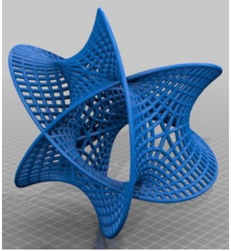
Machine Learning over Complete Intersection Calabi-Yau Manifolds



Workshop on Machine Learning Landscape

ICTP, Trieste, Italy

Challenger Mishra, ICMAT Madrid

based on 1806.03121, and upcoming

December 12, 2018

1. Physics Motivations
2. Calabi-Yau manifolds in String Theory
3. Machine Learning Calabi-Yau Geometries

Collaborators & Consultants

Humans

Yang-Hui He: Maths, City; Physics, NanKai

Vishnu Jejjala: Physics, Wits

Kieran Bull: Physics, Leeds

Yarin Gal: Computer Science, Oxford

Dvijotham Krishnamurthy: Google DeepMind

Machines

Hydra Computing Cluster: Oxford Physics

The Spam filter that discovered the Higgs Boson, or why ML is impressive

(even before the Higgs discovery at CERN)

Quarantined spam and bulk messages since Apr 11 09:00

quarantine@physics.ox.ac.uk

To: Challenger Mishra

Inbox

Tuesday, April 18,

```
The following messages were quarantined by Sophos because they appear to be
either spam or bulk messages. To request that a message be released from the
quarantine and delivered to you, click the message ID and send the request.
If your mail client does not support HTML, reply to this message and delete
lines that correspond to messages you do not want approved. To release all
messages in the list, simply reply to this message.
```

ID	Type	Time	From	Subject
[=0]	Spam High	09:24:36	dg.office@cern.ch	The Higgs boson does not exist

!

```
To change your anti-spam settings and preview quarantined emails please login using your physics department u
at the following web address. Only give your username, PHYSICS\ is not accepted as a prefix.

https://quarantine.physics.ox.ac.uk/

This is an automated email on behalf of Ian McArthur, IT Manager
```

Unification and String Theory

String theory is the only known consistent theory of quantum gravity.

- Postulates extra-dimensions of space.
- Relies on a fundamental symmetry between matter particles and force carriers, called supersymmetry (SUSY).

Unification and String Theory

String theory is the only known consistent theory of quantum gravity.

- Postulates extra-dimensions of space.
- Relies on a fundamental symmetry between matter particles and force carriers, called supersymmetry (SUSY).

String theory is (also) an organising principle for mathematics.

String Compactification

String theory unifies gravity and QM and reduces to the Standard Model (SM) in the low energy limit, via an intermediate Grand Unified Theory (GUT)[1].

$$\text{String Theory} \longrightarrow \text{GUT} \longrightarrow \text{SM}$$

This is called string 'compactification' where the low energy theory, SM, is recovered by hiding away or compactifying over the extra-dimensions of space.

This places severe geometrical constraints on the extra-dimensions of string theory.

[1] compactifications without an intermediate GUT also possible

String Phenomenology

The Holy grail: Embed the Standard Model (SM) of particle physics in its full glory within the framework of string theory.

1. Reproduce the particle content, coupling constants, masses of particles of the SM.
2. Explain the origin of discrete symmetries of SM that help explain unobserved couplings, the long lifetime of the proton, etc.
3. Other challenges: Explain fine tuning, moduli stabilisation, supersymmetry breaking.
4. No such model till date, but there has been considerable progress. Only a handful of string-derived Sandard Models until c. 2010[2]. Since then there are have been tens of thousands! This is primarily due to innovative mathematical constructions, and increased computational prowess.

[2] heterotic CY compactifactions

Discrete Symmetries in particle physics

- Discrete Symmetries are hypothesised in the 4 dimensional theory (SM) to explain the occurrence or absence of certain physical phenomena.
- Example 1: The discrete symmetry group
$$\Delta(27) := (\mathbb{Z}_3 \times \mathbb{Z}_3) \rtimes \mathbb{Z}_3 \subset SU(3)$$
is often invoked to explain the structure of the mismatch of quantum states in a flavor-changing weak process in the SM involving quarks (CKM) or neutrinos (PMNS).
- Example 2: An R-symmetry is often invoked to explain why the proton is stable and does not decay in a MSSM.
- But the origin of such hypothesised symmetries is not understood! In superstring theory they are thought to descend from isometries of the compactification space.

Discrete Symmetries and String theory

- Since most known CYs are simply-connected, most quasi-realistic string models are built over the quotient of a CY manifold by a freely acting discrete symmetry group.
- Flux lines around the irreducible paths of the manifold allow breaking of the GUT gauge group to the Standard Model gauge group, which may not be possible using a simply-connected CY.

$$\text{String Theory} \quad \longrightarrow \quad \text{GUT} \quad \longrightarrow \quad \text{SM}$$

- In addition, if the CY quotient manifold on which the string model is built, has any remnant discrete symmetry, such a symmetry might survive the gauge group breaking above, to appear as symmetries of the low energy SM, explaining in part, the origin of such discrete symmetries!

