Machine-Learning Mathematical Structures

YANG-HUI HE

London Institute of Mathematical Sciences, Royal Institution

Dept of Mathematics, City, **University of London**

Merton College, **University of Oxford**

School of Physics, **NanKai University**

Workshop on Black Holes, BPS and Quantum Information;

Lisboa (Virtual) Sept 2021

The String Landscape

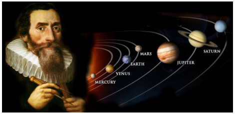
- The Geometric Origin of our Universe
  - String Theory (a unified ToE) exists in (9+1)-D; We live in (3+1)-D
  - Extra dimension X = 6-dimensional manifold $(10 - 4 = 6)$;
  - Geometry (algebraic/differential geometry/topology) of X determines physics of our world.
- The Vacuum Degeneracy Problem TOO MANY possibilities for X
  - e.g., X Calabi-Yau 3-folds (topology/combinatorics tends to grow exponentially)

Complex Dim	1	2	3	...
CY$_n$	T^2	$T^4, K3$	$> 10^{10}$	?

Reflexive Polytopes

  - see: YHH, *The CY Landscape: from geometry, to physics, to ML*, Springer LNM, 2021.

The Typical Calculation: Inevitability of Algebraic Geometry

- Each X (+ bundles, discrete symmetries) gives a (3+1)-d universe
  - particles and interactions $\sim$ cohomology theory; masses $\sim$ metric; Yukawa $\sim$ Triple intersections/integral of forms over X
  - $X \sim$ algebraic variety $\sim$ degree/coefficient data, numerical metrics
- 

  Ubi materia, ibi geometria

  – Johannes Kepler (1571-1630)

- Our Universe: (1) probabilistic/anthropic?
  (2) Sui generis/selection rule?
  (3) one of multi-verse ?

  cf. *Exo-planet/Habitable Zone search*

- cf. exact SM: Braun-YHH-Ovrut (2005), Candelas-de la Ossa-YHH-Szendroi (2009), Anderson-Gray-Lukas-Ovrut-Palti (2012), ...

A Wild Question: NN Doesn't Care/Know about maths

- Typical Problem in String Theory:

$$\underset{\boxed{\text{(integer, real) tensor}}}{INPUT} \longrightarrow \underset{\boxed{\text{integer}}}{OUTPUT}$$

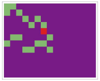
- Q: Can we "machine-learn the landscape?"

- [YHH 1706.02714] Deep-Learning the Landscape, *PLB 774, 2017*

  (*Science*, Aug, vol 365 issue 6452)

  - Hodge Number of a Complete Intersection CY is the association rule, e.g.

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}, \qquad h^{1,1}(X) = 8 \quad \rightsquigarrow$$



$$\longrightarrow 8$$

CICY is 12×15 integer matrix with entries $\in [0, 5]$ is simply represented as a 12×15 pixel image of 6 colours Proper Way ; ML in matter of seconds/minutes

Progress in String Theory

Major International Annual Conference Series

1986- First "Strings" Conference

2002- First "StringPheno" Conference

2006 - 2010 String Vacuum Project (NSF)

2011- First "String-Math" Conference

2014- First String/Theoretical Physics Session in SIAM Conference

2017- First "String-Data" Conference

YHH (1706.02714), Seong-Krefl (1706.03346), Ruehle (1706.07024), Carifio-Halverson-Krioukov-Nelson (1707.00655) A host of activity ;

Machine Learning Mathematical Structures

Why stop at string/geometry?

q.v. Review Paper: YHH 2101.06317

How does one *DO* mathematics, I ?

Russell-Whitehead *Principia Mathematica* [1910s] programme (since at least Frege, even Leibniz) to axiomatize mathematics, but . . .

