Machine Learning Nonlocal Correlations

INTERNATIONAL INSTITUTE OF PHYSICS
Federal University of Rio Grande do Norte

QiQM

Quantum Information and Quantum Matter Group
www.iip.ufrn.br/qiqm

Rafael Chaves

***QML+, Innsbruck,* September 2018**

serrapilheira

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

JOHN TEMPLETON FOUNDATION

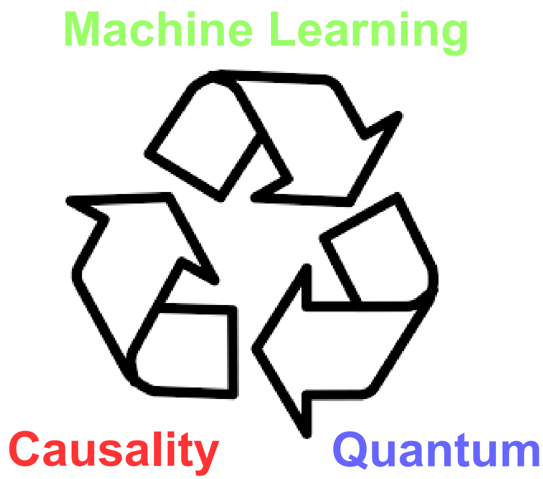
Ministério da Educação
GOVERNO FEDERAL
BRASIL

Few words about Natal...

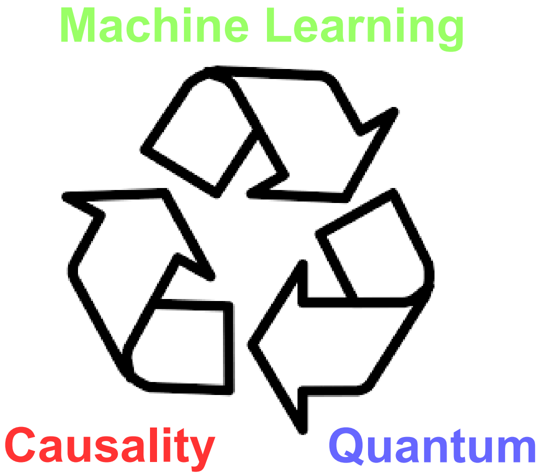
- 10 - 12 events every year

  - Quantum Thermo (2019)
  - Quantum Correlations (2019)
  - Quantum Info & Gravity (2020)
  - Causality & Machine Learning (2020)

What is this talk about?

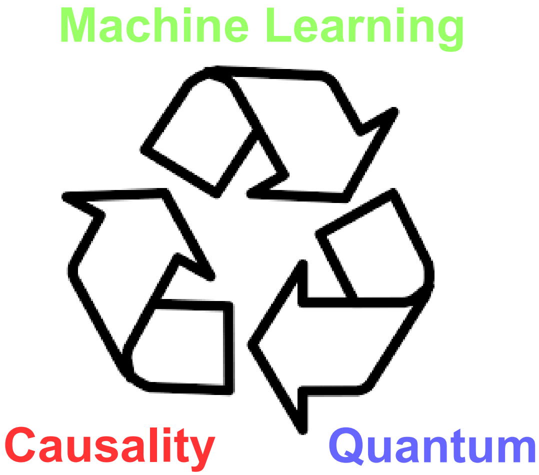


- Generalizations of Bell's theorem
  [Fritz NJP 2012], [Chaves PRL 2016]...
- Quantum advantages in causal problems
  [Ried et al NatPhys 2015], [Chaves et al NatPhys 2018]...
- Tools from one field applied in the other
  [Chaves et al UAI 2014], [Chaves & Budroni PRL 2016], [Lee & Spekkens arxiv 2017]
- Revisiting foundational problems
  [Rossi PRA 2018], [Chaves, Lemos, Pienaar PRL 2018]
- Machinery to derive physical principles/Quantum limits
  [Chaves, Majenz, Gross NatComm 2015], [Chaves, Brask, Brunner PRL 2015]



- Generalizations of Bell's theorem
  [Fritz NJP 2012], [Chaves PRL 2016]...
- Quantum advantages in causal problems
  [Ried et al NatPhys 2015], [Chaves et al NatPhys 2018]...
- Tools from one field applied in the other
  [Chaves et al UAI 2014], [Chaves & Budroni PRL 2016], [Lee & Spekkens arxiv 2017]
- Revisiting foundational problems
  [Rossi PRA 2018], [Chaves, Lemos, Pienaar PRL 2018]
- Machinery to derive physical principles/Quantum limits
  [Chaves, Majenz, Gross NatComm 2015], [Chaves, Brask, Brunner PRL 2015]

- Causal inference as a ML problem
- Counterfactual reasoning (AI)
  [Pearl "Causality" 2009]
- ML as a tool to discover causal relations from data (generative ML)
  [Goudet et al arxiv 2017]

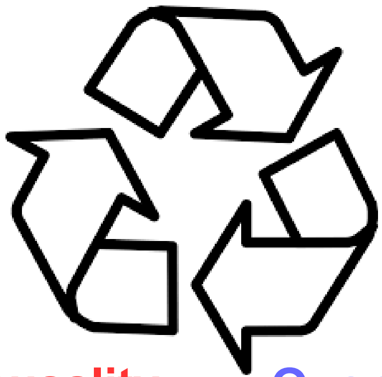


- Generalizations of Bell's theorem
  [Fritz NJP 2012], [Chaves PRL 2016]...
- Quantum advantages in causal problems
  [Ried et al NatPhys 2015], [Chaves et al NatPhys 2018]...
- Tools from one field applied in the other
  [Chaves et al UAI 2014], [Chaves & Budroni PRL 2016], [Lee & Spekkens arxiv 2017]
- Revisiting foundational problems
  [Rossi PRA 2018], [Chaves, Lemos, Pienaar PRL 2018]
- Machinery to derive physical principles/Quantum limits
  [Chaves, Majenz, Gross NatComm 2015], [Chaves, Brask, Brunner PRL 2015]

- Causal inference as a ML problem
- Counterfactual reasoning (AI)
  [Pearl "Causality" 2009]
- ML as a tool to discover causal relations from data (generative ML)
  [Goudet et al arxiv 2017]

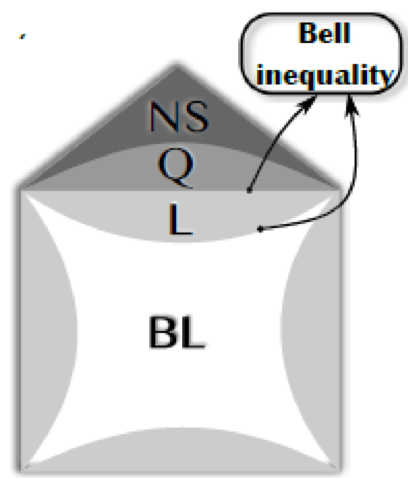
- Quantum machine learning algorithms
  See for instance [Biamonte, Wittek, Pancotti, Rebentrost, Wiebe & Lloyd Nature 2017]
  [Dunjko & Briegel arXIv 2017]
- Machine learning hard quantum problems
  See for instance [Carleo & Troyer Science 2017], [Carrasquila & Melko NatPhys 2017]

Machine Learning



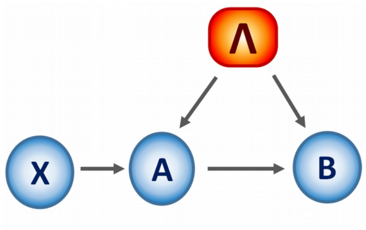
Causality Quantum

- Our aim here is to combine all 3 ingredients: use the mathematical theory of causality and machine learning to witness the classical or quantum (even post-quantum) behaviour of correlations.