Calabi-Yau manifolds in String theory

Calabi-Yau Compactifiactions of the Heterotic String

- CY compactifications of the Heterotic String is one of the most promising avenues for string model building.

- The space-time for the effective field theory is the direct product: $\mathcal{M}_4 \times X_6$, where $\mathcal{M}_4$ is a maximally symmetric space.

- If X_6 is Riemannian, irreducible and we demand $N = 1$ supersymmetry in the 4-dimensional theory (SM), then $\mathrm{Hol}(X_6) = SU(3)$. Do such manifolds exist?

- Calabi conjecture (proved by Yau): An n-dimensional complex Kähler manifold with vanishing first Chern class admits a metric with $SU(n)$ holonomy. This leads us to the class of Calabi-Yau manifolds. Thus X_6 is a CY threefold.

Calabi-Yau Geometry: Generalities

A Calabi-Yau manifold of complex dimension n is a compact Kähler[3] manifold (X, J, g) with

- vanishing first Chern class, or,
- holonomy group $SU(n)$, or ,
- a globally defined and nowhere vanishing holomorphic n-form.

where, J is the complex structure, and g is the metric.

[3] Hermitian manifold with a closed (1,1) form.

Moduli space of Calabi-Yau threefolds

The total parameter space of a CY manifold consists of parameters related to its structure as a complex manifold and parameters related to the deformations of its Kähler metric.

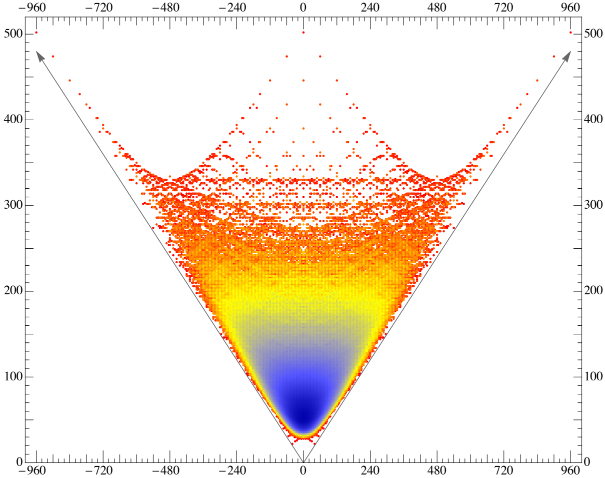
1. $h^{1,1}(M) = \dim H^{1,1}(M)$ is intimately related to the dimension of the Kähler structure moduli space of M.
2. $h^{2,1}(M) = \dim H^{1,2}(M)$ is intimately related to the dimension of the Complex structure moduli space of M.
3. Calabi-Yau threefolds come in mirror pairs, (M, W), such that $H^{2,1}(W) \cong H^{1,1}(M)$ and $H^{1,1}(W) \cong H^{2,1}(M)$. Roughly speaking, the complex structure moduli is exchanged with the Kähler structure moduli. This is the basic idea behind mirror symmetry.

Calabi-Yau threefold Geometry: Hodge Numbers

$h^{p,q} = \dim\ H^{p,q}(M)$:

$$
\begin{array}{ccccc}
 & & h^{m,m} & & \\
 & h^{m,m-1} & \vdots & h^{m-1,m} & \\
h^{m,0} & \cdots & & \cdots & h^{0,m} \\
 & h^{1,0} & \vdots & h^{1,0} & \\
 & & h^{0,0} & &
\end{array}
\qquad
\begin{array}{ccccccc}
 & & & 1 & & & \\
 & & 0 & & 0 & & \\
 & 0 & & h^{1,1} & & 0 & \\
1 & & h^{2,1} & & h^{2,1} & & 1 \\
 & 0 & & h^{1,1} & & 0 & \\
 & & 0 & & 0 & & \\
 & & & 1 & & &
\end{array}
$$

