



# City Research Online

## City St George's, University of London

**Citation:** Bao, J., He, Y-H., Hirst, E. & Pietromonaco, S. (2021). Lectures on the Calabi-Yau Landscape. Fields Institute Monographs, 2293, pp. 1-203. doi: 10.1007/978-3-030-77562-9

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/23685/>

**Link to published version:** <https://doi.org/10.1007/978-3-030-77562-9>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

# Lectures on the Calabi-Yau Landscape

Jiakang Bao<sup>1,2\*</sup>, Yang-Hui He<sup>1,3,4†</sup>, Edward Hirst<sup>1,2‡</sup>, Stephen Pietromonaco<sup>5§</sup>

<sup>1</sup>*Department of Mathematics, City, University of London, EC1V 0HB, UK*

<sup>2</sup>*Department of Physics, Imperial College London, SW7 2AZ, UK*

<sup>3</sup>*Merton College, University of Oxford, OX14JD, UK*

<sup>4</sup>*School of Physics, NanKai University, Tianjin, 300071, P.R. China*

<sup>5</sup>*Department of Mathematics, University of British Columbia, V6T 1Z2, Canada*

## Abstract

In these lecture notes, we survey the landscape of Calabi-Yau threefolds, and the use of machine learning to explore it. We begin with the compact portion of the landscape, focusing in particular on complete intersection Calabi-Yau varieties (CICYs) and elliptic fibrations. Then we examine non-compact Calabi-Yau manifolds which are manifest in Type II superstring theories. They arise as representation varieties of quivers, used to describe gauge theories in the bulk familiar four dimensions. Finally, given the huge amount of Calabi-Yau data, whether and how machine learning can be applied to algebraic geometry and string landscape is also discussed. These notes are directed to the beginning graduate student interested in mathematics and in physics, and are based on lectures given by the 2<sup>nd</sup> author at the 2019 PIMS Summer School on Algebraic Geometry in High-Energy Physics at the University of Saskatchewan.

---

\*[jiakang.bao18@imperial.ac.uk](mailto:jiakang.bao18@imperial.ac.uk)

†[hey@maths.ox.ac.uk](mailto:hey@maths.ox.ac.uk)

‡[Edward.Hirst@city.ac.uk](mailto:Edward.Hirst@city.ac.uk)

§[spietro@math.ubc.ca](mailto:spietro@math.ubc.ca)

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>I</b>	<b>Compact Calabi-Yau Landscape</b>	<b>5</b>
<b>2</b>	<b>Calabi-Yau Geometry in Math and Physics</b>	<b>5</b>
2.1	Topological Data . . . . .	7
2.2	String Compactifications . . . . .	9
<b>3</b>	<b>C.I.C.Y.</b>	<b>9</b>
3.1	Cyclic Calabi-Yau Threefolds . . . . .	9
3.2	CICY Calabi-Yau Threefolds . . . . .	11
<b>4</b>	<b>Elliptically Fibered Calabi-Yau Threefolds</b>	<b>13</b>
<b>5</b>	<b>Additional Regions of the Compact Landscape</b>	<b>15</b>
<b>II</b>	<b>Non-compact Calabi-Yau Landscape</b>	<b>16</b>
<b>6</b>	<b>String Theory Structures</b>	<b>16</b>
6.1	D-branes . . . . .	16
6.2	Quivers . . . . .	17
6.3	An Orbifold Example: $\mathbb{C}^3/\mathbb{Z}_3$ . . . . .	18
6.4	McKay Correspondence . . . . .	19
<b>7</b>	<b>Algebraic Geometry Viewpoint</b>	<b>20</b>
7.1	Brane Tilings . . . . .	20
7.2	Dessin d'Enfants . . . . .	22
<b>8</b>	<b>Non-compact Calabi-Yau Summary</b>	<b>24</b>
<b>III</b>	<b>Machine-Learning the Landscape</b>	<b>26</b>
<b>9</b>	<b>Performance Measures: Hypersurfaces in <math>WP^4</math></b>	<b>26</b>
<b>10</b>	<b>Learning CICYs</b>	<b>28</b>
10.1	Distinguishing Elliptic Fibrations . . . . .	31
<b>11</b>	<b>A Digression: Group Theory</b>	<b>33</b>
11.1	Learning Cayley Tables . . . . .	33

---

11.2 Learning Finite Simple Groups . . . . .	33
<b>12 Summary and Outlook</b>	<b>34</b>
<b>A Some Complex Geometry</b>	<b>35</b>
A.1 Kähler Manifolds . . . . .	36
A.2 Chern Classes . . . . .	36
<b>B Toric Varieties</b>	<b>37</b>
<b>C Introduction to Machine Learning</b>	<b>39</b>
C.1 Text recognition . . . . .	39
C.2 Neural Networks . . . . .	41
C.3 Support Vector Machines . . . . .	43
C.4 Decision Trees . . . . .	46
C.5 Types of Machine Learning . . . . .	48
<b>References</b>	<b>49</b>

# 1 Introduction

Superstring theories demand our spacetime dimension to be 10, which means we should reduce them to an effectively 4-dimensional theory. The standard solution of string compactification, as a generalization of Kaluza-Klein compactification, renders the extra six dimensions Calabi-Yau (CY). Thus, the study of Calabi-Yau and algebraic geometry has entered the field of theoretical physics.

In order to avoid an excess of symmetries in our observed 4-dimensional universe, isometries in our geometry, which leads to extra graviphotons, is not allowed [1]. This leaves us the only option of manifolds of complex dimension 3, which requires Kähler structure and vanishing first Chern classes ( $c_1 = 0$ ). As will be explained in §2.2, we also want the manifold to be Ricci-flat. However, given a Kähler manifold with zero  $c_1$ , the existence of a (unique) Kähler metric in the same Kähler class with vanishing Ricci form is not self-evident. Followed by the work of Calabi [2] and Yau [3, 4], mathematicians reached a great success in studying CY manifolds. Later, physicists realized the crucial role CY manifolds play in fundamental physics as aforementioned. Discoveries in physics enabled people to reconstruct the Standard Model from compactifications and also led to the mirror symmetry which is now a focused interface of mathematics and physics [5]. More details and discussions on the physical predictions from CY manifolds can be found in [1]. Nowadays, thanks to the information age, we are able to let machines help us learn the structure of CY manifolds due to the large volume of data which has been compiled since the mid-1980s by physicists and mathematicians. This even brings computer science and data science into this interdisciplinary area.

The outline is organized as follows. In Part I, we mainly focus on compact CY landscape. We start with a background on Calabi-Yau geometry. We also pay our attention to the complete intersection Calabi-Yaus (CICYs). Then we contemplate the non-compact case in Part II. In this part, more physics and mathematics, such as quivers and toric varieties, and their relations are discussed. Finally, we apply machine learning to the study of CY landscape in Part III. Along with a quick introduction to machine learning, we perform this technique to different topics in mathematics. In the appendices, some prerequisites are provided.

## Part I

# Compact Calabi-Yau Landscape

Some basic topological or geometric facts are given in Appendix A. For far more detailed treatment on what follows, we refer the reader to [6–9].

## 2 Calabi-Yau Geometry in Math and Physics

The story of Calabi-Yau manifolds originates in the mid-1950s with the following conjecture of Eugenio Calabi.

**Conjecture 2.1. (*The Calabi Conjecture*)** *Let  $(X, g, \omega)$  be a compact Kähler manifold, and fix  $R \in \Omega^{1,1}(X)$  such that  $[R] = c_1(T_X) \in H^{1,1}(X)$ . Then there exists a unique Kähler metric  $\tilde{g}$  with Kähler form  $\tilde{\omega}$  such that  $[\omega] = [\tilde{\omega}]$ , and*

$$R = \text{Ric}(\tilde{\omega})$$

where  $\text{Ric}(\tilde{\omega})$  is the Ricci form of  $\tilde{\omega}$ .

The power of this conjecture is that it describes complicated geometric data (curvature) in terms of simpler topological data (Chern classes). For example, in complex dimension 1, this conjecture reduces to the Gauss-Bonnet theorem for Riemann surfaces, which says that the curvature is determined completely by the genus. In higher dimensions, the conjecture is that the curvature is controlled by the first Chern class (of the tangent bundle).

Calabi himself proved the uniqueness part of his conjecture, but the existence remained an open problem for 20 years before Shing-Tung Yau completed the proof, for which he received the Fields Medal in 1982.

**Theorem 2.2. (*Yau*)** *The Calabi conjecture holds.*

We will be primarily interested in the special case of  $R = 0$ , in which we say that  $X$  admits a *Ricci-flat* metric. In general relativity, Riemannian manifolds with Ricci-flat metrics are vacuum solutions of Einstein’s equations (that is, solutions without matter and energy). We are therefore interested in such manifolds which are Kähler. This leads us to the definition of a Calabi-Yau manifold<sup>1</sup>.

**Definition 2.3.** Let  $X$  be a compact Kähler manifold with  $\dim_{\mathbb{C}}(X) = n$ . We say  $X$  is a Calabi-Yau  $n$ -fold if it admits a Ricci-flat metric<sup>2</sup> of strictly  $\text{SU}(n)$  holonomy.

<sup>1</sup>In fact, the word “Calabi-Yau” was coined by physicists later [5] for Ricci-flat Kähler manifolds.

<sup>2</sup>Yau’s proof of the Calabi conjecture was not constructive, and to-date, there is not a single compact Calabi-Yau manifold where the Ricci-flat metric is known explicitly (outside of trivial cases of tori). This is an important open problem.

Let us give some low-dimensional examples of Calabi-Yau manifolds:

1. The only Calabi-Yau manifold of (complex) dimension 1 is an elliptic curve. Thus, there is a single topological type.
2. The Calabi-Yau manifolds of complex dimension 2 are called K3 surfaces. A simple construction is as a smooth quartic hypersurface in  $\mathbb{P}^3$ . All K3 surfaces are simply connected, and diffeomorphic to one another; so there is only one topological type. (Note that 4-dimensional tori are indeed Ricci flat, but they do not satisfy the condition on the holonomy group in the definition.)

**Proposition 2.4.** *For  $X$  as in the definition, the following are equivalent<sup>3</sup>:*

1.  $X$  is a Calabi-Yau  $n$ -fold.
2. The first Chern class of  $X$  vanishes;  $c_1(T_X) = 0$ .
3. There exists a covariantly constant spinor on  $X$ .
4. There exists a non-vanishing holomorphic  $n$ -form on  $X$ .
5.  $X$  is a smooth projective algebraic variety with trivial canonical line bundle  $\omega_X \cong \mathcal{O}_X$ , where  $\omega_X = \bigwedge^n T_X^*$ , and which additionally satisfies  $H^k(X, \mathcal{O}_X) = 0$  for  $0 < k < n$ .

The final characterization in the proposition is clearly the preferable one in algebraic geometry. We can remove the hypothesis of projectivity, which results in non-compact Calabi-Yau manifolds, of interest to us in Part II. We could also allow for mild singularities, which inevitably arise when studying families of Calabi-Yau manifolds.

**Remark 2.5.** One must beware of mildly different definitions of Calabi-Yau. Our definition excludes all tori (in particular, abelian varieties) and, for example, the threefold  $K3 \times E$ ; the product of a K3 surface and an elliptic curve. These spaces admit Ricci-flat metrics, though of holonomy strictly contained in  $SU(n)$ . In physics, this will translate into the low-energy theory having enhanced supersymmetry. Both abelian threefolds and  $K3 \times E$  are of interest in enumerative geometry.

---

<sup>3</sup>There are some subtleties in these propositions. The second one is actually weaker. For instance, complex tori with dimension greater than one have vanishing first Chern classes, but they fail to satisfy the fifth one. On the other hand, people often count these as Calabi-Yaus as they have trivial holonomies and infinite fundamental groups. Moreover, we also have non-algebraic K3 surfaces that fail the fifth condition even though they are simply connected with holonomy  $SU(2)$  [10]. Anyway, people adopt different definitions in different literature. This won't be an issue in our applications.

## 2.1 Topological Data

One can assign to a complex manifold  $X$  the Hodge cohomology groups

$$H^{p,q}(X) := H^q(X, \Omega_X^p)$$

with Hodge numbers  $h^{p,q}$  the corresponding dimensions. If  $X$  is compact and Kähler, the topological Euler characteristic is given by

$$\chi(X) = \sum_{p,q=1}^{\dim X} (-1)^{p+q} h^{p,q}. \quad (2.1)$$

If  $X$  is a compact Calabi-Yau threefold, then due to various symmetries [7–9] the only relevant Hodge numbers are  $h^{1,1}$  and  $h^{2,1}$ . By Proposition 2.4,  $X$  is a smooth projective variety with vanishing  $h^{1,0}, h^{2,0}$  and therefore by the Hodge decomposition

$$H^{1,1}(X) \cong H^2(X, \mathbb{C}).$$

We can choose an integral basis  $\{J_k\}_{k=1, \dots, h^{1,1}}$  of  $H^2(X, \mathbb{C})$  such that the Kähler cone is  $\mathcal{K} = \{ \sum_k t_k J_k \mid t_k \in \mathbb{R}_{>0} \}$ . In other words, the quantity  $h^{1,1}$  measures the number of Kähler classes on  $X$  (or by dualizing, the number of curve classes). Using the Calabi-Yau condition, we similarly have

$$H^{2,1}(X) \cong H^1(X, T_X).$$

The cohomology group on the right encodes the infinitesimal deformations of the complex/algebraic structure of  $X$ .

Therefore on a Calabi-Yau threefold, the Hodge number  $h^{2,1}$  measures the dimension of the space of complex/algebraic deformations, while  $h^{1,1}$  measures the dimension of the Kähler cone. The two Hodge numbers determine the topological Euler characteristic via (2.1)

$$\chi(X) = 2(h^{1,1} - h^{2,1}). \quad (2.2)$$

Using the chosen basis of  $\mathcal{K}$  we define the triple intersection form of  $X$

$$d_{rst} = \int_X J_r \wedge J_s \wedge J_t.$$

This integral can be hard to compute in general, but we can use the following result [6, Thm. 1.3]. If we have an embedding  $f : X \hookrightarrow A$  with  $A$  a smooth projective variety of dimension  $m + 3$ , then for all  $\omega \in H^k(A)$

$$\int_X \omega|_X = \int_A \omega \wedge \eta$$

where  $\eta$  is a  $(m, m)$ -form which when restricted to  $X$  is the top Chern class of the normal bundle  $\mathcal{N}_{X/A}$ . For our purposes,  $A$  will be a simpler space than  $X$  itself; for example, a projective space or product of projective spaces.

For any Kähler threefold, the total Chern class can be written in the chosen basis of  $\mathcal{K}$  as

$$c(T_X) = 1 + \sum_{r=1}^{h^{1,1}} [c_1(T_X)]_r J_r + \sum_{r,s=1}^{h^{1,1}} [c_2(T_X)]_{rs} J_r \wedge J_s + \sum_{r,s,t=1}^{h^{1,1}} [c_3(T_X)]_{rst} J_r \wedge J_s \wedge J_t.$$

Moreover, the topological Euler characteristic of a Kähler manifold  $X$  is the integral over  $X$  of the top Chern class of  $T_X$ . Using the triple intersection form, we can therefore express

$$\chi(X) = \sum_{r,s,t=1}^{h^{1,1}} d_{rst} [c_3(T_X)]_{rst}.$$

For a Calabi-Yau threefold, of course  $c_1(T_X) = 0$ , so that leaves  $c_2(T_X)$  to be independently specified.

**Theorem 2.6 (Wall).** *The topological type of a compact Calabi-Yau threefold is completely determined by the Hodge numbers  $h^{p,q}$ , the triple intersection form  $d_{rst}$ , and the second Chern class  $c_2(T_X)$ .*

It is convenient to contract  $c_2$  with  $d$  by defining  $[c_2(T_X)]_r := \sum_{s,t} [c_2(T_X)]_{rs} d_{rst}$ . It suffices to record this contraction instead of the individual components  $[c_2(T_X)]_{rs}$ . Therefore, by Theorem 2.6, the data determining the topological type of a Calabi-Yau threefold is:

$$\left\{ (h^{1,1}, h^{2,1}), [c_2(T_X)]_r, d_{rst} \right\} \quad r, s, t = 1, \dots, h^{1,1}. \quad (2.3)$$

Recall from Section 2 that for Calabi-Yau manifolds of dimensions 1 and 2, there is respectively a single topological type. Does this pattern persist in dimension 3? Spectacularly, no. The lower bound on the number of topological types of Calabi-Yau threefolds is currently around 500,000,000! But there is the following conjecture.

**Conjecture 2.7 (Yau).** *The number of topological types of Calabi-Yau threefolds is finite<sup>4</sup>.*

In other words, there are finite possibilities for the values in the data set (2.3).

**Remark 2.8.** Beware that even after fixing the topological type of the Calabi-Yau, there is still generally a moduli of algebraic/complex structures on the variety of fixed type. This is typical of moduli problems: specify as much discrete data as possible, which fixes the topological type, and then study families of complex structures.

---

<sup>4</sup>In fact, this conjecture is made for any CY  $n$ -folds. It is certainly true for  $n = 1, 2$ .

## 2.2 String Compactifications

Calabi-Yau threefolds entered physics through string theory in the late 80s. The consistency of the physical string theories (Type I, Type IIA, Type IIB, and the Heterotic theories) remarkably requires that the (real) dimension of spacetime be 10. So we obviously have to contend with the fact that we only observe 4 dimensions. The idea behind *string compactifications* is to decompose the 10-dimensional spacetime  $M_{10}$  as

$$M_{10} = M_4 \times X \tag{2.4}$$

where  $M_4$  is our 4-dimensional spacetime, and  $X$  is a compact 6-dimensional manifold. The vague intuition should be that the extra 6 dimensions of  $X$  are tightly curled-up and unobservable at small energies.

If  $X$  is a complex threefold, then it has real dimension 6. But why do we want  $X$  to additionally be Calabi-Yau? It is because Calabi-Yau manifolds are those admitting Ricci-flat metrics. In general relativity Ricci-flat manifolds correspond to a vacuum configuration of spacetime, i.e. a universe without matter or energy. Therefore, compactifying on a Calabi-Yau threefold  $X$ , as in (2.4), models a string theory vacuum.

