The world in a Grain of Sand:

Condensing the String Vacuum Degeneracy

YANG-HUI HE

London Institute for Mathematical Sciences, Royal Institution

Dept of Mathematics, City, *University of London*

Merton College, *University of Oxford*

School of Physics, *NanKai University*

String_Data 2021 @ (Virtual) Jo'burg

Based on: 2111.04761, YHH, **Shailesh Lal**, M. Zaid Zas

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}
- Basically: Combinatorial geometry usually tends exponentially

  e.g., Kreuzer-Skarke (2000s): Reflexive polytopes up to $SL(n;\mathbb{Z})$:

  1, 16, 4319, 473800776, ???

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

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

Searching the Standard Model

SM places some constraints but still not enough:

- Braun-YHH-Ovrut; Bouchard-Cvetic-Donagi (2005): exact MSSM particles
- Gmeiner-Blumenhagen-Honecker-Lüst-Weigand (2005):1 in 10^9 in D-brane MSSM modles
- Candelas-de la Ossa-YHH-Szendroi (2007): Triadophilia $\Rightarrow$ "des res"?
- Anderson-Gray-Lukas-Palti (2012-3): Het line bundle MSSM: 200 in 10^{10}

Recent estimates

- Constatin-YHH-Lukas; Deen-YHH-SJ Lee-Lukas (2018-9) MSSM from heterotic line bundles: 10^{23} from CICYs; 10^{723} from KS
- Cvetic-Halverson-Lin-Liu-Tian (2019): 10^{15} F-theory MSSMs

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

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

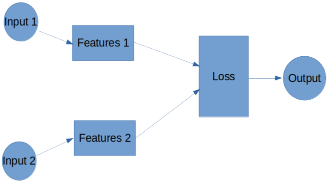
- Beginning of `String_Data`
- How can ML and modern data-science help with the vacuum degeneracy problem??
- Meanwhile . . . Sophia becomes a "human" citizen (in Saudi Arabia)

One-Shot Learning

Fei-Fei Li et al. (2002 -)

- Estimated that by 6, a child has learnt all $10 \sim 30 \times 10^3$ object categories
  NOT done by sampling % of cases in each category
- could not have supervise learnt everything in the standard way!
- Knowledge Transfer: having seen lots of horses and a single bird, would recognize a chicken is closer to a bird than to a horse
- a SINGLE representative in a category suffices, or at most a handful $\rightsquigarrow$ Few-Shot Learning

Siamese Neural Networks (SNN)



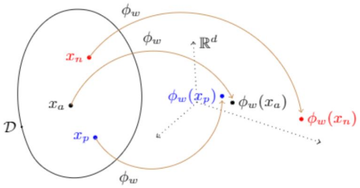
Loss $= \mathcal{L}(w) :=$

$$\max\{d_w(x_a, x_p) - d_w(x_a, x_n) + 1, 0\}$$

$$d_w(x_1, x_2) := (\phi_w(x_1) - \phi_w(x_2))^2$$

ϕ representation by features network (FN)

FN: represents the data by mapping to $\mathbb{R}^3$, say: $\phi : \mathcal{D} \to \mathbb{R}^3$:



a anchor point for the class;

p close-by; n far-apart

FN some appropriately chosen NN

SNN returns a similarity score $\in [0, \infty)$

where 0 means identical

Concrete Model of the Landscape: CICY D-folds

Earliest data in algebraic geometry (Candelas et al. 1980s);

first to be subjected to a (naive) experiment in ML (YHH, 2017)

$$(q^i_j) = \left[\begin{array}{c|cccc} n_1 & q^1_1 & q^2_1 & \ldots & q^K_1 \\ n_2 & q^1_2 & q^2_2 & \ldots & q^K_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_m & q^1_m & q^2_m & \ldots & q^K_m \end{array}\right]_{m \times K}$$

- Complete Intersection Calabi-Yau D-fold
- K eqns of multi-degree $q^i_j \in \mathbb{Z}_{\geq 0}$ embedded in $\mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_m}$
- $\sum\limits_{r=1}^{m} n_r = k + D$
- $c_1(X) = 0 \rightsquigarrow \sum\limits_{j=1}^{K} q^j_r = n_r + 1$
- M^T also CICY

Rmk: can uniformize configuration by

(1) -1 padding; (2) wrap padding; or (3) bilinear interpolation

CICY3

- classified by Candelas, Dale, Green, Hubsch, Lutken (1988-9) q.v. Hubsch's **Bestiary**
- successful supervised ML experiments: YHH (1706.02714) Bull-YHH-Jejjala-Mishra (1806.03121, 1903.03113), Krippendorf-Syvaeri (2003.13679) Erbin-Finotello (2007.13379; 2007.15706), Larfors-Lukas-Ruehle-Schneider (2111.01436)
- 7890 configurations, $h^{1,1} \in [1, 19]$; $h^{2,1} \in [15, 101]$
  (m, K) ranges from $(1, 1)$ to $(12, 15)$