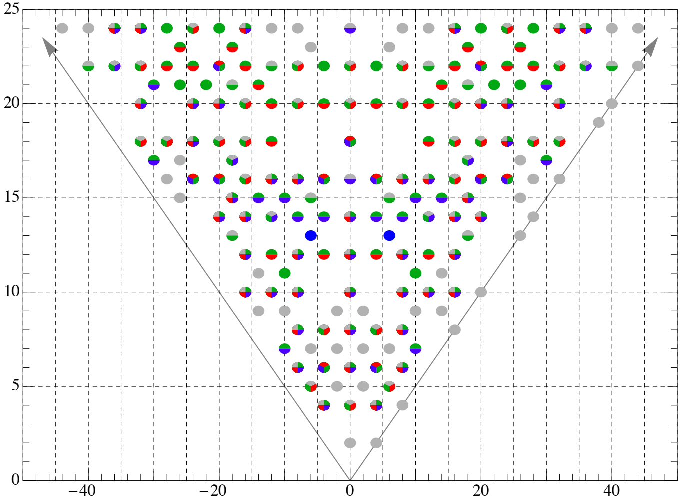
Many possible Calabi-Yau geometries: The Hodge Plot



x-axis: Euler Characteristic, y-axis: 'Height' $(h^{1,1} + h^{2,1})$

473,800,776 data points

Many possible Calabi-Yau geometries: Tip of the Hodge Plot



Calabi-Yau Threefolds With Small Hodge Numbers: Candelas, Constantin, CM, *Fort. der. Physik* (2018), 1602.06303

Constructing Calabi-Yau Manifolds

1. Submanifolds of $\mathbb{C}^m$ are not very interesting: a connected compact analytic submanifold of $\mathbb{C}^m$ is a point!
2. $\mathbb{CP}^m$ is compact; all its closed complex submanifolds are also compact.
3. Theorem due to Chow states that all such submanifolds of $\mathbb{CP}^m$ can be realized as the zero locus of a finite number of homogeneous polynomial equations, e.g., the Fermat quintic defined as a hypersurface in $\mathbb{CP}^4$ below:

$$\text{Fermat Quintic: } \{x \in \mathbb{CP}^4 \ \mid \ \sum_{a=0}^{4} x_a^5 = 0\}$$

Complete Intersection Calabi-Yau Manifolds

Taking cue from the Fermat quintic, one can construct Complete Intersection Calabi-Yau Manifolds $\subset \mathbb{CP}^{n_1} \times \ldots \times \mathbb{CP}^{n_m}$.

$$X = \begin{array}{c} \mathbb{CP}^{n_1} \\ \vdots \\ \mathbb{CP}^{n_m} \end{array} \begin{bmatrix} q_1^1 & \ldots & q_K^1 \\ \vdots & \ddots & \vdots \\ q_1^m & \ldots & q_K^m \end{bmatrix}, \sum_a q_a^r = n_r + 1, \forall r \in \{1, \ldots, m\}$$

X denotes the family of CY-threefolds defined by the vanishing locus of K polynomials. q_a^r is the multi-degree of the a^{th} polynomial in the r^{th} projective space $\mathbb{CP}^{n_r}$.

Example: $X = \mathbb{CP}^4[5]$: $X = \{x \in \mathbb{CP}^4 \mid p(x) = 0\}$, where p is the most general degree 5 polynomial in the 5 homogeneous co-ordinates of $\mathbb{CP}^4$.

The list of Complete Intersection Calabi-Yau Threefolds

$$X = \begin{array}{c} \mathbb{CP}^{n_1} \\ \vdots \\ \mathbb{CP}^{n_m} \end{array} \left[\begin{array}{ccc} q_1^1 & \dots & q_K^1 \\ \vdots & \ddots & \vdots \\ q_1^m & \dots & q_K^m \end{array} \right]_{\chi}^{h^{1,1}, h^{2,1}} , \sum_a q_a^r = n_r + 1, \forall r \in \{1, \dots, m\}$$

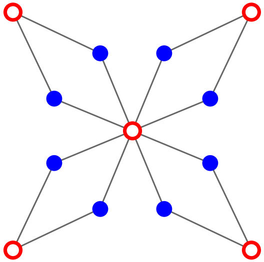
$K = N_1 + N_a + 3$, $N_1 \leq 9$, $N_a \leq 6$ $N_1 = \#\, \mathbb{CP}^1$ factors, $N_a = \#$ other factors

- 7890 CY threefold families in the CICY list.
- At least 2590 are known to be distinct as classical manifolds.
- Only 266 distinct pairs $(h^{1,1}, h^{2,1})$ of Hodge numbers.
- $0 \leq h^{1,1} \leq 19$, $0 \leq h^{2,1} \leq 101$.
- $\chi \in [-200, 0]$ and is computable from the config matrix.
- For comparison, there are 921,497 CICY fourfold configuration matrices, most of which correspond to elliptically fibered Calabi-Yaus. For these,

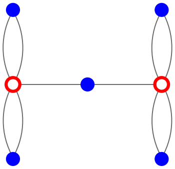
$$4h^{1,1} - 2h^{1,2} + 4h^{1,3} - h^{2,2} + 44 = 0.$$

Complete Intersection Calabi-Yau Manifolds: Examples[4]

$$\begin{array}{c} \mathbb{CP}^1 \\ \mathbb{CP}^1 \\ \mathbb{CP}^1 \\ \mathbb{CP}^1 \\ \mathbb{CP}^7 \end{array} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^{5,37}_{-64}$$



$$\begin{array}{c} \mathbb{P}^4 \\ \mathbb{P}^4 \end{array} \begin{bmatrix} 2 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 2 \end{bmatrix}^{12,\ 28}_{-32}$$



[4] Note the bipartite graph representation.

