

## Machine learning class numbers of real quadratic fields

Malik Amir

*Microsoft Research New England  
One Memorial Drive, Cambridge, MA 02140, USA  
malik.amir.math@gmail.com*

Yang-Hui He

*London Institute for Mathematical Sciences  
Royal Institution, London W1S 4BS, UK*

*Department of Mathematics, City  
University of London, EC1V 0HB, UK*

*Merton College, University of Oxford  
OX14JD, UK*

*School of Physics, NanKai University  
Tianjin, 300071, P. R. China  
hey@maths.ox.ac.uk*

Kyu-Hwan Lee

*Department of Mathematics  
University of Connecticut, Storrs, CT 06269, USA  
khlee@math.uconn.edu*

Thomas Oliver\*

*University of Westminster, New Cavendish Street  
London, W1W6UW, UK  
T.Oliver@westminster.ac.uk*

Eldar Sultanow

*Capgemini Germany, Bahnhofstraße 30  
90402 Nuremberg, Germany  
eldar.sultanow@capgemini.com*

Received 6 October 2022  
Accepted 18 October 2023  
Published 9 December 2023

We implement and interpret various supervised learning experiments involving real quadratic fields with class numbers 1, 2 and 3. We quantify the relative difficulties in

\*Corresponding author.

separating class numbers of matching/different parity from a data-scientific perspective, apply the methodology of feature analysis and principal component analysis, and use symbolic classification to develop machine-learned formulas for class numbers 1, 2 and 3 that apply to our dataset.

*Keywords:* Real quadratic fields; class numbers; supervised machine learning.

Mathematics Subject Classification 2020: 11R29, 11R80, 62R99

## 1. Introduction

The class number of a real quadratic field, or of a more general number field, measures how far its ring of integers is from being a unique factorization domain (UFD). In particular, if the class number is 1, then the ring of integers is a UFD. The *Gauss class number problem* for real quadratic fields concerns whether or not there are infinitely many real quadratic fields with class number 1. This fundamental question remains one of the central open questions in number theory to this day.

In a recent paper [4], it was observed that certain supervised learning classifiers could be used to distinguish real quadratic fields of class number 1 from those of class number 2. That paper was the second in a series, featuring also [3, 5], in which the unifying theme was the application of machine learning algorithms to arithmetic objects presented by finite lists of coefficients in certain Dirichlet series called  $\zeta$ -functions (or  $L$ -functions). Subsequently, this series of papers has been enhanced by [6], which documents a first glimpse of unexpected aggregate phenomena amongst the coefficients. In all of these papers, the methodology is motivated by the general philosophy that arithmetic objects may be classified through these functions; in particular, various interesting arithmetic invariants appear in their Laurent expansions.

In this paper, we seek to formalize, generalize, and interpret the aforementioned supervised learning experiment for real quadratic fields. In particular, we will train classifiers on data connected to real quadratic fields with class numbers 1, 2 and 3. We will implement different classifiers to those in [4], and once again observe that a machine can learn to distinguish between class numbers 1 and 2 when trained on finitely many  $\zeta$ -coefficients (i.e. Dirichlet coefficients of  $\zeta$ -functions). On the other hand, in order to accurately distinguish between class numbers 1 and 3, we incorporate some, but not necessarily all, additional features inspired by the analytic class number formula for real quadratic fields, that is

$$\lim_{s \rightarrow 1} (s-1)\zeta_d(s) = \frac{2R_d h_d}{\sqrt{D}}, \quad (1.1)$$

where  $\zeta_d(s)$  is the Dedekind zeta function of  $\mathbb{Q}(\sqrt{d})$ ,  $D$  is the discriminant,  $R_d$  is the regulator, and  $h_d$  is the class number. Inspired by experimental observations, we investigate structures in  $\zeta$ -coefficients that may account for the machine's success, or lack thereof. This investigation incorporates viewpoints that are, at times purely arithmetic, at times purely data scientific, and ultimately, a mixture of the two. The basic arithmetic concept is the *genus field*, which explain

some features present in the data and allow one to develop human strategies for distinguishing between class numbers 1 and 2 based on  $\zeta$ -coefficients alone (see Sec. 2.2). In particular, in the constrained settings of our experiments, observing occurrences of the value 1 is often a useful technique for distinguishing class numbers. In Sec. 4.2, we will see that symbolic classification produces a formula which recovers the same insights that we derived from genus theory. Strategies developed from genus theory break down when considering class numbers 1 and 3. In Sec. 3.3, we will introduce a cost function, the magnitude of which corresponds to the relative difficulty in separating two class numbers. Using a search algorithm, we heuristically compute the minimum of the cost function and so quantify a sense in which it is harder to separate class numbers 1 and 3 than it is to separate 1 and 2 (see Sec. 3.5).

In the rest of this introduction, we overview the subsequent sections. We begin by reviewing some mathematical theory, including the genus field. In Sec. 2, we review Dedekind  $\zeta$ -functions and establish various results in genus theory that explain several statistical observations about the class numbers of real quadratic fields. In Sec. 3, we explore the pairwise separation of real quadratic fields with different class numbers using finitely many coefficients of their Dedekind  $\zeta$ -functions. More precisely, we introduce a cost function which quantifies the separability, and may be heuristically computed and optimized using the so-called *bubble algorithm*. The cost function is built from certain counting functions, which enumerate square-free  $d$  such that  $\mathbb{Q}(\sqrt{d})$  has specified ramification properties. Section 3 is complemented by Appendix A.1, in which we explore separation of  $\zeta$ -coefficient data using principal component analysis (PCA). In Sec. 4, we investigate the binary classification of real quadratic fields with class numbers 1 and 2 using gradient boosting tree based learning algorithms (specifically, LightGBM and CatBoost) and genetic programming (specifically, symbolic classification). In particular, we undertake a supervised learning experiment in the style of [4], that is, using finite lists of  $\zeta$ -coefficients as features, but place greater emphasis on alternative methodologies and feature analysis. Furthermore, rediscover some results from genus theory, first stated in Sec. 2, concerning the parity of class numbers. Section 4 is complemented by Appendix A.2, in which we record additional plots and metrics for the experiments.

In Sec. 5, we investigate the binary classification of class numbers 1 and 3, using LightGBM. Having previously observed that  $\zeta$ -coefficients are not sufficient for high accuracy classifiers in this case, we also incorporate various combinations of features consisting connected to Eq. (1.1). In particular, we will involve the ramified primes, and some partial sums related to the Dedekind zeta function and the Dirichlet  $L$ -function. Applying the symbolic classifier, we are lead to two approximate formulas for the class number, the first of which is essentially Eq. (1.1), and the second of which looks somewhat different but nevertheless suggests that  $h_d$  is proportional (respectively, inversely proportional) to  $\sqrt{D}$  (respectively,  $R_d$ ).

