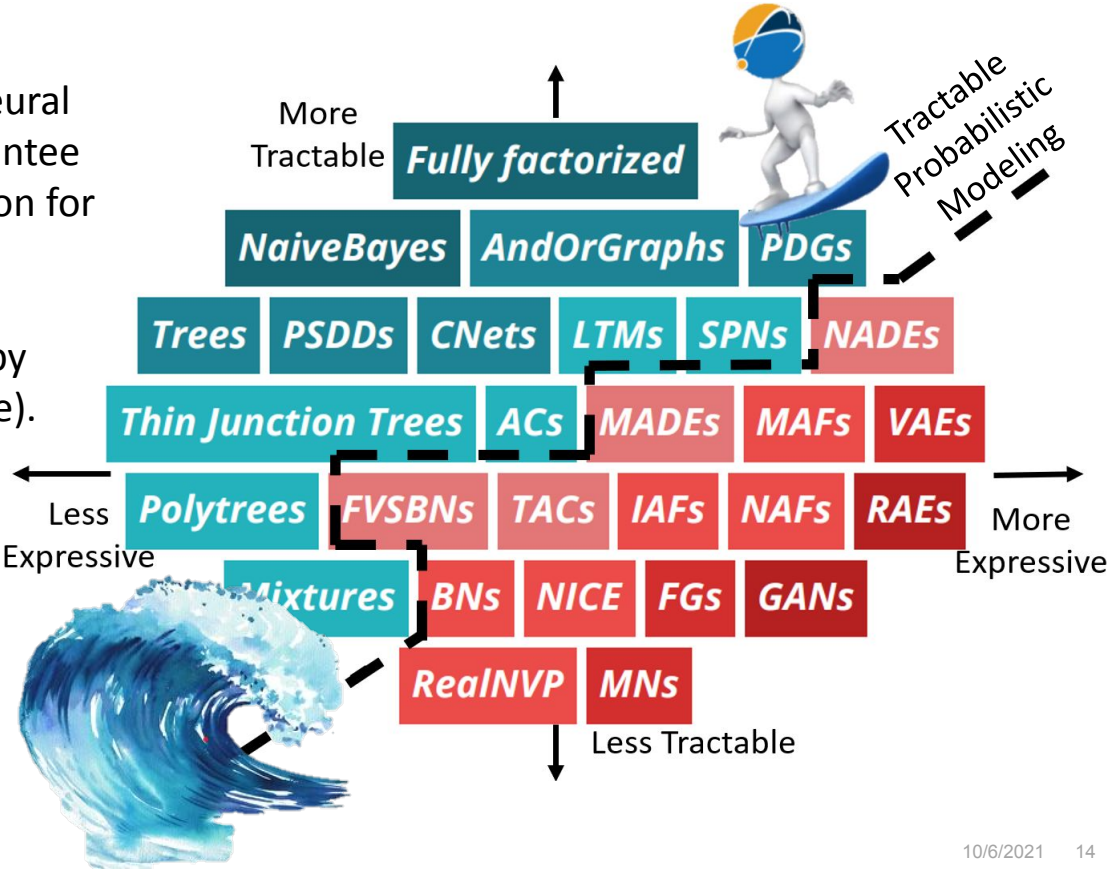
Group Fairness by Probabilistic Modeling with Latent Fair Decisions, YooJung Choi, Meihua Dang, and Guy Van den Broeck, AAAI, 2020.

re-arrest ≠ recidivism

Use a latent "fair label" that enforces demographic parity

Tractable Probabilistic Models

- Tractable Probabilistic Models are Neural Networks with constraints that guarantee an accurate probabilistic interpretation for particular queries.
- Query probabilities are computable by design (trustworthy and interpretable).
- A best of both worlds:
  Probabilities (Bayesian Networks)
  Flexibility (Deep Learning)
- The 3rd Wave of Machine Learning…



1. Accurate interpretable models already exist!

- Why Are We Using Black Box Models in AI When We Don't Need To?
  A Lesson From An Explainable AI Competition, C. Rudin, Harvard Data Science Review., 2019

- Density Estimation Benchmarks
  Probabilistic Circuits: A unifying framework for tractable probabilistic models, Guy Van den Broeck, 2021.