Favourability of CICYs

A CICY is favourable if its entire second cohomology descends from that of the ambient space. Favourable CICYs are especially amenable to the construction of stable holomorphic vector and monad bundles, leading to quasi-realistic heterotic string models.

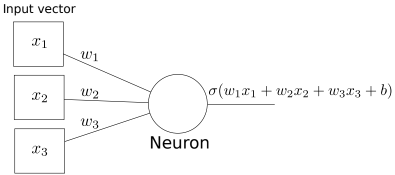
$\sim 62\%$ of all CICYs are favorable creating a balanced dataset.

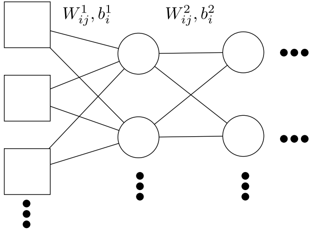
All but 48 CICY configuration matrices can be 'made' favourable. The remaining can be seen to be favourably embedded in a product of del Pezzo surfaces.

$$\begin{array}{c} \mathbb{P}^4 \\ \mathbb{P}^4 \end{array} \left[\begin{array}{ccccc} 2 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 2 \end{array} \right]^{12,28}_{-32} \subset dP_4 \times dP_4$$

Can *favorability* of CICYs be learnt by ML tools?

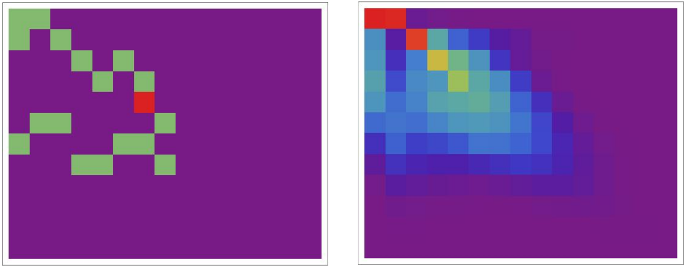
Machine Learning Tools: Neural Network





Schematic representation of feedforward neural network. The top figure denotes the perceptron (a single neuron), the bottom, the multiple neurons and multiple layers of the neural network.

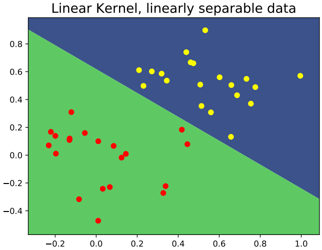
Complete Intersection Calabi-Yau Manifolds: (in visual form)

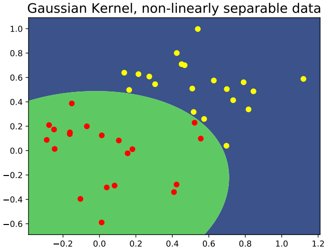


A typical and an average Complete Intersection CY manifold, borrowed from *Deep-Learning the Landscape, 1706.02714, Yang-Hui He*.

Machine Learning Tools: Support Vector Machines

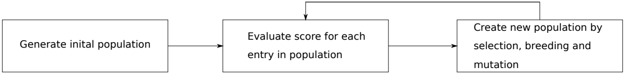
- The simplest SVM is a binary classifier for linearly separable data.
- The classification is performed by finding an optimal hyperplane that can separate clusters of points from the two classes in the feature space.
- This can be extended to tackle non-linearly separable data (using the so called kernel trick) and data that have multiple classes.
- An SVM regressor chooses the flattest line which fits the data within an allowed residue ϵ.





SVM separation boundary (calculated using our cvxopt implementation with a randomly generated data set.)

Neural Net and SVM Architecture



- A genetic algorithm fixes optimal hyperparameters for the Neural Network such as number of hidden layers, number of neurons in each, activation functions, and dropout[5].

- We use the quadratic programming Python package Cvxopt to solve the SVM optimization problem. We employ a Gaussian kernel. The hyperparameters (standard deviation, cost variable[6], and residue[7]) are selected by hand.

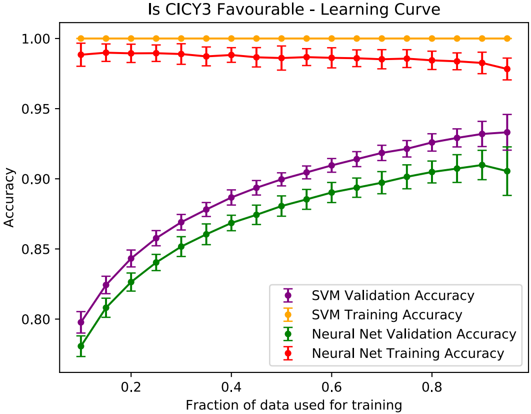
- Keras Python package with TensorFlow backend to implement the Neural Network. Performed on a Lenovo Y50 laptop, i7-4700HQ, 2.4 GHz quad core with 16 GB RAM.