Let us tie this back in with our exploration of the Calabi-Yau landscape. The vague principle one should keep in mind is:

As  $X$  varies over the compact Calabi-Yau landscape, the physics observed in  $M_4$  changes. In other words, the topology and geometry of  $X$  dictates physical phenomena in spacetime.

## 3 Complete Intersections in Products of Projective Spaces (CICYs)

In this section we begin constructing our first examples of compact Calabi-Yau threefolds. The simplest (and most famous) Calabi-Yau threefold is the quintic, and more generally, the cyclic manifolds. Subsuming these examples, is the important class of complete intersection in products of projective spaces, or CICY for short. After constructing these geometries, we show how certain crucial topological information is encoded into the defining equations.

### 3.1 Cyclic Calabi-Yau Threefolds

Let us now construct the most straightforward example of a Calabi-Yau manifold in each dimension. Let  $f(x_0, \dots, x_n)$  be a homogeneous degree  $d$  polynomial, or equivalently, a section of the line bundle  $\mathcal{O}_{\mathbb{P}^n}(d)$ . The vanishing locus of the section defines a degree  $d$  hypersurface  $X$  in the projective space  $\mathbb{P}^n$ .

**Theorem 3.1. (The Adjunction Formula)** *Let  $X \subset \mathbb{P}^n$  be a smooth, closed subvariety of codimension  $m$ . The canonical bundle of  $X$  is given by*

$$\omega_X = \Lambda^m \mathcal{N}_{X/\mathbb{P}^n} \otimes_{\mathcal{O}_X} \mathcal{O}_{\mathbb{P}^n}(-n-1)|_X \quad (3.1)$$

where  $\mathcal{N}_{X/\mathbb{P}^n}$  is the normal bundle of  $X$  in  $\mathbb{P}^n$  [11].

Since  $X$  is a divisor cut out by a section of  $\mathcal{O}_{\mathbb{P}^n}(d)$ , the normal bundle is the line bundle  $\mathcal{N}_{X/\mathbb{P}^n} = \mathcal{O}_{\mathbb{P}^n}(d)|_X$ . Therefore, the canonical bundle will be trivial if and only if  $d = n+1$ . By the Lefschetz hyperplane theorem,  $\pi_1(X)$  is trivial. We have therefore shown the following.

**Proposition 3.2.** *A homogeneous polynomial of degree  $n+1$  in the  $n+1$  projective coordinates on  $\mathbb{P}^n$  defines a compact Calabi-Yau  $n$ -fold as a divisor  $X \subset \mathbb{P}^n$ .*

Since we are interested in dimension 3, of most importance here will be the *the quintic* Calabi-Yau threefold constructed from a quintic polynomial in  $\mathbb{P}^4$ . For example, the Fermat quintic is the vanishing locus of

$$f(x_0, x_1, x_2, x_3, x_4) = x_0^5 + x_1^5 + x_2^5 + x_3^5 + x_4^5. \quad (3.2)$$

**Remark 3.3.** Note that saying “the” quintic is somewhat misleading, as we actually get a family of Calabi-Yau threefolds, by varying the coefficients in the quintic polynomial. However, these correspond to various complex structures on the same underlying topological type. It is conventional to refer to the entire family as “the quintic.” Similarly, note that certain quintic polynomials give singular varieties. Unless mentioned otherwise, we will assume to be working with a smooth member of the family, for example (3.2).

How can we generalize the quintic Calabi-Yau? The quintic is a hypersurface, and the most immediate generalization of a hypersurface is a complete intersection  $X \subset \mathbb{P}^n$ , which means the codimension of  $X$  equals the number of polynomials cutting it out. This is the most ideal intersection, though is quite rare in the world of varieties.

Suppose we have  $k$  homogeneous polynomials  $\{f_i\}_{i=1,\dots,k}$  on  $\mathbb{P}^n$  with  $q_i \in \mathbb{Z}_{\geq 0}$  the degree of  $f_i$ . The vanishing locus of the  $f_i$  produces a compact Calabi-Yau threefold as a complete intersection in  $\mathbb{P}^n$  if

$$\begin{aligned} k &= n - 3 && \text{(Complete intersection condition)} \\ n + 1 &= \sum_{i=1}^k q_i && \text{(Generalization of Adjunction)} \end{aligned} \quad (3.3)$$

One can show the fundamental group is trivial using a generalization of the Lefschetz hyperplane theorem [6, Thm. 1.4]. We call such a manifold a *cyclic* Calabi-Yau threefold. A notation which will prove helpful in the following section is to denote a collection of degrees as

$$M = [ q_1 \quad q_2 \quad \cdots \quad q_k ]$$

with  $X_M$  the corresponding cyclic Calabi-Yau. Note that  $n$  can be recovered from the condition  $n = k + 3$ .

Clearly, (3.3) defines a rather constrained combinatorial problem, and it turns out there are only 5 solutions. In the notation above, these are:

$$[5], \quad [2 \ 4], \quad [3 \ 3], \quad [3 \ 2 \ 2], \quad [2 \ 2 \ 2 \ 2].$$

The first example is the quintic, the second example is the complete intersection of a quadric and a quartic in  $\mathbb{P}^5$ , the third example is the complete intersection of two cubics in  $\mathbb{P}^5$ , etc.

## 3.2 CICY Calabi-Yau Threefolds

We can achieve a far greater generalization of the quintic by considering complete intersections in not just the ambient space  $\mathbb{P}^n$ , but rather in a product of projective spaces

$$A = \mathbb{P}^{n_1} \times \dots \times \mathbb{P}^{n_m}.$$

Suppose we have  $k$  multi-homogeneous polynomials  $\{f_i\}_{i=1,\dots,k}$  on  $A$ , with multi-degrees  $q_j^i \in \mathbb{Z}_{\geq 0}$  where  $i = 1, \dots, k$  and  $j = 1, \dots, m$ . In words,  $q_j^i$  is the degree of the  $i$ -th polynomial on the  $j$ -th factor of  $A$ . Generalizing the notation for cyclic Calabi-Yau threefolds, we package the data into the configuration matrix

$$M = \begin{bmatrix} q_1^1 & q_1^2 & \cdots & q_1^k \\ q_2^1 & q_2^2 & \cdots & q_2^k \\ \vdots & \vdots & \ddots & \vdots \\ q_m^1 & q_m^2 & \cdots & q_m^k \end{bmatrix} \quad (3.4)$$

We define  $X_M \subset A$  to be the vanishing locus of the  $\{f_i\}_{i=1,\dots,k}$ . The projective variety  $X_M$  is a Calabi-Yau threefold if the following conditions hold

$$\begin{aligned} k &= \sum_{i=1}^m n_i - 3 && \text{(Complete intersection condition)} \\ n_j + 1 &= \sum_{i=1}^k q_j^i, \text{ for all } j = 1, \dots, m && \text{(Generalization of Adjunction)} \end{aligned} \quad (3.5)$$

Such a  $X_M$  is called a *CICY*, which refers to a Calabi-Yau threefold realized as a complete intersection in products of projective space.

One is faced with the following combinatorial problem:

**Problem 3.4.** Can we classify all configuration matrices (3.4) up to equivalence and redundancies?

This represents one of the earliest big-data problems in the world of pure mathematics and physics. It was undertaken in the late 1980s by Candelas, Lutken, Schimmrigk and others [12]. Let us briefly survey the landscape of CICYs that were discovered:

- There are **7890** CICYs corresponding to 7890 inequivalent configuration matrices. The smallest matrix is  $1 \times 1$  (corresponding to the quintic) and they reach a maximum of 12 rows or 15 columns.
- $q_j^i \in [0, 5]$  for all  $i, j$ .
- There are 266 distinct Hodge pairs  $(h^{1,1}, h^{2,1})$ .
- There are 70 distinct Euler characteristics  $\chi \in [-200, 0]$ .
- The transpose of a configuration matrix is again a configuration matrix.
- The 5 cyclic Calabi-Yau threefolds are the only ones with a single row. In other words, there are only 5 complete intersection Calabi-Yau threefolds in a single projective space.

**Example 3.5.** Consider the following configuration matrix

$$S = \begin{bmatrix} 1 & 1 \\ 3 & 0 \\ 0 & 3 \end{bmatrix} \quad (3.6)$$

From the conditions (3.5), it is straightforward to check that  $S$  corresponds to a compact Calabi-Yau threefold  $X_S$  which is cut out of  $\mathbb{P}^1 \times \mathbb{P}^2 \times \mathbb{P}^2$  by two equations of multi-degrees  $(1, 3, 0)$  and  $(1, 0, 3)$ , respectively. This is a CICY, which we call the *Schöen manifold*, after Chad Schöen [13]. The two relevant Hodge numbers are  $h^{2,1} = h^{1,1} = 19$ , and therefore,  $\chi(X_S) = 0$ . In the next section we will see that  $X_S$  is also an elliptic fibration.

The transpose of the matrix (3.6)

$$TY = \begin{bmatrix} 1 & 3 & 0 \\ 1 & 0 & 3 \end{bmatrix} \quad (3.7)$$

of course also corresponds to a CICY, one called the Tian-Yau manifold  $X_{TY}$ . The Hodge numbers are  $h^{1,1} = 14, h^{2,1} = 23$  and therefore,  $\chi(X_{TY}) = -18$ . The Tian-Yau manifold carries a free  $G = \mathbb{Z}/3\mathbb{Z}$  action which preserves the Calabi-Yau structure. As a result, the quotient  $X_{TY}/G$  is a smooth compact Calabi-Yau threefold (though not a CICY) which has a special Euler characteristic  $\chi = -6$ , see [7–9]. At the time, this quotient was taken seriously as a candidate for the geometry of the universe! Unfortunately, it has some problems in its matter content.

In general, it is difficult to compute the Hodge numbers for the CICY dataset (in the above example, we gave them without proof). We present this topological data, along with the Euler characteristic for CICYs in Figure 1. The Hodge numbers are presented as frequency plots. Interestingly, the distribution of  $h^{1,1}$  is somewhat Gaussian while  $h^{2,1}$  is somewhat Poisson.

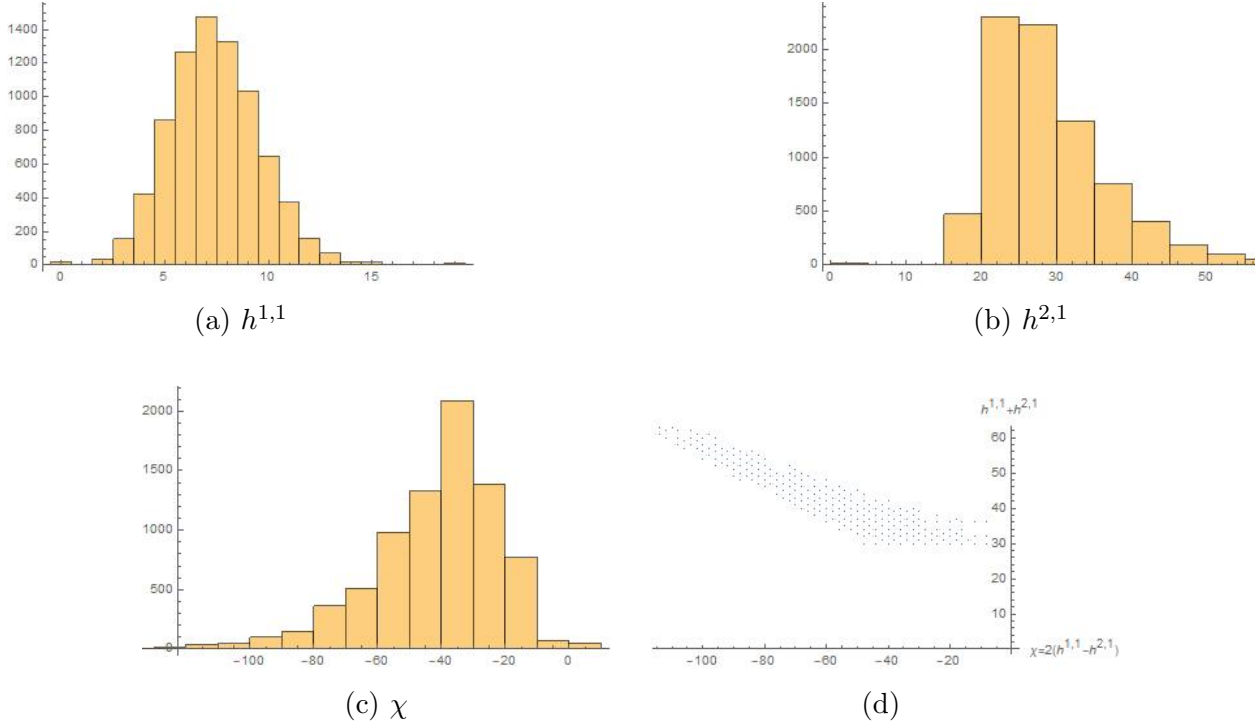


Figure 1: CICY topological data

All CICYs have non-positive Euler characteristic. One weak form of the mirror symmetry conjecture is that compact Calabi-Yau threefolds come in pairs with opposite Euler characteristics. Therefore, if one put too much stock in the CICY dataset, they might wrongly convince themselves that all Calabi-Yau manifolds have negative Euler characteristic! We clearly have to venture further in the landscape to encounter the mirror partners of the CICYs.

## 4 Elliptically Fibered Calabi-Yau Threefolds

Elliptic curves are among the most beautiful objects in mathematics. They provide a link between the fields of geometry, number theory, algebra, and even physics. In fact, as we saw in Section 2, an elliptic curve is the unique Calabi-Yau manifold in dimension 1. The notion of an *elliptic fibration* should be thought of as elliptic curves moving in a family. To understand this vague intuition, let us start with some basics.

Let  $\Lambda \subset \mathbb{C}$  be a full-rank lattice. Topologically, the quotient space  $\mathbb{C}/\Lambda$  is a complex torus, or a Riemann surface of genus 1. The following important proposition says that all such Riemann surfaces arise from cubic curves in the projective plane, i.e cubic plane curves.

**Proposition 4.1.** *Riemann surfaces of genus 1 are in bijection with smooth cubic hypersurfaces in  $\mathbb{P}^2$ , i.e. smooth vanishing loci of homogeneous degree 3 polynomials in 3 variables [14].*

By the degree-genus formula for plane curves, any smooth cubic hypersurface in  $\mathbb{P}^2$  has genus 1. Conversely, given a complex torus of the form  $\mathbb{C}/\Lambda$ , the Weierstrass  $\wp$ -function  $\wp(\tau, z)$  associated to  $\Lambda$  gives an embedding into  $\mathbb{P}^2$ . And the differential equation satisfied by  $\wp(\tau, z)$  implies that the image satisfies a cubic equation.

Consider the complex threefold  $X \subset \mathbb{P}^2 \times \mathbb{P}^2$  defined by the vanishing locus of the following bi-homogeneous degree (1,3) polynomial

$$a_0x_0^3 + a_1x_1^3 + a_2x_2^3 = 0. \tag{4.1}$$

Here  $(a_0 : a_1 : a_2)$  are coordinates on the first factor of  $\mathbb{P}^2$  and  $(x_0 : x_1 : x_2)$  are coordinates on the second. Notice that for any point  $(a_0 : a_1 : a_2) \in \mathbb{P}^2$  the above equation becomes a cubic in (the second)  $\mathbb{P}^2$ . Therefore, the map  $\pi : X \rightarrow \mathbb{P}^2$  defined by projection onto the first factor, is surjective and all fibers are cubics in  $\mathbb{P}^2$ . This motivates the following definition.

**Definition 4.2.** An *elliptic fibration* is a morphism<sup>5</sup>  $\pi : X \rightarrow B$  between smooth algebraic varieties  $X, B$  such that a generic fiber of  $\pi$  is a smooth elliptic curve. We call  $X$  the total space and  $B$  the base.

An *elliptically fibered Calabi-Yau threefold*, is a Calabi-Yau threefold  $X$  together with the structure of an elliptic fibration  $\pi : X \rightarrow B$ .

One should think of an elliptic fibration  $\pi : X \rightarrow B$  as a *family* of elliptic curves parameterized by the base  $B$ . However, over certain loci in the base, the elliptic curves can degenerate to singular curves. In virtually all interesting fibrations in algebraic geometry, one has to allow for singular fibers. For example, looking back to (4.1) the fiber above the point  $(1 : -1 : 0) \in \mathbb{P}^2$  is

$$x_0^3 - x_1^3 = (x_0 - x_1)(x_0^2 + x_0x_1 + x_1^2)$$

which is not a smooth cubic: it is the union of a line and a conic.

**Example 4.3.** Let  $Y \subset \mathbb{P}^1 \times \mathbb{P}^2$  be the vanishing locus of the bi-homogeneous degree (1, 3) polynomial

$$a_0f(x_0, x_1, x_2) + a_1g(x_0, x_1, x_2) = 0$$

where  $f, g$  are generic homogeneous cubic polynomials. Since for any point  $(a_0, a_1) \in \mathbb{P}^1$ , the above equation becomes a cubic in  $\mathbb{P}^2$ , the projection onto the first factor  $\pi : Y \rightarrow \mathbb{P}^1$  defines an elliptic fibration called a *rational elliptic surface*.

**Example 4.4.** Recall from Example 3.5, the Schöen manifold  $X_S \subset \mathbb{P}^1 \times \mathbb{P}^2 \times \mathbb{P}^2$  is the vanishing locus of homogeneous polynomials of multi-degree (1, 3, 0) and (1, 0, 3) respectively. Let  $Y$  be a rational elliptic surface from Example 4.3. We can define a map

$$\pi : X_S \rightarrow Y \subset \mathbb{P}^1 \times \mathbb{P}^2$$

---

<sup>5</sup>Strictly speaking, we want the map  $\pi$  to be flat and proper. These are technical algebro-geometric conditions ensuring we have nice family of projective curves of arithmetic genus 1.

by projecting onto the vanishing locus of the multi-degree  $(1, 3, 0)$  polynomial. The fiber over a point in  $\mathbb{P}^1 \times \mathbb{P}^2$  is a cubic in  $\mathbb{P}^2$  since we have to impose the second equation defining  $X_S$ . Therefore,  $X_S$  is an elliptically fibered Calabi-Yau threefold.