dataset	best circuit	BN	MADE	VAE	dataset	best circuit	BN	MADE	VAE
nltcs	**-5.99**	-6.02	-6.04	**-5.99**	*dna*	**-79.88**	-80.65	-82.77	-94.56
msnbc	**-6.04**	**-6.04**	-6.06	-6.09	*kosarek*	**-10.52**	-10.83	-	-10.64
kdd	-2.12	-2.19	**-2.07**	-2.12	*msweb*	-9.62	-9.70	**-9.59**	-9.73
plants	**-11.84**	-12.65	-12.32	-12.34	*book*	-33.82	-36.41	-33.95	**-33.19**
audio	-39.39	-40.50	-38.95	**-38.67**	*movie*	-50.34	-54.37	-48.7	**-47.43**
jester	-51.29	**-51.07**	-52.23	-51.54	*webkb*	-149.20	-157.43	-149.59	**-146.9**
netflix	-55.71	-57.02	-55.16	**-54.73**	*cr52*	-81.87	-87.56	-82.80	**-81.33**
accidents	-26.89	**-26.32**	-26.42	-29.11	*c20ng*	-151.02	-158.95	-153.18	**-146.9**
retail	**-10.72**	-10.87	-10.81	-10.83	*bbc*	**-229.21**	-257.86	-242.40	-240.94
*pumbs**	-22.15	**-21.72**	-22.3	-25.16	*ad*	-14.00	-18.35	**-13.65**	-18.81

TPM or BN Deep Network

2. Accuracy and Interpretability Requirements (should) Drive Application Specific Model Development

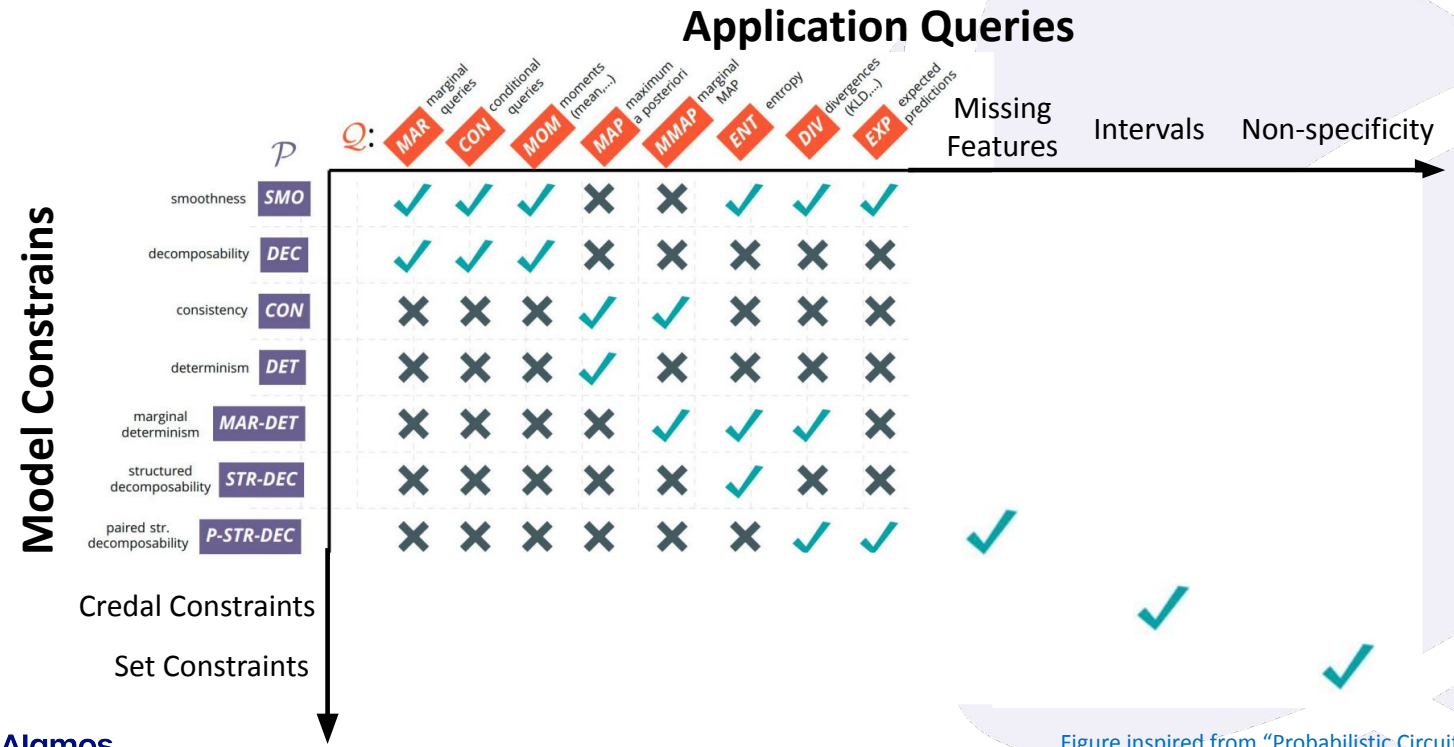
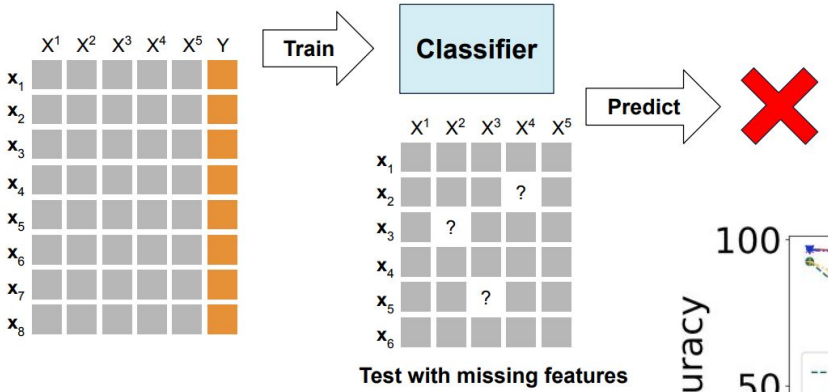