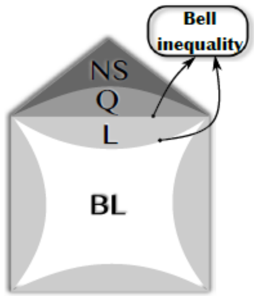


Outline

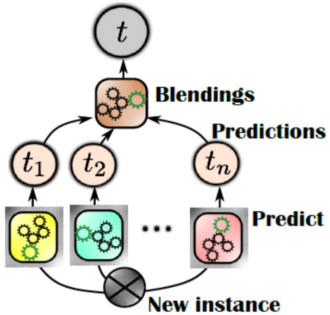
- **Causality and Bayesian Networks**



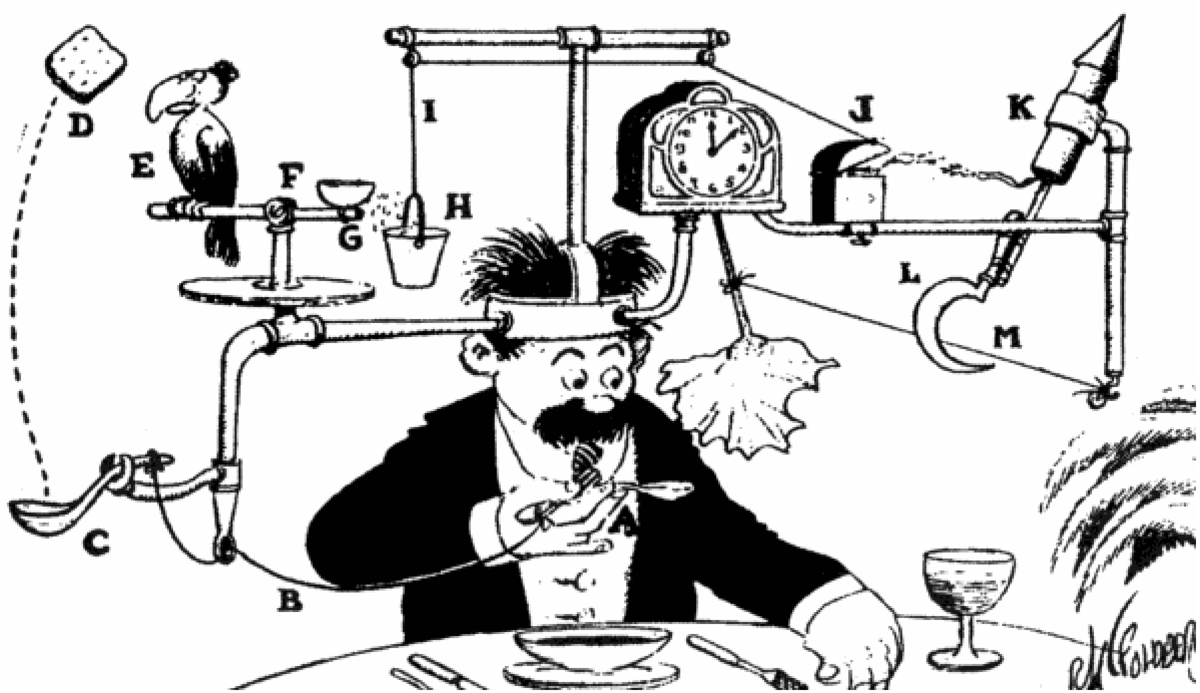
- **Bell’s theorem**



- **Machine learning non-local correlations**



Bayesian Networks: The language of causality



Reichenbach's principle: no correlation without causation.

Reichenbach‘s principle: no correlation without causation.

“If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“

Reichenbach's principle: no correlation without causation.

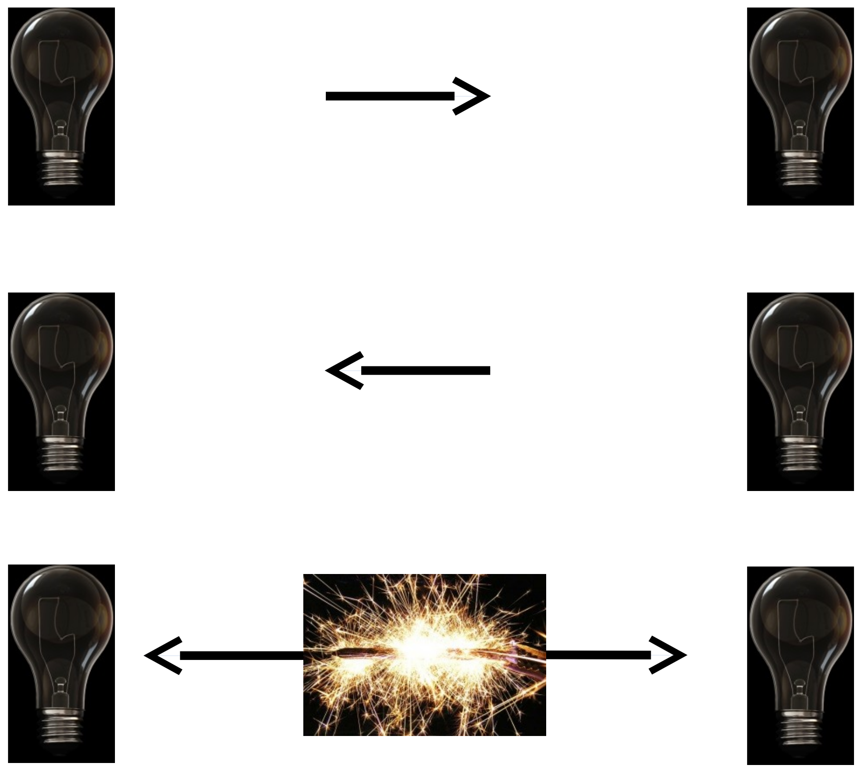
"If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause."

Reichenbach‘s principle: no correlation without causation.

