Warwick Business School
Gillmore Centre for Financial Technology
Investment AI and Machine Learning Symposium
3 April 2020

Continual Learning and Explainable AI through Knowledge Extraction from Deep Networks

Prof Artur d'Avila Garcez, FBCS
City, University of London
a.garcez@city.ac.uk

AI revolution mainly due to...

… deep learning (a form of ML): making predictions from data

State-of-the-art at image recognition, speech/audio analysis, games, language modeling and translation, and to some extent question answering and video understanding

- But with limitations (AI is more than ML): black box (trust), lack of robustness (extrapolation)

Neural-Symbolic AI

Systems that learn from data but also reason about what has been learned (data + knowledge)

Research since late 1990s c.f. www.neural-symbolic.org now applicable in practice

Combines neural networks with (rule-based) symbolic AI to achieve reasoning and explainable AI

Taking advantage of data-driven ML and knowledge-based AI (i.e. logical rules)

e.g. self-driving cars: 10 billion miles of driving data (Waymo) but there is also the highway code!

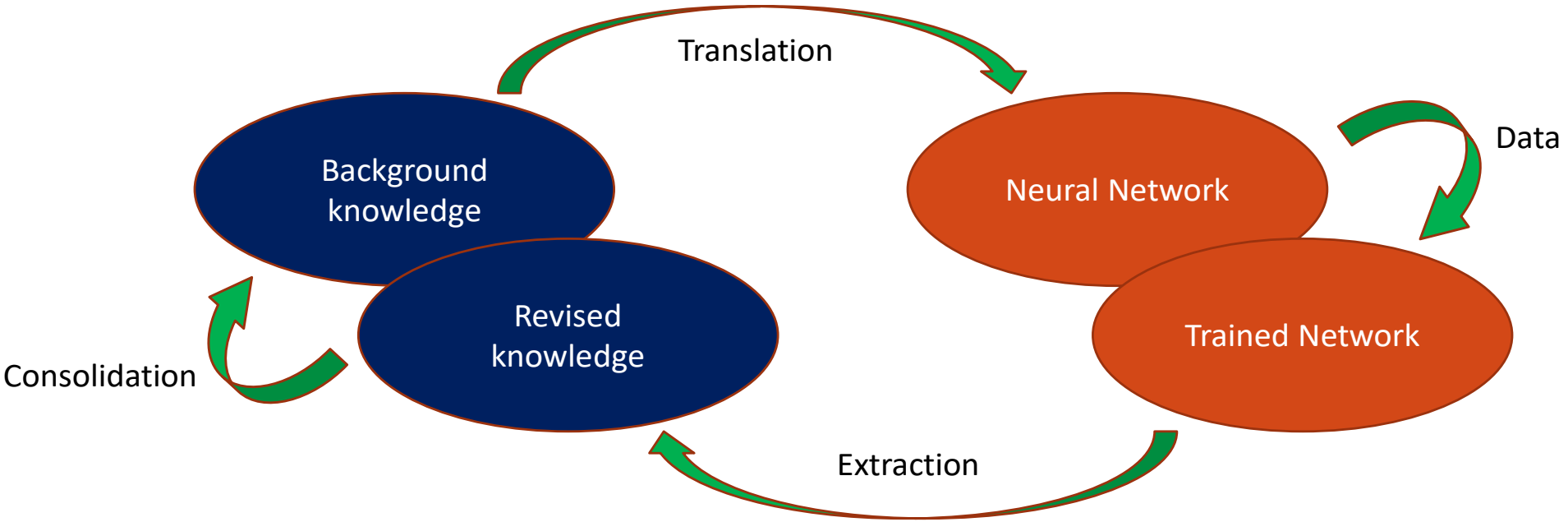
Two separate research communities: Brain/Mind dichotomy

