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**

M-Theory and Mathematics: classical and quantum aspects

NYU Abu Dhabi, Jan 2023

Enriching the Maths/Physics Dialogue

- Alg./diff. Geometry/topology - Rep. Theo : the right language for physics
  - Gravity $\sim$ Ricci 2-form of Tangent bundles;
  - Elementary Particles $\sim$ irred reps of the Lorentz group and sections of bundles with Lie structure group; Interactions $\sim$ Tensor products of sections ...
  - **String theory: brain-child of gauge-gravity geometrization tradition**
- A new exciting era for synergy with (pure & computational) geometry, group theory, combinatorics, number theory: *Sage, M2, GAP, LMFDB, GrDB* are becoming indispensible tools for physicists
- Interdisciplinary enterprise: cross-fertilisation of particle/string theory, phenomenology, pure mathematics, computer algorithms, data-bases, ...

Vacuum Degeneracy

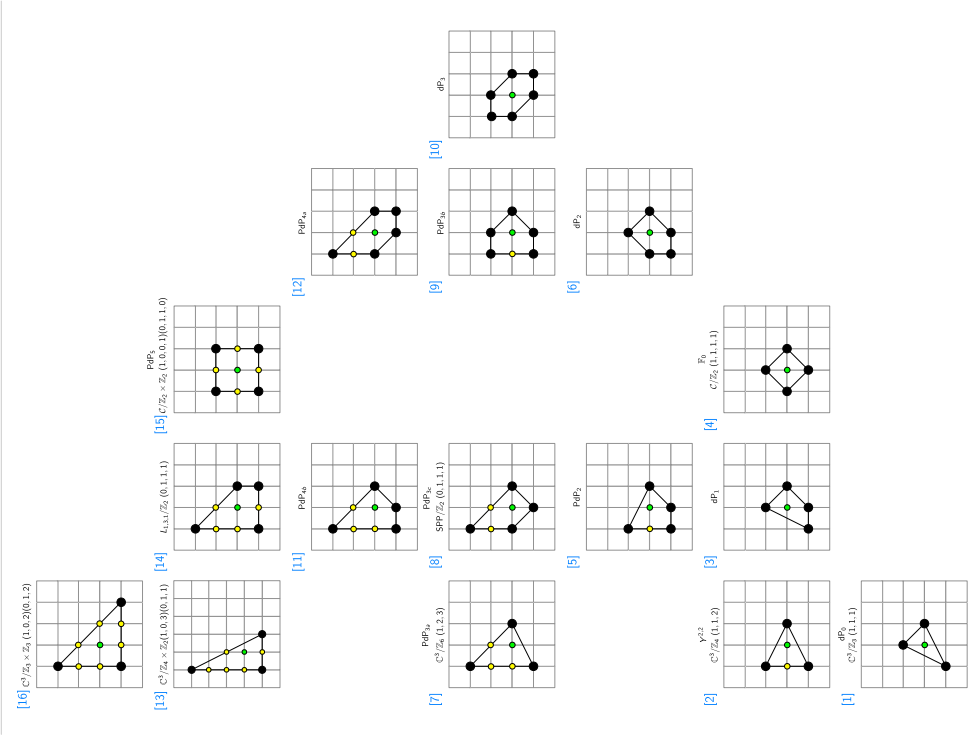
Perhaps the biggest theoretical challenge to string theory:

selection criterion??? metric on the landscape???

- Douglas (2003): Statistics of String vacua
- Kachru-Kallosh-Linde-Trivedi (2003): type II/CY estimates of 10^{500}
- Taylor-YN Wang (2015-7): F-theory estimates 10^{3000} to 10^{10^5}
- Basic Reason:

Algebraic Geometry $\rightsquigarrow$ Combinatorial Geometry $\rightsquigarrow$ Exponential Growth in dim

e.g., Borisov-Batyrev & Kreuzer-Skarke



- **Reflexive Polyhedra $\rightsquigarrow$ CY: anticanonical hypersurface in toric variety from Δ**
- **Dim 2: 16 up to $SL(2;\mathbb{Z})$ (Italian School 1890s)**
- **Dim 3: 4139 up to $SL(3;\mathbb{Z})$ (KS, 1999)**
- **Dim 4: 473800776 up to $SL(4;\mathbb{Z})$ (KS, 2000)**
- **Dim > 4: Open ??**