“If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“

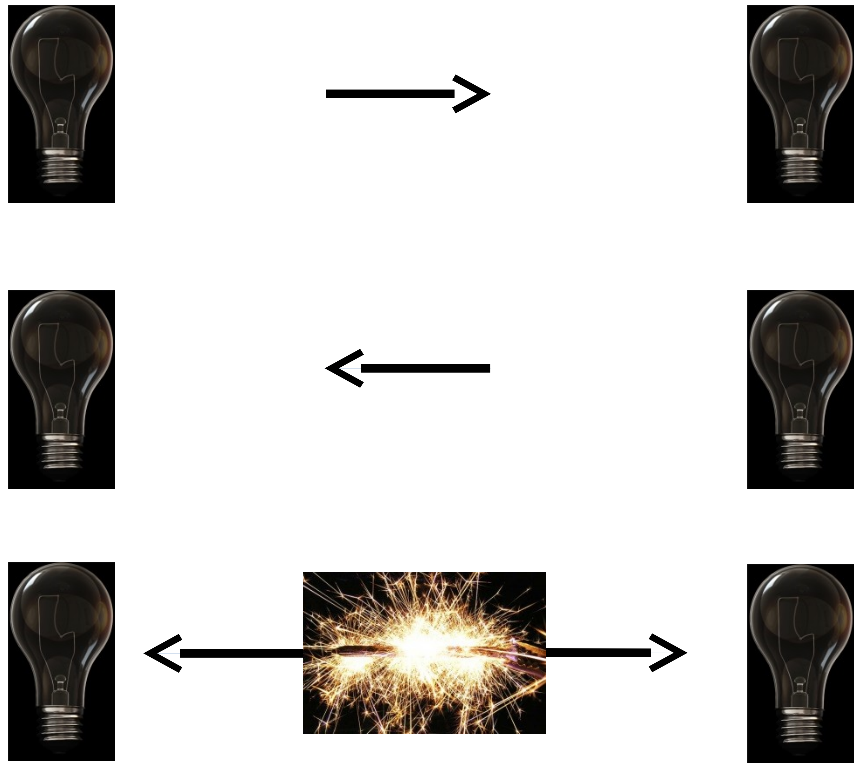
Reichenbach‘s principle: no correlation without causation.

“If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“



Reichenbach‘s principle: no correlation without causation.

“If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“



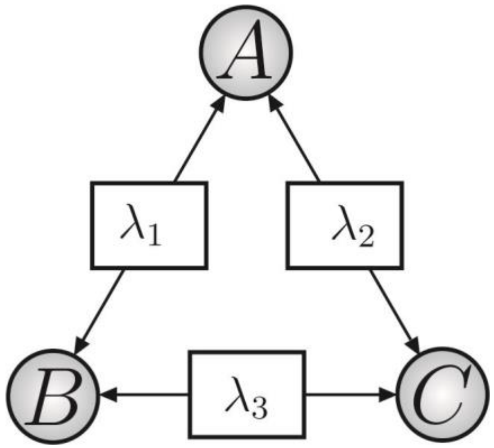
Task: Infer causal relationships from observational (statistical) data.

DAGs: Representing causal relations

- For n variables $X_1, \ldots, X_n$, the causal relationships are encoded in a causal structure, represented by a directed acyclic graph (DAG), with *i*th variable being a deterministic

  $\mathbf{x_i = f_i(pa_i, u_i)}$

  of its parents $\mathbf{pa}_i$ and jointly independent noise variables $\mathbf{u}_i$



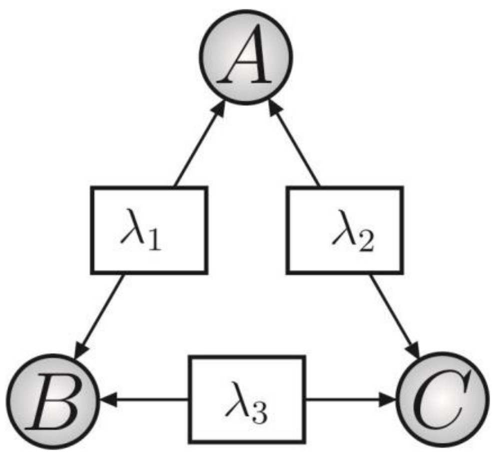
[See J. Pearl, ***Causality***]

DAGs: Representing causal relations

- For n variables $X_1, \ldots, X_n$, the causal relationships are encoded in a causal structure, represented by a directed acyclic graph (DAG), with *i*th variable being a deterministic

  $$\mathbf{x_i = f_i(pa_i, u_i)}$$

  of its parents $\mathbf{pa}_i$ and jointly independent noise variables $\mathbf{u}_i$



- Causal relationships are encoded in the conditional independencies (CIs) implied by the DAG

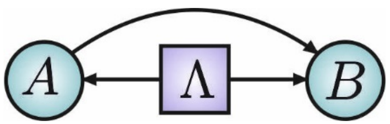
$$p(\lambda_1, \lambda_2) = p(\lambda_1)p(\lambda_2)$$
$$p(A, B|\lambda_1) = p(A|\lambda_1)p(B|\lambda_1)$$
...

Conditional independencies hold information about causation!

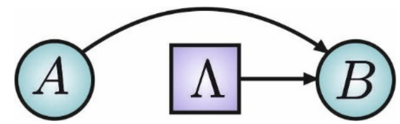
[See J. Pearl, *Causality*]

Interventions: Uncovering causal relations Part 1

Does A have some causal influence over B, or all the correlations between A and B are mediated via the common ancestor?



Intervention



$$p(b|a) = \sum_\lambda p(b|a,\lambda)p(\lambda|a)$$

$$p(b|\text{do}(a)) = \sum_\lambda p(b|a,\lambda)p(\lambda)$$

$$p(b|a) \neq p(b|\text{do}(a))$$

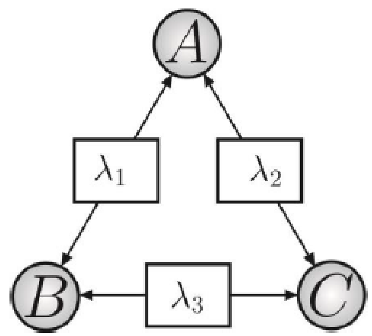
For various reasons, interventions are often not an option.

How to discover causal relations without interventions?

***For the quantum version of it see** [Ried, Agnew, Vermeyden, Janzing, Spekkens & Resch, Nat Phys 2015]
[Chaves, Carvacho, di Giulio, Agresti,Aolita, Giacomini, Sciarrino , Nat Phys 2018]

Is a given probability distribution compatible with a presumed *causal structure*?