The above example illustrates that there are CICYs which are also elliptically fibered Calabi-Yau threefolds. See Figure 2, where “S” denotes the Schöen manifold. According to [7–9], there is a common belief that “most” Calabi-Yau threefolds are elliptically fibered. It is an active area of research to determine precisely which Calabi-Yau threefolds are elliptically fibered.

## 5 Additional Regions of the Compact Landscape

Unfortunately, there are many important classes of compact Calabi-Yau threefolds which we cannot discuss in detail here. Most notably, the Calabi-Yau hypersurfaces in 4-dimensional toric varieties. This problem was undertaken in the late 1990s by Kreuzer-Skarke (KS), and resulted in one of the biggest datasets seen in pure mathematics. For details on the KS dataset, we refer the reader to [7–9].

In Figure 2 we summarize the portions of the Calabi-Yau landscape mentioned in this survey. The point marked “S” denotes the Schöen manifold, which is both an elliptic fibration and a CICY. The point marked “Q” is the quintic, which is both a CICY as well as a hypersurface in a toric variety. The points labelled “×” denote compact Calabi-Yau threefolds not falling into any of these groups.

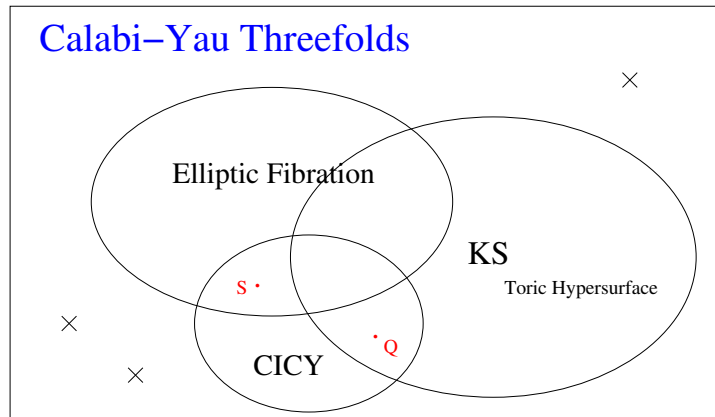


Figure 2: The compact Calabi-Yau threefold landscape

## Part II

## Non-compact Calabi-Yau Landscape

## 6 String Theory Structures

## 6.1 D-branes

D-branes occur in Type IIB Superstring theory as the Dirichlet boundary conditions of open strings. A D-brane is hence the hyperplane traced out by the allowed movement of the endpoint of an open string. The dimensionality of the D-brane in question defines the restriction on the directions the string endpoint can move in; such that a  $Dp$  brane only allows string endpoints to move in its  $(p+1)$ -dimensional world-volume. For example, a D0 brane is a spatial point moving through time, and fixes the endpoint of the string. Additionally, a D1 brane is a spatial line, forming a sheet as it is traced through time, and restricts the string endpoint to any position on this line for all time progression.

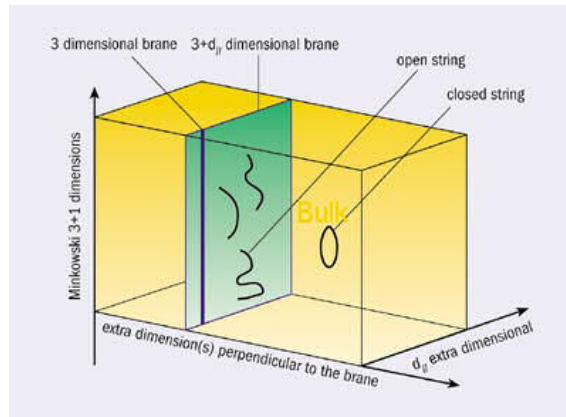


Figure 3: A graphic representation of a D-brane [15]. The vertical axis gives full Minkowski space,  $\mathbb{R}^{1,3}$ , such that a vertical line is the D3 brane considered in Superstring theory. Further theories may use higher dimensional branes indicated by the vertical line's extension into a plane along the  $d_{||}$  axis. The remaining dimensions of the theory are extra, and only endpoints of open strings are restricted to the D-brane as shown.

The D-branes world-volumes support a tensor form of dimension  $(p+1)$ , this can be integrated over the spatial dimensions to give a conserved charge, known as the Chan-Paton factor of the brane. The form in consideration connects the brane with a  $U(1)$ -bundle, such that enhanced gauge symmetry arises as the branes are stacked. In the stacking process,  $N$  D-branes' world-volumes are overlaid in spacetime at an infinitesimal limit, and the total brane gauge group enhances via:  $U(1)^N \mapsto U(N)$ . Here the gauge connection on the branes generalises to a higher rank tensor as the string endpoints can be connected across multiple branes in the stack. This becomes important in defining the quiver representation, which is used in the following machine-learning analysis.

D-branes are important in the brane-world physical interpretation of Type II Superstring theory. In the 10-dimensional spacetime of the Type IIB superstrings, the endpoints are restricted to exist on a D3 brane, whose world-volume is the familiar  $\mathbb{R}^{1,3}$  Minkowski space of general relativity and other theories. The remaining six dimensions form a non-compact Calabi-Yau space, such that  $X^{10} = \mathbb{R}^{1,3} \otimes X^6$ . The standard model exists on the D3 brane (or stack of  $N$  D3 branes), and only interacts with the  $X^6$  Calabi-Yau space via gravitation.

The simplest case of a non-compact Calabi-Yau 3-fold is  $\mathbb{C}^3$ , which is trivially Ricci-flat. Beyond that Orbifolds are a natural candidate. Orbifolds are formed from action of a discrete group quotient on a manifold. These manifolds are discussed further in Appendix B [7–9].

## 6.2 Quivers

A Quiver,  $\mathcal{Q}$ , is a multi-digraph, such that its set of nodes and arrows have finite cardinalities  $N_0$  and  $N_1$  respectively. The quiver represents a gauge theory, where each node has an associated  $U(N_i)$  gauge group. The product of all node gauge groups give the full gauge group of the theory. Each arrow is associated with a field,  $X_{ij}$ , in the bi-fundamental representation of the gauge groups associated with the nodes connected to the arrow. The fields transform according to the Young tableaux  $(\square, \bar{\square})$  for the nodes groups. The superpotential,  $W$ , of the theory the quiver represents leads to a set of polynomials,  $\{\partial_{X_{ij}} W = 0\}$ , which physically give the vacuum state of the theory.

Importantly, the representation variety of the quiver is the Vacuum Moduli Space of the gauge theory. A quiver’s representation variety is the gauge invariant quotient of the quiver’s representations, with relations from the superpotential, and quotiented by a product group of complex General Linear transformations. Geometric invariant theory (GIT) is generally used to construct moduli spaces by considering the quotients of groups on algebraic varieties.

This representation variety is an affine variety, such that the coefficients of the zero-locus of the variety’s polynomial set generates the corresponding prime ideal. Conversely, the Vacuum Moduli Space of a gauge theory is a geometric space with a vacuum state of the gauge theory associated to each point in the space. This moduli space often forms a manifold known as the vacuum manifold of the theory.

The space of quivers and superpotentials,  $(\mathcal{Q}, W)$ , produces a space of representation varieties, which hence give all the Vacuum Moduli Spaces of the gauge theory in question. Each of the Vacuum Moduli Spaces of a supersymmetric gauge theory is a non-compact Calabi-Yau manifold, and hence this is how the non-compact Calabi-Yau landscape naturally arises in Superstring theory. A simple example of a quiver is the “clover”, which represents

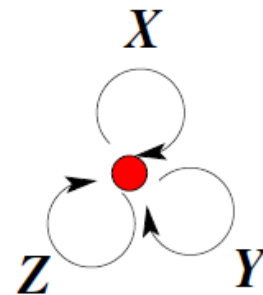


Figure 4: The quiver for  $\mathcal{N} = 4$  Super Yang-Mills theory, with three adjoint fields: X, Y, Z [7–9].

the famous  $\mathcal{N} = 4$  Super Yang-Mills theory, shown in figure 4. The superpotential for this example is  $W = \text{Tr}([X, Y]Z)$  which leads to the simplest Vacuum Moduli Space case of  $\mathbb{C}^3$  [7–9].

### 6.3 An Orbifold Example: $\mathbb{C}^3/\mathbb{Z}_3$

Here we consider a typical example of quiver gauge theory used commonly in association with AdS/CFT correspondence, as it is the worldvolume theory of a D3 brane in the bulk spacetime. This non-compact Calabi-Yau manifold examined is given by the toric variety:  $\mathbb{C}^3/\mathbb{Z}_3$ . This quotient structure makes the manifold an orbifold; where the algebraic geometry structure is explained further in Appendix B. The  $U(1)^3$  quiver in question is shown in figure 5 and shows 9 fields in the theory.

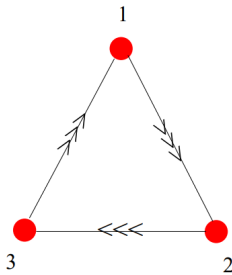


Figure 5: The  $U(1)^3$  quiver with 9 fields denoted by the 3 sets of 3 arrows [16].

Since 3 fields exist on each of the 3 edges, there are correspondingly  $3^3 = 27$  gauge invariant operators possible, associated with all the closed cycles in the quiver. In this theory, each of the gauge invariant operator terms appear in the superpotential as products of the fields in the corresponding cycle, giving

$$W = \sum_{\alpha, \beta, \gamma=1}^3 \varepsilon_{\alpha\beta\gamma} X_{12}^\alpha X_{23}^\beta X_{31}^\gamma, \quad (6.1)$$

for the totally antisymmetric rank 3 tensor  $\varepsilon_{\alpha\beta\gamma}$ , where the Greek indices run  $1 \mapsto 3$  for each of the 3 arrows between each pair of nodes. Each field has subscripts to denote the nodes it is in representations of. There are also 9 F-term equations of motion from the superpotential, which are

$$0 = \sum_{\beta, \gamma=1}^3 \varepsilon_{\alpha\beta\gamma} X_{23}^\beta X_{31}^\gamma = \sum_{\alpha, \gamma=1}^3 \varepsilon_{\alpha\beta\gamma} X_{12}^\alpha X_{31}^\gamma = \sum_{\alpha, \beta=1}^3 \varepsilon_{\alpha\beta\gamma} X_{12}^\alpha X_{23}^\beta, \quad (6.2)$$

where each term is 3 equations for each value of the uncontracted index. These equations arise under the action of  $0 = \partial_X W$  for each of the fields,  $X$ .

The 27 gauge invariant operators are redefined as dimensions of  $\mathbb{C}^{27}$ , denoted  $y_{\alpha\beta\gamma}$ . Elimination with the F term equations via low degree polynomial interpolation [16] leads to a system of 17 linear equations, and 27 quadratic equations. Further elimination via trivial

substitution with the 17 linear equations reduces the system to 27 equations in 10 variables, thus giving the  $\mathbb{C}^{10}$  space. These equations are recognised as the standard Veronese embedding:  $\mathbb{P}^2 \hookrightarrow \mathbb{P}^9$  which can be affined into a  $\mathbb{C}^3$  embedding in the  $\mathbb{C}^{10}$  found above. This embedding corresponds to their existing exactly 10 degree 3 monomials in 3 variables, such that each one then corresponds to a dimension in  $\mathbb{C}^{10}$ . These equations then give the degree 9 irreducible variety which defines the 3 dimensional orbifold  $\mathbb{C}^3/\mathbb{Z}_3$  [17].

The exponents of the 10 degree 3 monomials in 3 variables give vectors in the fan of the toric variety definition (noting that the orbifold being abelian makes it also toric). Taking the rays of this fan gives three coplanar vectors, which in the plane correspond to points which in turn define the toric diagram. These are  $\{(1, 0), (0, 1), (-1, -1)\}$ , which are plotted in figure 6, the dual of this diagram then gives the orbifold's toric diagram [7–9, 16].

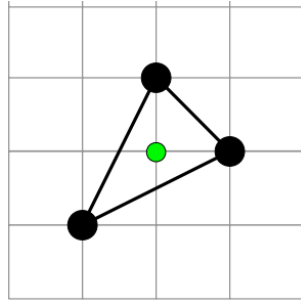


Figure 6: The toric diagram dual for the  $\mathbb{C}^3/\mathbb{Z}_3$  orbifold [7–9]. The origin is denoted in the diagram centre, and the toric diagram can be retrieved as this diagram's dual.

The corresponding brane tiling and dessin d'enfant can then be formed from the toric diagram; these objects are addressed in section 7 [18].

## 6.4 McKay Correspondence

McKay correspondence concerns a discrete finite subgroup  $G \subset SU(2)$ . Firstly taking the tensor product between the defining 2 complex dimensional rep of  $G$  and each irrep of  $G$ , and then taking the irrep decomposition of this tensor product makes the correspondence manifest. Whereby each decomposition coefficient is the square of the adjacency matrix for each of the simply-laced Dynkin diagrams.

Dynkin diagrams represent the root system of the gauge group's Lie algebra. To be simply-laced means there is only one edge between each node, which represents a restriction on the angles between the fundamental roots. Specifically, the simply-laced Dynkin diagrams are:  $A_n$ ,  $D_n$ ,  $E_n$  where the first two are series of diagrams for  $n \in \mathbb{Z}^+$ , whilst the  $E_n$  refers to three of the exceptional Lie algebras. The Dynkin diagrams in question are affine-extended, which is canonically achieved by central extension of the original Lie algebra. This amounts to introducing an additional imaginary root, which increases the dimensionality of the root system. These are hence represented with an additional node, and denoted:  $\tilde{A}_n$ ,  $\tilde{D}_n$ , and  $\tilde{E}_n$

respectively. In the special case of simply-laced, the Dynkin diagrams correspond exactly to their Coxeter diagrams, which represent Coxeter groups, defined by reflection symmetries.

This is relevant because the representation variety of the affine Dynkin diagrams formulated as quivers are Calabi-Yau 2-folds (a.k.a. K3-surfaces). We can then produce orbifolds from these described by McKay quivers such that they have the form  $\mathbb{C} \times (\mathbb{C}^2/G)$ . These orbifolds are hence also candidates for the extra dimensions in Superstring theory. However where  $\mathbb{C}^3$  leads to  $\mathcal{N} = 4$  Super Yang-Mills theory on the D3 brane, these orbifolds produce  $\mathcal{N} = 2$  supersymmetric QFTs.

When taking quotients to produce the orbifolds in question, relations between the invariants of the orbifold give rise to algebraic singularities. In  $\mathbb{C}^2$  these are the du Val singularities. Smoothing out these singularities through desingularisation requires the resolution map between the canonical bundle and canonical sheaf to be crepant. Meaning that no discrepancy divisor is needed with the resolution map. When this crepant resolution map is established, metrics and other physically relevant measures can be written explicitly for some special cases.

These crepant resolutions are key in generalising the quotient process to act on Calabi-Yau 3-folds (as  $\mathbb{C}^3/G$ ); introducing further orbifolds into the Calabi-Yau spectrum. However in this case the manifolds are related to one another by mirror symmetry and in particular flop transitions. These orbifolds correspond to  $\mathcal{N} = 1$  super-conformal gauge theories, hence extending also the practical applications of studying the Calabi-Yau landscape with respect to examining topological theories in physics.

The quotient product, and crepant resolution methods extend the landscape of non-compact Calabi-Yau manifolds from only  $\mathbb{C}^3$  to also include a plethora of orbifolds. Physicists interpret the manifold landscape as representation varieties of quivers, which indicate the equivalent gauge-field theories [7–9].

## 7 Algebraic Geometry Viewpoint

### 7.1 Brane Tilings

The method to connect the quivers of a gauge theory, with the toric diagram (see Appendix B) of the relevant Calabi-Yau that makes up the remaining dimensions in the full 10d superstring spacetime, exists for both directions [19, 20]. Deriving the toric diagram from the quiver is more straightforward and follows the clockwise process depicted in figure 7.

The converse, “geometric engineering”, toric diagram to quiver process was originally computationally demanding, with exponential time complexity. This process was streamlined by introducing the concept of brane tiling. This brane tiling concept was derived from noticing a consistent relation between the number of nodes, edges, and superpotential terms,  $(\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2)$  respectively,

$$\mathcal{N}_0 - \mathcal{N}_1 + \mathcal{N}_2 = 0. \tag{7.1}$$

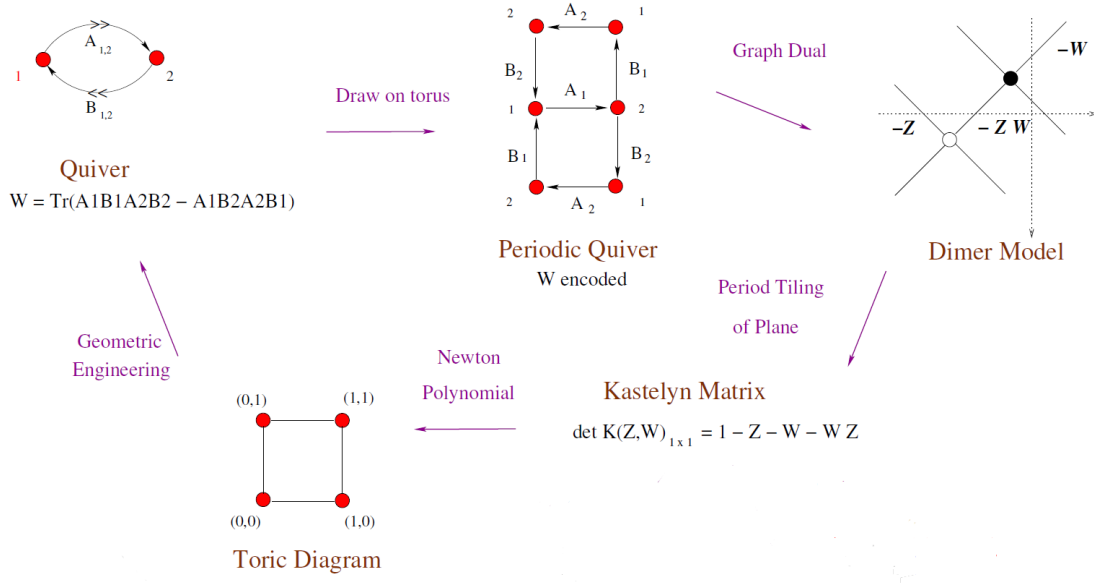


Figure 7: A pictorial representation of the process that links the quiver and superpotential  $(Q,W)$  to the Toric diagram of the equivalent non-compact Calabi-Yau manifold [7–9]. This specific example is for the conifold considered previously.