GrDB: Brown, Kaspryzyk, Nil, Kahle, ...`http://www.grdb.co.uk/`

Altman-Gray-YHH-Jejjala-Nelson (2014): brute-force: $\sim 10^6$ up to $h^{1,1} = 6$

Altman-Carifio-Halverson-Nelson (2018): estimated 10^{10^4} triangulations

Demirtas-Long-McAllister-Stillman (2019): all triang $240 \leq h^{1,1} \leq 491$

2017: String Theory enters the Machine-Learning Era

YHH (1706.02714); Krefl-Seong (1706.03346); Ruehle (1706.07024); Carifio-Halverson-Krioukov-Nelson (1707.00655)

Sophia: Hanson Robotics, HongKong

- Beginning of `String_Data` Annual conference series
- How can ML and modern data-science help with the vacuum degeneracy problem??

2017: String Theory enters the Machine-Learning Era

YHH (1706.02714); Krefl-Seong (1706.03346); Ruehle (1706.07024); Carifio-Halverson-Krioukov-Nelson (1707.00655)

Sophia: Hanson Robotics, HongKong

- Beginning of `String_Data` Annual conference series
- How can ML and modern data-science help with the vacuum degeneracy problem??
- Meanwhile . . . Sophia becomes a "human" citizen (in Saudi Arabia)

Progress in String Theory: Start Dates of Annual Series

1986- "Strings" Conference

2002- "StringPheno" Conference

2006 - 2010 String Vacuum Project (NSF)

2008 - ISGT Integrability in String/Gauge

2011- "String-Math" Conference (2020 - , M-theory & Maths Workshop)

2012- "Amplitudes"

2014- String/Theoretical Physics Session in SIAM Conference

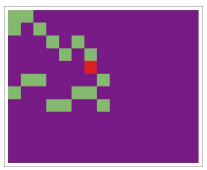
2017- "String-Data" Conference **(2022: Cambridge (Organizers Berman, YHH, Heyes, Hirst, Mishra)**

Algebraic Geometry as Image Processing

- A typical calculation:

$$X = \begin{pmatrix} 1&1&0&0&0&0&0&0 \\ 1&0&1&0&0&0&0&0 \\ 0&0&0&1&0&1&0&0 \\ 0&0&0&0&1&0&1&0 \\ 0&0&0&0&0&2&0&0 \\ 0&1&1&0&0&0&0&1 \\ 1&0&0&0&0&1&1&0 \\ 0&0&0&1&1&0&0&1 \end{pmatrix} \longrightarrow \boxed{\text{What Bourbaki teaches us}} \longrightarrow h^{2,1}(X) = 22$$

- Key to computational Algebraic Geometry: Gröbner basis, double-exponential complexity (unlike Gaussian elimination which is generalizes)
- [YHH 1706.02714] Deep-Learning the Landscape, *PLB 774, 2017*; (cf. Feature in *Science* , Aug, vol 365 issue 6452, 2019): think of it as an image processing problem



$\longrightarrow$ **What Machine-Learning teaches us** $\longrightarrow 22$

Machine Learning Mathematical Structures

Why stop at string/geometry?

q.v. Review Paper: YHH 2101.06317

Pattern Recognition: Human Eye

- $[0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, \ldots]$

Pattern Recognition: Human Eye

- $[0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, \ldots]$

  multiple of 3 or not.

- $[1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, \ldots]$

Pattern Recognition: Human Eye

- $[0,0,1,0,0,1,0,0,1,0,0,1,\ldots]$

  multiple of 3 or not.

- $[1,1,1,0,1,1,0,1,1,0,1,0,0,1,1,0,0,1,0,1,\ldots]$

  Prime or Not for odd integers.

- $[1,0,0,1,0,1,0,0,1,1,0,0,0,1,1,1,0,0,0,0,\ldots]$

Pattern Recognition: Human Eye

- $[0,0,1,0,0,1,0,0,1,0,0,1,\ldots]$

  multiple of 3 or not.