Symbolic AI: a symbol system has all that is needed for general intelligence

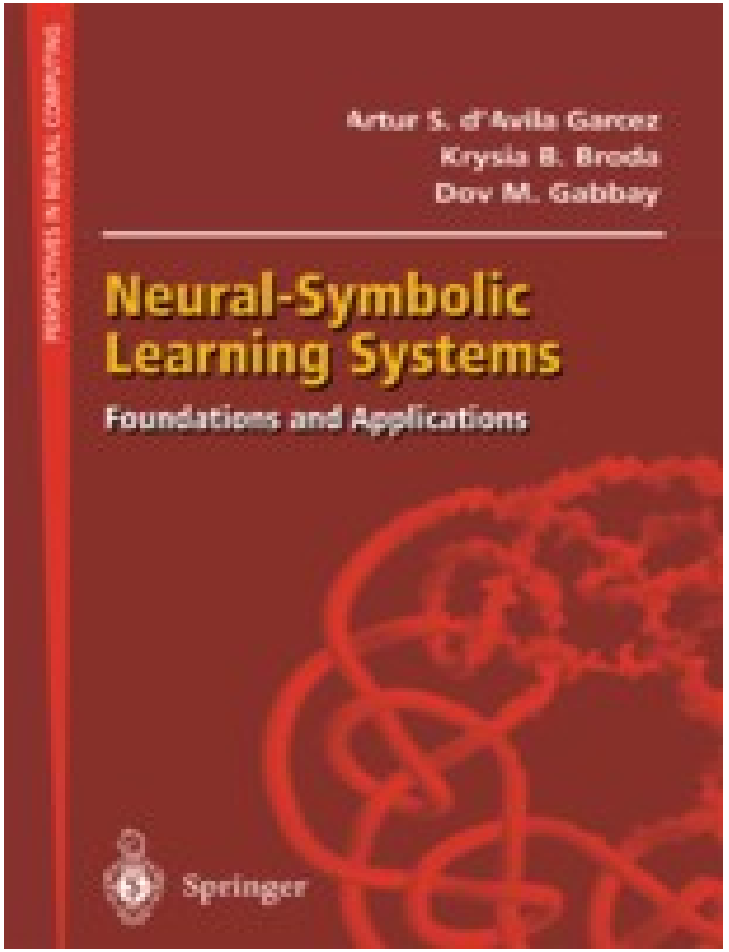
Sub-symbolic AI: intelligence emerges from the brain (neural networks)

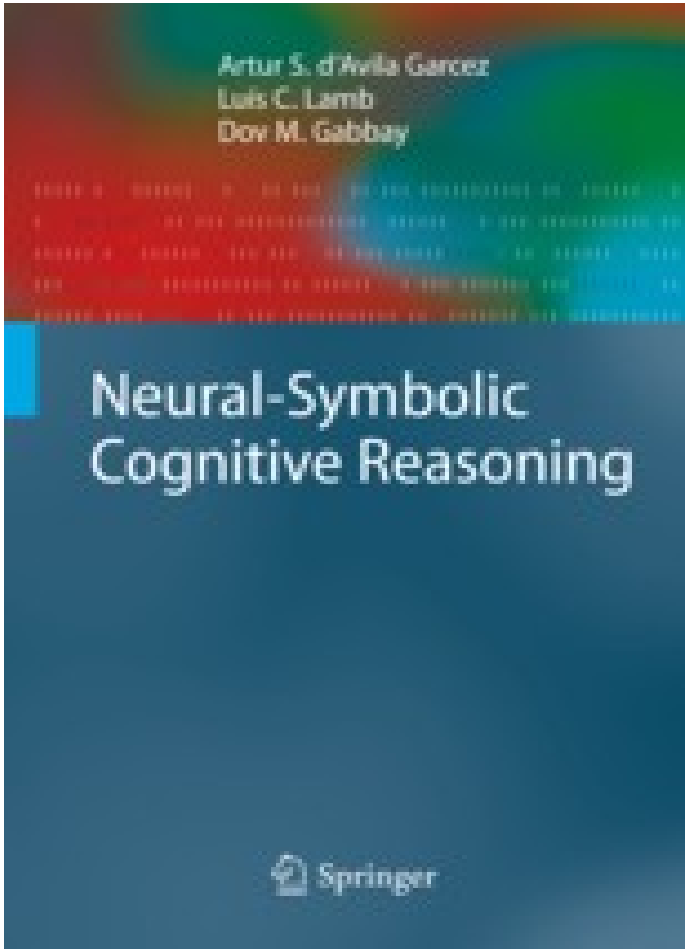
Neural-symbolic systems bring together the two traditions, e.g. a neural network that *knows when it doesn't know*