This applied for all quivers whose representation variety was a toric variety (as for those considered in string theory). The relation 7.1 was associated to the Euler characteristic for a torus, and this allowed the quiver and superpotential to be encoded as a bipartite graph tiling on a (genus,  $g = 1$ ) torus. The connection of brane tilings to quivers follows a simple algorithm. Whilst mapping from the toric diagrams to the brane tilings is epimorphism; with the orbit of tilings which are mapped to by the same toric diagram related by Seiberg duality [21].

Seiberg duality relates an “electric” and a “magnetic” theory, stating that under RG flow they both approach the same IR fixed point. Therefore they represent the same theory at lower energy densities. In our context it represents the relation between two quiver gauge theories, where some additional fields are integrated out/introduced, which graphically corresponds to contracting/expanding parts of the brane tilings. This concept is exemplified in figure 8.

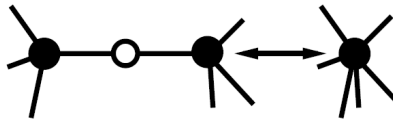


Figure 8: The contraction of part of a brane tiling [21], corresponding to integrating out a massive field to relate two quiver gauge field theories via Seiberg duality.

More mathematically, the Seiberg duality process corresponds to cluster mutation of the mathematical graph-theoretic quiver objects. Through a series of steps of reorienting and reassigning arrows associated with a node in the quiver, and adjusting the gauge group size by the number of fields, a different (dual) quiver is formed [22]. This cluster mutation process is a generalisation of the Seiberg duality. Multiple actions of the cluster mutation for different nodes creates “mutation classes” of quivers. Their equivalent brane tilings are connected by a process known as urban renewal, again a mathematical generalisation of the integrating out/introduction of fields in the physical application of Seiberg duality. These dual quivers are related, where their tilings correspond to the same toric diagram under the epimorphism previously mentioned. Tilings are an important step in the geometric engineering process.

The quiver duality concept may also be thought of as monodromy of wrapped 3-cycles in the dual theory via another duality known as mirror symmetry. Mirror symmetry connects mirror dual Calabi-Yau manifolds in different superstring theories, where they lead to the same resulting physics. In this case the D3 brane on one Calabi-Yau 3-fold is mirror dual to a D6 brane with 3 dimensions identified (3-cycle wrapping) on the dual Calabi-Yau 3-fold. This concept has been shown to be practical in Topological string theory where the mirror symmetry concept has been mathematically well defined [23].

Mirror symmetry allows calculation of certain complicated invariants by performing easier calculations in the dual theory. A key example is Gromov-Witten invariants, which arise in symplectic geometry which also satisfies the ‘almost complex’ structure requirements. The almost complex structure is a looser condition than Kähler geometry in that only the tangent space is required to be smooth linear complex, and not necessarily the underlying space. These invariants are calculated from pseudoholomorphic curves which are the symplectic equivalent of distances in Riemannian geometry. More general quantities are usually expressed in terms of the Gromov-Witten invariants, which are difficult to compute, but can be reduced to simpler integrals in the mirror dual theory [24, 25].

## 7.2 Dessin d’Enfants

Bipartite tilings are the algebraic geometry equivalent of Grothendieck’s Dessin d’Enfants from number theory. This interpretation can be useful for categorising the tilings, and hence the quiver gauge theories. Mathematically the dessins are interpreted using Belyi maps,  $\beta$ , which map from a smooth compact Riemann surface (described as a hyperelliptic curve of complex numbers),  $\Sigma$ , to projective space,  $\mathbb{P}^1$  such that [26]

$$\beta : \Sigma \mapsto \mathbb{P}^1. \tag{7.2}$$

A dessin is then formed from the preimage of a Belyi map which has three ramification points; where a ramification point is an element of  $\Sigma$  where the local Taylor expansion of  $\beta$  starts at order  $\geq 2$  and corresponds to degeneration of the map. Under the  $SL(2, \mathbb{C})$  symmetry of  $\mathbb{P}^1$ , the three ramification points are transformed to  $(0, 1, \infty)$ , and the dessin is formed by associating the preimages of 0 to black nodes, preimages of 1 to white nodes, and

preimages of the  $(0,1)$  interval to edges. Therefore the dessin is a bipartite graph drawn on the Riemann surface  $\Sigma$ , such that

$$\beta^{-1}(0) \rightarrow \bullet, \quad \beta^{-1}(1) \rightarrow \circ, \quad \beta^{-1}(0,1) \rightarrow -. \quad (7.3)$$

The dessins can be categorised by their passports, which is the collection of the ramification data, represented

$$[r_0(1), r_0(2), \dots, r_0(B) | r_1(1), r_1(2), \dots, r_1(W) | r_\infty(1), r_\infty(2), \dots, r_\infty(I)], \quad (7.4)$$

such that  $r_i(j)$  is the ramification value (order of the leading term in Taylor expansion) of the  $j$ th preimage of value  $i$  in the image of the Belyi map. The total number of preimage points are  $(B, W, I)$  for the ramification points  $(0, 1, \infty)$  respectively. Note also here that the Riemann-Hurwitz formula sets  $B = W$  for the genus 1 torus we are working on. The actual value of each ramification point then gives the valency of each node in the dessin.

The passport doesn't identify the dessins exactly, a more effective way of representing the dessins independently is combinatorically as permutation triples. Permutation triples encode the dessin information by creating elements of the symmetric group which are the products of all cycles containing either the white nodes,  $\sigma_W$ , or the black nodes,  $\sigma_B$ . An additional object,  $\sigma_\infty$ , in the symmetric group is defined also, such that:

$$\sigma_W \cdot \sigma_B \cdot \sigma_\infty = \mathbb{1}_d, \quad (7.5)$$

for  $\mathbb{1}_d$  the identity element of the symmetric group,  $S_d$ , of the  $d$  edges in the dessin. The group elements  $\sigma_\infty$  are then associated to cycles about faces of the dessin under this symmetric group [27].

In supersymmetric QFTs, R-symmetry connects the fields in the theory via their R-charge of the supersymmetric representations. For the tiling, each edge has an R-charge from the field it represents in the superpotential interpretation of the tiling. Under the symmetry, these charges must satisfy

$$\sum_{i \in \mathfrak{E}_n} R_i = 2, \quad \sum_{i \in \mathfrak{E}_f} (1 - R_i) = 2, \quad (7.6)$$

for  $\mathfrak{E}_n$  the edges bounding any node in question, and  $\mathfrak{E}_f$  the edges bounding any face in question. These relations in terms of the tiling are equivalent to Euler's relation as in equation 7.1.

Isoradial embedding is a method for constructing the tiling which automatically satisfies these required conditions on the R-charges. Nodes are organised on the circumferences of intersecting tessellated circles such that the angles subtended by triangles formed with adjacent nodes and the circle's centres satisfy  $\theta_i = \pi R_i/2$ . This causes the conditions in 7.6 to translate to basic geometric conditions on total angle around a point and total interior

angle of a polygon respectively. These conditions fix the nodes' positions up to rotation about the circles, this is then fixed by performing  $a$ -maximisation of the function

$$a(R_i) = \sum_{i \in \mathfrak{E}} (R_i - 1)^3, \quad (7.7)$$

for  $\mathfrak{E}$  the set of all edges in the tiling. Maximising this equation over the  $R_i$  partition is equivalent to minimising the conic base volume.

## 8 Non-compact Calabi-Yau Summary

The non-compact Calabi-Yau landscape makes itself of manifest importance in superstring theory through the interpretation of quiver gauge theories. Within this, the manifolds make up the additional dimensional space in the theories' brane world interpretation. Using generalizations of McKay Correspondence, the association from the manifolds to the quivers is made through their representation varieties. Orbifolds can then be introduced into the landscape using group quotients and crepant resolution.

Alternatively the manifolds may be considered more algebraically as toric varieties, generally defined using a fan structure on a lattice. The general toric variety construction allows formation of more Calabi-Yau manifolds, including the conifold. From this interpretation toric diagrams can be formed from the varieties which aid in manifold classification, especially in the context of brane tilings.

Brane tilings are a useful geometric interpretation of the quiver and superpotential properties; and particularly streamline the process of calculating the physical theories represented by toric diagrams. Seiberg duality and mirror symmetry also become important concepts in this consideration of forming physical quiver gauge theories.

Finally these tilings can be considered in parallel to dessin d'enfants from number theory. This interpretation in terms of Belyi maps and their ramifications offers some explanation of structure associated with the underlying theories' supersymmetry. This interconnection between these interpretations of the Calabi-Yau manifolds is well depicted in the example in figure 9 for the conifold.

Points of interest for further investigation include the interpretation of the Seiberg duality in terms of the dessin structure; and how the use of dessins may relate the absolute Galois group (important in the theory of dessins) into the physical theories of the tilings. Beyond these, the parallels between the physical, algebraic geometry, and number theoretic structures offers many sources for inspiration.

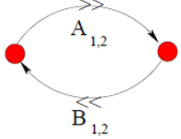
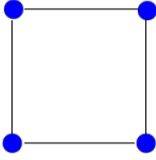
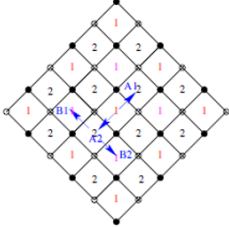
(a) Theory	(b) Toric Diag
 $W = \text{Tr}(\epsilon_{il}\epsilon_{jk}A_iB_jA_lB_k)$	 $\mathcal{M} \simeq \{uv - wz = 0\} \subset \mathbb{C}^4$
(c) Belyi Pair	(d) Dessin on $T^2$ (dimer)
$y^2 = x(x - 1)(x - \frac{1}{2})$ $\beta(x, y) = \frac{x^2}{2x-1}$	

Figure 9: The conifold Calabi-Yau manifold interpreted in terms of: (a) its underlying physical theory in terms of quiver and superpotential; (b) the representation variety's toric diagram (note it is the equivalent dual diagram that is shown); (c) the Belyi pair used to encode the dessin tiling structure; and (d) the brane tiling on a torus [7–9].

## Part III

# Machine-Learning the Landscape

As we have seen above, different areas in mathematics, including algebraic geometry, representation theory and even number theory, have appeared in our study of CY manifolds in theoretical physics. As the extra six dimensions are believed to be “wrapped” as a CY 3-fold under string compactification, people began to search for the possible  $CY_3$ ’s, and have so far collected a gigantic list of  $CY_3$ ’s from reflexive polytopes, estimated at order  $10^{10}$  [8, 9]. Furthermore, the number of string vacua in the landscape is astonishingly of order  $10^{500}$  for type IIB theory<sup>6</sup> [28]. Thus, the power of computers and algorithms is urgent for this interdisciplinary research. A different version of “WWJD” has now been raised: what would Jython do?

## 9 Performance Measures: Hypersurfaces in $WP^4$

In Appendix C, machine learning is briefly introduced. Whatever approach the machine adopts for the learning, we always need to know how well it performs. Let us quantify its performance using the following example. Recall the weighted complex projective space  $WP^4$  in (B.2)<sup>7</sup>. Our input for each hypersurface in  $WP^4$  is a 5-vector of co-prime positive integers which determines the space. Let us consider a simple query of whether the Hodge number  $h^{2,1} > 50$ . Geometrically, we are searching for  $CY_3$ ’s with a relatively large number of complex deformations. Our data  $D$  consists of 7555 5-vectors,  $x_i$ , each resulting in a binary output,  $y_i$ . For example,  $(\{1, 1, 1, 1, 1\} \rightarrow 1) \in \{(x_i \rightarrow y_i)\} = D$  as  $h^{2,1} = 101 > 50$  in this case. On the other hand, we have  $(\{2, 2, 3, 3, 5\} \rightarrow 0)$  since  $h^{2,1} = 43 < 50$  here. In [29], the Hodge numbers are computed using Landau-Ginzburg method. However, such procedure would take hours, and this is just a very simple query.

Now that our data  $D$  is fully known, we can then split our data into a training set  $T$  and a validation set  $V$ , viz,  $D = T \sqcup V$ . We can then establish a machine-learning algorithm so as to check how well it performs. This procedure is known as the *cross validation*. To quantify the accuracy, we make the following definitions.

**Definition 9.1.** Let  $V = \{(x_i \rightarrow y_i)\}$  where  $y_i$  is the actual correct output for input  $x_i$ , and let  $y_i^{\text{pred}}$  be the output predicted by the machine-learning model on  $x_i$  with  $i$  running from 1 to  $N$ . Then the *precision*  $p$  is the percentage that  $y_i^{\text{pred}}$  agrees with  $y_i$ :

$$p := \frac{1}{N} |\{y_i^{\text{pred}} = y_i\}| \in [0, 1]. \quad (9.1)$$

<sup>6</sup>For F-theory, the number of flux vacua arisen from elliptic fourfolds even rockets to at least  $10^{272000}$ .

<sup>7</sup>In terms of the notation in (B.2), this is  $\mathbb{C}P^{(a_0, a_1, a_2, a_3)}$ . For brevity, we will henceforth denote it as  $WP^4$ .

**Definition 9.2.** Consider  $y_i$  and  $y_i^{\text{pred}}$  as vectors  $\mathbf{y}$  and  $\mathbf{y}^{\text{pred}}$  respectively. Then the *cosine distance* is

$$d_C := \frac{\mathbf{y} \cdot \mathbf{y}^{\text{pred}}}{\|\mathbf{y}\| \|\mathbf{y}^{\text{pred}}\|} \in [-1, 1]. \quad (9.2)$$

If the cosine angle between the two vectors is 1, we have a complete agreement. If  $d_C$  is -1, then it is the worst fit. If  $d_C = 0$ , then it is a random correlation.

Now we take 2000 samples (out of 7555) from  $D$ , which is approximately 25%, to be our training data. Then we establish our MLP and test it using the remaining data. The detailed `Python` code can be found in [7–9]. It turns out that there are only 375 errors in our experiment, which gives  $p = (5555 - 375)/5555 \simeq 93.25\%$ , and the cosine distance  $d_C$  is 0.91. This is a quite impressive result with such a high accuracy. Remarkably, the running time is less than one minute on an ordinary laptop<sup>8</sup>!

To make sure that our machine-learning makes satisfying predictions, we need to introduce Matthews correlation coefficient (MCC). Firstly, we have:

**Definition 9.3.** Let  $\{(x_i \rightarrow y_i)\}$  be categorical data, where  $y_i \in \{1, 2, \dots, k\}$  takes value in  $k$  categories. Then the *confusion matrix* is a  $k \times k$  matrix where 1 is added to the  $(ab)^{\text{th}}$  entry if the actual value of  $y$  is  $a$  while the predicted  $y^{\text{pred}}$  is  $b$ .

As a result, we want the confusion matrix to be diagonal ideally. In our binary case, the confusion matrix is  $2 \times 2$ , and we have this table:

<div style="display: flex; justify-content: space-between;"> <span style="font-size: 0.8em;">Predicted \ Actual</span> </div>	True (1)	False (0)	(9.3)
True (1)	True Positive ( $tp$ )	False Positive ( $fp$ )	
False (0)	False Negative ( $fn$ )	True Negative ( $tn$ )	

Then we can define:

**Definition 9.4.** For binary classifications, the Matthews correlation coefficient is the square root of the normalized  $\chi$ -squared, that is,

$$\phi := \sqrt{\frac{\chi^2}{N}} = \frac{tp \cdot tn - fp \cdot fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}} \in [-1, 1]. \quad (9.4)$$

Such definition can also be generalized to  $k \times k$  confusion matrices [30, 31]. If the MCC returns 1, then we have a perfect prediction. If MCC is -1, then our fit is a complete disagreement. If  $\phi = 0$ , then it is a random prediction. It is crucial to notice that other measures such as  $p$  and  $d_C$ , unlike MCC, are not useful when the sizes of two classes differ too much, i.e., when we have *imbalanced* data. For example, if there is only 0.1% of the

---

<sup>8</sup>This can also be done using `Mathematica` with a high accuracy as well. In particular, `Mathematica` has machine learning built into its core operating system from version 11.2, and now `Mathematica` 12 has been released with detailed documentation on machine learning.

data to be classified as true. Then our algorithm would naively train a model predicting false for any input. Nevertheless, the accuracy  $p$  would still reach 99.9%<sup>9</sup>. In our case of hypersurfaces in  $W\mathbb{P}^4$ , we have  $\phi = 0.84$ , which gives a quite nice prediction.

Now one may wonder how well our NN will behave when we change the number of samples in the training data. This can be analyzed via *learning curves*:

**Definition 9.5.** Let  $D = \{(x_i \rightarrow y_i)\}$  have  $N$  data-points. We choose cross validation by taking  $\gamma N$  data-points randomly as training data  $T$ , with some  $\gamma \in (0, 1]$ . Then the remaining  $(1-\gamma)N$  data-points form the validation data  $V$ . The performance of the machine-learning algorithm, upon training on  $T$  and validated on  $V$ , is a function  $L(\gamma)$  measured by any goodness of fit as aforementioned. The learning curve is then the plot of  $L(\gamma)$  against  $\gamma$ .

In practice,  $\gamma$  is chosen discretely. Moreover, for each  $\gamma$ , we repeat random samples  $\gamma N$  a number of times for statistical stability, so there are error bars associated to our points on the curve.