## 2. Genus Theory

In [4], we saw that a random forest classifier trained with a dataset of coefficients for the Dedekind zeta function was able to distinguish between real quadratic fields of class numbers 1 and 2 to high accuracy. In order to explain mathematically what a machine learns from the dataset, we review the genus field of a real quadratic field and establish various constraints on its class number in terms of the number of ramified primes. In Sec. 3, we will quantify one way in which the analogous learning task for real quadratic fields with class numbers 1 and 3 is more challenging. In Sec. 4, we will use the results in this section to inform the feature analysis undertaken.

### 2.1. Dedekind zeta functions

Let  $K$  be a number field with ring of integers  $\mathcal{O}_K$ . The Dedekind zeta function of  $K$  is defined to be

$$\zeta_K(s) = \sum_{I \leq \mathcal{O}_K} N(I)^{-s} = \prod_{\mathfrak{p} \leq \mathcal{O}_K} (1 - N(\mathfrak{p})^{-s})^{-1}, \quad \text{Re}(s) > 1, \quad (2.1)$$

where  $N$  denotes the norm map and the sum (respectively, product) is over the non-zero (respectively, prime) ideals of  $\mathcal{O}_K$ .

For a square-free integer  $d$ , let  $K_d = \mathbb{Q}(\sqrt{d})$ . We say that  $K_d$  is real (respectively, imaginary) if  $d > 0$  (respectively,  $d < 0$ ). The discriminant  $D$  of  $K_d$  is given by

$$D = \begin{cases} d & \text{if } d \equiv 1 \pmod{4}, \\ 4d & \text{if } d \equiv 2, 3 \pmod{4}. \end{cases} \quad (2.2)$$

When  $K = K_d$ , we will write  $\zeta_K(s) = \zeta_d(s)$ . We note that

$$\zeta_d(s) = \zeta(s)L(s, \chi_D), \quad (2.3)$$

where  $\chi_D := \left(\frac{D}{\cdot}\right)$  is the Kronecker symbol attached to  $K_d$ ,  $L(s, \chi_D)$  is the associated Dirichlet  $L$ -function, and  $\zeta(s)$  is the Riemann zeta function. By  $h_d$ , we denote the class number of  $K_d$ .

We may write

$$\zeta_d(s) = \sum_{n=1}^{\infty} a_n n^{-s},$$

where  $a_n$  denotes the  $n$ th Dirichlet coefficient of  $\zeta_d(s)$ . We will refer to the sequence  $(a_n)_{n=1}^{\infty}$  as the  $\zeta$ -coefficients of  $K_d$ . Using Eq. (2.3), we deduce that

$$a_n = \sum_{m|n} \chi_D(m), \quad (2.4)$$

where the sum is over  $m$  dividing  $n$ . In particular, if  $n = p$  is prime, then Eq. (2.4) simplifies to

$$a_p = 1 + \chi_D(p). \quad (2.5)$$

Since  $\chi_D$  is a real quadratic character, we have  $\chi_D(p) \in \{-1, 0, 1\}$ . Thus, for  $p$  prime, Eq. (2.5) implies that  $a_p \in \{0, 1, 2\}$ . By construction, we have  $\chi_D(p) = 0$  if and only if  $p$  divides  $D$ . Subsequently, it follows from Eq. (2.5) that

$$a_p = 1 \Leftrightarrow p \text{ divides } D \Leftrightarrow p \text{ is ramified in } K_d. \quad (2.6)$$

## 2.2. Genus fields

By definition, the *genus field*  $E_d$  of  $K_d$  is the maximal unramified extension of  $K_d$  which is abelian over  $\mathbb{Q}$ , and the *extended genus field*  $E_d^+$  of  $K_d$  is the maximal extension of  $K_d$  which is unramified at all finite primes and abelian over  $\mathbb{Q}$ . Recall that a prime discriminant is a discriminant divisible by a single prime. Write  $D = d_1 \cdots d_t$  as a product of prime discriminants  $d_i$  with  $t = \omega(D)$ , the number of distinct primes dividing  $D$ .

We will use the following results on genus fields (see, e.g. [7, 10]):

**Proposition 2.1.** *We have*

$$E_d^+ = \mathbb{Q}(\sqrt{d_1}, \dots, \sqrt{d_t}), \quad E_d = E_d^+ \cap \mathbb{R}, \quad (2.7)$$

$$\text{Gal}(E_d^+/K_d) \cong C_d^+/2C_d^+ \cong (\mathbb{Z}/2\mathbb{Z})^{t-1}, \quad \text{Gal}(E_d/K_d) \cong C_d/2C_d \cong (\mathbb{Z}/2\mathbb{Z})^s, \quad (2.8)$$

where  $C_d^+$  denotes the narrow class group of  $K_d$  and  $C_d$  the class group of  $K_d$  and  $s = t - 1$  if  $d$  is a sum of two squares and  $s = t - 2$  otherwise.

Let  $n_d$  denote the number of rational primes ramified in  $K_d$ . Since the ramified primes in  $K_d$  are precisely those which divide  $D$ , we have  $n_d = t = \omega(D)$ . The above proposition implies a general result given below on the parity of class numbers  $h_d$ .

**Corollary 2.2.** *The class number  $h_d$  is odd if and only if either  $n_d = 1$  or  $D = d_1 d_2$  with prime discriminants  $d_1, d_2 < 0$ .*

In Sec. 4.1, we will investigate datasets in which  $h_d \leq 2$ . In such sets, we deduce from Corollary 2.2 that, if  $n_d = 1$  then  $h_d = 1$ . Such implications do not hold more generally, for example, if  $d = 229$ , then  $n_d = 1$  and  $h_d = 3$ .

**Proof.** Assume that  $h_d$  is odd. Then  $s = 0$  in (2.8) and so  $t = 1$  or  $t = 2$ . If  $t = n_d = 1$  then we are done. If  $t = 2$  then  $d$  cannot be a sum of two squares. Then  $d = p_1$ ,  $d = 2p_1$  or  $d = p_1 p_2$  with primes  $p_1, p_2 \equiv 3 \pmod{4}$ . In all three cases,  $D$  is a product of two negative prime discriminants. Conversely, if  $n_d = t = 1$  then  $s = 0$  and  $h_d$  is odd from (2.8), and if  $D = d_1 d_2$  with  $d_1, d_2 < 0$ , then  $t = 2$ ,  $s = t - 2 = 0$  and  $h_d$  is odd.  $\square$

**Lemma 2.3.** *If  $h_d = 1$  and  $a_p = 1$  for some prime  $p \equiv 1 \pmod{4}$ , then  $d = p$ .*

We cannot allow primes  $p \equiv 3 \pmod{4}$  in Lemma 2.3. For example,  $h_{14} = 1$  and  $a_7 = 1$  for  $K_{14}$ .