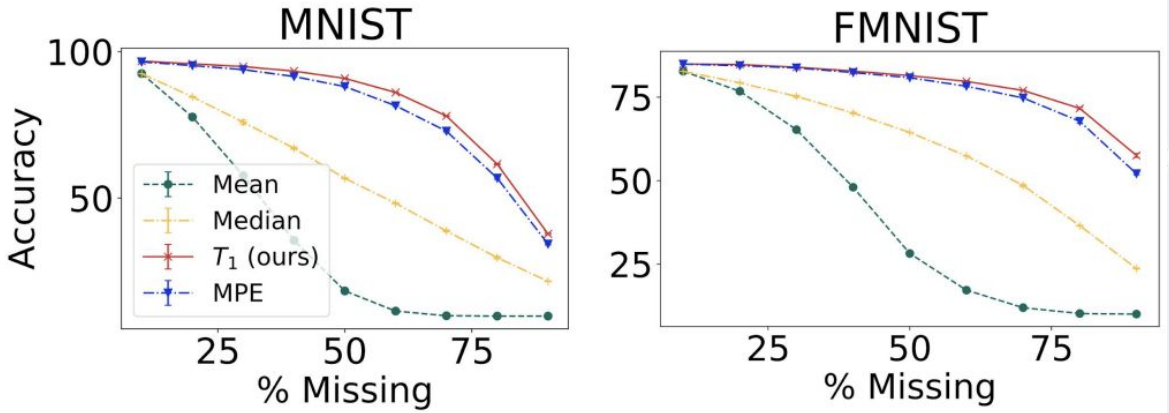
Figure inspired from "Probabilistic Circuits: A unifying framework for tractable probabilistic models", Guy Van den Broeck, 2021.

Example: Prediction with Missing Features

- Set up the model in such a way that it can compute Expectations of the classifier.
- The are few additional constraints over models used for learn the data distribution.