Neural-Symbolic Cycle

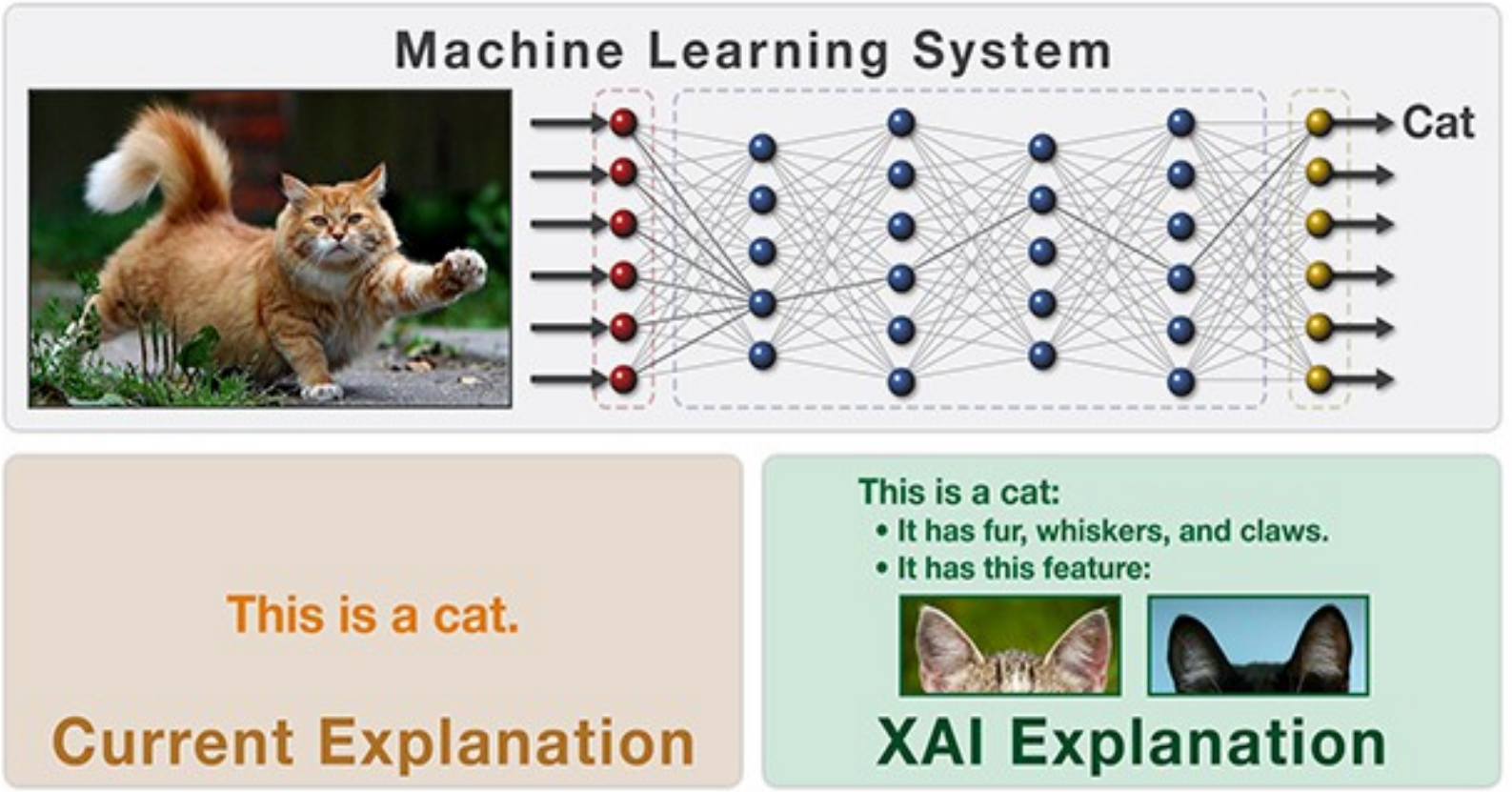


For more information...





Different forms of explanation e.g. DARPA's Explainable AI (XAI)



- XAI = Interpretable ML
- Explanation = knowledge extraction, not XAI

Practical Examples...