**Proof.** Since  $h_d = 1$ , we have  $E_d = E_d^+ \cap \mathbb{R} = K_d$ . Since  $a_p = 1$ , we know that  $p$  divides  $D$  and  $p$  is a prime discriminant. It follows from (2.7) that  $\sqrt{p} \in E_d^+ \cap \mathbb{R} = K_d$ . Then we have  $\mathbb{Q}(\sqrt{p}) = K_d$ .  $\square$

Now we establish a series of lemmas which will be utilized primarily in Secs. 3.1 and 4.2. In the former, we will rigorously evaluate certain counting functions attached to sets of real quadratic fields. In the latter, we will compare the consequences of our lemmas to the output of symbolic classification.

**Lemma 2.4.** *If  $h_p = 1$  for a prime  $p \equiv 1 \pmod{4}$  then  $h_{mp} \geq 2$  for square-free integers  $m \in \mathbb{Z}_{>1}$  not divisible by  $p$ .*

We cannot allow primes  $p \equiv 3 \pmod{4}$  in Lemma 2.4. For example,  $h_7 = h_{14} = 1$ .

**Proof.** Given square-free  $m \in \mathbb{Z}_{>1}$  and  $p \equiv 1 \pmod{4}$ , the field  $K_{mp}$  satisfies  $a_p = 1$ . Since  $mp \neq p$ , Lemma 2.3 already implies that  $h_{mp} \neq 1$ .  $\square$

**Remark 2.5.** For  $m \geq 1$ , the field  $K_{mp}$  satisfies  $a_p = 1$  by (2.6). If  $p \equiv 1 \pmod{4}$  and  $h_p = 1$ , then Lemma 2.4 implies the existence of infinitely many real quadratic fields with class number  $\geq 2$  such that  $a_p = 1$ . This will be relevant in Example 3.2.

**Lemma 2.6.** *If  $n_d \geq 3$ , then  $h_d \geq 2$ . In particular, if  $n_d \geq 3$  and  $h_d \leq 2$ , then  $h_d = 2$ . Similarly, if  $n_d \geq 4$ , then  $h_d \geq 4$ .*

Note that it is possible that  $K_d$  satisfies  $n_d = 2$  and  $h_d = 1$ . This occurs, for example, if  $d = 7$ , in which case  $h_d = 1$  and  $K_d$  is ramified at the primes 2 and 7.

**Proof.** The assertions follow from (2.8).  $\square$

**Lemma 2.7.** *Assume that  $n_d = 2$  and  $h_d \in \{1, 2\}$ . Let  $p_1$  be the smallest ramified prime in  $K_d$ . If  $p_1 \equiv 1 \pmod{4}$  (respectively,  $p_1 \equiv 3 \pmod{4}$ ), then  $h_d = 2$  (respectively,  $h_d = 1$ ).*

**Proof.** Write  $D = d_1 d_2$  with  $d_i$  prime discriminants. If  $p_1 = d_1 \equiv 1 \pmod{4}$  then  $d_2$  is also a prime  $\equiv 1 \pmod{4}$ . Further, if  $h_d = 1$ , then Lemma 2.3 yields a contradiction. Thus,  $h_d = 2$ .

If  $p_1 = -d_1 \equiv 3 \pmod{4}$  then  $d_2 = -p_2$  for a prime  $p_2 \equiv 3 \pmod{4}$ . We obtain from (2.7) that  $E_d^+ = \mathbb{Q}(\sqrt{-p_1}, \sqrt{-p_2})$  and  $E_d = \mathbb{Q}(\sqrt{p_1 p_2}) = K_d$ . By (2.8), we have  $h_d = 1$ .  $\square$

### 3. Separability of $\zeta$ -Coefficients

In this section, we explore the relative difficulty in distinguishing between different pairs of class numbers using only the associated  $\zeta$ -coefficients. This is in keeping with the strategy implemented in [4]. In Secs. 4 and 5, we will go beyond [4] and

incorporate other features inspired by the analytic class number formula into our dataset.

The separability of  $\zeta$ -coefficients will be quantified in terms of a cost function, which will be introduced in Sec. 3.3. The cost function is constructed in terms of certain counting functions explored in Sec. 3.1. The counting functions and the cost function may be heuristically computed and optimized using the so-called *bubble algorithm* introduced in Sec. 3.4. In Sec. 3.5, we will use the bubble algorithm to establish that the problem of distinguishing between class numbers 1 and 3 is more challenging than between class numbers 1 and 2. In Sec. 3.2, we will explore the role played by  $\zeta$ -coefficients with non-prime indices in the separation of  $\zeta$ -coefficients. In Appendix A.1, we document alternative approaches to the separation of  $\zeta$ -coefficients using PCA.

### 3.1. Counting functions

In this section, we introduce various functions connected to counting real quadratic fields with prescribed class number and splitting behavior at given primes. Roughly speaking, the idea is that when it is known that the class number can take one of two values, if the count is zero for one class number, and non-zero for the other, then observing the splitting behavior determines the class number. We will be interested in the effectiveness of implementing this strategy in Sec. 3.3.

**Definition 3.1.** For a positive real number  $X$ , a positive integer  $h \in \mathbb{Z}_{>0}$ , a vector of non-negative integers  $v \in \mathbb{Z}_{\geq 0}^3$ , and positive integers  $\ell, m, n \in \mathbb{Z}_{>0}$ , let  $f_h^v(\ell, m, n)$  be the number of square-free positive integers  $d < X$  such that  $h_d = h$  and  $(a_\ell, a_m, a_n) = v$ , where  $a_\ell$  (respectively,  $a_m, a_n$ ) is the  $\ell$ th (respectively,  $m$ th,  $n$ th) coefficient of  $\zeta_d(s)$ .

In symbols, we have

$$f_h^v : \mathbb{Z}_{>0}^3 \rightarrow \mathbb{Z}_{\geq 0}, \quad f_h^v(\ell, m, n) = \#\{d < X : h_d = h \text{ and } (a_\ell, a_m, a_n) = v\}. \quad (3.1)$$

The function  $f_h^v$  depends on  $X$ , though this fact is suppressed from the notation.

For rational primes  $p, q, r$ , if  $f_h^v(p, q, r) > 0$  then there exists a real quadratic field with class number  $h$  and certain splitting behavior at  $p, q, r$  prescribed by the vector  $v$ . Indeed, we have already mentioned in Sec. 2.1 that we have  $a_p, a_q, a_r \in \{0, 1, 2\}$  and a prime  $p$  ramifies in  $K_d$  if and only if the corresponding coefficient  $a_p$  of  $\zeta_d(s)$  is 1. Furthermore, we know that a prime is split (respectively, inert) if the corresponding coefficient is 2 (respectively, 0) (cf. [9, Eq. (4.7)]).