[5] Dropout provides a way to counter overfitting, by randomly dropping neurons along with their connections from the neural network during training.

[6] To counter overfitting in SVMs and allow better generalisation to unseen data, one can allow a few training points to be misclassified.

[7] for SVM regressors

Experiment 1: Machine Learning *Favourability*



	Accuracy	WLB	WUB
SVM Class	$\mathbf{0.933 \pm 0.013}$	0.867	0.893
NN Class	0.905 ± 0.017	**0.886**	**0.911**

Errors were obtained by averaging over 100 random cross validation splits.

High accuracy and speed. Can other CICY properties be learnt with such accuracies?

Computing Hodge Numbers

CICY threefolds:

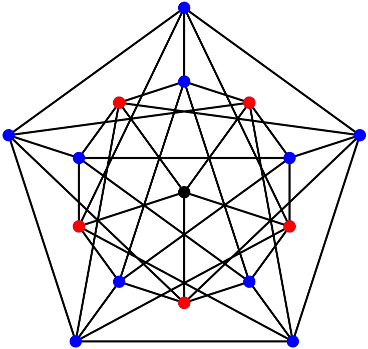
1. Complete Intersection Calabi-Yau Manifolds, Candelas, Dale, Lütken, Schimmrigk, Nuclear Physics B 298.3 (1988): 493-525

CICY Quotients:

2. New Calabi-Yau Manifolds with Small Hodge Numbers, Candelas, Davies, arXiv:0809.4681
3. Completing the Web of $\mathbb{Z}_3$ - Quotients of Complete Intersection Calabi-Yau Manifolds, Candelas, Constantin, arXiv:1010.1878
4. Hodge Numbers for CICYs with Symmetries of Order Divisible by 4, Candelas, Constantin, CM, arXiv:1511.01103
5. Calabi-Yau Threefolds With Small Hodge Numbers, Candelas, Constantin, CM, arXiv:1602.06303
6. Hodge Numbers for All CICY Quotients - Constantin and Lukas arXiv:1607.01830

Computing Hodge Numbers

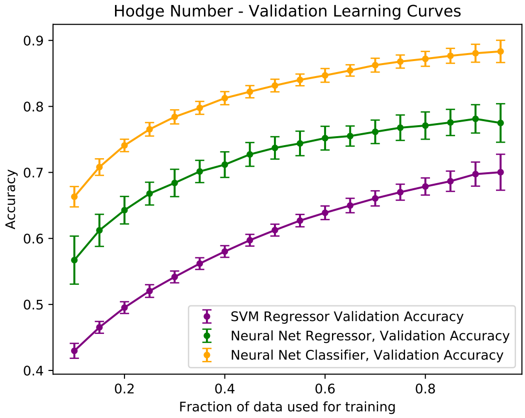
Computing the Hodge numbers is non-trivial and they have been painstakingly computed using computers whenever possible and often by understanding the algebraic-geometry of the manifold in all its detail (often more gratifying).



Hodge Numbers for CICYs with Symmetries of Order Divisible by 4, Candelas, Constantin, CM, arXiv:1511.01103

Experiment 2: Machine Learning Hodge number $h^{1,1}$

- $\chi = 2(h^{1,1} - h^{2,1})$ is computable directly from the CICY matrix.
- Choice between learning $0 \leq h^{1,1} \leq 19$ and $0 \leq h^{2,1} \leq 101$.



Experiment 2: Machine Learning Hodge number $h^{1,1}$

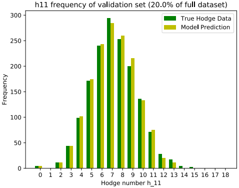
	Accuracy	RMS	R^2	WLB	WUB
SVM Reg	0.70 ± 0.02	$\mathbf{0.53 \pm 0.06}$	$\mathbf{0.78 \pm 0.08}$	0.642	0.697
NN Reg	0.78 ± 0.02	0.46 ± 0.05	0.72 ± 0.06	0.742	0.791
NN Class	$\mathbf{0.88 \pm 0.02}$	-	-	**0.847**	**0.886**

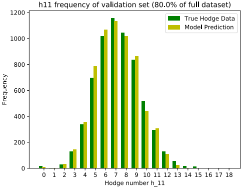
Errors were obtained by averaging over 100 different random cross validation splits using a cluster. The Neural Net classifier yields high accuracy.