Consumer protection (collaboration with Playtech plc)

Transaction data + regulatory framework

Predict whether a player might be at risk of harm and should take a break for a period of time

Anti money laundering (collaboration with Kindred)

Also a case of data + rules (e.g. KYC)

Data imbalance (matters a lot in practice); credit card fraud another well-known example

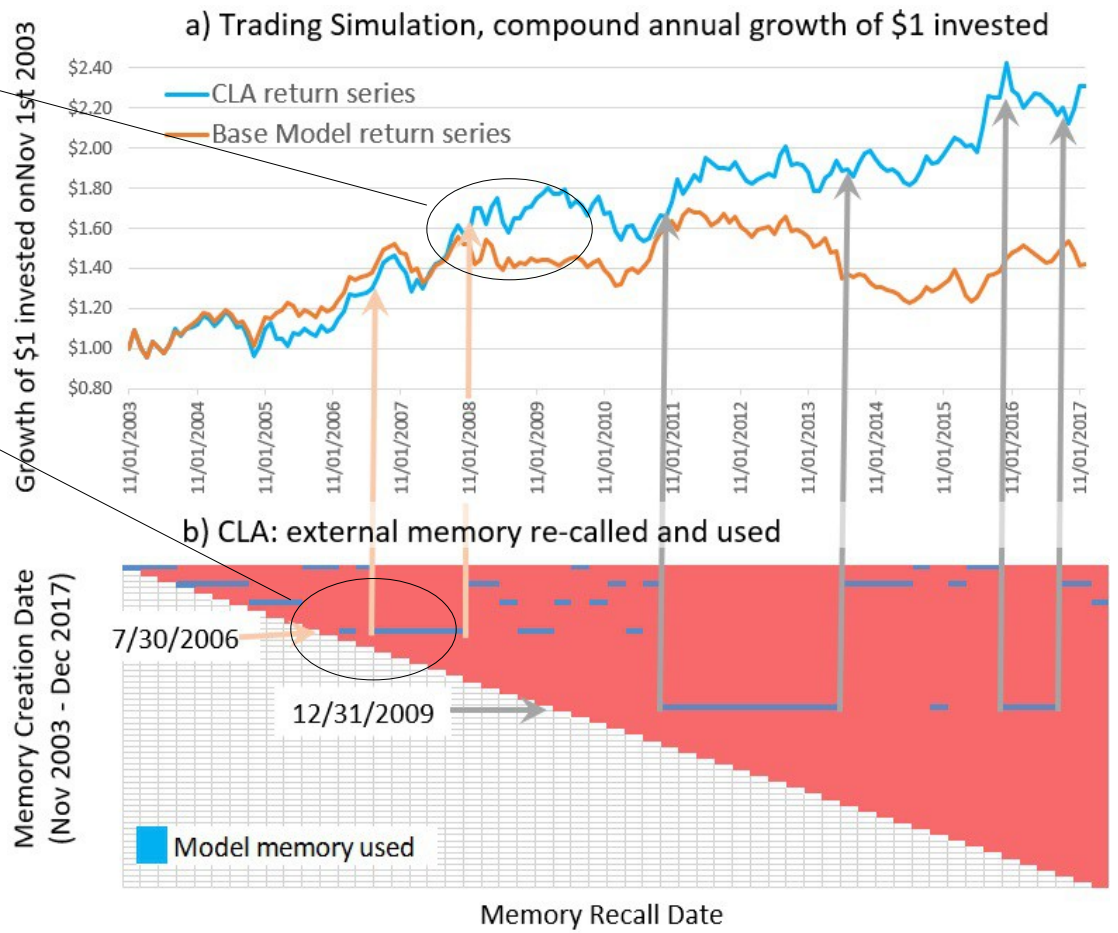
Rules can help adjust and understand relationships between false positive and false negative cases

C. Charitou, A. d'Avila Garcez and S. Dragicevic. Semi-supervised GANs for Fraud Detection. In Proc. IEEE International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, UK, July 2020.

Investment decisions (Continual Learning Augmentation)

Lehman Brothers 2008

Quant Quake?