We will refer to the inputs of  $f_h^v$  as triples, and use the notation  $[\ell, m, n]$  so as to distinguish the inputs from the indexing vectors  $v$ .

**Example 3.2.** Let  $h = 1$  and  $[\ell, m, n] = [3, 5, 7]$ . Then, for  $v = (0, 1, 0)$  and  $X > 5$ , we have  $f_1^v(3, 5, 7) = 1$ . For all the other  $v = (a, 1, c)$  for  $a, c \in \{0, 1, 2\}$  and

$(a, c) \neq (0, 0)$ , we have  $f_1^v(3, 5, 7) = 0$ . Indeed, if  $a_5 = 1$ , then  $K_d$  is ramified at 5, and so  $d = 5m$  for square-free  $m$  not divisible by 5. Since  $h_5 = 1$  and  $5 \equiv 1 \pmod 4$ , Lemma 2.4 implies that  $h_{5m} \geq 2$  for  $m > 1$ . When  $d = 5$ , we have  $a_3 = a_7 = 0$ , proving the claim.

**Definition 3.3.** For  $h \in \mathbb{Z}_{>0}$  and a triple  $[\ell, m, n]$ , let  $g_h(\ell, m, n)$  be the number of  $v = (v_1, v_2, v_3) \in \mathbb{Z}_{\geq 0}^3$  such that  $f_h^v(\ell, m, n) > 0$ , where  $f_h^v$  is as in Eq. (3.1). In symbols, we have

$$g_h : \mathbb{Z}_{>0}^3 \rightarrow \mathbb{Z}_{\geq 0}, \quad g_h(\ell, m, n) = \#\{v = (v_1, v_2, v_3) \in \mathbb{Z}_{\geq 0}^3 : f_h^v(\ell, m, n) > 0\}. \tag{3.2}$$

In order to gain some familiarity with these functions, we consider first the case that  $\ell, m, n$  are all primes (composites will appear in the sequel). Then, since  $a_p, a_q, a_r \in \{0, 1, 2\}$  for primes  $p, q, r$ , a trivial bound is given by the number of all possible vectors  $v$ , i.e.

$$g_h(p, q, r) \leq 3^3 = 27, \quad (h \in \mathbb{Z}_{>0}). \tag{3.3}$$

**Example 3.4.** For  $h \in \{1, 2, 3\}$  and  $[\ell, m, n] = [3, 5, 7]$ , we will calculate  $g_h(3, 5, 7)$ . In Table 1, we list the smallest  $d$  such that  $K_d$  has class number given by row and  $(a_3, a_5, a_7)$  given by column. An entry  $\times$  indicates that the vector does not occur, which may be verified using the statements in Sec. 2.2. Indeed, if  $a_5 = 1$  then it follows from Corollary 2.2 that  $h_d$  is odd only when  $d = 5$ . This accounts for 17 occurrences of  $\times$  in Table 1. In all the remaining three occurrences, we have  $a_3 = a_7 = 1$ . Again by Corollary 2.2, we have  $h_d$  is odd only when  $d = 21$ , which explains exactly the three occurrences of  $\times$ .

Table 1. Smallest  $d$  such that  $K_d$  has class number given by row and  $(a_3, a_5, a_7)$  given by column. An entry  $\times$  indicates that the vector does not occur.

$h_d$	(0, 0, 0)	(0, 0, 1)	(0, 0, 2)	(0, 1, 0)	(0, 1, 1)	(0, 1, 2)	(0, 2, 0)	(0, 2, 1)	(0, 2, 2)
1	17	77	19	5	$\times$	$\times$	41	14	11
2	122	182	218	185	35	65	26	119	74
3	257	2177	473	$\times$	$\times$	$\times$	761	2429	254
$h_d$	(1, 0, 0)	(1, 0, 1)	(1, 0, 2)	(1, 1, 0)	(1, 1, 1)	(1, 1, 2)	(1, 2, 0)	(1, 2, 1)	(1, 2, 2)
1	3	$\times$	57	$\times$	$\times$	$\times$	6	21	141
2	87	42	78	285	105	15	66	609	39
3	993	$\times$	1257	$\times$	$\times$	$\times$	321	$\times$	1101
$h_d$	(2, 0, 0)	(2, 0, 1)	(2, 0, 2)	(2, 1, 0)	(2, 1, 1)	(2, 1, 2)	(2, 2, 0)	(2, 2, 1)	(2, 2, 2)
1	13	7	22	$\times$	$\times$	$\times$	19	301	46
2	178	238	58	10	70	85	34	91	106
3	733	7273	142	$\times$	$\times$	$\times$	229	469	316



Using Table 1, we deduce

$$g_h(3, 5, 7) = \begin{cases} 18, & h = 1, X > 301, \\ 27, & h = 2, X > 609, \\ 16, & h = 3, X > 7273. \end{cases} \quad (3.4)$$

In particular, we see that there exists a real quadratic field of class number 2 for every possible combination of ramification at 3, 5, 7.

**Definition 3.5.** For distinct positive integers  $i, j \in \mathbb{Z}_{>0}$ , let  $g_{i,j}$  count the number of  $v$  so that both  $f_i^v$  and  $f_j^v$  are positive. That is, we define

$$g_{i,j}(\ell, m, n) = \#\{v \in \mathbb{Z}_{\geq 0}^3 : f_i^v(\ell, m, n) > 0 \text{ and } f_j^v(\ell, m, n) > 0\}. \quad (3.5)$$

Clearly, we have  $g_{i,j} \leq \min\{g_i, g_j\}$ .

**Example 3.6.** Using Table 1, we obtain

$$g_{1,j}(3, 5, 7) = \begin{cases} 18, & j = 2, X > 609, \\ 27, & j = 3, X > 7273. \end{cases} \quad (3.6)$$

### 3.2. Composite indices

In our initial investigation of the counting functions introduced in Sec. 3.1, we considered only prime indices. With composite indices, the coefficients  $a_n$  can take many more values than their prime counterparts. Indeed, each character value appearing in Eq. (2.4) is in  $\{-1, 0, 1\}$ , and so we see that  $a_n$  is an integer bounded by  $1 \pm \Omega(n)$ , where  $\Omega(n)$  denotes the number of prime factors dividing  $n$  (counted with multiplicity). On the other hand, the coefficient  $a_n$  counts the number of ideals of norm  $n$ , which is a non-negative integer. Combining these observations, we deduce

$$a_n \in \{0, 1, \dots, \Omega(n) - 1, \Omega(n), \Omega(n) + 1\}. \quad (3.7)$$

The number of values actually achieved by  $a_n$  depends on the multiplicity in the prime factorization of  $n$ .