Experiment 2: Machine Learning Hodge number $h^{1,1}$

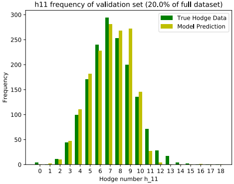
20% **80%**

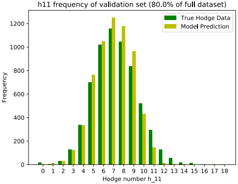
NN classifier



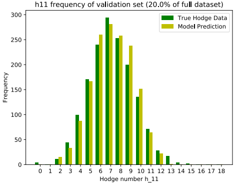


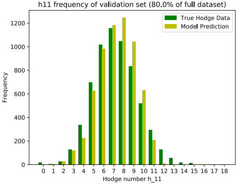
NN regressor





SVM regressor



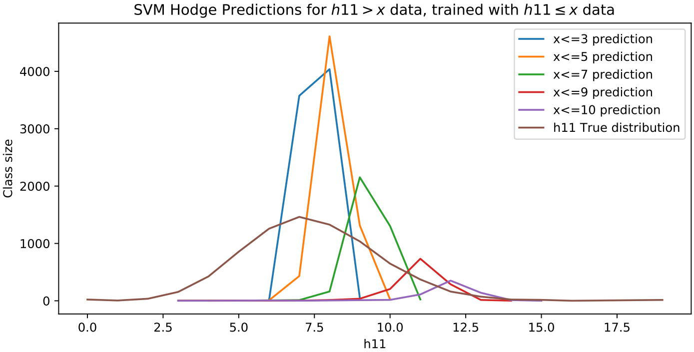


Experiment 2: Machine Learning $h^{1,1}$

- The methodology so far does not address the fundamental technical problem we encounter when studying Calabi-Yau compactification: the difficulty of a calculation increases with the Hodge numbers and the dimension. At the same time, any systematic survey of the string landscape is infeasible.
- All explicit Standard Model constructions are on manifolds with Hodge numbers of $O(1)$. Triangulating polytopes to populate the toric Calabi-Yau database stopped at $h^{1,1} = 6$.
- We would therefore like to develop techniques such that the training and validation sets are different in character.
- We aim to train with the easy cases and use the machine to predict solutions to harder problems for which the calculations are more intricate or where the answers could be unknown.
- We organize the CICY dataset into a low $h^{1,1}$ training set and a high $h^{1,1}$ validation set and provide proof of concept that such an extrapolation is possible.

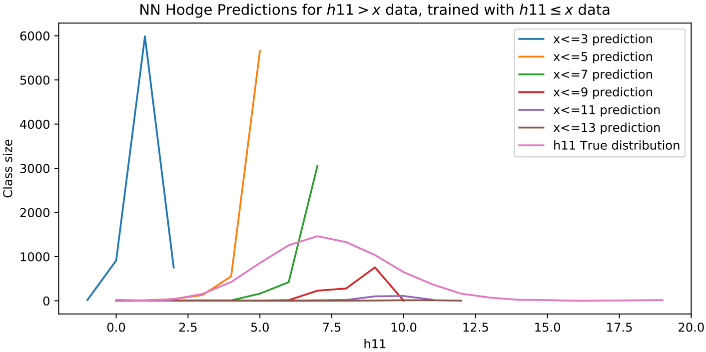
Experiment 2: Machine Learning $h^{1,1}$

SVM predictions of $h^{1,1}$ for CICY threefolds.
Bull, Hui-He, Jejjala, CM, *upcoming*.



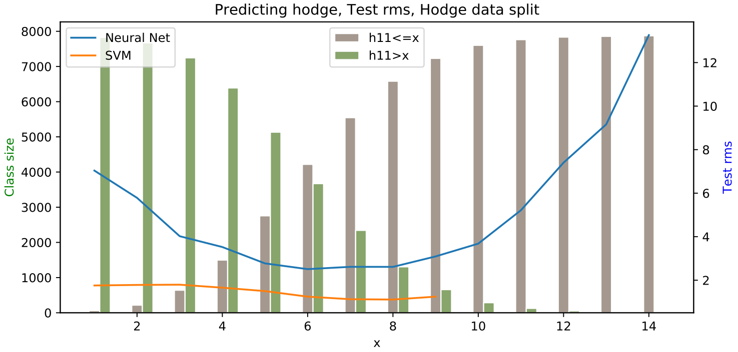
Experiment 2: Machine Learning $h^{1,1}$

Neural network regressor predictions of $h^{1,1}$ for CICY threefolds.
Bull, Hui-He, Jejjala, CM, *upcoming*.