- $[1,1,1,0,1,1,0,1,1,0,1,0,0,1,1,0,0,1,0,1,\ldots]$

  Prime or Not for odd integers.

- $[1,0,0,1,0,1,0,0,1,1,0,0,0,1,1,1,0,0,0,0,\ldots]$

  Even/Odd of number of prime factors (Liouville Lambda)

Pattern Recognition: Machine-Learning

- Binary Classification of a Binary Vector (sliding window of, say, length 100); supervised learning: predict next one, e.g., Prime/Not becomes:

$$
\begin{array}{lcl}
\{0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,\ldots,0\} & \longrightarrow & 1 \\
\{1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,\ldots,1\} & \longrightarrow & 0 \\
\{0,0,0,1,0,0,0,0,0,1,0,0,1,0,0,0,0,1,1,0,0,\ldots,0\} & \longrightarrow & 1 \\
\ldots & \ldots & \ldots
\end{array}
$$

Pattern Recognition: Machine-Learning

- Binary Classification of a Binary Vector (sliding window of, say, length 100); supervised learning: predict next one, e.g., Prime/Not becomes:

$$\begin{array}{lll} \{0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,\ldots,0\} & \longrightarrow & 1 \\ \{1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,\ldots,1\} & \longrightarrow & 0 \\ \{0,0,0,1,0,0,0,0,0,1,0,0,1,0,0,0,0,1,1,0,0,\ldots,0\} & \longrightarrow & 1 \\ \ldots & \ldots & \ldots \end{array}$$

- pass to standard classifiers: SVW, Bayes, Nearest Neighbour; NN of the form $\mathbb{R}^{100} \overset{\text{linear}}{\longrightarrow} \mathbb{R}^{20} \overset{\tanh}{\longrightarrow} \mathbb{R}^{20} \overset{\textbf{Round}\sum}{\longrightarrow} \mathbb{Z}$, your kitchen sink, ...
- take 50,000 samples, 20-80 cross-validation, record (precision, MCC)
- similar performance for most: Mod3: (1.0, 1.0); PrimeQ, after balancing: (0.8, 0.6); Liouville Λ: (0.5, 0.001)

Thank you! Since 2017-

my fantastic students

Jiakang Bao, Elli Heyes, Ed Hirst

Tejas Acharya, Daatta Aggrawal, Malik Amir, Kieran Bull, Lucille Calmon, Siqi Chen, Suvajit Majumder, Maks Manko, Toby Peterken, Juan Pérez-Ipiña, Max Sharnoff, Yan Xiao

my wonderful collaborators

Physics: **Guillermo Arias-Tamargo, David Berman, Heng-Yu Chen, Andrei Constantin, Sebastián Franco, Vishnu Jejjala, Seung-Joo Lee, Andre Lukas, Shailesh Lal, Brent Nelson, Diego Rodriguez-Gomez, Zaid Zaz**

Algebraic Geometry: **Anthony Ashmore, Challenger Mishra, Rehan Deen, Burt Ovrut**

Number Theory: **Laura Alessandretti, Andrea Baronchelli, Kyu-Hwan Lee, Tom Oliver, Alexey Pozdnyakov, Drew Sutherland, Eldar Sultanow**

Representation Theory: **Mandy Cheung, Pierre Dechant, Minhyong Kim, Jianrong Li, Gregg Musiker**

Combinatorics: **Johannes Hofscheier, Alexander Kasprzyk, Shiing-Tung Yau**

How does one *DO* mathematics, I ?

Russell-Whitehead *Principia Mathematica* [1910s] (Leibniz, Frege, ...) axiomatize maths, 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: subset of *Principia*
- Type Theory [1970s] Martin-Löf, Coquand, . . . Coq: 4-color (2005); Feit-Thompson Thm (2012); Lean (2013); Univalent Foundation / Homotopy Type Theory [2006-] Voevodsky

Buzzard: "Future of Maths" 2019, ICM 2022 Davenport: ICM 2018 "Computer Assisted Proofs" Szegedy: more extreme view, computers > humans @ chess (1990s); @ Go (2018); @ Proving theorems (2030)

How does one *DO* mathematics, I ?