**Example 3.7.** If  $n = p^2$  for some prime  $p$ , then, Eq. (2.3) implies that

$$a_{p^2} = 1 + \chi_D(p) + \chi_D(p)^2 \in \{1, 3\}. \quad (3.8)$$

In Eq. (3.8), we have  $a_{p^2} = 3$  (respectively,  $a_{p^2} = 1$ ) if and only if  $p$  is ramified (respectively, unramified). On the other hand, if  $n = pq$  is a product of two distinct primes, then Eq. (2.3) implies that

$$a_{pq} = 1 + \chi_D(p) + \chi_D(q) + \chi_D(p)\chi_D(q) \in \{0, 1, 2, 4\}. \quad (3.9)$$

In Eq. (3.9), we have  $a_{pq} = 0$  if and only if  $p$  or  $q$  is inert,  $a_{pq} = 1$  if and only if  $p$  and  $q$  are ramified,  $a_{pq} = 2$  if and only if  $K_d$  is ramified at one of  $p$  and  $q$  and splits at the other, and  $a_{pq} = 4$  if and only if  $p$  and  $q$  split.

Using Eq. (3.7), we deduce the following analogue of Eq. (3.3) for non-prime indices:

$$g_n(\ell, m, n) \leq (\Omega(\ell) + 2)(\Omega(m) + 2)(\Omega(n) + 2). \tag{3.10}$$

### 3.3. Cost function

In order to measure separability of class numbers 1, 2, 3 in our datasets of  $\zeta$ -coefficients, we will introduce a cost function which may be heuristically calculated using the searching algorithm described in Sec. 3.4.

The cost function is constructed so as to account for two key considerations. On one hand, the cost function favors triples which minimize  $g_{i,j}$ ; that is, we are interested in  $\operatorname{argmin}(g_{i,j}) \subset \mathbb{Z}_{>0}^3$ . This is natural, since  $g_{i,j}$  is a coarse measure of the extent to which the sets  $\{K_d : h_d = i\}$  and  $\{K_d : h_d = j\}$  may be separated by the associated  $\zeta$ -coefficients. Indeed, if  $g_{i,j}$  were to hypothetically take the value 0 at some triple  $[\ell, m, n]$ , then each  $v \in \mathbb{Z}_{\geq 0}^3$  would correspond to at most one class number.

On the other hand, the minimization of  $g_{i,j}$  needs to be taken in a relative way. Namely, if the union  $\{v : f_i^v(\ell, m, n) > 0\} \cup \{v : f_j^v(\ell, m, n) > 0\}$  is small, the size  $g_{i,j}$  of intersection would also tend to be small. Consequently, we define our cost function  $C_{i,j}$  to be the ratio of the intersection over the symmetric difference:

$$C_{i,j}(\ell, m, n) := \frac{g_{i,j}(\ell, m, n)}{g_i(\ell, m, n) + g_j(\ell, m, n) - 2g_{i,j}(\ell, m, n)} \tag{3.11}$$

for any pair of class numbers  $\{i, j\}$  and any triple  $[\ell, m, n] \in \mathbb{Z}_{>0}^3$ . Then we are searching for

$$\operatorname{argmin}(C_{i,j}). \tag{3.12}$$

We will find heuristic solutions to Eq. (3.12) using a searching algorithm described in the next section.

**Remark 3.8.** In our heuristic calculations of the counting functions and cost function, we will count only  $d$  whose discriminant  $D$  appears in the LMFDB. Though the LMFDB is complete for  $D < 2 \times 10^6$ , it includes some larger  $d$  and for the purposes of this section, we note that the largest  $d$  such that  $h_d = 1$  (respectively,  $h_d = 2$ , respectively,  $h_d = 3$ ) is 34, 554, 953 (respectively, 43, 723, 857, respectively, 35, 598, 713). The number of real quadratic fields in our dataset for each class number is given in Table 2.

Table 2. Number of real quadratic fields in our dataset of real quadratic fields extracted from the LMFDB.

$\#\{h_d = 1\}$	$\#\{h_d = 2\}$	$\#\{h_d = 3\}$
177159	183436	25201

**Example 3.9.** From the dataset available in the LMFDB, we find

$$g_{1,2}(3, 5, 7) = g_{1,2}(665, 740, 940) = g_{1,2}(520, 783, 991) = 18. \quad (3.13)$$

The value 18 is much lower than the tentative upper bound which comes from Eq. (3.10). Indeed, we have  $\Omega(3) = \Omega(5) = \Omega(7) = 1$ ,  $\Omega(665) = \Omega(740) = \Omega(940) = 3$ ,  $\Omega(520) = 5$ ,  $\Omega(783) = 4$ ,  $\Omega(991) = 1$ , and so Eq. (3.10) implies that

$$g_h(3, 5, 7) \leq 27, \quad g_h(665, 740, 940) \leq 125, \quad g_h(520, 783, 991) \leq 126 \quad (3.14)$$

for any positive integer  $h$ . The discrepancy between the value given in Eq. (3.13) and the hypothetical upper bound in Eq. (3.14) means that there could potentially be many triples  $[\ell, m, n] \in \mathbb{Z}_{\geq 0}^3$  that make  $g_i(\ell, m, n) - g_{i,j}(\ell, m, n)$  and  $g_j(\ell, m, n) - g_{i,j}(\ell, m, n)$  large and  $C_{i,j}(\ell, m, n)$  small.

### 3.4. The bubble algorithm for $h_d \in \{1, 2\}$

In order to find a solution to Eq. (3.12), we utilize a searching algorithm referred to as the *bubble algorithm*. More generally, the bubble algorithm may be used to evaluate all counting functions introduced so far. In this section, we will focus on  $\{i, j\} = \{1, 2\}$ . The generalization to other pairs of class numbers is straightforward.

The terminology “bubble” is motivated by certain visualizations (bubble charts) of the value distributions of  $\zeta$ -coefficients, such as Fig. 1. In Fig. 1 left (respectively, right), we see a cube with axes given by coefficient triples  $v = (a_3, a_5, a_7)$  (respectively,  $v = (a_2, a_3, a_5)$ ). At each integer vector  $v$ , we see a colored bubble whose size is determined by  $f_1^v(\ell, m, n) + f_2^v(\ell, m, n)$ . The presence of the color red (respectively, green) at  $v$  indicates that  $f_1^v(\ell, m, n) > 0$  (respectively,  $f_2^v(\ell, m, n) > 0$ ), and the size of the red (respectively, green) contribution is proportional to the value