Time series trained by distributed neural nets

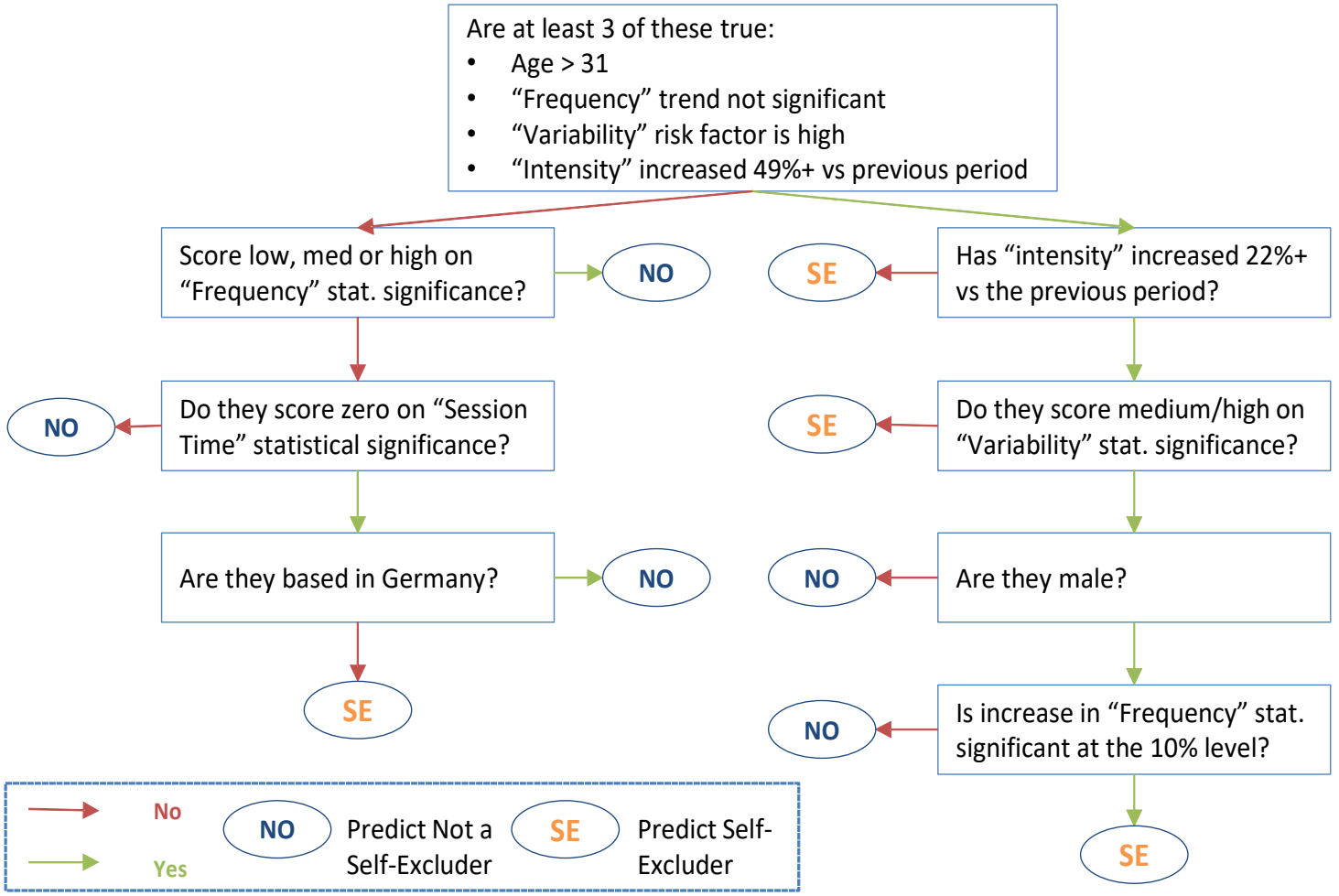
Memory recall explained by localist (if-then) rules

D. Philps, T. Weyde, A. d'Avila Garcez, R. Batchelor. Continual Learning Augmented Investment Decisions. NeurIPS 2018 Workshop on AI in Finance, Montreal, Dec 2018

Collaboration with Playtech on XAI

- Neural net or random forest give a higher accuracy than Bayesian net or regression model
- Predict self-exclude based on transaction data: frequency of play, betting intensity, variation, ratio of night play, etc.
- Self-exclusion data used as a proxy for harm
- Query the black box to extract a decision tree for explanation or to improve the prediction

Knowledge Extraction



C. Percy, A. S. d'Avila Garcez, S. Dragicevic, M. Franca, G. Slabaugh and T. Weyde. The Need for Knowledge Extraction: Understanding Harmful Gambling Behavior with Neural Networks, In Proc. ECAI 2016, The Hague, September 2016.

For every complex problem there is an answer that is clear, simple, and wrong

H. L. Mencken

Measuring the fidelity of the knowledge extraction to the original black box ML is key!