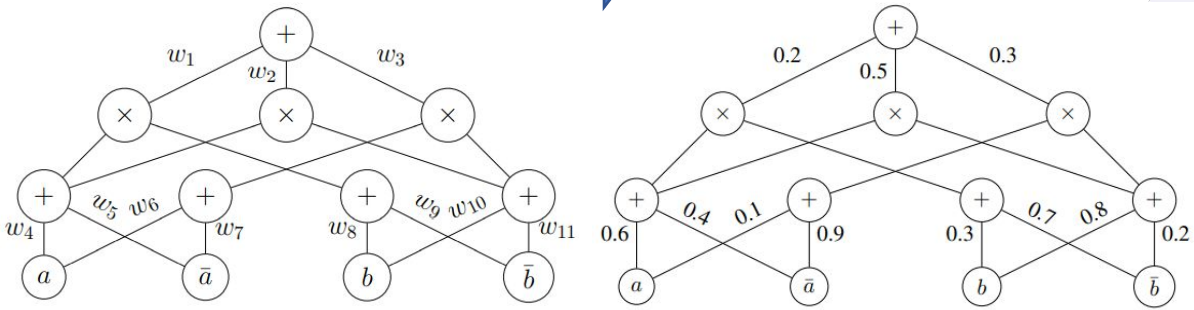


Guy Van den Broeck et. al. (2020). On Tractable Computation of Expected Predictions, ICML.



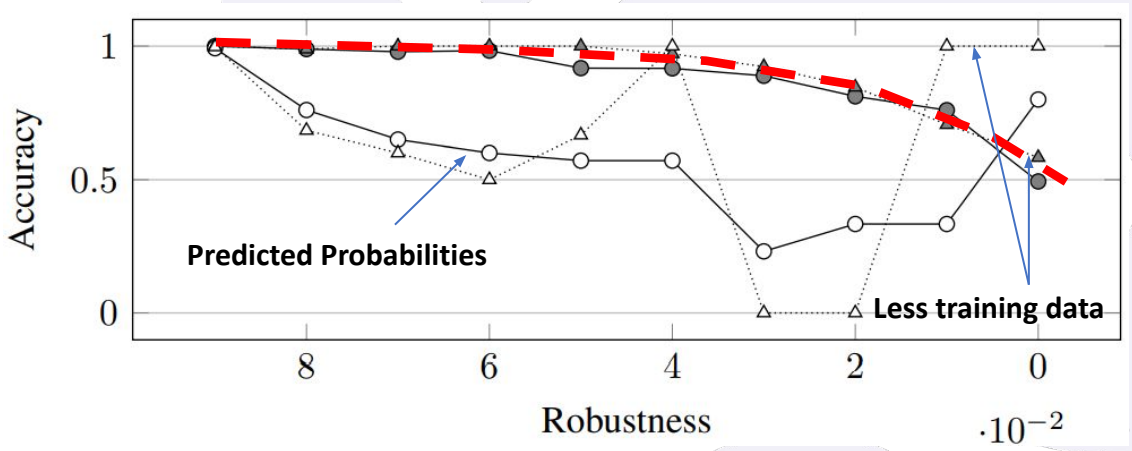
Example: Prediction with Model Uncertainty

Deep Network → Trained Parameters → $\in -admissability$



$$0.54 \leq w_4 \leq 0.64, \quad 0.36 \leq w_5 \leq 0.46,$$
$$0.09 \leq w_6 \leq 0.19, \quad 0.81 \leq w_7 \leq 0.91,$$
$$0.27 \leq w_8 \leq 0.37, \quad 0.63 \leq w_9 \leq 0.73,$$
$$0.72 \leq w_{10} \leq 0.82, \quad 0.18 \leq w_{11} \leq 0.28,$$

Model propagates intervals



Mauá, Denis D., et al. "Credal sum-product networks." Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications. PMLR, 2017.

3. Accurate interpretable models can be expensive to construct.

- Target applications that involve high stakes decisions and/or limited and imperfect data…

3. Accurate interpretable models can be expensive to construct.

- Target applications that involve high stakes decisions and/or limited and imperfect data… that's almost every LANL application!

3. Accurate interpretable models can be expensive to construct.

- Target applications that involve high stakes decisions and/or limited and imperfect data… that's almost every LANL application!

- Target applications that have funding!

3. Accurate interpretable models can be expensive to construct.

- Target applications that involve high stakes decisions and/or limited and imperfect data… that's almost every LANL application!
- Target applications that have funding!
- Data Analysis (CCS-3)

  Image and signal processing, text and scientific data analysis.
- Design of Experiments (CCS-6)

  Computational (simulations) and Scientific (data collection).

NA-22 new start project that has a need for both!

Robust exploration of multi-faceted morphologic signatures of actinide process materials for nuclear forensic science, Kari Sentz (PI) ~$1M/year for 3 years.

- Image and Data Analysis (Kari Sentz, Reid Porter, Ian Schwerdt, Cole Thompson)
- Design of Data Collection Experiments (Christine Anderson Cook, Tom Burr)

NA-22 new start project that has a need for both!