The learning curve of our hypersurfaces in  $W\mathbb{P}^4$  case is depicted in Fig. 10. As we can

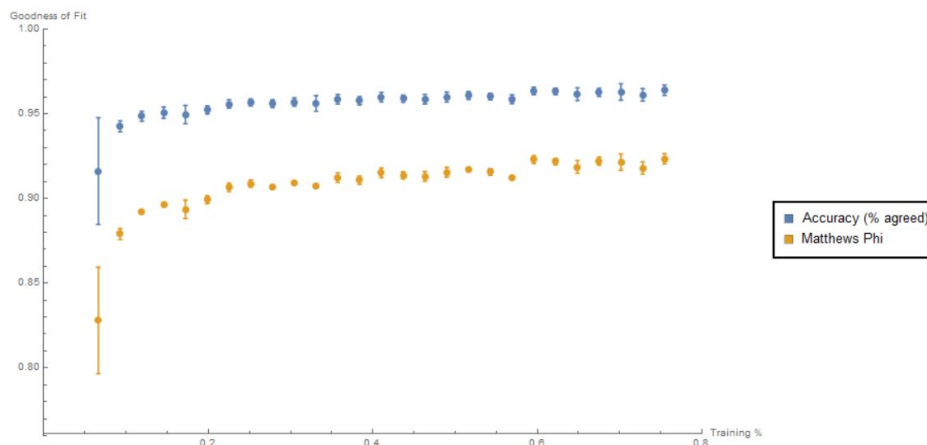


Figure 10: The learning curves for machine learning whether  $h^{2,1} > 50$  for a hypersurface in  $W\mathbb{P}^4$ . We repeat cross validation 10 times at each incremental interval of 5%.

see, there is a large error for training data less than 10% as the NN has not seen enough data for valid predictions. However, from 20%, our predictions become really well-behaved. The curve then ascends steadily as the (both) measures are approaching to 1.

## 10 Learning CICYs

Let us now focus on the CICY dataset of 7890 inequivalent complete intersection CY 3-folds in products of (unweighted) complex projective spaces. As discussed in §3.2, a CICY

<sup>9</sup>There are also other measures (especially required for imbalanced data) such as F-score. However, MCC is the most informative one as it includes all the four categories in confusion matrices [32].

is represented by a matrix, whose entries are 0 to 5, with number of rows ranging from 1 to 12 and number of columns ranging from 1 to 15. In terms of computer graphics, it is a  $12 \times 15$  pixelated image with 6 different colours (or 6 shades of grey in greyscale image). As an example, the CICY of 8 equations in  $(\mathbb{P}^1)^5 \times (\mathbb{P}^2)^3$  is the matrix in Fig. 11 such that we have the image as in Fig. 12.

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

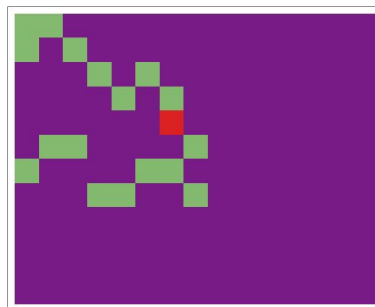


Figure 11: The matrix representing the CICY in  $(\mathbb{P}^1)^5 \times (\mathbb{P}^2)^3$ .

Figure 12: The corresponding image of Fig. 11 where purple pixels are 0, green, 1 and red, 2.

As a matter of fact, we need CNN to take advantage of pixelations of CICYS. Nevertheless, for this task, we are only using the graphic images to emphasize that even though our computers have no knowledge of algebraic geometry, they can still “learn” to make good predictions, and we will keep using MLP in our analysis.

Similar to §9, the machine learns the binary query of the Hodge number  $h^{1,1} > 5$ . The input would be  $12 \times 15$  matrices, and the output is again either 0 or 1. Now we take 4000 random samples ( $< 50\%$ ) as our training data, and test the remaining 3890 data-points as validation. The learning time only takes about 5 minutes while the performance is remarkable. The accuracy  $p$  is 97% and the cosine distance  $d_C$  reaches 0.98. For MCC, we have  $\phi = 0.87$ . Varying the numbers of samples in  $T$  yields the learning curves depicted in Fig. 13.

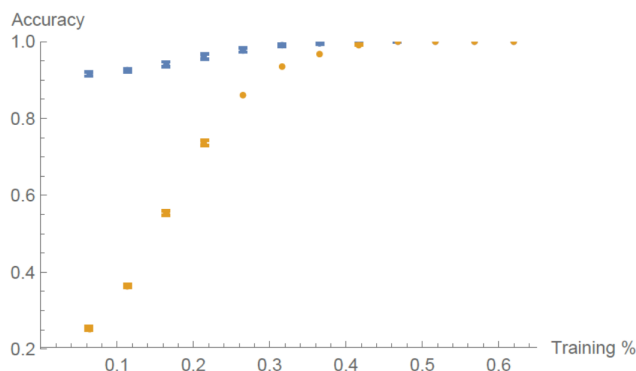


Figure 13: The learning curves for machine learning whether  $h^{1,1} > 5$  for CICYS.

We can see that there is a huge discrepancy between  $p$  and  $\phi$  for small  $\gamma$ 's. This is due to the great disparity between the sizes of two classes ( $h^{1,1} \leq 5$  and  $h^{1,1} > 5$ ). Indeed, Fig.

1 with the distribution of  $h^{1,1}$ 's verifies our argument. As aforementioned, MCC would be much more useful in this case. For larger  $\gamma$ 's, we do see the ascent of both curves, approaching to 1.

Let us now make our problem more sophisticated and compute the precise values of  $h^{1,1}$ . Here we try three different methods and compare their results:

- NN Classifier: As  $h^{1,1} \in [0, 19]$ , the output is a 20-channel classifier (cf. the 10-channel classifier in text recognition) with each neuron mapping to 0 or 1. The detailed architecture is described in [7–9, 33].
- NN Regressor: The output is some real number which means it is continuous. There are certain parameters known as *hyperparameters* that need to be optimized before training by hand, such as the number of hidden layers etc. This is discussed in detail in Appendix C in [33].
- SVM Classifier: The output is one of the possible values of  $h^{1,1}$ , that is, some integer between 0 and 19. The hyperparameter optimization is also discussed in [33].

We can now plot the three learning curves as in Fig. 14. We see that the NN classifier

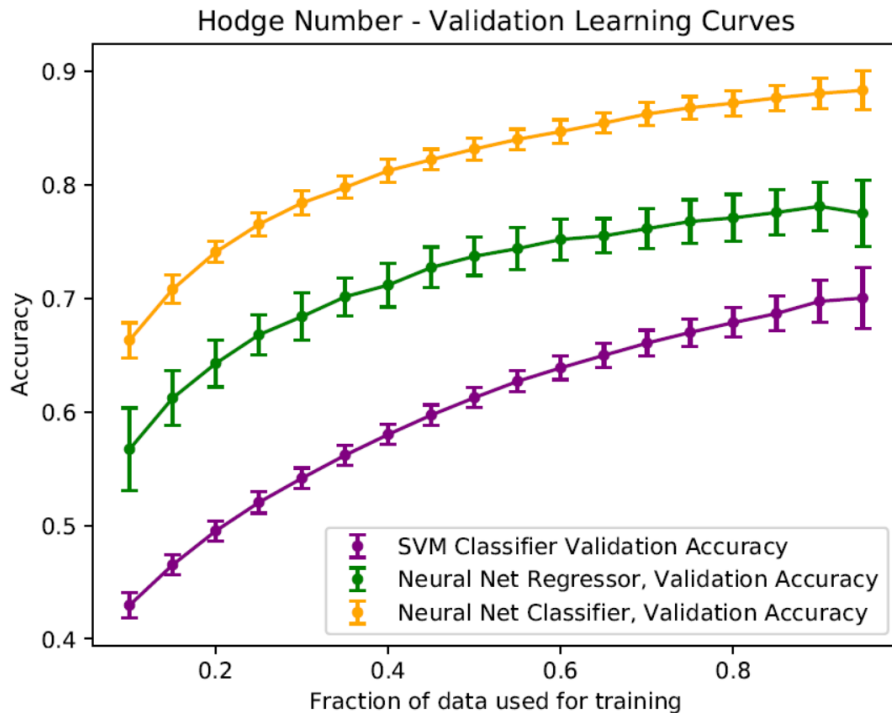


Figure 14: The learning curves generated by averaging over 100 different random cross validation splits.

performs best in the machine-learning. Again, it is impressive that such training on an ordinary laptop takes only about 10 minutes and the validation only takes a few seconds. For reference, we plot the histograms of frequencies of predicted and actual  $h^{1,1}$ 's with

validation sets of sizes 20% and 80% of the total data respectively for the three methods in Fig. 15.

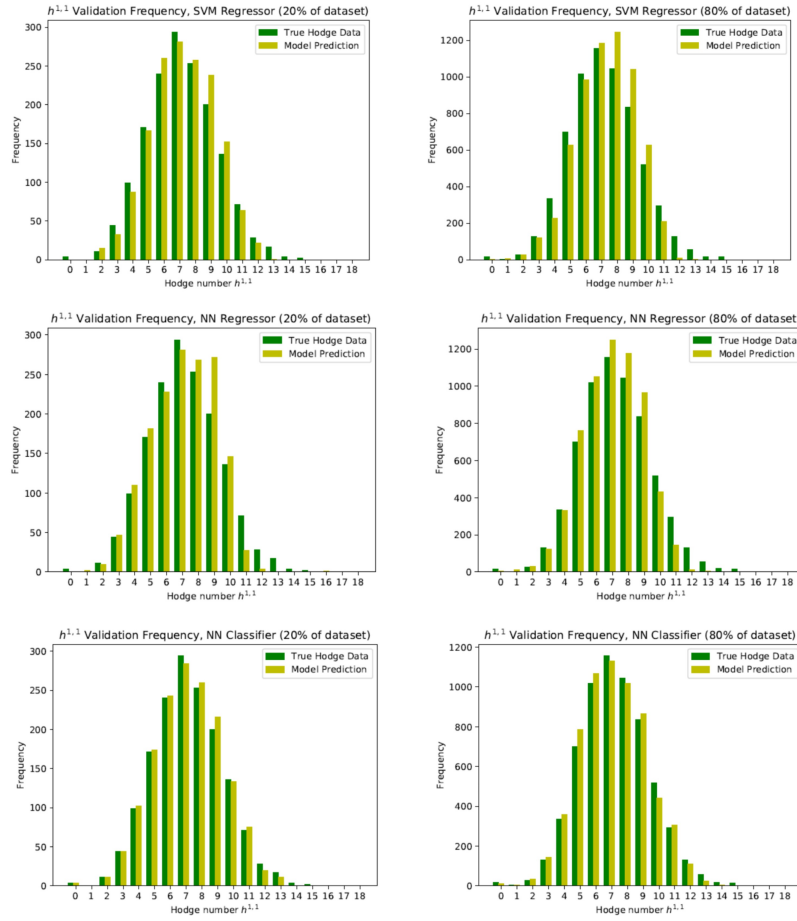


Figure 15: The frequencies of  $h^{1,1}$ 's.

## 10.1 Distinguishing Elliptic Fibrations

As CICYs may admit elliptic fibration (EF) structures, [34] machine-learnt these elliptic fibrations. Here we will contemplate the 643 CICY 3-folds with  $h^{1,1} \leq 4$  since all those with  $h^{1,1} > 4$  can be (obviously) elliptically fibred [35–38]. As we have an unbalanced dataset (with 53 non-elliptic and 590 elliptic), we would like to make an enhancement on 53. Notice that CICY configurations are the same up to row and column permutations. We can therefore take 10 random permutations (independently) of both rows and columns on each of the 53 configuration matrices, which yields  $10^2 \times 53 = 5300$  non-elliptic cases with output 0. We also perform 3 such permutations so that we have  $3^2 \times 590 = 5310$  elliptic cases with output 1. Moreover, as these CICYs can all be represented by configuration matrices with 6 rows and 7 columns, the input will be a  $6 \times 7$  matrix.

Now we are dealing with our familiar binary queries. Following the similar recipe as above, we can draw the learning curves in Fig. 16. The error bars for up to about 25%

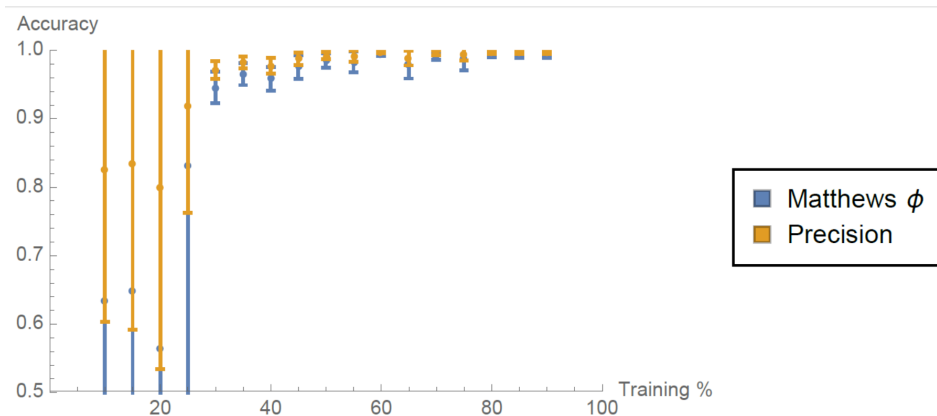


Figure 16: The learning curves for the (enhanced) ten thousand data-points on EFs.

looks ugly as there is inadequate training data. However, from 30% and above, machine-learning gives us a pretty nice result with high accuracy. Again, each training only takes a few seconds.

It is also worth noting that we can make a *control test* for this problem. We just arbitrarily choose 53 configuration matrices out of the 643 and assign 0 or 1 to these 53 matrices randomly. Then we have the learning curves depicted in Fig. 17. We see that the machine-

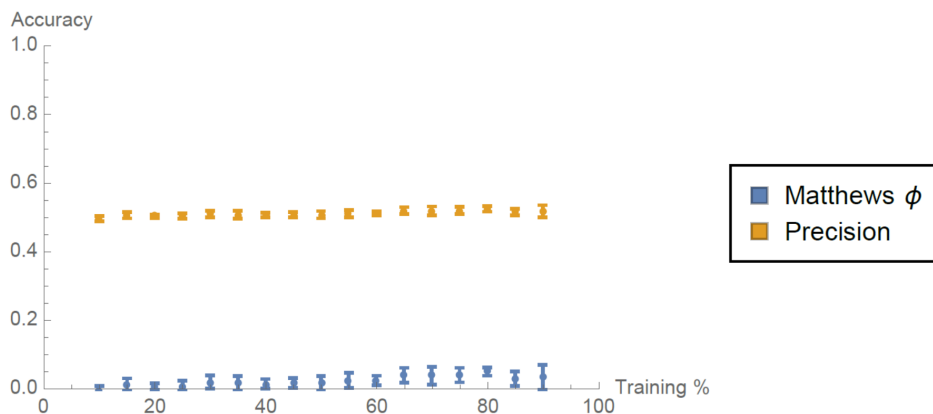


Figure 17: The learning curves on a control set of a randomly chosen property.

learning is poorly behaved (with  $\sim 50\%$  precision and  $\phi \sim 0$ ) which shows that there is no inherent pattern for the machine to find in the control test. In contrast, EF is truly not a random property.

## 11 A Digression: Group Theory

### 11.1 Learning Cayley Tables

Now we would like to apply machine-learning to more basic problems in mathematics. Let us first start with recognizing Cayley tables  $\mathcal{C}$  [39] out of Latin squares  $\mathcal{L}$ .

Allowing permutations of rows and columns, the number of Cayley tables of size  $n$  will be<sup>10</sup>  $\#\mathcal{C}_n = 1, 2, 6, 48, 120, 1440, \dots$ . These Cayley tables form a subset of Latin squares<sup>11</sup>  $\mathcal{L}$ . The number of Latin squares grows as:  $\#\mathcal{L}_n = 1, 2, 12, 576, 161280, 812851200, \dots$ . Compare  $\#\mathcal{C}$  with  $\#\mathcal{L}$ , we see that the probability of a Latin square being a Cayley table is *essentially* 0 from  $n$  as small as 5. This is important for our algorithm so that we can choose  $n \geq 5$ , and assign  $\mathcal{L} \rightarrow 0$  and  $\mathcal{C} \rightarrow 1$ . Thus, we are back to our familiar binary query. A more detailed descriptions of algorithms can be found in [40]. Here, we will just give out the learning curves as in Fig. 18. Both of the measures show that we have a perfect result when we have

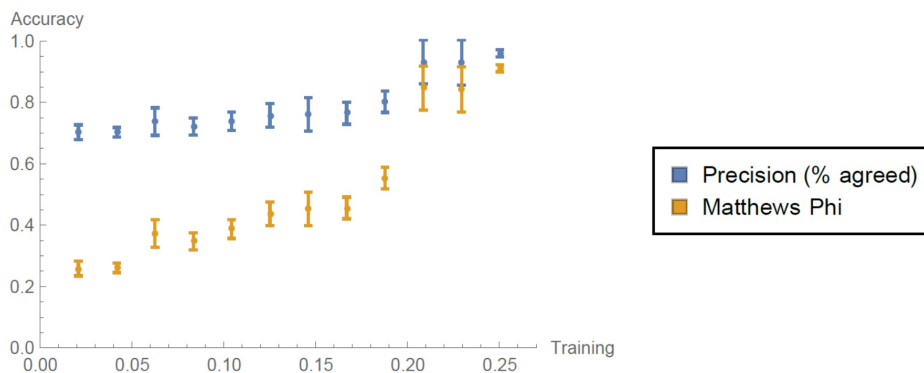


Figure 18: The learning curves for  $n = 8$  Latin squares. We vary the training size from 500 to 6000 in increments of 500.

$\sim 25\%$  out of the total data as training data.

### 11.2 Learning Finite Simple Groups

After recognizing the Cayley tables, one may wonder how machine-learning would perform when studying other group properties. We now focus on the problem of finite simple groups.

We know that it is not straightforward to determine whether a finite group is simple by just contemplating its Cayley table. However, there shouldn't be random properties in group

<sup>10</sup>If we naively consider all possible permutations, then  $\#\mathcal{C}_n$  should be  $(n!)^2 \times \#G$  (where  $\#G$  denotes the number of elements in group  $G$ ). However, Cayley tables always have degeneracies under permutations which leads to only  $(n!) \times \#G$  distinct matrices though this is not obvious for non-symmetric matrices (which corresponds to non-abelian groups).