Example: Is a given $p(\lambda_1, \lambda_2, \lambda_3, A, B, C)$ compatible with



$$p(\lambda_1, \lambda_2) = p(\lambda_1)p(\lambda_2)$$

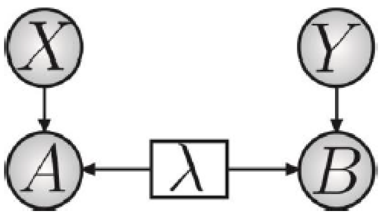
$$p(A, B|\lambda_1) \neq p(A|\lambda_1)p(B|\lambda_1)$$

- If the the full probability distribution (of all nodes in a DAG) is available, CIs hold all information required to solve the compatibility problem

However…

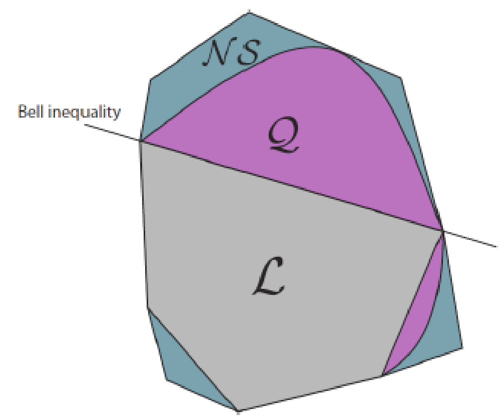
Bell Inequalities: Uncovering causal relations Part 3

- Usually and for a variety of reasons not all variables in a DAG are observable, i.e., not all CIs are available from empirical data



$$p(x, y, \lambda) = p(x)p(y)p(\lambda)$$
$$p(a|x, y, \lambda) = p(a|x, \lambda)$$
$$p(a, b|\lambda) = p(a|\lambda)p(b|\lambda)$$

$$\begin{aligned} p(a,b|x,y) &= \sum_{\lambda} p(a,b,\lambda|x,y) \\ &= \sum_{\lambda} p(a,b|x,y,\lambda)p(\lambda|x,y) \\ &= \sum_{\lambda} p(a|x,\lambda)p(b|y,\lambda)p(\lambda) \end{aligned}$$



Pic from **[Rev. Mod. Phys. 86, 419 (2014)]**

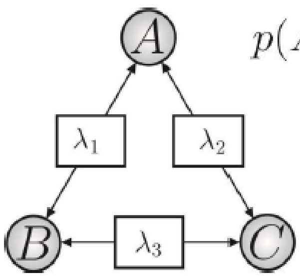
- CIs impose non-trivial constraints on the level of the observable variables, for example, Bell inequalities.

- In quantum mechanics non commuting observables cannot be jointly observed

Marginal scenario: subset of variables that are (jointly) observable

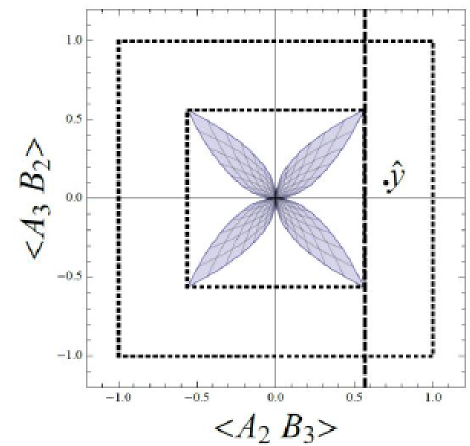
The challenge: Uncovering causal relations Part 4

- Describe marginals compatible with DAGs
- The observable probability dist. contains the full information required for that...
- ..very difficult, non-convex sets (algebraic geometry methods required, see for instance **[Geiger & Meek, UAI 1999]**)



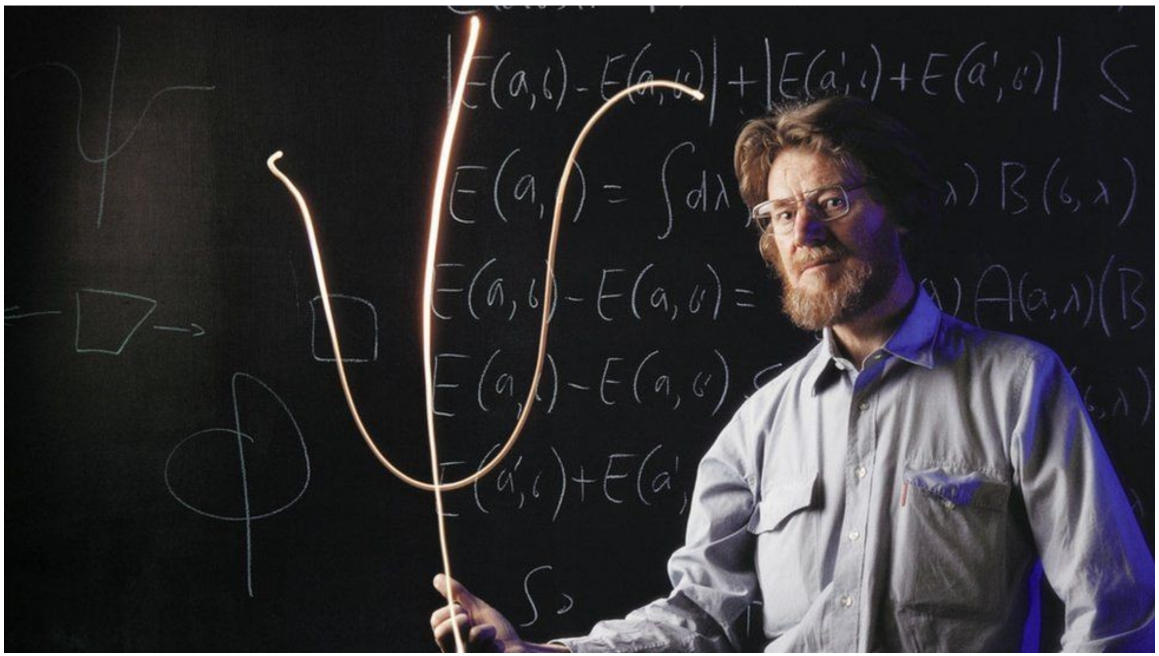
$$p(A,B,C) = \int d\lambda_1 d\lambda_2 d\lambda_3 p(\lambda_1)p(\lambda_2)p(\lambda_3) p(A|\lambda_1,\lambda_2)p(B|\lambda_1,\lambda_3)p(C|\lambda_2,\lambda_3)$$

Picture from **[Steeg & Galstyan, UAI 2011]**

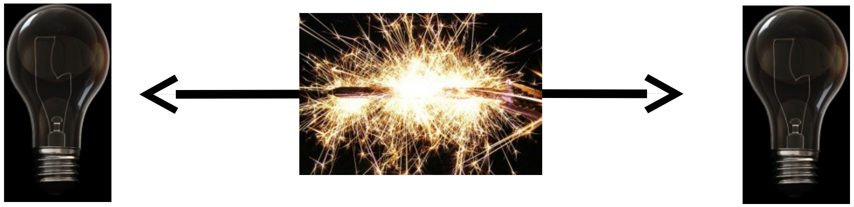


Can machine learning help us to characterize such complicated sets of correlations?