If the system is too complex and a global XAI method offers a low fidelity then using a local explanation approach may be the solution

The need for a measure of fidelity

- Sound/complete knowledge extraction may be intractable computationally
- In practice, efficient extraction may be unsound (and work more like a learning algorithm)
- Soundness/completeness needed if neural net is used for decision making in a safety-critical domain
- Very relevant to recent efforts on verification of neural networks (EPSRC project proposal in collaboration with JP Morgan)

So, Knowledge Extraction is needed

From a regulatory perspective,

But also to help improve system performance (*learn from your mistakes*), and

increase consumer confidence (in future, consumers will choose the systems that they can trust), and

for data/energy efficiency (*e.g. will the Waymo self-driving car require 10 billion miles worth of data to learn to drive in London?*)

Not to mention Ethics of AI (c.f. *Are they male?*)

Knowledge Extraction techniques

- Decision trees from feedforward networks
- Graphs from recurrent networks
- Counterfactual explanations (local explanations of an instance/case, as alternative to global explanations of the entire ML model)

  E.g.: if *your* salary were to increase by 20%, *you* would have been successful with the application for credit all else being equal

A. White and A. d'Avila Garcez, Measurable Counterfactual Local Explanations for Any Classifier, Proc ECAI 2020, Spain, September 2020. https://arxiv.org/pdf/1908.03020.pdf

Knowledge extraction enables...

- Continual learning
- Reasoning (about what has been learned)
- Explanation

Knowledge Extraction algorithms

Pedagogical: treat network as an oracle to query input/output patterns

Decompositional: inspect the internal structure of the network

Eclectic: consider doing both of the above

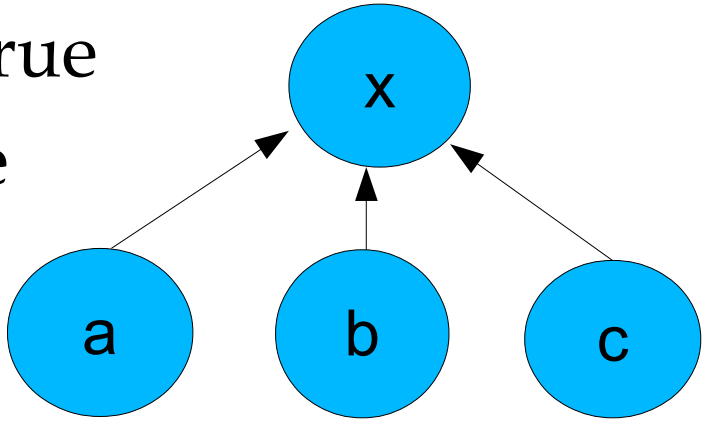
Reasoning: M of N rules

- Neural nets are very good at learning/ representing MofN rules:

  If 2 of (a,b,c) are True then x is True

  If 1 of (a,b) is False then x is True

  etc.