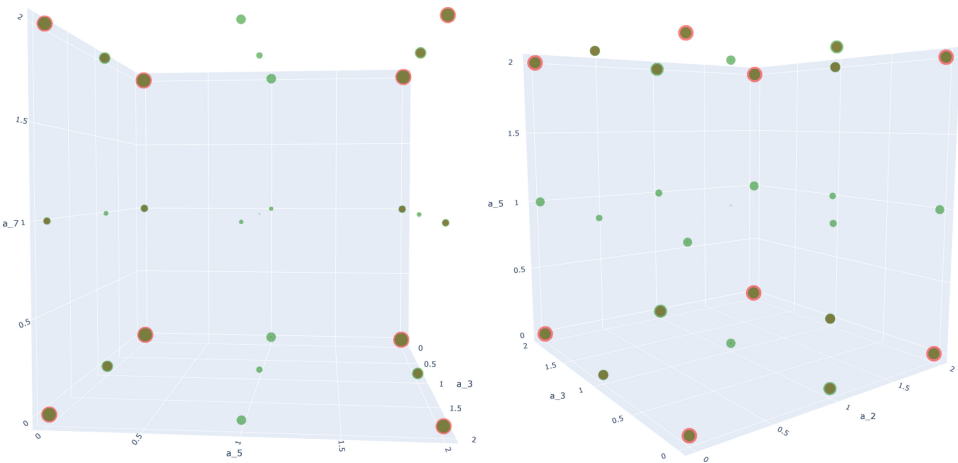


Fig. 1. (Color online) Value distribution for the triples  $(a_3, a_5, a_7)$  (left) and  $(a_2, a_3, a_5)$  (right) where red (respectively, green) bubbles correspond to class number 1 (respectively, 2) real quadratic fields.

of  $f_1^v(\ell, m, n)$  (respectively,  $f_2^v(\ell, m, n)$ ). Summarizing Sec. 3.3 in visual language, we are interested in minimizing the number of mixed bubbles, and maximizing the number of pure bubbles.

In order to compute the counting functions of Sec. 3.1, and hence the cost function of Sec. 3.3, we define a large matrix whose rows are indexed by real quadratic fields  $K_d$ . The first column contains  $d$ , the second column contains its class number  $h_d$ , and the remaining columns contain the values taken by the coefficients  $a_i$  of  $\zeta_d(s)$ . If the matrix contains  $c$  rows in which  $(a_\ell, a_m, a_n) = (x, y, z)$ , then we say that the triple  $[\ell, m, n]$  provides  $c$  collisions for the vector  $(x, y, z)$ .

**Example 3.10.** A very small section of this matrix is shown in Table 3. Looking at Table 3, we see that the triple  $[1, 2, 4]$  provides three collisions for the vector  $(1, 0, 1)$  and two collisions for the vector  $(1, 2, 3)$ . On the other hand, the triple  $[2, 4, 7]$ , yields no collisions at any vector.

For each triple  $[\ell, m, n]$ , we search the entire LMFDB to generate an output of the form

$$(g_1(\ell, m, n), g_{1,2}(\ell, m, n), g_2(\ell, m, n), C_{1,2}(\ell, m, n)).$$

**Example 3.11.** Consider the triple  $[\ell, m, n] = [665, 740, 985]$ . In this case, the output of the bubble algorithm reads as  $(18, 18, 57, 0.461538)$ . In other words,  $(a_{665}, a_{740}, a_{985})$  takes 18 different values for class number one fields, 18 common values for both class numbers, 57 different values for class number two fields, and has a cost of 0.461538. In particular, there is no value for  $(a_{665}, a_{740}, a_{985})$  taken by a class number 1 field which is not also taken by a class number 2 field.

Note that there may exist many pure bubbles of one color and very few or zero pure bubbles of the other. More precisely, the number of pure red (class number 1) bubbles is given by  $g_1(\ell, m, n) - g_{1,2}(\ell, m, n)$  and that of pure green (class number 2) ones by  $g_2(\ell, m, n) - g_{1,2}(\ell, m, n)$ . Thus, in Example 3.11, there are no pure red bubbles and  $57 - 18 = 39$  pure green bubbles for  $[\ell, m, n] = [665, 740, 985]$ .

Actually, in the entire dataset available at LMFDB, there exist triples  $[\ell, m, n]$  yielding 109 pure green bubbles with no pure red bubbles. In Table 4, we list the maximal number of pure green bubbles conditional on the constraint that there is a fixed small number of pure red bubbles.

Table 3. Sample of the matrix used to minimize  $C_{1,2}$ .

$d$	Class number	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_7$
5	1	1	0	0	1	1	0
33	1	1	2	1	3	0	0
61	1	1	0	2	1	2	0
10	2	1	1	2	1	1	0
15	2	1	1	1	1	1	2
65	2	1	2	0	3	1	2
1309	2	1	0	2	1	2	1

Table 4. Maximum number of pure green bubbles (class number 2), conditional on a prescribed small number of pure red bubbles (class number 1).

# of pure red bubbles	0	1	2	3	4
max # of pure green bubbles	109	80	60	48	33

Table 5. Triples  $[\ell, m, n]$  whose value distribution yields 1 pure red bubble and 80 pure green bubbles. In all cases, note that  $g_1(\ell, m, n) - g_{1,2}(\ell, m, n) = 1$  and  $g_2(\ell, m, n) - g_{1,2}(\ell, m, n) = 80$ .

$\ell$	$m$	$n$	$g_1$	$g_{1,2}$	$g_2$	$C_{1,2}$
589	637	720	48	47	127	0.580247
637	720	989	48	47	127	0.580247
585	620	931	64	63	143	0.777778
372	931	975	83	82	162	1.012346
804	931	975	83	82	162	1.012346
775	819	987	84	83	163	1.024691

Given 3 pure red bubbles, there is a unique triple achieving the maximum of 48 pure green bubbles. Specifically, the triple is  $[691, 693, 850]$ . In general, a triple achieving the maximum need not be unique. For example, given 1 pure red bubble, there are 6 triples achieving the maximum of 80 pure green bubbles, as listed in Table 5.

### 3.5. Challenges in the case $h_d \in \{1, 3\}$

The bubble algorithm naturally generalizes to other binary classification problems for class numbers in which the features are given by  $\zeta$ -coefficients. When applied to the dataset of  $h_d \in \{1, 3\}$ , the best performing triples  $[\ell, m, n]$  computed by the bubble algorithm are summarized in Table 6. Note that the minimal value of achieved by the cost function (which is  $5/3$ ) is much larger than the value previously seen for  $h_d \in \{1, 2\}$  in Example 3.11. This quantifies one way in which the classification of  $h_d \in \{1, 3\}$  is a fundamentally more challenging problem using only  $\zeta$ -coefficients. Furthermore, when  $h_d \in \{1, 2\}$ , we saw that the value distribution for the triple  $[3, 5, 7]$ , which consists of prime numbers, yields some pure green bubbles (cf. Fig. 1). On the other hand, in the case that  $h_d \in \{1, 3\}$ , the bubble algorithm does not yield a single such triple of prime numbers. Instead, the optimal values of the function  $C_{1,3}$  listed in Table 6 are taken by triples of composite numbers. In Sec. 5, we will circumvent these challenges by introducing training sets with additional features.