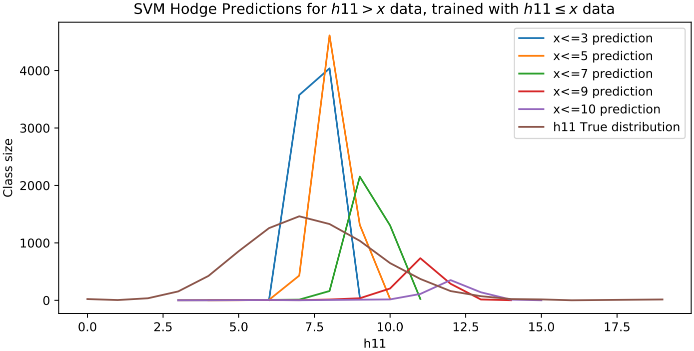
Experiment 2: Machine Learning $h^{1,1}$

Accuracy of predictions of $h^{1,1}$ for CICY threefolds.
Bull, Hui-He, Jejjala, CM, *upcoming*.



- Brown bars: size of training set; Green: size of validation set.
- The rms decreases with increasing x, as expected, but starts increasing after a certain point, since the problem becomes very unbalanced.

Experiment 2: Machine Learning $h^{1,1}$



- This analysis shows that the algorithms are capable of predicting trends in the distribution of Hodge numbers from the limited data.
- Both algorithms seem to predict a lot of values below the x in query, which is natural.
- The SVM performs much better than the Neural Net. Achieves an RMS error of 1 when only seeing data with $h^{1,1} \leq 7$.

Experiment 3: Machine Learning CY symmetries

Fantastically symmetric Calabi-Yaus and where to find them.

Experiment 3: Machine Learning CY symmetries

The datasets:

- Candelas, Davies, Braun (2011): Only 2.5% of all CICYs admit group action by a freely acting group (G_f). A highly imbalanced dataset.
- Lukas, CM (2017): Of these manifolds, 25% have residual (non-freely acting) discrete symmetries (G_Y), acting trivially on the complex structure moduli space, of which 30% are R-symmetries (useful for ruling out proton decay channels). A more balanced dataset but much smaller in size.

$$G_Y \in \{\mathbb{Z}_2,\ \mathbb{Z}_3,\ \mathbb{Z}_4,\ \mathbb{Z}_2^2,\ \mathbb{Z}_2^3,\ \mathbb{D}_8,\ \mathbb{Z}_2^4,\ \mathbb{Z}_2\times\mathbb{D}_8,\ (\mathbb{Z}_3\times\mathbb{Z}_3)\rtimes\mathbb{Z}_2\}\ .$$

Experiment 3: Machine Learning CY symmetries

Exciting new observations:

- Candelas, CM (2017): At special points in the complex structure moduli space, there are enhanced symmetries, while still preserving the generality of a large number of complex structure moduli.
- Candelas, Lukas, CM (upcoming): We report large discrete symmetry groups in CY threefolds. We find a group of order 1944 containing $\Delta(27)$, (possibly $\Delta(27) \rtimes \mathbb{Z}_3 \rtimes SL_{2,3}$) in a CY on which there is a 3 generation SM. This is also quite possibly largest discrete symmetry group on a smooth Calabi-Yau threefold ever found (to our knowledge!)
- Distinct possibility of such symmetries appearing in the 4d theory to explain structure of mixing matrices.

Experiment 3: Machine Learning CY symmetries

$$\text{CICY} \quad X \quad \hookrightarrow \quad \mathcal{A} = \mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_m}$$

$$\boxed{\begin{array}{ccccccc} \mathbf{G_f} & \hookrightarrow & \mathbf{N^{\star}_{G}(G_f)} & \hookrightarrow & N_G(G_{\mathrm{f}}) & \hookrightarrow & G = \mathrm{Aut}_{\mathrm{L}}(\mathcal{A}) \\ & & & & \big\uparrow & & \\ & & & & C_G(G_{\mathrm{f}}) & & \\ & & & & \big\uparrow & & \\ & & & & \mathbf{C^{\star}_{G}(G_f)} & & \end{array}}$$

$$N_G(G_{\mathrm{f}})/C_G(G_{\mathrm{f}}) \ \subset \ \mathbf{Aut(G_f)}$$

$$\mathbf{G_f} \ \lhd \ N_G(G_{\mathrm{f}}),\ \mathbf{N^{\star}_{G}(G_f)}; \quad \mathbf{G_Y} = \mathbf{N^{\star}_{G}(G_f)/G_f}$$

Experiment 3: Machine Learning CY symmetries

Given a CICY configuration, can we predict if the CICY admits any freely acting group?

A binary classification problem, but very unbalanced!