Bell’s theorem: Beyond classical causal models



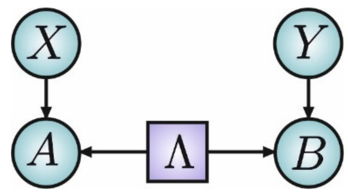
Bell´s theorem from a causal perspective



Bell´s theorem from a causal perspective

- Alice and Bob measure two possible observables each: A_0, A_1, B_0, B_1

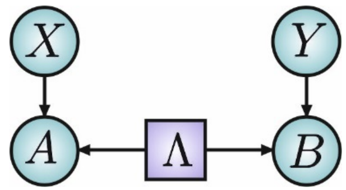
Bell´s theorem from a causal perspective



- Alice and Bob measure two possible observables each: A_0, A_1, B_0, B_1

- After sufficiently many repetitions they can estimate statistical quantities. The experiment can be described in terms of $p(a,b|x,y)$.

Bell´s theorem from a causal perspective



- Alice and Bob measure two possible observables each: A_0, A_1, B_0, B_1

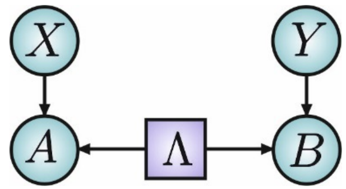
- After sufficiently many repetitions they can estimate statistical quantities. The experiment can be described in terms of $p(a,b|x,y)$.

- There are two causal assumptions entering in Bell‘s theorem.

Measurement Independence: $p(x,y,\lambda) = p(x)p(y)p(\lambda)$

Locality: $p(b|a,x,y,\lambda) = p(b|y,\lambda)$

Bell´s theorem from a causal perspective



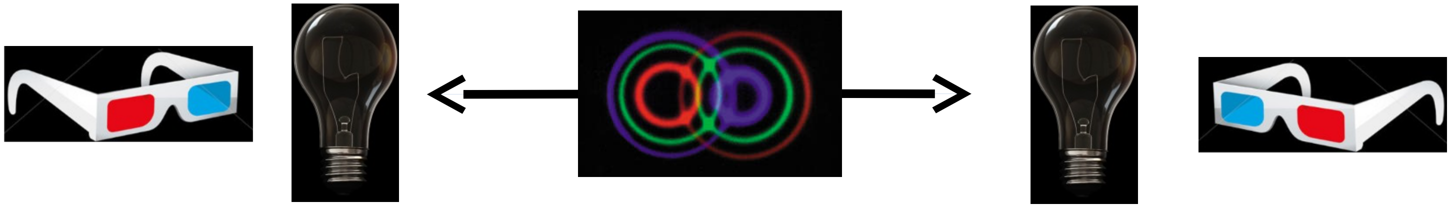
The causal assumptions of a LHV model impose constraints on the possible observed distributions. Those can be tested via Bell inequalities.

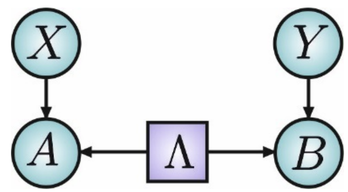
- E.g., LHV model must respect the CHSH inequality:

$$\mathrm{CHSH} = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2$$

$$\langle A_x B_y \rangle = \sum_{a,b} (-1)^{a+b} p(a,b|x,y)$$

Bell´s theorem from a causal perspective





The causal assumptions of a LHV model impose constraints on the possible observed distributions. Those can be tested via Bell inequalities.

- E.g., LHV model must respect the CHSH inequality:

$$\mathrm{CHSH} = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2$$

$$\langle A_x B_y \rangle = \sum_{a,b} (-1)^{a+b} p(a,b|x,y)$$

- Quantum mechanically we can violate this bound:

$$\mathrm{CHSH}_{\mathrm{QM}} \leq 2\sqrt{2}$$

Quantifying Non-locality: The violation of a Bell inequality

- Intuitively, the more we violate a Bell inequality the more non-local the correlation should be...

Some issues...

- The number of Bell inequalities increases very quickly

  Example: 2 parties, 2 outcomes, 5 measurements already more than half million inequalities (actually, no precise characterization is known)

  What does it mean to violate more a given (very specific) Bell Inequality

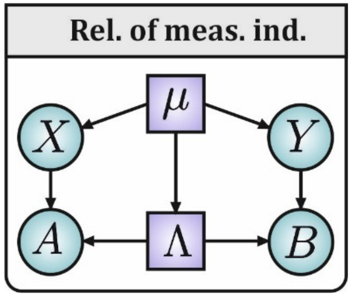
- Maximally entangled states do not necessarily violate more a given Bell inequality...

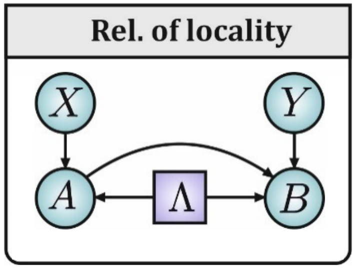
  [Method & Scarani, QIC 2007]

Quantifying Non-locality: Causal approach

- The simulation of nonlocal correlations by classical causal models must give up on the assumption of no fine-tuning (faithfulness).

[Wood & Spekkens, NJP 2015]



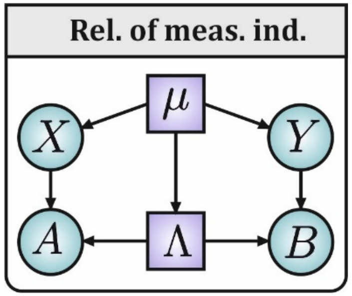


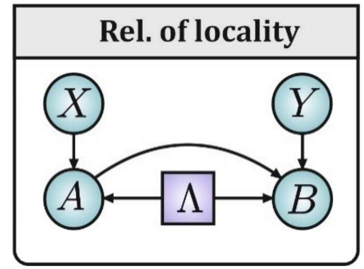
By how much we have to relax these causal assumptions?

Quantifying Non-locality: Causal approach

- The simulation of nonlocal correlations by classical causal models must give up on the assumption of no fine-tuning (faithfulness).

[Wood & Spekkens, NJP 2015]





By how much we have to relax these causal assumptions?