Gödel [1931] Incompleteness ; Church-Turing [1930s] Undecidability

Automated Theorem Proving (ATP) The practicing mathematician hardly ever worries about Gödel

- Newell-Simon-Shaw [1956] Logical Theory Machine: proved subset of *Principia* theorems
- Type Theory [1970s] Martin-Löf, Coquand, . . . Coq interactive proving system: 4-color (2005); Feit-Thompson Thm (2012); Lean (2013)
- Univalent Foundation / Homotopy Type Theory [2006-] Voevodsky

We can call this Bottom-up Mathematics

How does one *DO* mathematics, I ?

Russell-Whitehead *Principia Mathematica* [1910s] programme (since at least Frege, even Leibniz) to axiomatize mathematics, but ...

Gödel [1931] Incompleteness ; Church-Turing [1930s] Undecidability

Automated Theorem Proving (ATP) The practicing mathematician hardly ever worries about Gödel

- Newell-Simon-Shaw [1956] Logical Theory Machine: proved subset of *Principia* theorems
- Type Theory [1970s] Martin-Löf, Coquand, ... Coq interactive proving system: 4-color (2005); Feit-Thompson Thm (2012); Lean (2013)
- Univalent Foundation / Homotopy Type Theory [2006-] Voevodsky

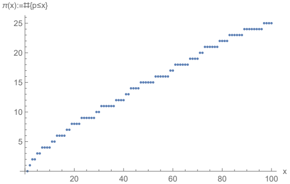
We can call this Bottom-up Mathematics

How does one do mathematics, II ?

- Late C20th - increasing rôle of computers: 4-color [Appel-Haken-Koch 1976]; Classif. Finite Simple Groups [Galois 1832 - Gorenstein et al. 2008] ...
- Buzzard: "Future of Maths" 2019: already plenty of proofs unchecked (incorrect?) in the literature, MUST use computers for proof-checking; XenaProject, Lean establish database of mathematical statements
- Davenport: ICM 2018 "Computer Assisted Proofs".
- Hale & Buzzard: Foresee within 10 years AI will help prove "early PhD" level lemmas, all of undergrad-level maths formalized;
- Szegedy: more extreme view, computers > humans @ chess (1990s); @ Go (2018); @ Proving theorems (2030)

How does one *DO* mathematics, III ?

- Historically, Maths perhaps more Top-Down: practice before foundation
  - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra ...
  - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16

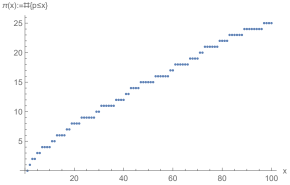


(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

- BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

How does one *DO* mathematics, III ?

- Historically, Maths perhaps more Top-Down: practice before foundation
  - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra ...
  - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16

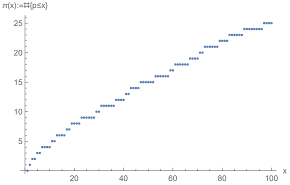


(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

- BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

How does one *DO* mathematics, III ?

- Historically, Maths perhaps more Top-Down: practice before foundation
  - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra ...
  - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16



(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

- BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena . . . how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena …how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena ... how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2 \\ 5&0&4&5&2&4&4&2&2&4 \\ 1&1&2&2&0&4&1&4&5&0 \\ 5&0&1&1&0&2&0&5&0&1 \\ 2&5&0&1&1&3&2&3&0&3 \\ 3&2&2&3&0&0&2&2&1&0 \\ 2&2&5&1&4&4&0&0&1&2 \\ 5&0&0&0&4&5&0&4&1&1 \\ 4&3&4&3&3&1&0&0&2&5 \\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2 \\ 5&0&4&5&2&4&4&2&2&4 \\ 1&1&2&2&0&4&1&4&5&0 \\ 5&0&1&1&0&2&0&5&0&1 \\ 2&5&0&1&1&3&2&3&0&3 \\ 3&2&2&3&0&0&2&2&1&0 \\ 2&2&5&1&4&4&0&0&1&2 \\ 5&0&0&0&4&5&0&4&1&1 \\ 4&3&4&3&3&1&0&0&2&5 \\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena . . . how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Mathematical Data: perfect for mining