Russell-Whitehead *Principia Mathematica* [1910s] (Leibniz, Frege, ...) axiomatize maths, 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: subset of *Principia*
- Type Theory [1970s] Martin-Löf, Coquand, . . . Coq: 4-color (2005); Feit-Thompson Thm (2012); Lean (2013); Univalent Foundation / Homotopy Type Theory [2006-] Voevodsky

Buzzard: "Future of Maths" 2019, ICM 2022 Davenport: ICM 2018 "Computer Assisted Proofs" Szegedy: more extreme view, computers > humans @ chess (1990s); @ Go (2018); @ Proving theorems (2030)

We can call this Bottom-up Mathematics

How does one *DO* mathematics, II ?

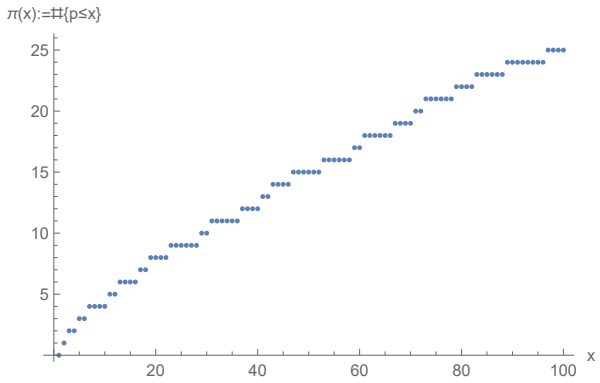
- Historically,

How does one *DO* mathematics, II ?

- 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?

How does one *DO* mathematics, II ?

- 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

Example I: 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

Example II: Combinatorics

[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); ?$\exists$ Hamilton cycles (0.8, 0.6); ?$\exists$ 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);

Example III: Quivers, Clusters, Brane setups, . . .

- [Bao-Franco-Hirst-Musiker, 2006.10783, Dechant-YHH-Heyes-Hirst 2203.13847] Recognition of mutation types (> 0.9)
- [Hirst-YHH-Peterken 2004.05218]: adjacency+permutation triple of dessin d'enfants; predicting transcendental degree > 0.9
- [Arias-Tamargo, YHH, Heyes, Hirst, Rodriguez-Gomez 2202.05845] Recognition of equivalence ($SL(2;\mathbb{Z})$, Seiberg, Hanany-Witten) of brane-webs
- [Cheung-Dechant-YHH-Heyes-Hirst-Li 2212.09771] learning Young tableaux representation of variables in Grassmannian cluster algebras (> 0.99)

Example IV: Number Theory

Arithmetic, A Classical Reprobate?

- [YHH 1706.02714, 1812.02893:]
  - Predicting primes $2 \to 3,\ 2,3 \to 5,\ 2,3,5 \to 7$; no way
  - PrimeQ: (0.7, 0.8); Sarnak's Challenger of Liouville Lambda (0.5, 0.001)
- [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
- Arithmetic Geometry: A Modern Hope? YHH-KH Lee-Oliver
  - 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]

Clearly useful for maths and physics

looking for new conjectures e.g.,

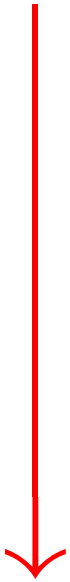
- '19 YHH-Kim: separating hyperplane - simple/non-simple groups; open
- '19 Brodie-Constantin-Lukas: exact formulae for cohomo surf.; proved.
- '20 YHH-Lee-Oliver: L-coefs and integer pt./torsion on ell; Known.
- '20 Craven-Jejjala-Par: Jones poly best-fit function; open
- '22 DeepMind-Oxford-Sydney,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

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. improves Donaldson 10-100 times

An Inherent Hierarchy?