Table 6. Triples  $[\ell, m, n]$  minimizing the cost function in the case that  $\{i, j\} = \{1, 3\}$ . In all cases, we have  $C_{1,3}(\ell, m, n) = 5/3$ .

$\ell$	$m$	$n$	$g_1(\ell, m, n)$	$g_{1,3}(\ell, m, n)$	$g_3(\ell, m, n)$
62	904	1120	14	10	12
65	1166	1868	18	15	21
65	1300	1604	19	15	20
65	1316	1820	18	15	21
258	1456	1784	19	15	20
258	1580	1784	17	15	22
262	1324	1616	12	10	14
262	1280	1844	14	10	12
269	1436	1844	21	15	18
274	1316	1820	12	10	14
274	1324	1576	12	10	14
289	1436	1576	19	15	20
1364	1568	1913	26	20	26
1374	1444	1664	20	15	19
1468	1802	1984	20	15	19

#### 4. Class Numbers 1 and 2

In this section, we investigate the binary classification of real quadratic fields with class numbers 1 and 2 using gradient-boosting tree-based learning algorithms and genetic programming. In Sec. 4.1, we apply the LightGBM and CatBoost machine learning algorithms to finite lists of  $\zeta$ -coefficients. We report all metric scores in the form of tables and figures, including the list of most important features used for the predictions in each model. In Sec. 4.2, we use a genetic programming algorithm called symbolic classification to the ramified primes to obtain an optimal approximation for the class number formula, and subsequently recover some results about the parity of the class number first presented in Sec. 2. We maintain the notation from Sec. 2.

##### 4.1. Learning $h_d \in \{1, 2\}$ from the prime index coefficients of $\zeta_d(s)$

To each square-free  $d \in \mathbb{Z}_{>0}$ , we attach the vector

$$v(d) = (a_p)_p^{\text{prime}} \in \mathbb{Z}^{168}, \quad (4.1)$$

$p \leq 1000$

where  $a_p$  is the  $p$ th Dirichlet coefficient of  $\zeta_d(s)$ . The dimension 168 is the number of primes  $\leq 1000$ . Using Eq. (2.5), we observe that  $v(d) \in \{0, 1, 2\}^{168}$ . Using the vectors in Eq. (4.1), we introduce the labeled dataset

$$\mathcal{D}_{1,2} = \{v(d) \rightarrow h_d\}, \quad (4.2)$$

where  $d$  varies over square-free positive integers such that  $h_d \in \{1, 2\}$  and  $d$  appears in the LMFDB [11]. As was noted in Remark 3.8, the dataset is complete for  $d$  such

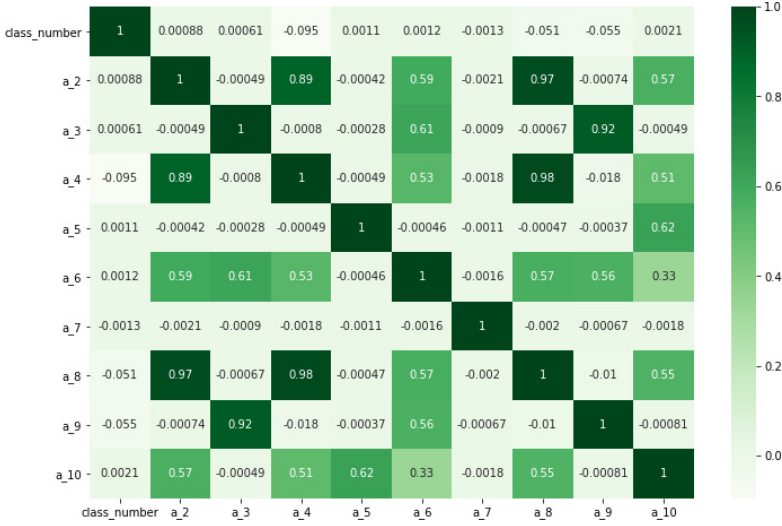


Fig. 2. Correlation matrix of the first ten coefficients of the Dedekind zeta functions  $\zeta_d(s)$  of real quadratic fields with  $h_d \in \{1, 2\}$ .

that  $D < 2 \times 10^6$ , and contains some larger  $d$  (the largest  $d$  such that  $h_d \in \{1, 2\}$  being 43, 723, 857).

The labeled dataset in (4.2) is different to that in [4, Sec. 6.1], which also incorporated the  $\zeta$ -coefficients with composite indices. Our choice to include only prime indices seems intuitively reasonable, since all coefficients can be recovered from those of prime index via (2.1). Whilst it was observed in Sec. 3.2 (in particular, Example 3.9) that  $\zeta$ -coefficients with composite indices yield greater separation between real quadratic fields of different class numbers, prime indices are nevertheless sufficient for the high accuracy classifiers in the context of this section. We furthermore note that the exclusion of composite indices can also be motivated by the correlation matrix in Fig. 2, which shows that composite indices sharing similar prime decomposition are highly correlated. Excluding the composite indices therefore allows for faster training, better generalization, and more reliable (permutation) feature importance.

To build our LightGBM and CatBoost supervised learning models, we will use the automated machine learning library AutoMLjar [12]. Our choice to use LightGBM and CatBoost was motivated in large part by Fig. 3, which shows that LightGBM and CatBoost are, in addition to being fast to train, among the best performing models that can be built using AutoMLjar for our experiment. Furthermore, they often represent state-of-the-art models on tabular data. We will consider a training/testing split of 70/30 for  $\mathcal{D}_{1,2}$ , together with a 10-fold cross validation performed on the training set, since other splits including 30/70 produce similar results. All the folders generated by AutoMLjar, including codes, datasets and figures for all experiments presented in this paper can

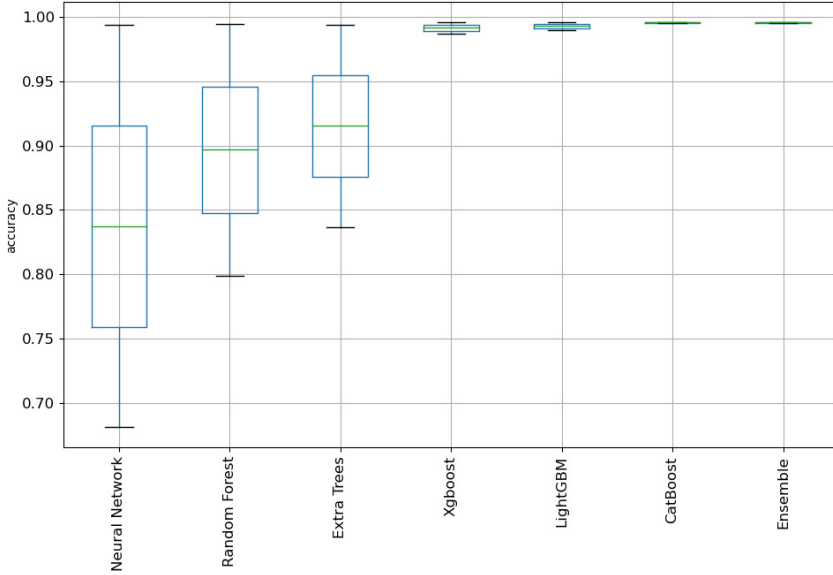


Fig. 3. Comparison of the LightGBM and CatBoost learning models against other AutoMLjar models on a 70/30 split for the binary classification task  $h_d = 1$  versus  $h_d = 2$ .