CICY4

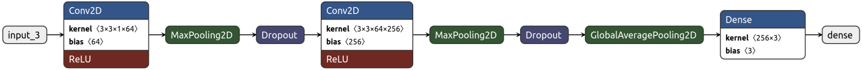
- classified by Gray, Haupt, Lukas (2013-4)
- successful supervised ML experiments: YHH-Lukas (2009.02544), Erbin-Finotello-Schneider-Tamaazousti (2108.02221)
- 905684 configurations, $h^{1,1} \in [1, 24]$; $h^{2,1} \in [1, 33]$; $h^{3,1} \in [20, 426]; h^{2,2} \in [204, 1752]$
  (m, K) ranges from $(1, 1)$ to $(16, 20)$

Methodology

Labelled Data of the form $(q_j^i) \longrightarrow h^{1,1}$ where similarity is

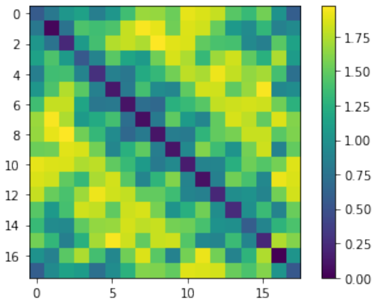
$$q^{(A)} \sim q^{(B)} \quad \text{iff} \quad h^{(A)} = h^{(B)}$$

- Represent each CICY as pixelated image (after normalization), and use CNN as FN (tried other architectures like Inception and MLP):

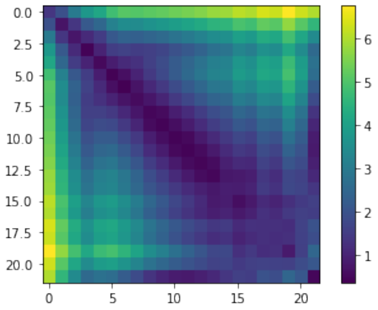


- trained on 3% of CICY3 and 0.6% of CICY4 (mostly just few per class of $h^{1,1}$): Few-Shot ML hundreds to extrapolate to hundreds of thousands
- Standard ADAM optimizer @ learning-rate of 0.01

Mean Similarity Scores on Pairs

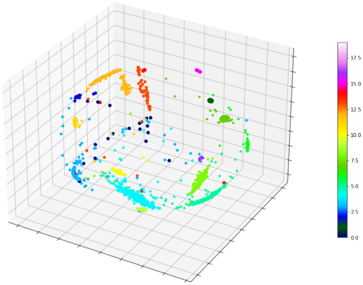


$CICY3$

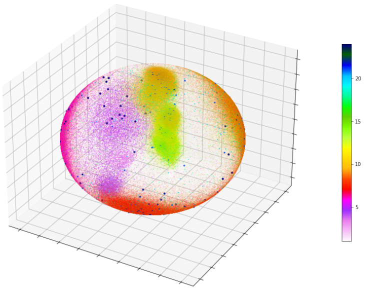


$CICY4$

Clustering of CICY by $h^{1,1}$



$CICY3$



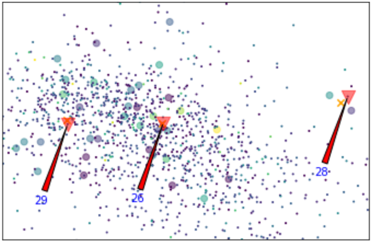
$CICY4$

Typicality

Zoom in onto a particular $h^{1,1}$ for CICY3

- e.g. take $h^{1,1} = 7$, we have $h^{2,1} \in [23, 49]$ using k-means clustering

  (cf. Otuska-Takemoto 2020, supervised clustering of heterotic line bundles)



so *typical* complex structure are 26, 27 and 28

Conclusion & Outlook

- Two-birds with one stone
  1. Few-shot ML of the landscape
  2. The similarity score gives a distance measure on the landscape
- This reduction + distance: a step toward a vacuum selection principle given the complexity of the landscape
- concrete baby example of CICYs; clearly should try to see about string standard models

Another Fun Landscape: Arithmetic Geometry

- Alessandretti-Baronchelli-YHH 1911.02008: initial ML/TDA@Birch-Swinnerton-Dyer
- Hirst-YHH-Peterken 2004.05218 Grothendieck's dessin d'enfants: predicting transcendental degree (0.92, 0.9)
- YHH-KH Lee-Oliver arithmetic 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)

ML BSD

- 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:])

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

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

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