- In decreasing precision/increasing difficulty:



numerical

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

algebra

string theory $\to$ combinatorics

analytic number theory

Please submit

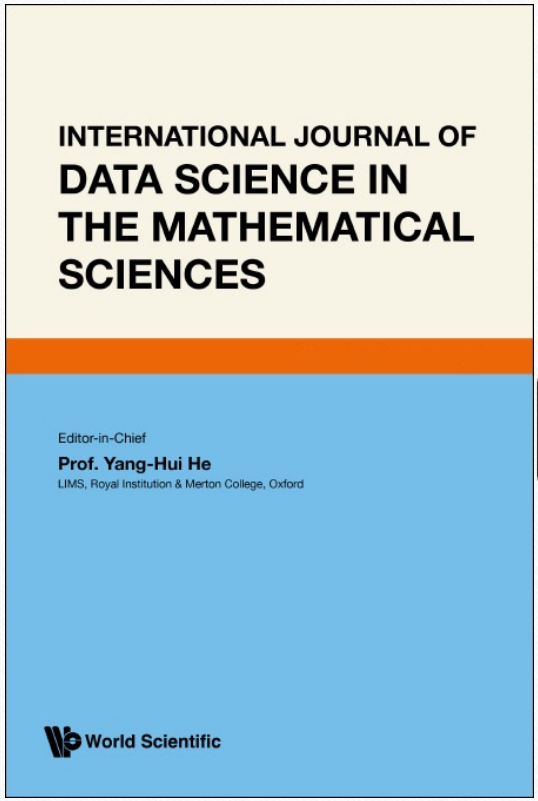
Launching in 2023

IJDSMS

Calling for Papers

Editor-in-Chief
Yang-Hui He
London Institute for Mathematical Sciences
& Merton College, University of Oxford
email: hey@maths.ox.ac.uk

More Information:
https://www.worldscientific.com/worldscinet/ijdsms



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

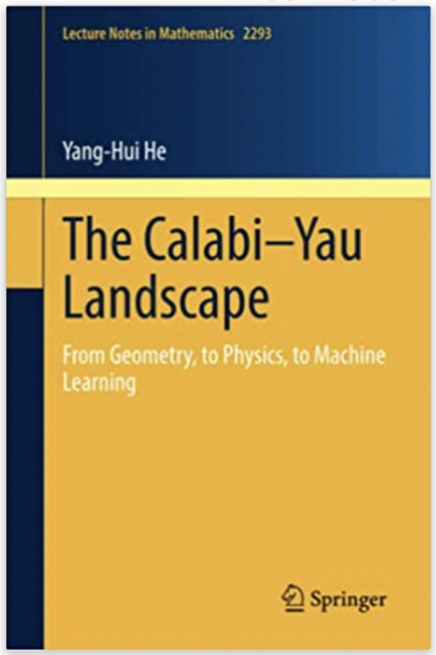
  holography + quantum + string + ads = extremal-black-hole

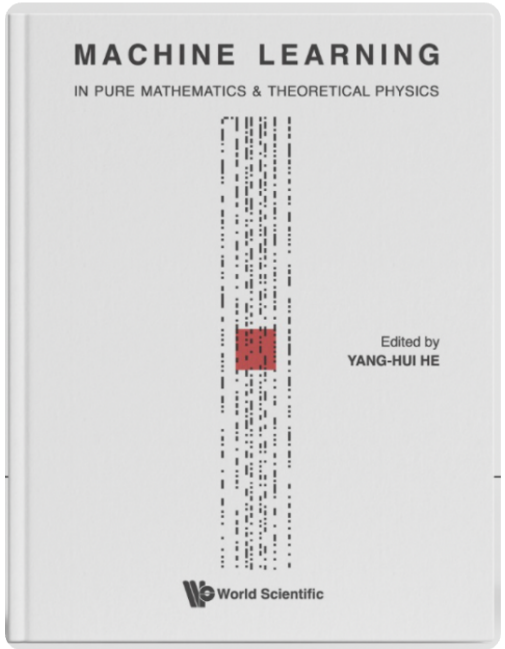
  string-theory + calabi-yau = m-theory + g2