Average causal effect: observable measure of causal influence

$$\mathrm{ACE}_{A\rightarrow B} = \sup_{b,y,a,a'} |p(b|do(a),y) - p(b|do(a'),y)|$$

If this is the underlying model by intervening in A we should see changes in the prob of B.

$$\mathrm{CHSH} = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2$$

$$\min \mathrm{ACE}_{A\rightarrow B} = \max\left[0, \left(\mathrm{CHSH} - 2\right)/2\right]$$

[Chaves, Kueng, Brask & Gross, PRL 2015]

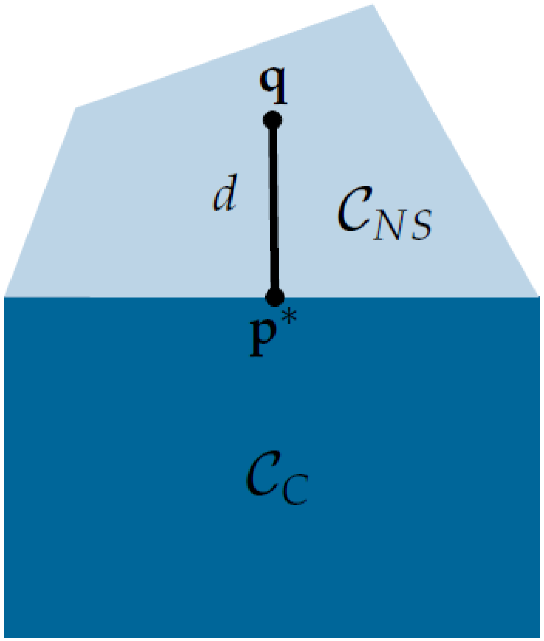
[Ringbauer, Giarmatzi, Chaves, Costa, Fedrizzi & White, Science Advances 2016]

Quantifying Non-locality: Geometric approach

- Trace distance to the set of local correlations

[Brito, Amaral & Chaves, PRA 2018]

$$\mathrm{NL}(\mathbf{q}) = \frac{1}{|x||y|} \min_{\mathbf{p}\in\mathcal{C}_C} D(\mathbf{q},\mathbf{p})$$
$$= \frac{1}{2|x||y|} \min_{\mathbf{p}\in\mathcal{C}_C} \sum_{a,b,x,y} |q(a,b|x,y) - p(a,b|x,y)|$$

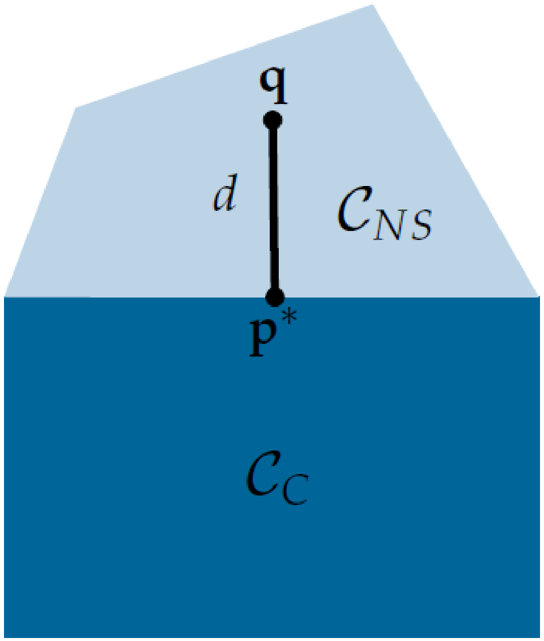


Quantifying Non-locality: Geometric approach

- Trace distance to the set of local correlations

[Brito, Amaral & Chaves, PRA 2018]

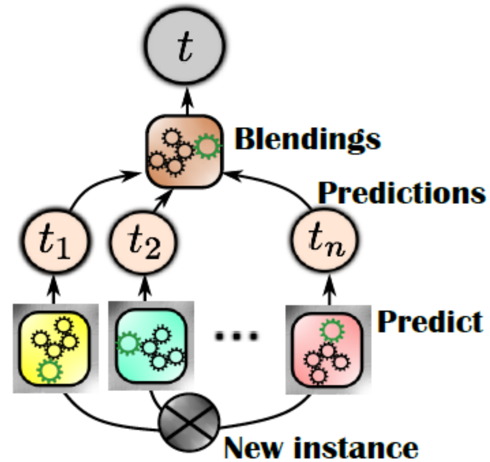
$$\mathrm{NL}(\mathbf{q}) = \frac{1}{|x||y|} \min_{\mathbf{p}\in\mathcal{C}_C} D(\mathbf{q},\mathbf{p})$$
$$= \frac{1}{2|x||y|} \min_{\mathbf{p}\in\mathcal{C}_C} \sum_{a,b,x,y} |q(a,b|x,y) - p(a,b|x,y)|$$



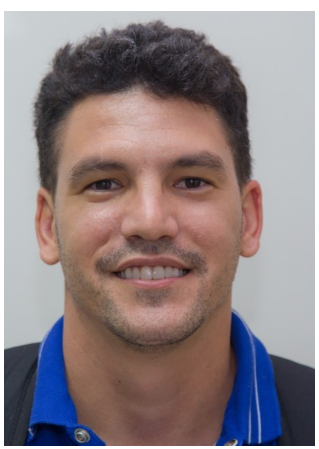
$$\mathrm{CHSH} = q_{AB}^{0,0} + q_{AB}^{0,1} + q_{AB}^{1,0} - q_{AB}^{1,1} - q_A^0 - q_B^0 \leq 0$$

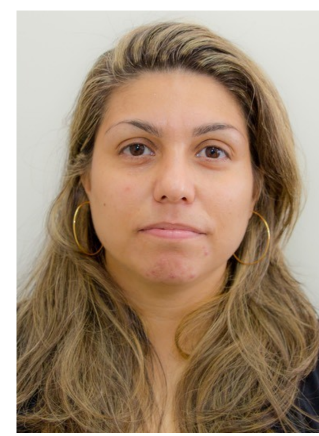
$$\mathrm{NL}(\mathbf{q}) = \frac{1}{2} \max\left[0, \Pi(\mathrm{CHSH})\right]$$

Machine Learning Nonlocal Correlations



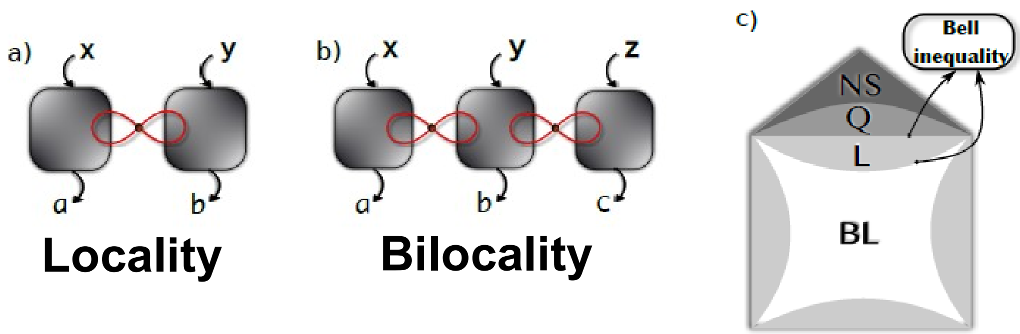
Machine Learning Non-local Correlations,
A. Canabarro, S. Brito, RC
arXiv:1808.07069





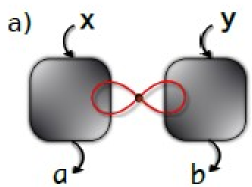
Training the machine

- Two scenarios have been considered: bipartite and tripartite (entanglement swapping)