- Mathematical Data is more structured than "real world" data, much less susceptible to noise; Outliers even more interesting, e.g. Sporadics, Exceptionals, . . .
- Last 10-20 years: large collaborations of computational mathematicians, physicists, CS (cf. SageMATH, GAP, Bertini, MAGMA, Macaulay2, Singular, Pari, Wolfram, . . .) computed and compiled vast data
  - links
  - Generic computation *HARD*
  - mining provides some level of "intuition" & is based on "experience"

Methodology

Bag of Tricks Hilbert's Programme of *Finitary Methods*, Landau's *theoretical minimum*, Migdal's *Mathmagics* ...

IMO Grand Challenge (2020-) Good set of concrete problems to try on AI

Standard ML Regressor & Classifiers (w/ NO KNOWLEDGE of the maths) UAT

- NN: MLPs; CNNs; RNNs, ... (gentle tuning of architecture and hyper-parameters)
- SVM, Bayes, Decision Trees, PCA, Clustering, ...
- ML: emergence of complexity via connectivity $\rightsquigarrow$ Intution (?)

This Talk: Status Report of Experiments in the last couple of years

- focus on supervised ML ("knows where to get to")
- all standard methods $\simeq$ same performance
- $\sim$ 20-80 split; training on 20 (precision, Matthews' ϕ or R^2)

Representation/Group Theory

- ML Algebraic Structures (GAP DB) [YHH-MH. Kim 1905.02263,]
  - When is a Latin Square (Sudoku) the Cayley (multiplication) table of a finite group? Bypass quadrangle thm (0.95, 0.9)
  - Can one look at the Cayley table and recognize a finite simple group?
    - bypass Sylow and Noether Thm; (0.97, 0.95) rmk: can do it via character-table T, but getting T not trivial
    - **SVM:** space of finite-groups (point-cloud of Cayley tables) seems to exist a hypersurface separating simple/non-simple
- ML Lie Structure Chen-YHH-Lal-Majumder [2011.00871] Weight vector $\to$ length of irrep decomp / tensor product: (0.97, 0.93); (train on small dim, predict high dim: (0.9, 0.8))

Combinatorics, Graph/Quivers

- [YHH-ST. Yau 2006.16619] Wolfram Finite simple graphs DB
  - ML standard graph properties:

    ?acyclic (0.95, 0.96); ?planar (0.8, 0.6); ?genus $>, =, < 0$ (0.8, 0.7); ?∃ Hamilton cycles (0.8, 0.6); ?∃ Euler cycles (0.8, 0.6)

    (Rmk: NB. Only "solving" the likes of traveling salesman stochastically)
  - spectral bounds ($R^2 \sim 0.9$) ...
  - Recognition of Ricci-Flatness (0.9, 0.9) (todo: find new Ricci-flat graphs);
- [Bao-Franco-YHH-Hirst-Musiker-Xiao 2006.10783]: categorizing different quiver mutation (Seiberg-dual) classes (0.9 - 1.0, 0.9)

Number Theory: A Classical Reprobate?

Arithmetic (prime numbers are Difficult!)

- [YHH 1706.02714, 1812.02893:]
  - Predicting primes $2 \to 3,\ 2,3 \to 5,\ 2,3,5 \to 7$; no way
  - fixed (or $x/\log(x)$-scaled) window of $(\text{yes/no})_{1,2,\ldots,k}$ to $(\text{yes/no})_{k+i}$ for some i (in binary); ML PRIMES problem (0.7, 0.8) NOT random! (prehaps related to AKS algorithm [2002], PRIMES is in P)
  - Sarnak's challenge: same window $\to$ Liouville Lambda (0.5, 0.001) Truly random (no simple algorithm for Lambda)
- [Alessandretti-Baronchelli-YHH 1911.02008]

  ML/TDA@Birch-Swinnerton-Dyer Ш and Ω ok with regression & decision trees: RMS < 0.1; Weierstrass $\to$ rank: random

Number Theory: A Modern Hope?