  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

Please buy





BUY THESE BOOKS... ☺☺☺

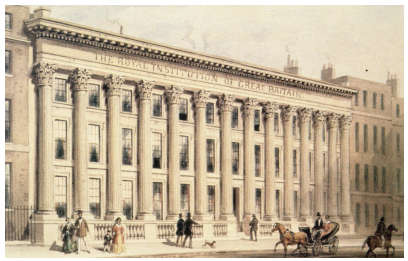
The London Institute for Mathematical Sciences

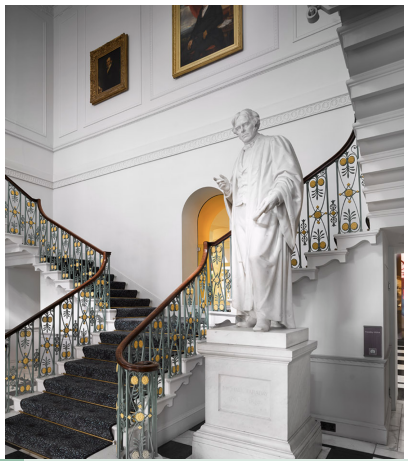
- UK's only independent research institute for maths; modelled after IAS, Princeton
- Founded in 2011 by Dr. Thomas Fink
- Housed in the Faraday Suites of the Royal Institution of Great Britain
- 1 of 23 themes: AI for Maths Discovery

  https://lims.ac.uk/event/ai-assisted-maths-discovery/
- Just established:

  Arnold Felowships

  Landau Fellowships





THANK YOU!

The Proper Way $\mathcal{O}(e^{e^d})$

- 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:

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

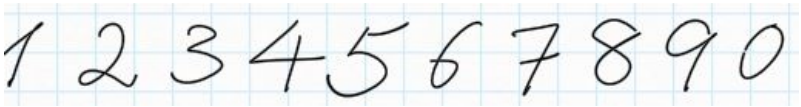
- Induces long exact sequence in cohomology:

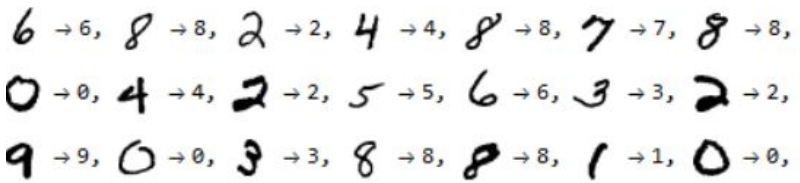
$$\begin{array}{ccccccc} 0 & \to & \cancel{H^0(X, T_X)}^{\,0} & \to & H^0(X, T_A|_X) & \to & H^0(X, N_X) \quad \to \\ & \to & \boxed{H^1(X, T_X)} & \stackrel{d}{\to} & H^1(X, T_A|_X) & \to & H^1(X, N_X) \quad \to \\ & \to & H^2(X, T_X) & \to & \cdots & & \end{array}$$

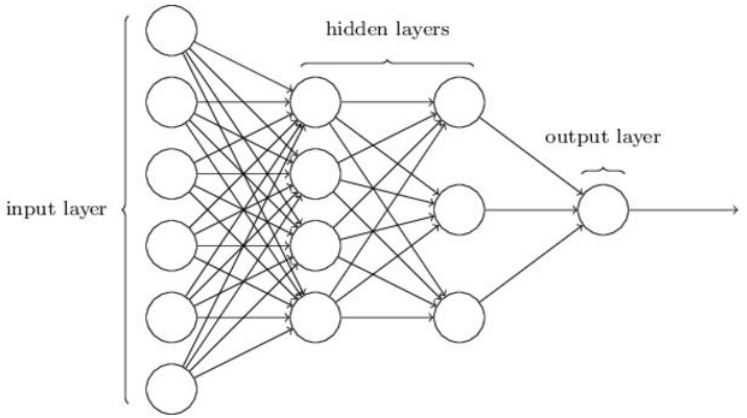
- Need to compute Rk(d), cohomology and $H^i(X, T_A|_X)$ (Cf. Hübsch)

Back to Alg Geo

The Neural Network Approach

- Bijection from  to $\{1, 2, \ldots, 9, 0\}$?
- Take large sample, take a few hundred thousand (e.g. NIST database)