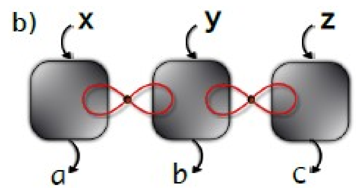


Training the machine

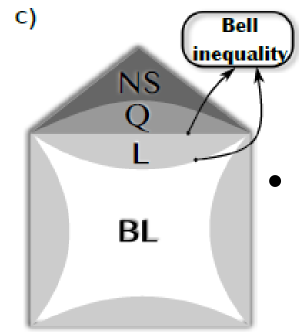
- Two scenarios have been considered: bipartite and tripartite (entanglement swapping)



Locality



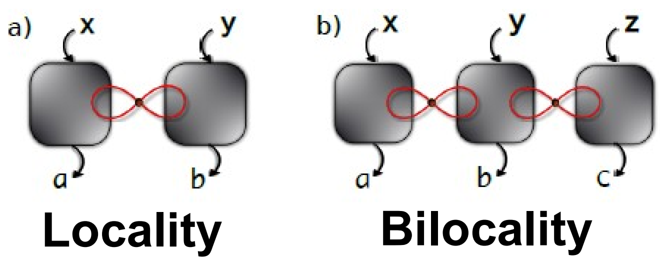
Bilocality

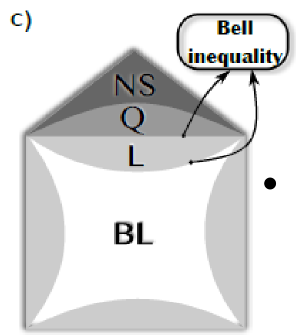


- For each scenario we randomly sample over the NS distributions and compute its trace distance to the local or bilocal set

Training the machine

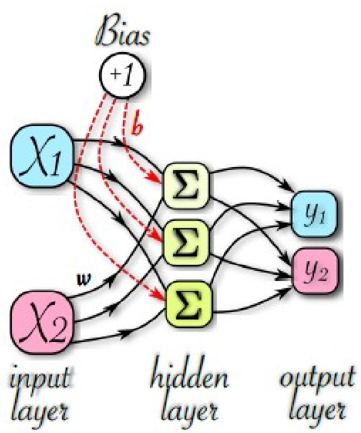
- Two scenarios have been considered: bipartite and tripartite (entanglement swapping)



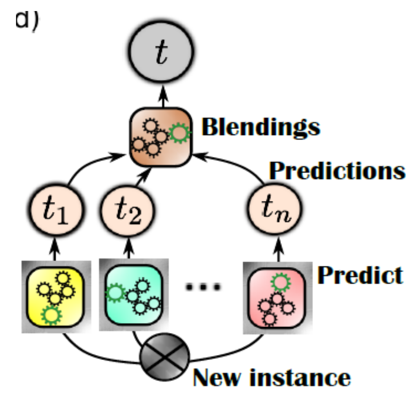


- For each scenario we randomly sample over the NS distributions and compute its trace distance to the local or bilocal set

- Use this data to train the machine (75% training + cross-validation)

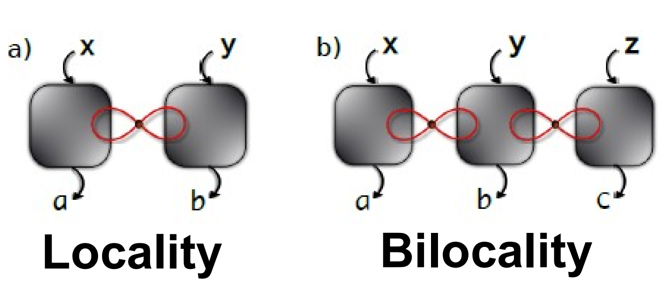


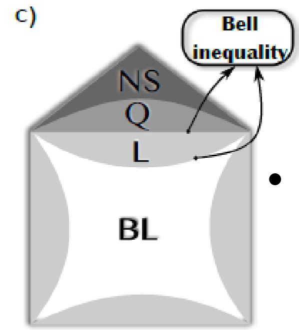
- We use a blending of multilayer perceptrons with different number of layers an neurons



Training the machine

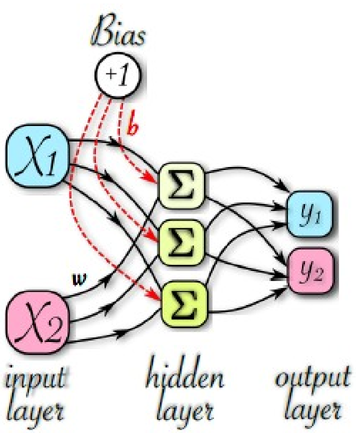
- Two scenarios have been considered: bipartite and tripartite (entanglement swapping)



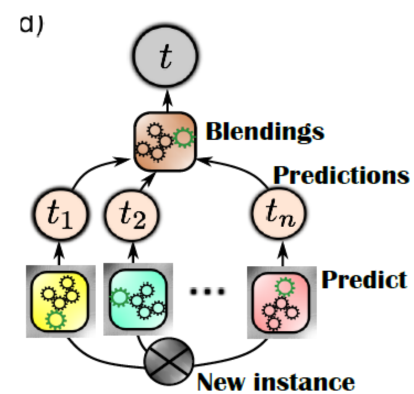


- For each scenario we randomly sample over the NS distributions and compute its trace distance to the local or bilocal set

- Use this data to train the machine (75% training + cross-validation)



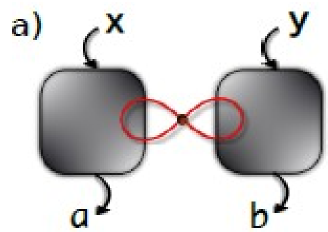
- We use a blending of multilayer perceptrons with different number of layers an neurons



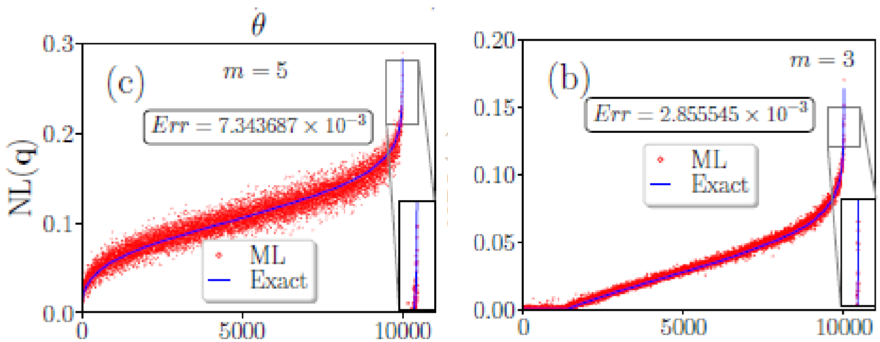
- The rest of data is used as testing set (25%)