Robust exploration of multi-faceted morphologic signatures of actinide process materials for nuclear forensic science, Kari Sentz (PI) ~$1M/year for 3 years.

- A Big Space, Small Data problem:

Data Collection Space (simplified)

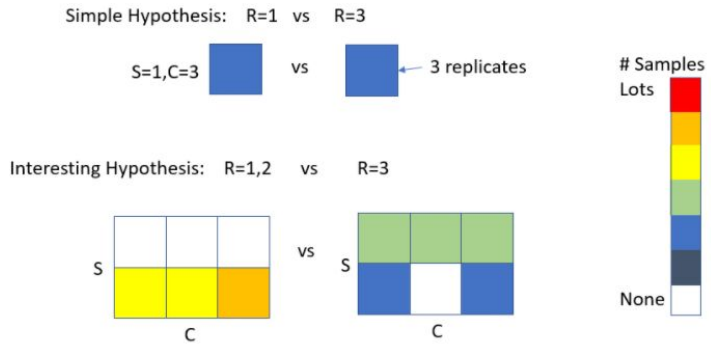
R: Route	1	2	3	4	5
C: Calcination	1	2	3		
A: Aging	1	2	3		
S: Strike	1	2			

Hypothesis Space (simplified)

Do sample sets 1 and 2 come from the same distribution?

Simple Hypothesis	More Interesting Hypothesis
Set 1: { R=1 , S=1 , C = 3 }	Set 1: { R=1:2 , S=1:2 , C = 1:3 }
Set 2: { R=3 , S=1 , C = 3 }	Set 2: { R=3 , S=1:2 , C = 1:3 }

Sample Histograms (w.r.t. Hypothesis Space)

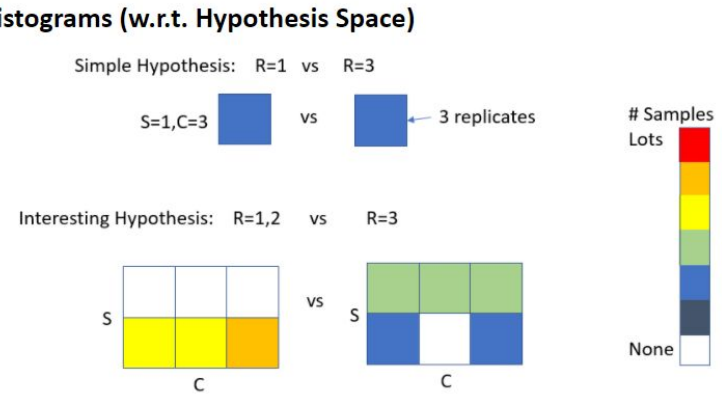


NA-22 new start project that has a need for both!

Robust exploration of multi-faceted morphologic signatures of actinide process materials for nuclear forensic science, Kari Sentz (PI) ~$1M/year for 3 years.

- A Big Space, Small Data problem:

- **What hypotheses are the most interesting in terms of forensics?**
  Not all hypothesis sets are interesting
  Are there relationships in the data collection space that can be exploited?
- **What hypothesis space has the most "support" in terms of data?**
  It might not be that interesting, but we can at least be confident in the answers!
- **What data to collect next? (aka Design of Experiments)**



- TPMs provide the methodology to develop Accurate Interpretable models.

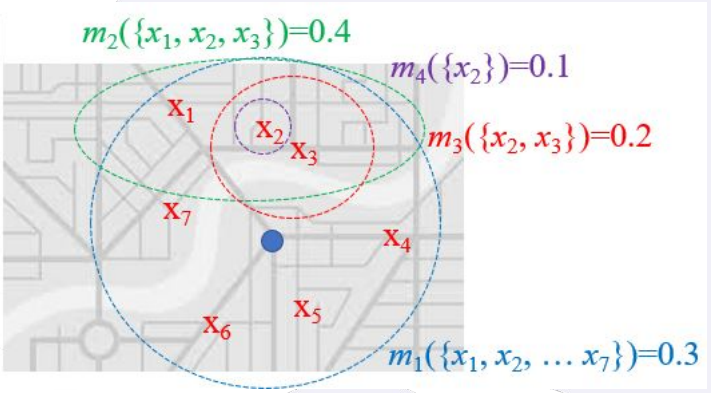
Application Development for High Stakes Decisions

Interactive Geospatial-Temporal Reasoning for Robust Strategic Deterrence, Kari Sentz (PI), Directors Strategic Research Initiative proposal, Sept. 2021.