Arithmetic Geometry (Surprisingly Good)

- [Hirst-YHH-Peterken 2004.05218]: adjacency+permutation triple of dessin d'enfants (Grothendieck's Esquisse for $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$) ; predicting transcendental degree (0.92, 0.9)
- YHH-KH Lee-Oliver arithmetic of curves
  - 2010.01213: Complex Multiplication, Sato-Tate $(0.99 \sim 1.0, 0.99 \sim 1.0)$
  - 2011.08958: Number Fields: rank and Galois group (0.97, 0.9)
  - 2012.04084: BSD from Euler coeffs, integer points, torsion (0.99, 0.9); Tate-Shafarevich Ш (0.6, 0.8) [Hardest quantity of BSD]

Elliptic Curves: case study

- E an elliptic curve, local zeta-function & L-function:

$$Z(E/\mathbb{F}_p;T) = \exp\left(\sum_{k=1}^{\infty} \frac{\#E(\mathbb{F}_{p^k})T^k}{k}\right) = \frac{L_p(E,T)}{(1-T)(1-pT)} \ ;$$

$$L_p(E,T) = 1 - a_p T + pT^2; \quad a_p = p + 1 - \#E\left(\mathbb{F}_p\right).$$

Fix N and define vector $v_L(E) = (a_{p_1}, \ldots, a_{p_N}) \in \mathbb{Z}^N$;

$\sim 10^5$ balanced data from `www.lmfdb.org`; 50-50 cross validation.

- Labeled data: $v_L(E) \longrightarrow$ rank, torsion, . . . ([Birch-Swinnerton-Dyer:])

$$L(E,s) := \prod_p L^{-1}(E, T := p^{-s}); \quad \frac{L^{(r)}(E,1)}{r!} \overset{???}{=} \frac{|Ш|\Omega \mathrm{Reg} \prod_p c_p}{(\#E(\mathbb{Q})_{\mathrm{tors}})^2},$$

r=rank; Ш=Shafarevich group; Reg=regulator; c_p=Tawagama; tors=Torsion

- Try generic ML algorithms on the data, record naive precision and Matthew's correlation coefficient/F1-Score

elliptic curves: results

report (naive precision, Matthew's Correlation = χ^2)

Rank 0 or 1 $N = 300$, conductor $\in [1, 10^4]$, Logistic regression: (0.991, 0.982)

(Goldfield-Katz-Sarnak Conjecture: r=0 and 1 at 50% each)

Torsion Order = 1 or 2 $N = 500$, conductor $\in [1, 3 \times 10^4]$, naive Bayes: (0.9997, 0.9995) (Mordell-Weil, Faltings: max torsion = 16, but in LMFDB mostly 1 or 2)

$\exists$ $\mathbb{Z}$-points (not just $\mathbb{Q}$), $N = 500$, conductor range $[1, 3 \times 10^4]$, naive Bayes: (0.999, 0.998) (Siegel Thm: finite # integer points.)

Tate-Shafarevich group nothings gets better than 0.6; hardest part of BSD

Clearly useful for maths and physics

looking for new conjectures e.g.,

- '19 YHH-Kim: separating hyperplane - simple/non-simple groups; open
- '19 Brodie-Constantin-Lukas: exact formulae for cohomo surf.; proved.
- '20 YHH-Lee-Oliver: L-coefs and integer pt./torsion on ell; Known.
- '20 Craven-Jejjala-Par: Jones poly best-fit function; open
- ...

speed up computations and accuracies e.g.,

- computing/estimating (top.inv., charges, etc) MUCH FASTER
- '19 Ashmore-YHH-Ovrut: speed up Donaldson alg@CY metric 10-100
- '20 Douglas et al., Anderson et al. accuracy improvement on Donaldson 10-100 times
- ...

The other Extreme (?) View-Point

On the other hand, what is analyticity?