Experiment 3: Machine Learning CY symmetries

We need different benchmarks for unbalanced data such as F-values, AUC.
Confusion matrix:

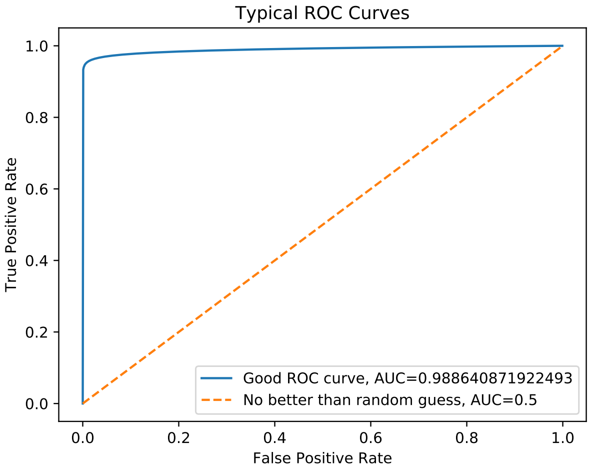
		Actual	
		True	False
Predicted Classification	True	True Positive (tp)	False Positive (fp)
	False	False Negative (fn)	True Negative (tn)

$$\text{TPR (recall)} := \frac{tp}{tp+fn}, \qquad \text{FPR} := \frac{fp}{fp+tn},$$

$$\text{Accuracy} := \frac{tp+tn}{tp+tn+fp+fn}, \qquad \text{Precision} := \frac{tp}{tp+fp}.$$

- $F := \frac{2}{\frac{1}{\text{Recall}}+\frac{1}{\text{Precision}}}, \quad 0 \leq F \leq 1.$
- AUC, or, Area Under ROC (Receiver Operating Characteristic). ROC plots TPR against FPR; $0.5 \leq \text{AUC} \leq 1$.

Experiment 3: Machine Learning CY symmetries



Typical ROC curves. The points above the diagonal represent classification results which are better than random.

Experiment 3: Machine Learning CY symmetries

SMOTE	SVM AUC	SVM max F	NN AUC	NN max F
0	0.77 ± 0.03	0.26 ± 0.03	0.60 ± 0.05	0.10 ± 0.03
100	0.75 ± 0.03	0.24 ± 0.02	0.59 ± 0.04	0.10 ± 0.05
200	0.74 ± 0.03	0.24 ± 0.03	0.71 ± 0.05	0.22 ± 0.03
300	0.73 ± 0.04	0.23 ± 0.03	0.80 ± 0.03	0.25 ± 0.03
400	0.73 ± 0.03	0.23 ± 0.03	0.80 ± 0.03	0.26 ± 0.03
500	0.72 ± 0.04	0.23 ± 0.03	0.81 ± 0.03	0.26 ± 0.03

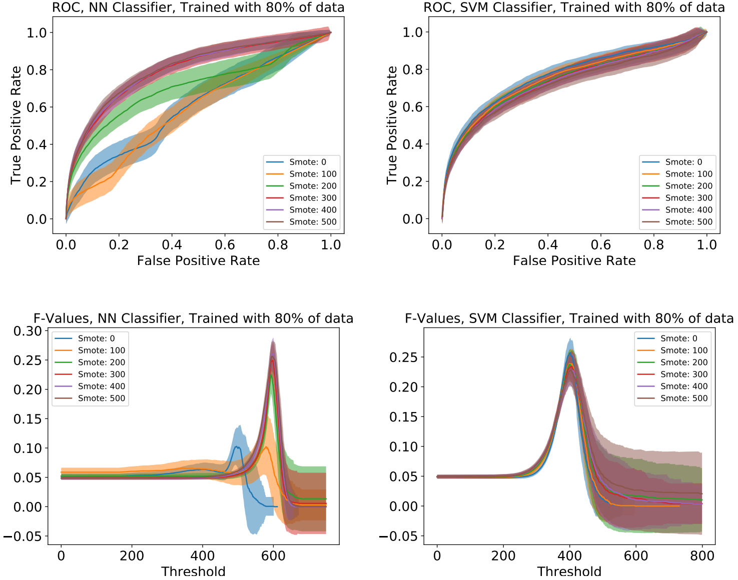
Metrics for predicting freely acting symmetries. Errors were obtained by averaging over 100 random cross validation splits using a cluster.

- SMOTE helps NN slightly, but not SVM.
- Very challenging to predict whether a CICY admits a freely acting symmetry!

Possbile Directions

- The same analysis could be applied to the KS dataset and more naturally to CICY fourfolds. Compare with existing results.
- This would require creation of further datasets, e.g. discrete symmetry dataset for CICY fourfolds.
- Explore further ML techniques to extrapolate (even better) to complex geometries by training only with simpler geometries.
- Keep pushing the boundaries of our stringy understanding of nature with the newly acquired ally that is Machine Learning!

Grazie!



ROC and F-curves generated for both SVM and neural network for several SMOTE values