- Quantify the accuracy by the average trace distance error

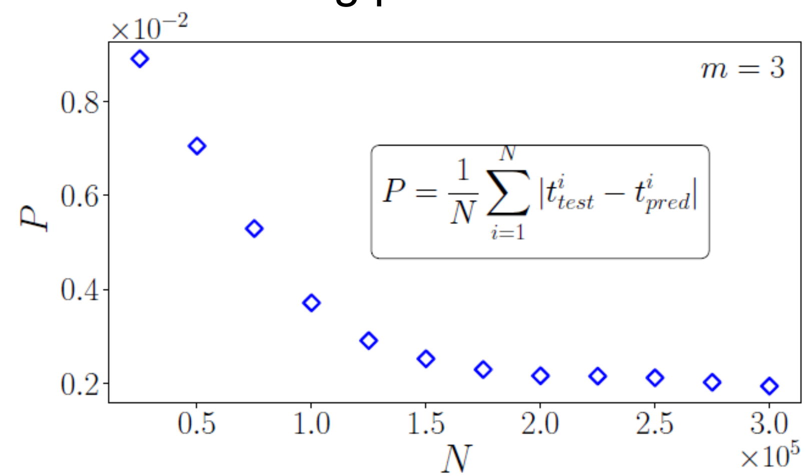
Results: Part 1



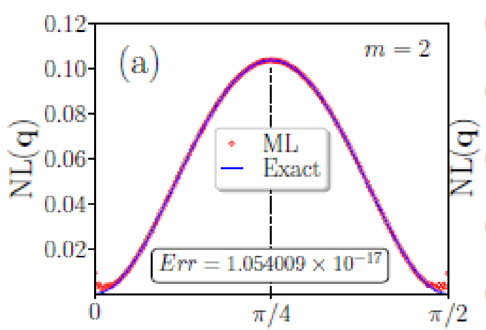
Technique \ Scenario	$m = 2$	$m = 3$	$m = 4$	$m = 5$
Typical MLP ($\times 10^{-3}$)	0.46	2.20	7.75	8.50
Blending ($\times 10^{-3}$)	0.05	1.54	6.78	7.31



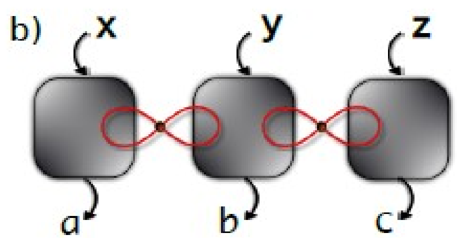
- Learning plateau

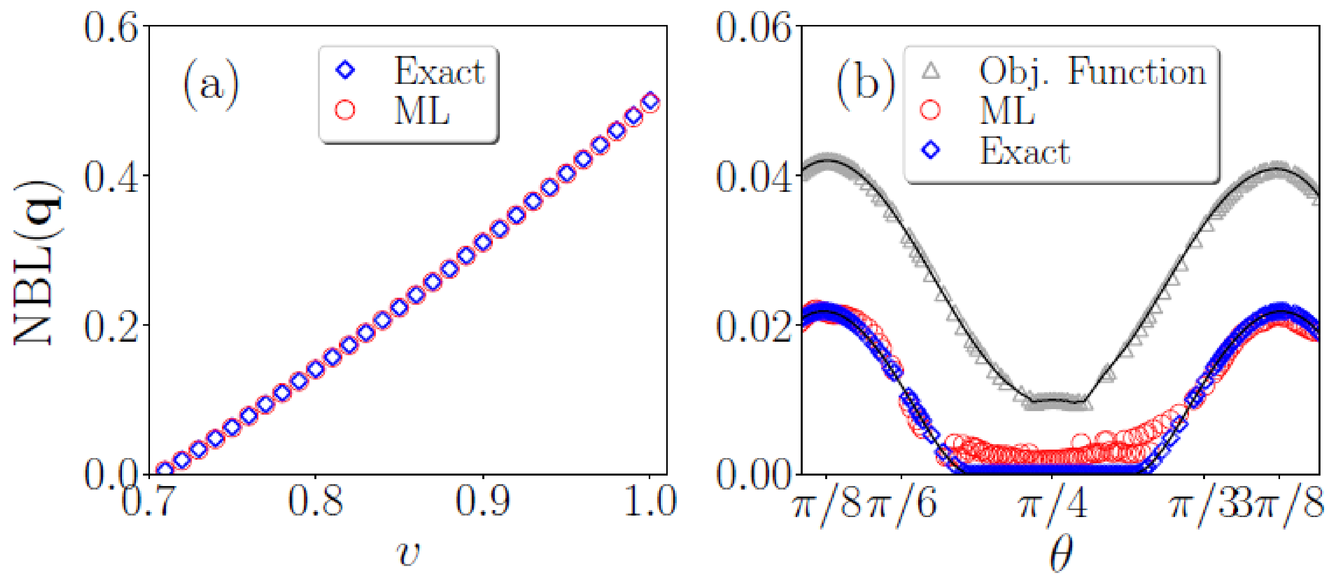


- Good performance on NS and quantum correlations



Results: Part 2

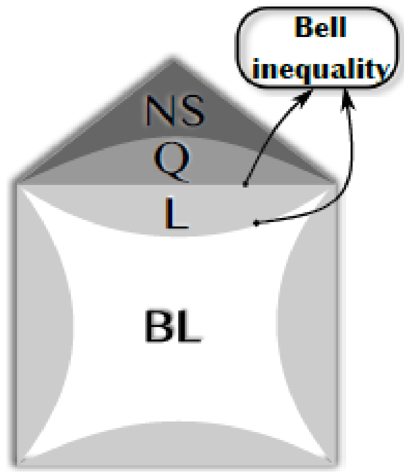




- Good performance on NS and quantum correlations
- 10^5 improve in time (in comparison with the best known method)
- New candidate correlations that cannot be detected by any known inequality

Results: Part 3

- Distinguishing classical, quantum and post-quantum correlations



Predictions / True Class	Local	Quantum	Post-quantum
Local	33436	96	0
Quantum	41	33173	236
Post-quantum	0	136	32882

Discussion/Future Investigations

What to remember (if anything):

- i) Deep learning provide a very accurate (black box) description of high-dimensional complicated sets of correlations.
- ii) We can also learn from the machine (new candidate non-classical correlations that hardly could be found by other means).
- iii) this was only the first step

What to remember (if anything):

- **i) Deep learning provide a very accurate (black box) description of high-dimensional complicated sets of correlations.**
- **ii) We can also learn from the machine (new candidate non-classical correlations that hardly could be found by other means).**
- **iii) this was only the first step**

What's next?

- **Open the black box?**
- **Machine generated certificates? (Generative ML)**
- **Applications to other marginal problems?**
- **ML trained to other purposes (e.g. dog x cat)? (Transfer Learning)**