<sup>11</sup>A Latin square is a  $n \times n$  matrix filled by  $n$  symbols (here  $1, 2, \dots, n$ ), each of which appears exactly once in each row and in each column.

theory, so we would like to see how machines can behave in this task. The detailed treatment can be found in [40]. Here, we report the learning curves as in Fig. 19. We see that even at

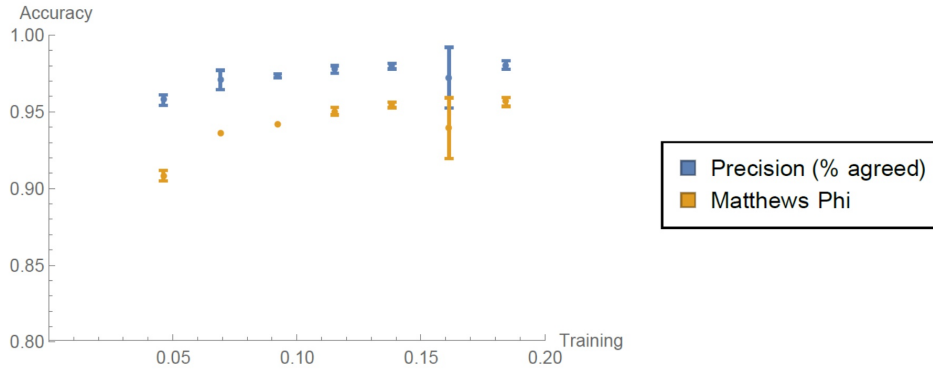


Figure 19: The learning curves for identifying whether a finite group of order  $n \leq 70$  is simple. There are 20 simple groups out of the total 602 groups.

a low percentage of training data, the machine can still make very good predictions directly from Cayley tables without knowing Sylow theorems.

More examples of studying group properties and algebraic structures via machine-learning can be found in [40]. Recently, similar explorations have also been done in number theory [41].

## 12 Summary and Outlook

As argued in [8, 9], any computational algebraic geometry problem is machine-learnable as it is in essence a finite number of steps finding kernels and cokernels of integer matrices. Thus, machines can be pretty well-behaved although they know nothing about algebraic geometry. On the other hand, this also shows that our properties in algebraic geometry are not random. Otherwise, our machine-learning would fail, just like the control case in §10.1.

Hence, we expect that machine so far is not able to learn number theory in which elusive prime numbers play a key role. As a sanity check, when we input a bunch of prime numbers and train our machine to valid larger primes in our data, we only achieve a terrible 0.1% accuracy. After all, AI does not stand for all-powerful incredibility. Machine learning is good at matrix/tensor manipulations (and this is why `TensorFlow` is given such a name).

Anyway, algebraic geometry is the area where machine learning can show its power as we have seen above. Despite our success in machine-learning, we still do not know why this works. Unlike other disciplines in science, we do not know what is going on among neurons and their connections, and we can still get good predictions from them regardless of the theoretical intractability. An *almost* 1 accuracy certainly cannot satisfy mathematicians, but machine learning still bypasses the expensive steps for practical purposes.

After decades of research, string landscape now solidly resides in the era of big data. CY manifolds is only a small portion of the heterotic landscape. There is plenty of room at the bottom where NNs can act as classifiers or predictors for generalized Kähler geometry, for

stable holomorphic bundles, for quivers and brane tilings and so forth. The landscape is still on the threshold of benefiting from data science.

## Acknowledgement

We would like to thank Thomas Creutzig and Steve Rayan for organizing the ‘‘PIMS - USaskatchewan Summer School on Algebraic Geometry in High-Energy Physic’’ which provided a wonderful atmosphere for mathematicians and physicists, students and experts, to interact and collaborate. YHH would like to thank STFC for grant ST/J00037X/1. EH would like to thank STFC for the PhD studentship.

## A Some Complex Geometry

Let us consider a complex manifold  $\mathcal{M}$  of complex dimension  $m$ . Then the set of  $(p,q)$ -forms,  $\Omega^{p,q}(\mathcal{M})$ , obviously form an abelian group under addition. As  $\mathcal{M}$  is also a  $2m$ -dimensional real manifold, we can decompose the familiar exterior derivative into two pieces:  $d = \partial + \bar{\partial}$ , such that  $\partial$  acts on the holomorphic part of a  $(p,q)$ -form while  $\bar{\partial}$  acts on the antiholomorphic part, that is,  $\partial : \Omega^{p,q}(\mathcal{M}) \rightarrow \Omega^{p+1,q}(\mathcal{M})$  and  $\bar{\partial} : \Omega^{p,q}(\mathcal{M}) \rightarrow \Omega^{p,q+1}(\mathcal{M})$ . Followed from  $d^2 = 0$ , we get  $\bar{\partial}^2 = 0$  (as well as  $\partial^2 = 0$  and  $\partial\bar{\partial} + \bar{\partial}\partial=0$ ). We may therefore construct the *cochain complex*:

$$0 \xrightarrow{\bar{\partial}} \Omega^{p,0}(\mathcal{M}) \xrightarrow{\bar{\partial}} \Omega^{p,1}(\mathcal{M}) \xrightarrow{\bar{\partial}} \dots \xrightarrow{\bar{\partial}} \Omega^{p,m}(\mathcal{M}) \xrightarrow{\bar{\partial}} 0. \quad (\text{A.1})$$

Then

**Definition A.1.** The *Dolbeault cohomology group* is defined as

$$H_{\bar{\partial}}^{p,q}(\mathcal{M}) := \frac{\ker(\bar{\partial} : \Omega^{p,q}(\mathcal{M}) \rightarrow \Omega^{p,q+1}(\mathcal{M}))}{\text{Im}(\bar{\partial} : \Omega^{p,q-1}(\mathcal{M}) \rightarrow \Omega^{p,q}(\mathcal{M}))}. \quad (\text{A.2})$$

The dimensions of the Dolbeault cohomology groups are known as the *Hodge numbers*,  $h^{p,q} := \dim H_{\bar{\partial}}^{p,q}(\mathcal{M})$ .

Not all the Hodge numbers are independent. For any complex manifold, Kodaira-Serre duality yields [14]

$$h^{p,q} = h^{m-p,m-q}. \quad (\text{A.3})$$

In particular, we have

$$h^{0,0} = h^{m,m} = 1. \quad (\text{A.4})$$

Also, one can prove that the Dolbeault cohomology group is always of finite dimension, viz,  $h^{p,q} < \infty$ .

## A.1 Kähler Manifolds

As we will see, Kähler structure would put more constraints on the Hodge numbers. First, we need to introduce

**Definition A.2.** A *Hermitian metric*  $g$  on the complex manifold  $\mathcal{M}$  with complex structure  $J$  is a Riemannian metric satisfying  $g(Jv, Jw) = g(v, w)$  for any vector fields on  $\mathcal{M}$ . Equivalently, in terms of (complex) components,  $g_{\alpha\beta} = g_{\bar{\alpha}\bar{\beta}} = 0$ . Then the *Hermitian form* is the 2-form defined by  $\omega(v, w) := g(Jv, w)$ . Equivalently, in terms of (complex) components<sup>12</sup>,  $\omega_{ab} = ig_{\alpha\bar{\beta}} - ig_{\bar{\alpha}\beta}$ . Therefore,  $\omega$  is also a (1,1)-form.

As the pure holomorphic and antiholomorphic components of the Hermitian metric vanish, it is not hard to see that  $\omega_{ab} = -\omega_{ba}$ .

Now we are able to define

**Definition A.3.** The Hermitian metric  $g$  is *Kähler* if  $d\omega = 0$ , and  $\omega$  is called a *Kähler form*. A complex manifold is a *Kähler manifold* if it admits a Kähler metric.

In some literature, the Kähler manifold is defined as a complex manifold having a symplectic form (being bilinear, non-degenerate and antisymmetric). In fact, bilinearity and non-degeneracy come from the Riemannian metric, and antisymmetry follows the Hermiticity of the metric as mentioned above.

It is worth remarking that since  $d\omega = 0$  (which is the same as  $\partial_\alpha g_{\beta\bar{\gamma}} = \partial_\beta g_{\alpha\bar{\gamma}}$  along with its conjugate equation), then equivalently we can write  $\omega = i\partial\bar{\partial}K$  for some real scalar function  $K$  known as the *Kähler potential*.

For Kähler manifolds, the Hodge numbers further satisfy

$$h^{p,q} = h^{q,p}, \quad (\text{A.5})$$

$$h^{p,p} \geq 1. \quad (\text{A.6})$$

Then (A.3) and (A.5) yield

$$h^{p,q} = h^{m-q, m-p}. \quad (\text{A.7})$$

If the manifold is further Calabi-Yau, one can show that  $h^{m,0} = h^{0,m} = 1$ , and  $h^{m,p} = h^{p,m} = h^{0,p} = h^{p,0} = 0$  for  $0 < p < m$  [6].

Kähler geometry is ubiquitous in physics. We care about Kähler manifolds since they preserve holomorphicity under parallel transportations of vectors, and a Kähler structure is essential for a manifold being Calabi-Yau.

## A.2 Chern Classes

Another important concept for Calabi-Yau manifolds is the *Chern classes*.

---

<sup>12</sup>We use Latin indices for real coordinates and Greek ones for complex coordinates.

**Definition A.4.** Given the complex vector bundle  $E$  over the complex manifold  $\mathcal{M}$  of complex dimension  $m$  and the gauge group (aka structure group)  $G$ , let  $F = dA + A \wedge A$  be the strength field (aka curvature 2-form) of the gauge potential (aka connection)  $A$ . We define the *total* Chern class as<sup>13</sup>

$$c(E) = \det \left( I + \frac{i}{2\pi} F \right). \quad (\text{A.8})$$

Using  $\det \left( I + \frac{i}{2\pi} F \right) = \exp \left( \text{tr} \left( \log \left( I + \frac{i}{2\pi} F \right) \right) \right)$ , we may expand the total Chern class as

$$c(E) = c_0(E) + c_1(E) + c_2(E) + \cdots + c_m(E), \quad (\text{A.9})$$

where  $c_k(E)$  is the  $k^{\text{th}}$  Chern class<sup>14</sup>. As  $F$  is a 2-form,  $c_k(E)$  is a  $2k$ -form and vanishes for  $k > m$ . There are some explicit formulae for the Chern forms such as

$$c_0(E) = 1, \quad (\text{A.10})$$

$$c_1(E) = \frac{i}{2\pi} \text{tr}(F), \quad (\text{A.11})$$

...

$$c_m(E) = \left( \frac{i}{2\pi} \right)^m \det(F). \quad (\text{A.12})$$

For more on Chern classes (including Chern characters, Todd classes etc.), see [43].

## B Toric Varieties

Toric varieties are a generalisation of complex weighted projective vector spaces. Complex projective space,  $\mathbb{C}\mathbb{P}^m$ , is the  $\mathbb{C}^{m+1}$  complex space with the origin removed, and quotiented out by the identification:

$$(z_0, z_1, \dots, z_m) \sim (\lambda z_0, \lambda z_1, \dots, \lambda z_m) \quad \forall \lambda \in \mathbb{C} \setminus \{0\}; \quad (\text{B.1})$$

such that all points along lines through the origin are identified. This concept is generalised to a weighted complex projective space, denoted  $\mathbb{C}\mathbb{P}^{(a_0, a_1, \dots, a_m)}$ , where instead the identification quotiented with is:

$$(z_0, z_1, \dots, z_m) \sim (\lambda^{a_0} z_0, \lambda^{a_1} z_1, \dots, \lambda^{a_m} z_m) \quad \forall \lambda \in \mathbb{C} \setminus \{0\}, \quad (\text{B.2})$$

with the  $a_i$  constants acting as powers on  $\lambda$ . Both weighted and unweighted complex projective spaces are types of toric variety, however more toric varieties can be formed through

<sup>13</sup>Although we use the 2-form  $F$  in our definition, the Chern class should be independent of the choice of  $F$  [42].

<sup>14</sup>Strictly speaking, they are *Chern forms*, and the Chern classes are the cohomology classes of the Chern forms. When  $E$  is the holomorphic line bundle  $T^{1,0}\mathcal{M}$ , we also say that  $c_k(E)$  is the Chern class of the manifold  $\mathcal{M}$  and denote it as  $c_k(\mathcal{M})$ .

more general removal of a subset of the space,  $\mathcal{U}$ , and quotienting by a more general algebraic torus  $(\mathbb{C} \setminus \{0\})^p$ . Thereby a toric variety is defined:

$$\mathcal{M} = (\mathbb{C}^m \setminus \mathcal{U}) / (\mathbb{C} \setminus \{0\})^p; \tag{B.3}$$

where there are  $p$  identifications to quotient out by with  $p$  coefficient sets of non-zero complex numbers. Note for unweighted complex planes all coefficients in the set are the same [44].

Toric varieties can also be defined in terms of fans. Whereby a fan is a collection of cones such that all cone faces are themselves cones in the fan, and cones intersect at mutual faces. A toric variety is then defined using the one-dimensional cones in a fan defined on a vector space obtained from a lattice  $(\mathbb{Z}^n \otimes_{\mathbb{Z}} \mathbb{R})$ . First a complex homogeneous coordinate is associated to each generator of a one-dimensional cone such that a fan with  $k$  generators, for the  $k$  one-dimensional cones, corresponds to  $\mathbb{C}^k$ . Next, all points in the space which correspond to combinations of these generators which are not contained within cones in the fan are removed from the space. Finally the remaining space is quotiented by equivalence relations that correspond to these generator combinations outside of the fan. More mathematically this defines the full toric variety as:

$$\mathcal{M} = \{\mathbb{C}^k \setminus \mathcal{U}\} / \mathcal{G}; \tag{B.4}$$

for  $\mathcal{U}$  as the set of generator combinations that are not contained within the fan; and  $\mathcal{G}$  the algebraic torus quotiented by (potentially with some additional finite abelian group) to give the equivalence relations, usually  $(\mathbb{C} \setminus \{0\})^p$ .

This construction method can be more practical, as singularities in a manifold correspond to singularities in its corresponding toric variety; and singularities in a toric variety can be resolved by introducing further cones into the variety's fan in a specific way.

A general toric variety is Calabi-Yau if its generators exist in an affine hyperplane of the lattice. This makes identifying Calabi-Yau manifolds very easy from the fan structure. The Calabi-Yau property can alternatively be identified from the charges of the equivalences quotiented out by in the variety definition. If the sum of the coefficients ("charges") for each equivalence relation is zero for all the relations, then the space is also Calabi-Yau. Since a Toric variety is compact if its fan fills the lattice space; by definition all Calabi-Yau manifolds formed from toric varieties are thus non-compact, as their generators do not span the lattice (but exist in a codimension 1 hyperplane of it).

Toric diagrams, which are useful for manifold classification and physical interpretation, can then be constructed. Considering the hyperplane containing the Calabi-Yau generators in the lattice space, a graph can be formed by connecting the points where this hyperplane intersects with the fan's generators. The dual of this graph is defined to be the manifold's "toric diagram", and it encodes the degeneration of the fibres of the manifold.

The toric diagram in figure 20 represents a new branch of Calabi-Yau surfaces, known as conifolds. These permit conic singularities in their description, and were key in deriving mirror symmetry which connects Calabi-Yau 3-folds. The charges defining this conifold are

$Q_i = (1, 1, -1, -1)$ , since these sum to zero the space is thus Calabi-Yau. Generators are created that all exist in the same hyperplane (of the form  $(v_i, 1)$  for  $v_i$  a 2-dimensional vector), and satisfy the relation  $\sum_i Q_i \cdot v_i = 0$ . This gives the  $v_i$ 's as the points labelled in figure 20, where they are drawn as points in the  $x_3 = 1$  hyperplane.

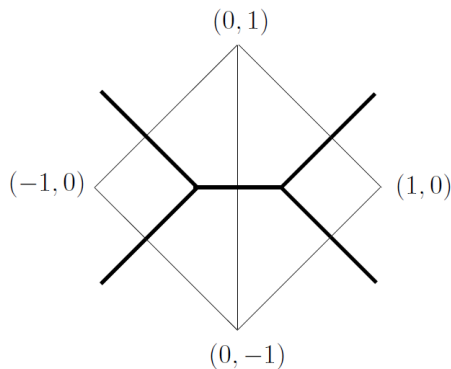


Figure 20: An example of a toric diagram (in bold) for the resolved conifold [42]. The points give the intersection of the one-dimensional cone generators of the fan with the hyperplane they all exist in under the Calabi-Yau condition. Connecting the points gives the graph, whose dual is the toric diagram.

The full  $(v_i, 1)$  vectors define the 1-dimensional cone generators of the fan in the lattice space. In the corresponding complex space, where each of these generators has a coordinate associated to it, the above conifold example is thus defined by the relation:

$$\sum_i Q_i \cdot |z_i|^2 = |z_1|^2 + |z_2|^2 - |z_3|^2 - |z_4|^2 = t, \tag{B.5}$$

for a parameter  $t$  known as the Kähler parameter (which counts the number of Kähler forms on the manifold). Considering the conifold as a fibration of a base space with a  $T^3$  torus, this relation along with the other boundary relations  $|z_i|^2 = 0$  define the base space of the conifold total space. The intersections of the hyperplanes defined by these equations gives an equivalent version of the toric diagram, and the full conifold space is the base with its fibration. This is where the conic idea behind the “conifold” name comes from [42].

## C Introduction to Machine Learning

### C.1 Text recognition

To make a start, let us first contemplate a prototypical example of text recognition [7–9]. Given the 10 handwritten digits:

$$1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 0, \tag{C.1}$$

we would want the computers to recognize them. Notice that the inputs, which are images, are essentially  $m \times n$  matrices representing the pixels in our 2-dimensional grid. The entries are either 0 or 1, encoding the black-white information for our binary images<sup>15</sup>. The outputs are simply the ten integers from 0 to 9 (aka *10-channel outputs*).