- prime generator $= \left\lfloor \frac{n! \bmod (n+1)}{n} \right\rfloor (n-1) + 2$ (not efficient)
- bundle-cohomology:

Bott for Projective space:

$$h^q(\mathbb{P}^n, (\wedge^p T\mathbb{P}^n) \otimes \mathcal{O}(k)) = \begin{cases} \binom{k+n+p+1}{p}\binom{k+n}{n-p} & q = 0 \quad k > -p-1, \\ 1 & q = n-p \quad k = -n-1, \\ \binom{-k-p-1}{-k-n-1}\binom{-k-n-2}{p} & q = n \quad k < -n-p-1, \\ 0 & \text{otherwise} \end{cases}$$

e.g. $(2,4)$-CY3 hypersurface:

$$h^q(X, \mathcal{O}_X(-k,m)) = \begin{cases} (k+1)\binom{m}{3} - (k-1)\binom{m+3}{3} & q = 0 \quad k < \frac{(1+2m)(6+m+m^2)}{3(2+3m(1-m))} \\ (k-1)\binom{m+3}{3} - (k+1)\binom{m}{3} & q = 1 \quad k > \frac{(1+2m)(6+m+m^2)}{3(2+3m(1-m))} \\ 0 & \text{otherwise} \end{cases}$$

- ...
- better suited for a computer programme any way

An Inherent Hierarchy?

- In decreasing precision/increasing difficulty:

numerical

string theory $\rightarrow$ algebraic geometry over $\mathbb{C}$ $\sim$ arithmetic geometry

algebra

string theory $\rightarrow$ combinatorics

analytic number theory

- **Categorical Theory**
  - suggested by & in prog. w/ B. Zilber, Merton Prof. of Logic, Oxford
  - major part of Model Theory: Morley-Shelah Categoricity Thm
  - Hart-Hrushovski-Laskowski Thm: 13 classes (levels) of iso-classes $I(T, k)$ of a theory T in first order logic over some cardinality k.

Please submit

- Special Collection in AACA, Birkhäuser, *Dechant, YHH, Kaspryzyk, Lukas, ed*:
  `https://www.springer.com/journal/6/updates/18581430`
- Special Volume in JSC, Springer, *Hauenstein, YHH, Kotsireas, Mehta, Tang, ed.*
  `https://www.journals.elsevier.com/journal-of-symbolic-computation/call-for-papers/algebraic-geometry-and-machine-learning`
- ML in theoretical physics & pure maths, Book, WS, *YHH, ed.*
- Int. J. Data Science in the Mathematical Sciences, WS, *YHH et al., ed.*

Meta-mathematics/physics?

[YHH-Jejjala-Nelson] "hep-th" 1807.00735

- Word2Vec: [Mikolov et al., '13] NN which maps words in sentences to a vector space **by context** (much better than word-frequency, quickly adopted by Google); maximize (partition function) over all words with sliding window ($W_{1,2}$ weights of 2 layers, C_α window size, D # windows)

$$Z(W_1, W_2) := \frac{1}{|D|} \sum_{\alpha=1}^{|D|} \log \prod_{c=1}^{C_\alpha} \frac{\exp([\vec{x}_c]^T \cdot W_1 \cdot W_2)}{\sum_{j=1}^{V} \exp([\vec{x}_c]^T \cdot W_1 \cdot W_2)}$$

- We downloaded all $\sim 10^6$ titles of hep-th, hep-ph, gr-qc, math-ph, hep-lat from ArXiv since the beginning (1989) till end of 2017 Word Cloud
(rmk: Ginzparg has been doing a version of linguistic ML on ArXiv)
(rmk: abs and full texts in future)

Subfields on ArXiv has own linguistic particulars

- Linear Syntactical Identities

  bosonic + string-theory = open-string

  holography + quantum + string + ads = extremal-black-hole

  string-theory + calabi-yau = m-theory + g2

  space + black-hole = geometry + gravity . . .