- Data = Training Data ⊔ Validation Data

  Test trained NN on validations data to see accuracy performance

Universal Approximation Theorems

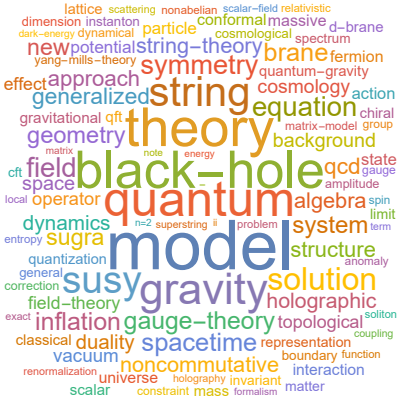
Large Depth Thm: (Cybenko-Hornik) For every continuous function $f : \mathbb{R}^d \to \mathbb{R}^D$, every compact subset $K \subset \mathbb{R}^d$, and every $\epsilon > 0$, there exists a continuous function $f_\epsilon : \mathbb{R}^d \to \mathbb{R}^D$ such that $f_\epsilon = W_2(\sigma(W_1))$, where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

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 $\int_{\mathbb{R}^n} |f(x) - F(x)| dx < \epsilon$.

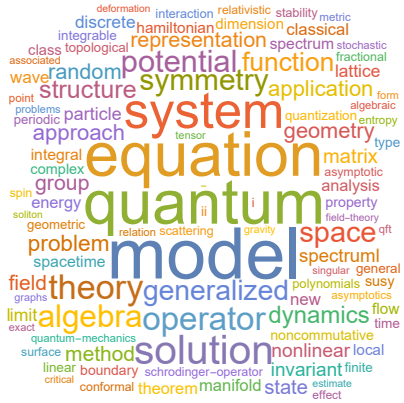
Back to NN@Alg Geo

ArXiv Word-Clouds

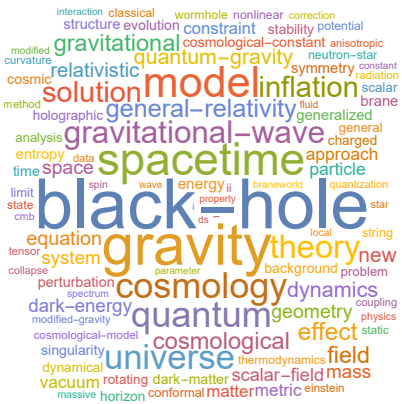


hep-th

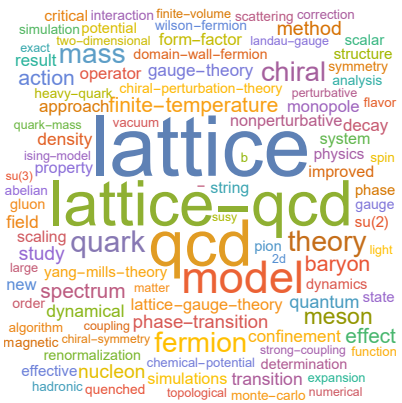
hep-ph



math-ph



gr-qc



hep-lat

Back to Word2Vec

Classifying Titles

Compare, + non-physics sections, non-science (Times), pseudo-science (viXra)

Actual \ Word2Vec + SVM	1	2	3	4	5
1	40.2	6.5	8.7	24.0	20.6
2	7.8	65.8	12.9	9.1	4.4
3	7.5	11.3	72.4	1.5	7.4
4	12.4	4.4	1.0	72.1	10.2
5	10.9	2.2	4.0	7.8	75.1

1 : hep-th
2 : hep-ph
3 : hep-lat
4 : gr-qc
5 : math-ph

Actual \ NN	1	2	3	4	5	6	7	8	9	10
viXra-hep	11.5	47.4	6.8	13.	11.	4.5	0.2	0.3	2.2	3.1
viXra-qgst	13.3	14.5	1.5	54.	8.4	1.8	0.1	1.1	2.8	3.

6: cond-mat, 7: q-fin, 8: stat, 9: q-bio, 10: Times of India

Back to Main