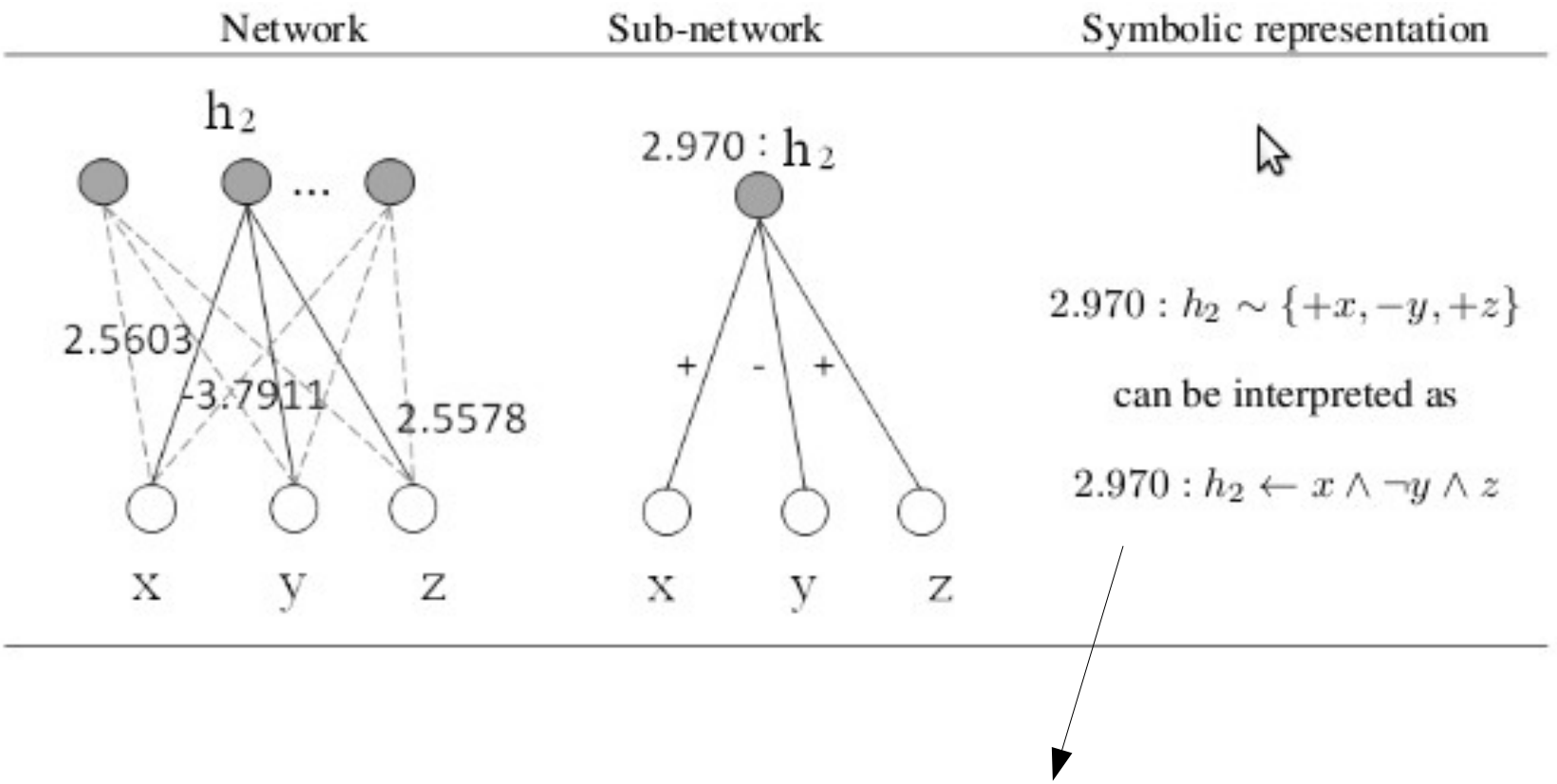


Knowledge-based artificial neural networks, G. Towell and J. Shavlik, AIJ, 1994

Symbolic knowledge extraction from trained neural networks: A sound approach, A. d'Avila Garcez, K. Broda, D. Gabbay, AIJ, 2001.

Reasoning: Confidence Rules

Knowledge extraction from RBMs (building block of Hinton's deep networks)



Each rule has a confidence value ∑||w||/n

Probabilistic MofN

We can improve the accuracy of rules extracted from RBMs by extracting MofN rules

Search values for M given extracted rules, e.g. M=0,1,2,3 in

$$2.970 : h_2 \leftarrow \text{M of } \{x, \sim y, z\}$$

- Simon Odense and Artur S. d'Avila Garcez. Extracting M of N Rules from Restricted Boltzmann Machines, ICANN 2017.
- S. Tran and A. S. d'Avila Garcez. Deep Logic Networks: Inserting and Extracting Knowledge from Deep Belief Networks. IEEE Transactions NNLS, November 2016