- binary classification (Word2Vec + SVM) of formal (hep-th, math-ph, gr-qc) vs phenomenological (hep-ph, hep-lat) : 87.1% accuracy (5-fold classification 65.1% accuracy). ArXiv classifications

- Cf. **Tshitoyan et al.**, "Unsupervised word embeddings capture latent knowledge from materials science literature", **Nature** July, 2019: 3.3. million materials-science abstracts; uncovers structure of periodic table, predicts discoveries of new thermoelectric materials years in advance, and suggests as-yet unknown materials

Merci!

Syntax		Semantics
Alpha Go	$\to$	Alpha Zero
ML Patterns	$\to$	Auto Thm Pf&Chk

- Renner et al., PRL/Nature News, 2019: ML (*SciNet*, *autoencoder*)
- Lample-Charton, 2019: ML Symolic manipulations in mathematics
- Tegmark et al., 2019 AI Feynman, symb regressor
- Raayoni et al. 2020 Ramanujan-Machine
- Barbaresco-Nielson 2021 Infor Geom/ML

Sophia (Hanson Robotics, HK)

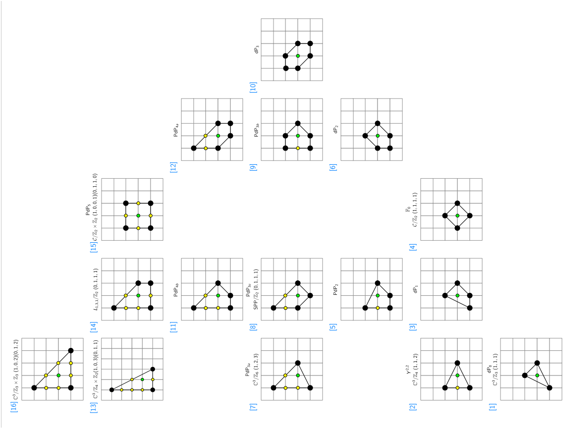
1st non-human citizen (2017, Saudi)

1st non-human with UN title (2017)

1st String Data Conference (2017)

16 Reflexive Polygons

Back to Reflexives



classify convex lattice polytopes with single interior point and all faces are distance 1 therefrom (up to $SL(n;\mathbb{Z})$)

Kreuzer-Skarke: 4319 reflexive polyhedra, 473,800,776 reflexive 4-polytopes,

Skarke: next number is at least 185,269,499,015.

Computing Hodge Numbers $\mathcal{O}(e^{e^d})$

- Recall Hodge decomposition $H^{p,q}(X) \simeq H^q(X, \wedge^p T^\star X) \rightsquigarrow$

$$H^{1,1}(X) = H^1(X, T_X^\star), \qquad H^{2,1}(X) \simeq H^{1,2} = H^2(X, T_X^\star) \simeq H^1(X, T_X)$$

- Euler Sequence for subvariety $X \subset A$ is short exact:

$$0 \to T_X \to T_M|_X \to N_X \to 0$$

- Induces long exact sequence in cohomology:

$$\begin{array}{cccccccc}
0 & \to & \cancel{H^0(X, T_X)} \overset{0}{\nearrow} & \to & H^0(X, T_A|_X) & \to & H^0(X, N_X) & \to \\
 & \to & \boxed{H^1(X, T_X)} & \overset{d}{\to} & H^1(X, T_A|_X) & \to & H^1(X, N_X) & \to \\
 & \to & H^2(X, T_X) & \to & \ldots & & &
\end{array}$$

- Need to compute Rk(d), cohomology and $H^i(X, T_A|_X)$ (Cf. Hübsch)