For physicists and mathematicians, it is natural to think of the hardcore approach to solve this problem by finding the Morse function via detecting the different critical points for different digits. We are able to do so because the shapes vary from one digit to another. However, there will be two cons: the variation of different hand-writings and the too expensive computation.

This is pretty much the situation we have in algebraic geometry. For instance, the Gröbner basis is way too expensive for computation, and the input may also vary in configuration [7–9].

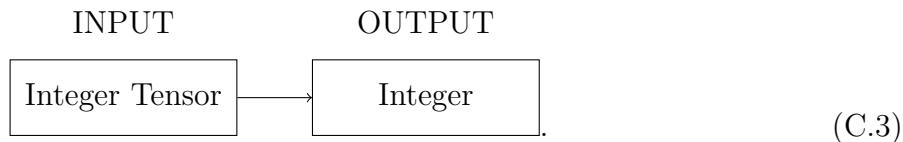
Computer scientists and data scientists tell us that we can machine-learn this problem as the following steps:

1. Data Acquisition: We collect adequate known cases (input→output), which are called *training data*. For instance, the National Institute of Standards and Technology (NIST) database [45] has  $\sim 10^6$  samples in the form:

$$\begin{array}{l} 6 \rightarrow 6, 8 \rightarrow 8, 2 \rightarrow 2, 4 \rightarrow 4, 8 \rightarrow 8, 7 \rightarrow 7, 8 \rightarrow 8, \\ 0 \rightarrow 0, 4 \rightarrow 4, 2 \rightarrow 2, 5 \rightarrow 5, 6 \rightarrow 6, 3 \rightarrow 3, 2 \rightarrow 2, \\ 9 \rightarrow 9, 0 \rightarrow 0, 3 \rightarrow 3, 8 \rightarrow 8, 8 \rightarrow 8, 1 \rightarrow 1, 0 \rightarrow 0, \dots \end{array} \quad (\text{C.2})$$

2. Neural Network (NN) Setup and Training: This is the core machine-learning part that we will fixate on in §C.2. If the NN is sufficiently complex, we will call it deep learning.
3. Validation: After our machine/AI has “learnt” the training data, we can feed it with unseen data named *validation data* in the same form as training data. This test will then reflect how the machine performs after training.
4. Prediction: If the NN passes our validation test, then it can be used in applications.

Similar to the example above, a typical problem in string theory and computational algebraic geometry has the format:



As a FLOSS (Free/Libré and Open Source Software) with pseudo-code nature, `Python` is simple and popular. Hence, `Python` becomes our preference<sup>16</sup>. There are also standard

<sup>15</sup>If we want colour images, then the entries of these matrices would range from 0 to 1, indicating the percentage of RGB values.

<sup>16</sup>Beginners can refer to [46] which is a well-known free book on `Python`.

softwares for our Python programming language such as SageMath (aka Sage or SAGE) [47] and TensorFlow [48]. Perhaps, the only shortage of Python is that it is not as fast as C or Java. Fortunately, we are allowed to run Python on the Java platform thanks to Jpython/Jython<sup>17</sup>.

## C.2 Neural Networks

Motivated by biological neural networks, (*artificial*) *neural networks* ((A)NNs) are essentially a set of algorithms designed to find patterns/relationships of our input data. They cluster and label the data we feed them. In the end, they return numerical results interpreting the patterns they recognize.

The basic components of NN are units/nodes called *neurons* analogous to human brains. As we will see below, a neuron in the network is a *function* that collects and classifies data. The neurons are often organized in *layers*, and there may be multiple connections among different layers. A large collection of neurons then gives rise to an NN. The more layers or the greater the complexity of the inter-connectivity we have, the deeper our learning is.

**Definition C.1.** A single-layer *perceptron* (SLP) is an archetypal neuron which is a function called *activation function*,  $f(x_i)$ , of some input vector  $x_i$ . The activation function is typically taken to be binary (i.e. Heaviside function), or a sigmoid function such as the hyperbolic tangent or the logistic function<sup>18</sup>. The activation function is set to contain real parameters of the form  $f(\sum_i w_i x_i + b)$ , where  $w_i$ 's and  $b$  are called *weights* and *bias* respectively.

Another widely used function is the famous rectified linear unit (ReLU) activation function:

$$f(x) := \begin{cases} x, & x > 0; \\ 0, & x \leq 0. \end{cases} \quad (\text{C.4})$$

As a result, this is less computationally expensive than sigmoid functions. The sparsity of ReLU makes it behave more like a biological brain. Moreover, it has a better gradient propagation compared to sigmoid functions who has tiny gradients towards the ends. If the gradient is too small or even vanished, then the NN would learn rather slowly or even refuse to learn further. Although the gradient of ReLU will vanish for negative  $x$ , no activation function is always perfect for every case, and we need to choose the one fits best.

Now given some training data

$$T = \{(x_i^{(j)}, d^{(j)})\}, \quad (\text{C.5})$$

---

<sup>17</sup>We also have Cython as a compiler which compiles our Python codes to C. Notice, however, Jython and Cython are different in principle. In particular, speeding up is pretty much the goal for the latter while it is not for Jython which mainly aims to import and use Java classes. Anyway, both of them make our codes run faster and get rid of the heavy codes in the meantime.

<sup>18</sup>We choose sigmoid functions because they lie between 0 to 1, hence useful in probability predictions. Besides, they are differentiable with relatively simple derivatives proportional to themselves.

where  $x_i^{(j)}$ 's are inputs and  $d^j$ 's are *known* outputs with labelling  $j$ . Then we want our error, the standard deviation

$$SD = \sum_j \left( f \left( \sum_i w_i x_i^{(j)} + b \right) - d_i \right)^2, \quad (\text{C.6})$$

to be minimized with respect to  $w_i$ 's and  $b$  by the method of steepest descent<sup>19</sup>. After such training, we can now proceed to validation against unseen data. As we can see, this is basically the (non-linear) regression for model function  $f$ .

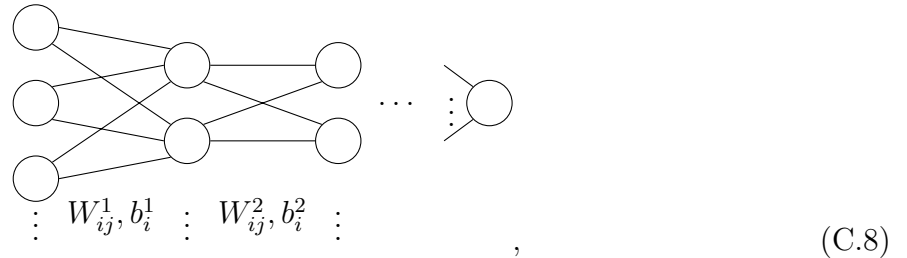
As aforementioned, we can have many layers for deep learning. This leads to

**Definition C.2.** A *multi-layer perceptron* (MLP) is a sequence of layers where the output of the previous layer is the input of the next layer, with different weights and biases. The output of the  $i^{\text{th}}$  neuron in the  $n^{\text{th}}$  layer is

$$f_i^n = f(W_{ij}^n f_j^{n-1} + b_i^n). \quad (\text{C.7})$$

Notice that we have promoted the weights  $w_j$ 's to a (layer-wise) weight matrix  $W^n$  such that the  $W_{ij}^n$  denotes the weight  $w_j$  connected from neuron  $j$  in the  $(n-1)^{\text{th}}$  layer to neuron  $i$  in the  $n^{\text{th}}$  layer. Likewise, the bias  $b$  has become a bias vector  $b_i$ . Our input would be  $f_i^0 = x_i$ .

Thus, the MLP is depicted as



where the left-most layer is the input layer while the right-most layer is the output layer. The layers between them are called hidden layers. This is just the simplest NN which only involves forward propagation from left to right. NNs can allow backward propagation and cycles as well. Besides MLP, we also have other NNs such as convolutional neural network (CNN) [7–9, 49]. Based on discussions before, we have

**Definition C.3.** A *convolutional (neural) network* is a neural network with general matrix multiplications replaced by convolutions in at least one of its layers.

For a simplest case with no hidden layers, the output (aka *feature map*)  $s$  is

$$s = (x * w)(T) = \int x(t)w(T-t)dt, \quad (\text{C.9})$$

<sup>19</sup>This is often done numerically due to the large number of parameters.

where  $x$  is the input and  $w$  is the weighting function called *kernel* that restricts our input. For instance, we would like to learn how a stretch of coastline varies over a period of time. Then our input would be the position of the coastline with variable  $T$  being the time we make each measurement. Moreover, measurements at later times (indicated by  $t$ ) are more relevant, hence acquire more weights controlled by the kernel.

This is basically how CNNs use filters when scanning images. Instead of detecting images pixel by pixel, CNNs, just like what humans do, “look at” an area known as *receptive field* of an image. The movement of filter is measured by *stride*. If the stride is 2, then the filter will move 2 pixels each time. Hence, CNNs are very powerful dealing with graphic inputs. Thus, for a 2-dimensional image as our input  $x$ , we want to use convolutions over both of the two axes. Then

$$S(i, j) = (x * w)(i, j) = \sum_m \sum_n x(m, n)w(i - m, j - n), \quad (\text{C.10})$$

where the kernel  $w$  is now of dimension 2. As the dummy variables often varies less, we can *flip* the kernel due to commutativity of convolutions:

$$S(i, j) = (w * x)(i, j) = \sum_m \sum_n w(m, n)x(i - m, j - n). \quad (\text{C.11})$$

Quite often, people do not bother flipping the kernel and use the *cross-correlation*<sup>20</sup> instead:

$$S(i, j) = (w * x)(i, j) = \sum_m \sum_n x(i + m, j + n)w(m, n). \quad (\text{C.12})$$

Besides the convolutional layers (and ReLU layers), a CNN also has *pooling* layers that combine the outputs of neurons in one layer into a single neuron in the next layer and fully connected layers to connect every neuron in one layer to every neuron in the next layer.

For more detailed stuff on NN, there are good texts such as [50–52].

### C.3 Support Vector Machines

Besides NN approach to machine learning, we also have support vector machines (SVMs), decision trees,  $k$ -nearest neighbours ( $k$ -NNs) etc. Here we are going to introduce the most widely used SVMs. SVMs, which take a more geometric approach compared to NNs, can act as both classifiers and regressors<sup>21</sup>.

---

<sup>20</sup>It is worth noting that some literature call cross-correlations convolutions as well.

<sup>21</sup>In brief, programmes are asked to specify which categories the inputs belong to in classifications, and programmes need to predict a numerical value given some input in regressions. Hence, for classifiers, the function maps from input variables  $x_i$  to *discrete* output variables  $y_i$  while for regressors, the function maps from input variables  $x_i$  to *continuous* output variables  $y_i$ . Besides classifications and regressions, there are other tasks in machine learning as well. See [49].

When an SVM acts as a classifier, it establishes an optimal hyperplane in the  $n$ -dimensional *feature space* where the input  $n$ -vectors live in, indicating the binary feature as required in a classification predictive modelling. We can define the hyperplane as

$$\{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = 0\}, \quad (\text{C.13})$$

where  $\mathbf{w}$  is the normal vector to the hyperplane. Then

**Definition C.4.** The *support vectors* are the points in the feature space lying closest to the hyperplane on either side denoted as  $\mathbf{x}_i^\pm$ . The *margin*  $M$  is the distance between these two vectors projected along  $\mathbf{w}$ , viz,

$$M := \mathbf{w} \cdot (\mathbf{x}_i^+ - \mathbf{x}_i^-) / |\mathbf{w}|. \quad (\text{C.14})$$

Our optimal hyperplane will maximize the margin. The reason is that we want our points in different classes as far from the separating hyperplane as possible since points close to the boundary can be easily misclassified. As the hyperplane is invariant under rescaling ( $f = 0 \rightarrow \alpha f = 0$ ), we can rescale  $\mathbf{w}$  such that  $f(\mathbf{x}_i^\pm) = \pm 1$  and  $M = 2/|\mathbf{w}|$ . Thus, maximizing the margin is realized via minimizing  $|\mathbf{w}|$ . Moreover, as a classifier, the SVM gives  $y_i=1$  if  $\mathbf{w} \cdot \mathbf{x}_i + b \geq 1$  while  $y_i=-1$  if  $\mathbf{w} \cdot \mathbf{x}_i + b \leq -1$ . Equivalently, this yields  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$  for each  $i$ . Therefore, our problem is actually to minimize  $|\mathbf{w}|$  with the constraints  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$ . This can be solved using Lagrange multipliers<sup>22</sup>:

$$L = \frac{1}{2}|\mathbf{w}|^2 - \sum_i \alpha_i (y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1). \quad (\text{C.15})$$

Hence,

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0; \\ \frac{\partial L}{\partial b} &= - \sum_i \alpha_i y_i = 0. \end{aligned} \quad (\text{C.16})$$

Likewise, the (linear) SVM regressor (aka support vector regression, SVM) follows the same principles, only with the difference of the output being real which make it difficult to make predictions. Now we have a tolerance of errors that allows  $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$  to deviate from the actual result  $y_i$  at most  $\epsilon$ . Meanwhile, we want to keep  $f$  as flat as possible. As  $|\nabla f|^2 = |\mathbf{w}|^2$ , our problem boils down to minimize  $|\mathbf{w}|^2/2$  subject to the condition  $|y_i - (\mathbf{w} \cdot \mathbf{x}_i + b)| \leq \epsilon$ . This can be solved using Lagrange multipliers as well:

$$L = \frac{1}{2}|\mathbf{w}|^2 - \sum_i \alpha_i (y_i - (\mathbf{w} \cdot \mathbf{x}_i + b) + \epsilon) + \sum_i \alpha_i^* (y_i - (\mathbf{w} \cdot \mathbf{x}_i + b) - \epsilon). \quad (\text{C.17})$$

---

<sup>22</sup>We use  $|\mathbf{w}|^2$  rather than  $|\mathbf{w}|$  here so as to get rid of the square root in the norm to make our life easier.

Hence,

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_i (\alpha_i - \alpha_i^*) \mathbf{x}_i = 0; \\ \frac{\partial L}{\partial b} &= -\sum_i (\alpha_i - \alpha_i^*) = 0.\end{aligned}\tag{C.18}$$

Notice that we actually require that such  $f$  does exist with precision  $\epsilon$  for all data-points. Sometimes, this may not be feasible. We introduce the *slack variables*  $\xi_i$  and  $\xi_i^*$  for each point such that our problem becomes minimizing  $\left(\frac{1}{2}|\mathbf{w}|^2 + C \sum_i (\xi_i + \xi_i^*)\right)$  with the constraints

$$\begin{aligned}y_i - (\mathbf{w} \cdot \mathbf{x}_i + b) &\leq \epsilon + \xi_i; \\ (\mathbf{w} \cdot \mathbf{x}_i + b) - y_i &\leq \epsilon + \xi_i^*; \\ \xi_i, \xi_i^* &\geq 0.\end{aligned}\tag{C.19}$$

The constant  $C$  is the *box constraint* which determines the trade-off between the flatness of  $f$  and the amount up to which deviations larger than  $\epsilon$  are tolerated [53]. This is depicted in Fig. 21. Now we need to add

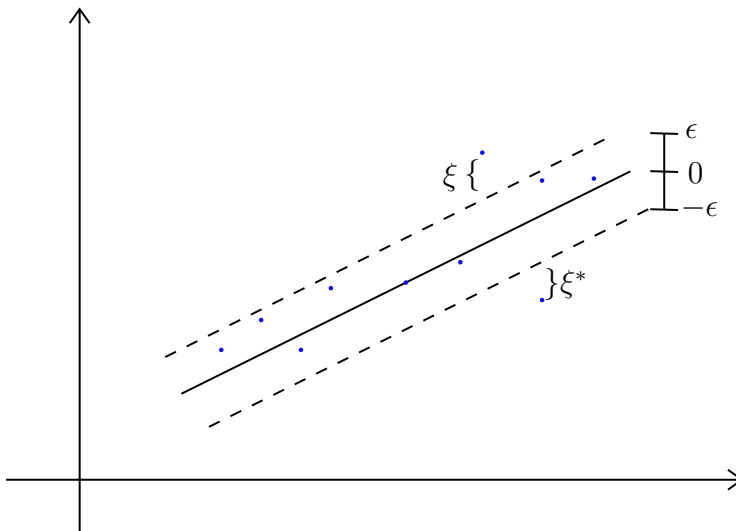


Figure 21: The “soft margin” setting for linear SVR.

$$C \sum_i (\xi_i + \xi_i^*) - \sum_i (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_i \alpha_i \xi_i - \sum_i \alpha_i^* \xi_i^*\tag{C.20}$$

to (C.17) and

$$\begin{aligned}\frac{\partial L}{\partial \xi_i} &= C - \eta_i - \alpha_i; \\ \frac{\partial L}{\partial \xi_i^*} &= C - \eta_i^* - \alpha_i^*\end{aligned}\tag{C.21}$$

to (C.18).

Furthermore, one can easily imagine that we cannot always separate the points linearly<sup>23</sup>. This then requires the nonlinear SVMs. We will not discuss such cases here.

## C.4 Decision Trees

As the name suggests, this is a tree-like model for making decisions. Decision tree learning can be used to cope with both categorical and numerical data, and therefore comes the term CART (classification and regression tree). There are also other advantages of decision trees. For instance, they are white box models so that the results can be explained by Boolean logic, unlike NNs.