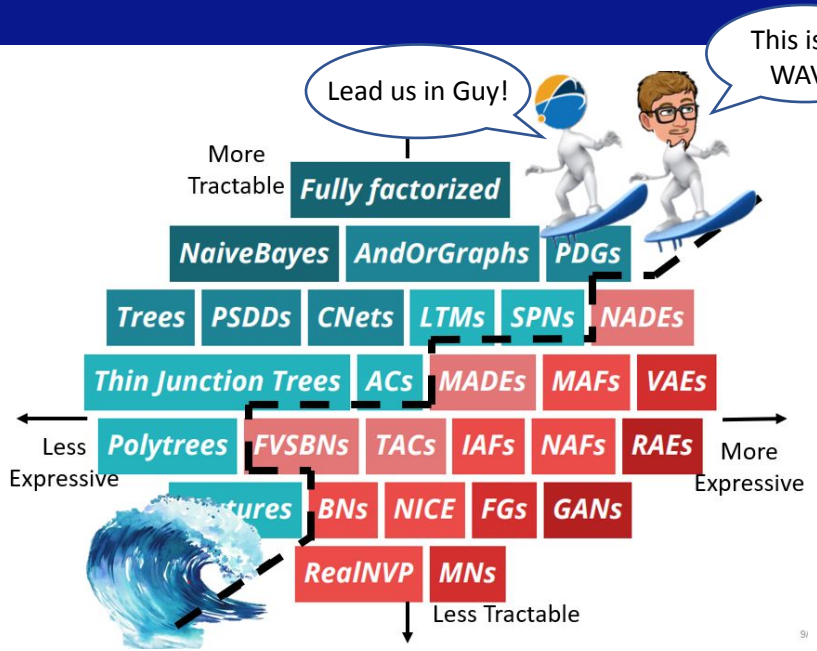
- Deterrence has unique challenges in non-specificity and uncertainty.
- TPMs provide the methodology to tailor the model to the domain:
  - Flexibility of Deep Networks to maximize the data available.
  - Additional constraints to meet interpretable and verifiable requirements.

Order of evidence	Subset of S	m	$\underline{P}$	$\overline{P}$	$\overline{P} - \underline{P}$	p
1	$\{x_1, x_2, \ldots x_7\}$	0.3	0.3	1	0.7	
2	$\{x_1, x_2, x_3\}$	0.4	.7	1	0.3	
3	$\{x_2, x_3\}$	0.2	.9	1	0.1	
4	$\{x_2\}$	0.1	1	1	0	0.38*

* Assumes a uniform probability distribution across sets



Positioning LANL for the 3rd wave of Machine Learning



Tractable Probabilistic Models (TPMs) push the black boundary to the right and provide a general methodology to produce accurate interpretable models that can solve Verifiable AI challenges.

PI: Reid Porter, Kari Sentz
Total Project Budget: $40k
ISTI Focus Area: Artificial Intelligence

Project Description

Assessment of current TPM capabilities with respect to LANL applications and verifiable AI challenges. Establish collaborations with researchers at the forefront of TPM research. Identify LANL applications that stand to benefit the most from TPMs (the funded ones). Increase awareness through proposals and presentations.

Project Outcomes

Propaganda: Expanded versions of this talk.

Proposals:

- "Robust exploration of multi-faceted morphologic signatures of actinide process materials for nuclear forensic science", Kari Sentz (PI), NA-22 new start, June 2021, ~$1M/year for 3 years (ROI).
- "Interactive Geospatial-Temporal Reasoning for Robust Strategic Deterrence", Kari Sentz PI, Directors Strategic Research Initiative Proposal, Sept. 2021.
- "Statistically Defensible Deep Learning", Kari Sentz PI, LDRD-DR Preproposal #202200027DR (not funded).

Outreach:

- Established CCS-3/6 collaboration on design of experiments.
- ISTI Seminar Series: Guy Van den Broeck (UCLA), Fabio Gozman (Sao Paulo), Chris Tosh (Columbia), TBD Cynthia Rudin (Duke).

END