Back to ML

String/Algebraic Geometry: 2017-

- CICY configuration → Hodge Numbers: YHH (1706.02714) Bull-YHH-Jejjala-Mishra (1806.03121, 1903.03113), Krippendorf-Syvaeri [2003.13679] Erbin-Finotello (2007.13379; '21) : $(0.99, 0.9)$ YHH-Lukas [2009.02544] CICY4: $(0.98, 0.9)$
- Elliptic fibrations (from CICYs): YHH-SJ Lee (1904.08530) $(0.99, 0.9)$
- Distinguishing Heterotic SMs from the sum-line-bundle database (Anderson-Constantin-Gray-Lukas-Palti) and extrapolating beyond Deen-YHH-Lee-Lukas (2003.13339): $(0.98, 0.99)$
- Calabi-Yau metric: improves Donaldson alg. for numerical CY metric by 10-100 times Ashmore–YHH–Ovrut '19, q.v. Anderson, Gray, Krippendorf, Raghuram, Ruehle; Douglas–Lakshminarasimhan–Qi, '20, Jejjala-Pena-Kaloni-Mishra, '21

- q.v., Bundle Cohomology (Ruehle, Brodie-Constantin-Lukas, Larfors-Schneider, Otsuka-Takemoto, Klaewer-Schlechter)
- q.v., Kreuzer-Skarke Dataset (Halverson, Long, Nelson; McCallister-Stillman)
- q.v., Calabi-Yau volumes in AdS/CFT (Krefl-Seong)
- q.v., MSSM from orbifold models (Parr-Vaudrevange-Wimmer)
- q.v. Particle Masses Gal-Jejjala-Pena-Mishra . . .
- q.v. Knot invariants: Jejjala-Kar-Parrikar, Craven-Jejjala-Kar Gukov-Halverson-Ruehle-Sułkowski, using NLP
- YHH-Jejjala-Nelson NLP on ArXiv sections
- q.v. DEEP CONNECTIONS K. Hashimoto: AdS/CFT = Boltzmann Machine; Halverson-Maiti-Stoner: QFT = NN; de Mello-Koch: NN = RG; Vanchurin 2008: Universe = NN. Back to ML Maths YHH, 2011.14442 Review

Various Databases

- Kreuzer-Skarke: http://hep.itp.tuwien.ac.at/~kreuzer/CY/
  - new PALP: Braun-Walliser: ArXiv 1106.4529
  - Triang: Altmann-YHH-Jejjala-Nelson: http://www.rossealtman.com/
- CICYs: resurrected Anderson-Gray-YHH-Lukas, http://www-thphys.physics.ox.ac.uk/projects/CalabiYau/cicylist/index.html
- q.v. other databases of interesting to the math/physics community:

  Graded Rings/Varieties: Brown, Kasprzyk, et al. http://www.grdb.co.uk/

  Finite Groups/Rings: GAP https://www.gap-system.org/

  Modular Forms: Sutherland, Cremona et al. https://www.lmfdb.org/

  Knots & Invariants: KnotAtlas http://katlas.org/ Return

  ...

Universal Approximation Theorems

Large Depth Thm: (Cybenko-Hornik) For every continuous function $f : \mathbb{R}^d \to \mathbb{R}^D$, every compact subset $K \subset \mathbb{R}^d$, and every $\epsilon > 0$, there exists a continuous function $f_\epsilon : \mathbb{R}^d \to \mathbb{R}^D$ such that $f_\epsilon = W_2(\sigma(W_1))$, where σ is a fixed continuous function, $W_{1,2}$ affine transformations and composition appropriately defined, so that $\sup\limits_{x \in K} |f(x) - f_\epsilon(x)| < \epsilon$.

Large Width Thm: (Kidger-Lyons) Consider a feed-forward NN with n input neurons, m output neuron and an arbitrary number of hidden layers each with $n + m + 2$ neurons, such that every hidden neuron has activation function φ and every output neuron has activation function the identity. Then, given any vector-valued function f from a compact subset $K \subset \mathbb{R}^m$, and any $\epsilon > 0$, one can find an F, a NN of the above type, so that $|F(x) - f(x)| < \epsilon$ for all $x \in K$.

ReLU Thm: (Hanin) For any Lebesgue-integral function $f : \mathbb{R}^n \to \mathbb{R}$ and any $\epsilon > 0$, there exists a fully connected ReLU NN F with width of all layers less than $n + 4$ such that $\int_{\mathbb{R}^n} |f(x) - F(x)| dx < \epsilon$.

Back to Landscape