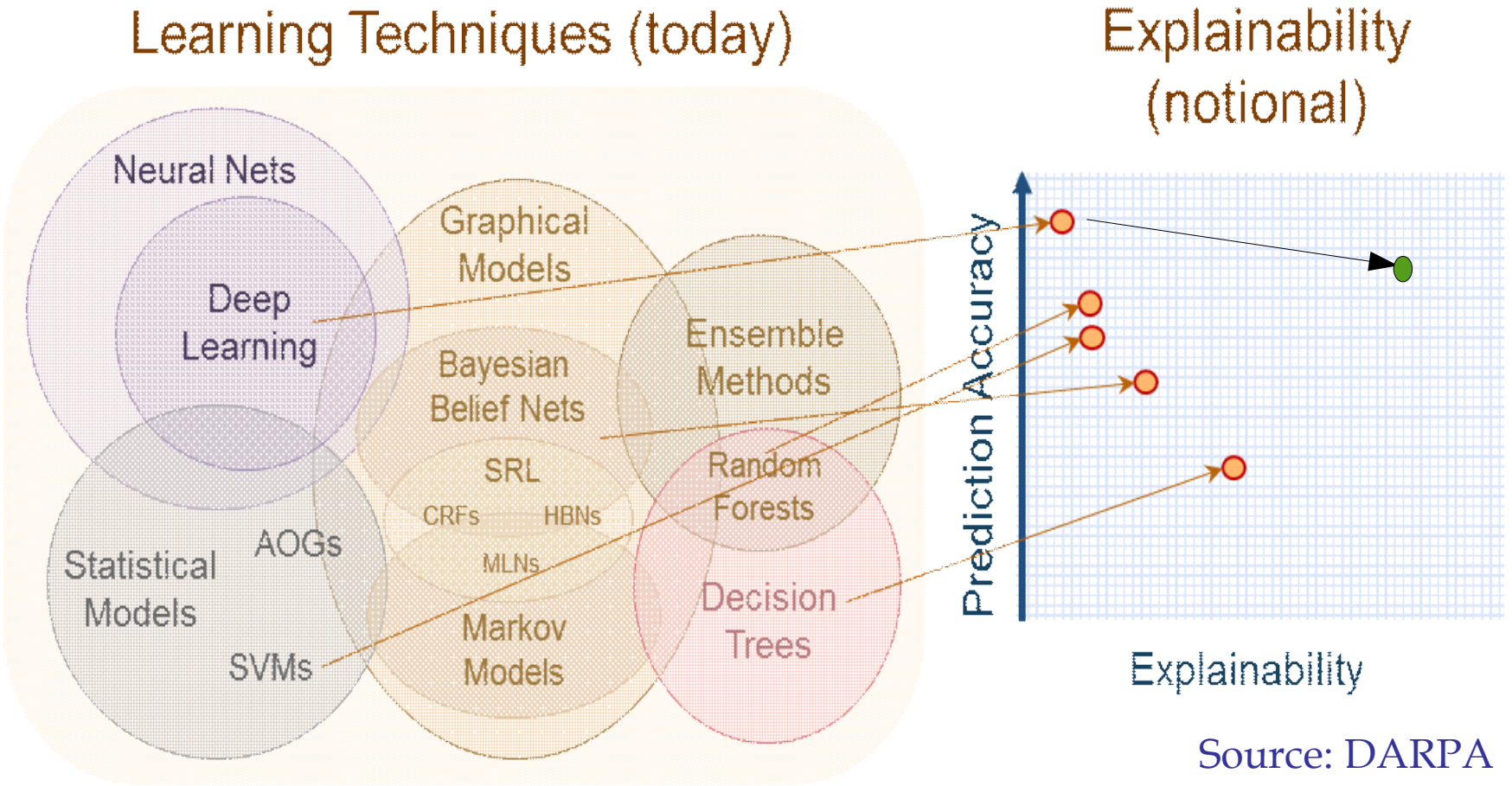
Risk curves + Interventions

Make this sharp drop smoother?

Different interventions by different stakeholders (e.g. hospital or insurance company)

- S. Dragicevic, A. d'Avila Garcez, C. Percy and S. Sarkar. Understanding the Risk Profile of Gambling Behaviour through Machine Learning Predictive Modelling and Explanation. In Proc NeurIPS 2019, KR Meets ML Workshop, Vancouver, Canada, Dec 2019.

Explainable AI = ML + KR



Source: DARPA

The City Data Science Institute

https://www.city-data-science-institute.com/

- Machine Learning (Computer Science)
- Computer Vision (Engineering)
- Complex Networks (Mathematics)
- Data Visualization (Computer Science)
- Finance (CASS Business School)
- Healthcare (School of Health Science)
- Economics (School of Arts and Social Sciences)
- Ethics (The City Law School)

Partners: NHS, BBC, British Library, Imperial College Data Science, Societe Generale, Chinese Academy of Sciences, Delta Capita, Rothko, Telefonica Alpha

Conclusion: Neural-Symbolic Systems

To study the statistical nature of learning and the logical nature of reasoning.

To provide a unifying foundation for robust learning and efficient reasoning.

To develop effective computational systems for AI applications.

Thank you!