Suppose we are working at CERN and want to understand what particles are created in our experiment. We let the jet traverse the CMS so that different particles would stop in different detectors. The decision tree is then depicted in Fig. 22. This tree is drawn upside

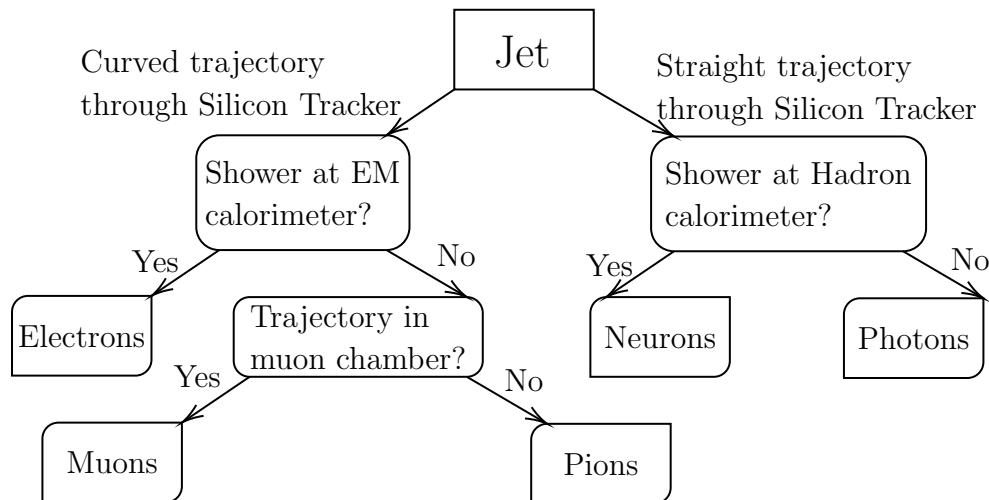


Figure 22: Particles in CMS.

down with the *root node* at the top. Internal nodes are known as *decision nodes*, and the nodes at the end are called *leaves* or *terminal nodes*. We call the process of dividing a node (*parent node*) into sub-nodes (*child nodes*) a *splitting*. A subsection of a tree is known as a *branch*.

In general, we make a collection of rules based on our variables to get a best split of our data set. This yields the child nodes, and the same process is acted on each children node. As a result, this is a recursion procedure. Finally, the splitting will stop when no further gain can be made or we have some stopping rule preset.

The algorithms we use in decision trees are essentially greedy algorithms since a best choice is made at each decision. Different algorithms have different ways to measure the quality of the choice/splitting.

<sup>23</sup>Examples can be found in Figure 6.1 in [54].

Often appeared in algorithms such as ID3 and C4.5, *information gain* measures how “best” our splitting is<sup>24</sup>.

**Definition C.5.** An *attribute* (aka *feature*) is an individual measurable property or characteristic of a phenomenon being observed [55]. For a dataset  $D$  after split on an attribute  $A$ , the *information gain* (aka *mutual information*)  $I(D, A)$  is

$$I(D, A) = S(D) - S(D|A), \quad (\text{C.22})$$

where  $S(D) = -\sum_{j \in J} p(j) \log_2 p(j)$  is the Shannon entropy. We denote the set of classes as  $J$ , and  $p(j)$  is the percentage of the number of elements in class  $j$  in the number of total elements in  $D$  such that  $\sum_{j \in J} p(j) = 1$ . Moreover,  $S(D|A)$  is the conditional entropy such that

$$S(D|A) = -\sum_a p(a) \sum_j p(j|a) \log_2 p(j|a) \quad (\text{C.23})$$

where  $p(a)$  is the proportion of  $A = a$  and  $p(j|a)$  is the proportion  $p(j)$  constrained by  $A = a$  (cf. conditional probability).

As entropy is used to quantify the uncertainty (or equivalently, our knowledge) of our information, the information gain, as the name suggests, measures the difference of entropy between before and after the dataset  $D$  is split on attribute  $A$ . It shows how much uncertainty is reduced (or equivalently, how much knowledge we gain) after splitting. Now that we are making best choices, we often want the information gain to be maximized at each step.

In our particle jet example in Fig.22, there are four attributes: charge (charged or uncharged, determined by Sicicon tracker), (primarily) electromagnetic interaction (yes or no, determined by EM calorimeter), (primarily) nuclear interaction (yes or no, determined by hadron calorimeter) and absorption by calorimeters (yes or no, determined by muon chamber). For instance, let us compute the information gain of the first splitting based on charge. The Shannon entropy of the five types/classes of particles is<sup>25</sup>

$$S(D) = -5 \times \left( \frac{1}{5} \log_2 \frac{1}{5} \right) = \log_2 5. \quad (\text{C.24})$$

There are three types of particles carrying charge, and two being neutral. Thus, the conditional entropy is

$$S(D|\text{Charge}) = -\frac{3}{5} \times 3 \times \left( \frac{1}{3} \log_2 \frac{1}{3} \right) - \frac{2}{5} \times 2 \times \left( \frac{1}{2} \log_2 \frac{1}{2} \right) = \frac{3}{5} \log_2 3 + \frac{2}{5}. \quad (\text{C.25})$$

Hence, the information gain is

$$I = \log_2 5 - \left( \frac{3}{5} \log_2 3 + \frac{2}{5} \right) = 0.97. \quad (\text{C.26})$$

---

<sup>24</sup>Other algorithms may use Gini impurity, variance reduction and so forth.

<sup>25</sup>Here we only care about different kinds of particles, rather than the percentage of particle numbers of each kind in the jet.

Decision trees are frequently used in machine learning and data mining. A main problem of decision trees is *overfitting*. The algorithm may build a tree too close to the data, which leads to an overfitting tree. This is often over-complex and may give poor performance during validation and prediction. A technique known as *pruning* is then used to reduce the size of learning trees. For more details, one is referred to [54, 56].

## C.5 Types of Machine Learning

So far we have been mainly talking about various models used in machine learning, there are different machine learning algorithms based on different types. Here, we will quickly introduce the three basic machine learning paradigms:

- **Supervised Learning:** In supervised learning, our data is split into training and validation data. These data contain both inputs and outputs. Then we will train the machine with training data and examine its learning result using validation data. We are mainly using this approach in the next few sections, and a more detailed instruction is given in 9 by an example. The models of algorithms aforementioned (NNs, SVMs, decision trees etc.) are all commonly used in supervised learning. It is worth mentioning that the *no free lunch theorem*<sup>26</sup> tells us no learning algorithm would always work best on all supervised learning problems.
- **Unsupervised Learning:** In unsupervised learning, the algorithm is asked to find unknown patterns of the data without labels. Hence, the data only contains inputs. One usual approach in unsupervised learning is cluster analysis where the machine groups the data-points with similar properties. There is a category in machine learning which hybridize supervised and unsupervised learnings known as semi-supervised learning as well. Further discussions can be found in Part 3 of [54].
- **Reinforcement Learning (RL):** A reinforcement learning is usually modelled as a *Markov decision process* (MDP), where

**Definition C.6.** A Markov decision process is a tuple  $(S, A, P, R, \gamma)$  where  $S(\ni s)$  is a finite set of states and  $A(\ni a)$  is a set of actions. Then  $P$  is the *state transition probability matrix* such that  $P_{ss'}^a = \text{Prob}(S_{t+1} = s' | S_t = s, A_t = a)$  with  $t$  labelling the time-step, and  $R$  is the *reward function* such that  $R_s^a = E(R_{t+1} | S_t = s, A_t = a)$  with  $E$  being the expectation value and  $R_t$  being the reward at  $t$ . Moreover,  $\gamma \in [0, 1]$  is the *discount factor*.

---

<sup>26</sup>This may reminds some readers of the famous saying by Alan Guth: “The universe is the ultimate free lunch.” Apparently, this refers to different contexts and topics from here. Although not everyone is a fan of inflation, we all agree that vacuum is full of fluctuations. It is kind of interesting to see a proverb quoted in various areas with different meanings.

As the MDP is built on the Markov chain, it inherits the memorylessness property. To “reinforce” our learning, we not only give rewards to the *agent* (which is the component making decisions of what action to take), but also accumulate the rewards:

**Definition C.7.** A *return*  $G_t$  is the total discounted reward from  $t$ , viz,  $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$ .

We see that discount factor is added since a reward received at present is more worthy than delayed rewards<sup>27</sup>. Therefore, the agent is able to know the value of being in a state  $s$  in long term:

**Definition C.8.** The *state-value function* is  $v_{\pi}(s) = E(G_t | S_t = s)$ , and the *action-value function* is  $q_{\pi}(s, a) = E(G_t | S_t = s, A_t = a)$ , where  $\pi$  is the *policy* such that  $\pi(a|s) = \text{Prob}(A_t = a | S_t = s)$ , which defines the behaviour of our agent.

RL is now used not only in machine learning, but also in fields such as game theory, information theory, statistics and so forth. Such reinforcement learning with rewards reminds us of the Skinner box—we wish to explore this point in future works. Recently, RL is also applied to studying the string vacua as in [57].

With different algorithms and categories in machine learning, what we need is an appropriate way to measure how well our machine learns. This is discussed in §9.

## References

- [1] M. R. Douglas, “Calabi-Yau metrics and string compactification,” *Nucl. Phys.* **B898** (2015) 667–674, [arXiv:1503.02899 \[hep-th\]](#).
- [2] E. Calabi, “On Kähler manifolds with vanishing canonical class,” *Algebraic Geometry and Topology. A Symposium in Honor of S. Lefschetz* (1957) 78–89.
- [3] S.-T. Yau, “Calabi’s conjecture and some new results in algebraic geometry,” *A Proc. Natl. Acad. Sci. USA* (1977) 1798–1799.
- [4] S.-T. Yau, “On Ricci curvature of a compact Kähler manifold and complex Monge-Ampère equation I,” *Comm. Pure and App. Math* (1979) 339–411.
- [5] P. Candelas, G. Horowitz, A. Strominger, and E. Witten, “Vacuum configurations for superstrings,” *Nucl. Phys. B* (1985) 46–74.
- [6] T. Hübsch, *Calabi-Yau Manifolds: A Bestiary For Physicists*. Wspc, Singapore, Mar., 1992.

---

<sup>27</sup>In fact, creatures also seem to prefer immediate rewards.

- [7] Y.-H. He, *The Calabi-Yau Landscape: from Geometry, to Physics, to Machine-Learning*. 2018. [arXiv:1812.02893 \[hep-th\]](#).
- [8] Y.-H. He, “Deep-Learning the Landscape,” [arXiv:1706.02714 \[hep-th\]](#).
- [9] Y.-H. He, “Machine-learning the string landscape,” *Phys. Lett. B* **774**, 564 (2017) .
- [10] D. Joyce, “Lectures on calabi-yau and special lagrangian geometry,” [arXiv:math/0108088](#).
- [11] R. Hartshorne, *Algebraic Geometry*. Graduate Texts in Mathematics. Springer New York, 2013.
- [12] P. Candelas, A. M. Dale, C. A. Lutken, and R. Schimmrigk, “Complete Intersection Calabi-Yau Manifolds,” *Nucl. Phys.* **B298** (1988) 493.
- [13] C. Schöen, “On fiber products of rational elliptic surfaces with section,” *Math. Z.* **197** 177199 (1988) .
- [14] P. A. Griffiths and J. Harris, *Principles of algebraic geometry*. Wiley classics library. Wiley, New York, NY, 1994.
- [15] Universe-Review, *Theory of Superstring, and M Theory*. Available from: <http://universe-review.ca/R15-26-CalabiYau02.htm>.
- [16] J. Hauenstein, Y.-H. He, and D. Mehta, “Numerical elimination and moduli space of vacua,” *JHEP* **09** (2013) 083, [arXiv:1210.6038 \[hep-th\]](#).
- [17] J. Gray, Y.-H. He, V. Jejjala, and B. D. Nelson, “Exploring the vacuum geometry of N=1 gauge theories,” *Nucl. Phys.* **B750** (2006) 1–27, [arXiv:hep-th/0604208 \[hep-th\]](#).
- [18] S. Franco, Y.-H. He, C. Sun, and Y. Xiao, “A Comprehensive Survey of Brane Tilings,” *Int. J. Mod. Phys.* **A32** no. 23n24, (2017) 1750142, [arXiv:1702.03958 \[hep-th\]](#).
- [19] A. Hanany and K. D. Kennaway, “Dimer models and toric diagrams,” [arXiv:hep-th/0503149 \[hep-th\]](#).
- [20] S. Franco, A. Hanany, D. Martelli, J. Sparks, D. Vegh, and B. Wecht, “Gauge theories from toric geometry and brane tilings,” *JHEP* **01** (2006) 128, [arXiv:hep-th/0505211 \[hep-th\]](#).
- [21] M. Yamazaki, “Brane Tilings and Their Applications,” *Fortsch. Phys.* **56** (2008) 555–686, [arXiv:0803.4474 \[hep-th\]](#).

- [22] S. Fomin, L. Williams, and A. Zelevinsky, *Introduction to Cluster Algebras. Chapters 1-3*. 2016. [arXiv:1608.05735](#).
- [23] E. Witten, *On the structure of the topological phase of two-dimensional gravity*. Nuclear Physics B. 340 (23): 281332, 1990.
- [24] E. Zaslow, *Mirror Symmetry*. In Gowers, Timothy (ed.). The Princeton Companion to Mathematics. ISBN 978-0-691-11880-2, 2008.
- [25] B. Feng, Y.-H. He, K. D. Kennaway, and C. Vafa, “Dimer models from mirror symmetry and quivering amoebae,” *Adv. Theor. Math. Phys.* **12** no. 3, (2008) 489–545, [arXiv:hep-th/0511287](#) [[hep-th](#)].
- [26] V. Jejjala, S. Ramgoolam, and D. Rodriguez-Gomez, “Toric CFTs, Permutation Triples and Belyi Pairs,” *JHEP* **03** (2011) 065, [arXiv:1012.2351](#) [[hep-th](#)].
- [27] Y.-H. He, “Calabi-Yau Varieties: from Quiver Representations to Dessins d’Enfants,” [arXiv:1611.09398](#) [[math.AG](#)].
- [28] W. Taylor and Y.-N. Wang, “The F-theory geometry with most flux vacua,” *JHEP* **12** (2015) 164, [arXiv:1511.03209](#) [[hep-th](#)].
- [29] P. Candelas, X. de la Ossa, and S. H. Katz, “Mirror symmetry for Calabi-Yau hypersurfaces in weighted  $P^3$  and extensions of Landau-Ginzburg theory,” *Nucl. Phys.* **B450** (1995) 267–292, [arXiv:hep-th/9412117](#) [[hep-th](#)].
- [30] B. Matthews, “Comparison of the predicted and observed secondary structure of t4 phage lysozyme,” *Biochimica et Biophysica Acta (BBA) - Protein Structure* **405** no. 2, (1975) 442 – 451.
- [31] J. Gorodkin, “Comparing two k-category assignments by a k-category correlation coefficient,” *Computational Biology and Chemistry* **28** no. 5, (2004) 367 – 374.
- [32] D. Chicco, “Ten quick tips for machine learning in computational biology,” *BioData Mining* (2017) .
- [33] K. Bull, Y.-H. He, V. Jejjala, and C. Mishra, “Machine Learning CICY Threefolds,” *Phys. Lett.* **B785** (2018) 65–72, [arXiv:1806.03121](#) [[hep-th](#)].
- [34] Y.-H. He and S.-J. Lee, “Distinguishing Elliptic Fibrations with AI,” [arXiv:1904.08530](#) [[hep-th](#)].
- [35] J. Gray, A. S. Haupt, and A. Lukas, “Topological Invariants and Fibration Structure of Complete Intersection Calabi-Yau Four-Folds,” *JHEP* **09** (2014) 093, [arXiv:1405.2073](#) [[hep-th](#)].

- [36] L. B. Anderson, X. Gao, J. Gray, and S.-J. Lee, “Tools for CICYs in F-theory,” *JHEP* **11** (2016) 004, [arXiv:1608.07554 \[hep-th\]](#).
- [37] L. B. Anderson, X. Gao, J. Gray, and S.-J. Lee, “Multiple Fibrations in Calabi-Yau Geometry and String Dualities,” *JHEP* **10** (2016) 105, [arXiv:1608.07555 \[hep-th\]](#).
- [38] L. B. Anderson, X. Gao, J. Gray, and S.-J. Lee, “Fibrations in CICY Threefolds,” *JHEP* **10** (2017) 077, [arXiv:1708.07907 \[hep-th\]](#).
- [39] A. Cayley, “On the theory of groups, as depending on the symbolic equation  $\theta^n = 1$ ,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **7** (1854) .
- [40] Y.-H. He and M. Kim, “Learning Algebraic Structures: Preliminary Investigations,” [arXiv:1905.02263 \[cs.LG\]](#).
- [41] L. Alessandretti, A. Baronchelli, and Y.-H. He, “Machine Learning meets Number Theory: The Data Science of Birch-Swinnerton-Dyer,” [arXiv:1911.02008 \[math.NT\]](#).
- [42] V. Bouchard, “Lectures on complex geometry, Calabi-Yau manifolds and toric geometry,” [arXiv:hep-th/0702063 \[HEP-TH\]](#).
- [43] M. Nakahara, *Geometry, topology and physics*. Taylor & Francis, 2003.
- [44] H. Skarke, “String dualities and toric geometry: An Introduction,” *Chaos Solitons Fractals* **10** (1999) 543, [arXiv:hep-th/9806059 \[hep-th\]](#).
- [45] “The MNIST database of handwritten digits.” <http://yann.lecun.com/exdb/mnist/>.
- [46] S. C. Swaroop, “A Byte of Python.” [https://www.ibiblio.org/swaroopch/byteofpython/files/120/byteofpython\\_120.pdf](https://www.ibiblio.org/swaroopch/byteofpython/files/120/byteofpython_120.pdf).
- [47] The Sage Developers, *SageMath, the Sage Mathematics Software System (Version 9.0)*, 2019. <https://www.sagemath.org>.
- [48] M. Abadi *et al.*, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. <http://tensorflow.org/>. Software available from tensorflow.org.
- [49] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [50] R. G. P. J. Herz, A. Krough, *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.
- [51] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*. MIT Press, 1995.

- 
- [52] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Macmillan, New York, 2nd ed., 1999.
- [53] A. J. Smola and B. Schölkopf, “A Tutorial on Support Vector Regression.” <https://alex.smola.org/papers/2003/SmoSch03b.pdf>.
- [54] P. Harrington, *Machine Learning in Action*. Manning Publications, 2012.
- [55] C. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [56] I. Witten, E. Frank, and M. Hall, *Data Mining*. Morgan Kaufmann, 2011.
- [57] J. Halverson, B. Nelson, and F. Ruehle, “Branes with Brains: Exploring String Vacua with Deep Reinforcement Learning,” *JHEP* **06** (2019) 003, [arXiv:1903.11616](https://arxiv.org/abs/1903.11616) [[hep-th](https://arxiv.org/abs/1903.11616)].