Table 7. Performance metrics over the training set of the LightGBM and CatBoost model for the binary classification task  $h_d = 1$  versus  $h_d = 2$  with split 70/30. As is standard, AUC is the “area under the receiver operating characteristic curve”, F1 is the F1-score and MCC, the Matthews correlation coefficient. All these quantities need to be close to 1 for a good model. LogLoss is the log of the loss function and needs to be close to 0.

Model	Logloss	AUC	F1	Accuracy	MCC
LightGBM	0.0687	0.99776	0.9902	0.9902	0.9804
CatBoost	0.0323	0.9992	0.9952	0.9951	0.9903

be found in the GitHub repository [1]. Tables 7 summarizes various performance scores for the LightGBM and CatBoost models, and Fig. 4 summarizes the KS statistics.

Figure 5 summarizes the permutation feature importance. Looking at Fig. 5, we observe that the small prime index coefficients are the most important features for the prediction of the class number. In particular, the triple  $(a_2, a_3, a_5)$  is the most important for all models. Table 8 gives the value distribution of this triple, from which we can clearly see an unequal distribution depending on whether  $h_d = 1$  or  $h_d = 2$ . This unequal distribution of the data is much more pronounced in the case where  $a_2 = a_3 = a_5 = 1$  and it is visualized in Fig. 1.

This points towards the distribution of ones in the sequences of prime coefficients, namely, towards the number  $n_d$  of primes that ramify in  $K_d$  and that can



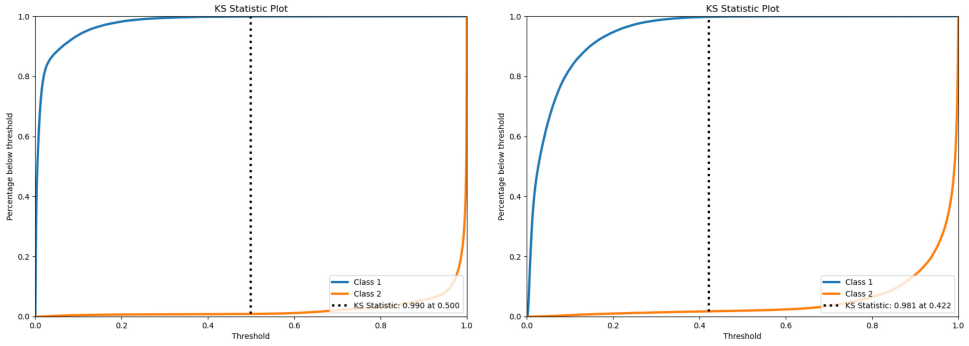


Fig. 4. KS statistic plot over the training set of the LightGBM (right) and CatBoost model (left) for the binary classification task  $h_d = 1$  versus  $h_d = 2$  with split 70/30.

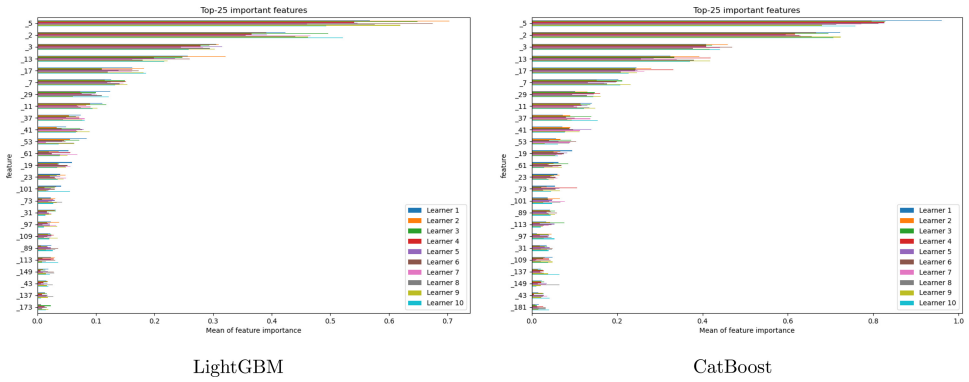


Fig. 5. Permutation feature importance for the LightGBM (left) and Catboost (right) model for the binary classification task  $h_d = 1$  versus  $h_d = 2$  on a 70/30 split.

Table 8. Number of real quadratic fields with class number  $h_d \in \{1, 2\}$  in dataset  $\mathcal{D}_{1,2}$  for specified values of the  $\zeta$ -coefficients  $a_2, a_3, a_5$ .

$h_d$	$a_2 = 0$	$a_3 = 0$	$a_5 = 0$	$a_2 = 1$	$a_3 = 1$	$a_5 = 1$	$a_2 = 2$	$a_3 = 2$	$a_5 = 2$
1	74868	76965	88786	27804	23650	1	74487	76544	88372
2	60392	69501	75225	62773	44673	33013	60271	69262	75198

Table 9. Number of real quadratic fields with class number  $h_d \in \{1, 2\}$  in dataset  $\mathcal{D}_{1,2}$  for specified number  $n_d$  of ramified primes.

$h_d$	$n_d = 1$	$n_d = 2$	$n_d = 3$
1	64522	112637	0
2	0	52451	130985

Table 10. Detected ramified primes in the first 1000 coefficients of the function  $\zeta_d(s)$  for fields  $K_d$  with class number  $h_d = 1, 2$  from dataset  $\mathcal{D}_{1,2}$ .

$h_d$	# of detected ramified primes	# of fields with this property
1	0	65468
	1	108527
	2	3164
2	0	806
	1	49595
	2	101978
	3	31057

be captured in the first 1000 coefficients of the function  $\zeta_d(s)$ . This is in line with the lemmas in Sec. 2. Table 9 shows the distribution of ramified primes against the class number and Table 10 represents the distribution of ramified primes appearing in the first 1000 coefficients of the function  $\zeta_d(s)$ .

#### 4.2. Learning $h_d \in \{1, 2\}$ with symbolic classification

In this section, we will construct a learning model obtained from an genetic programming algorithm called symbolic classification [8]. We will explore how symbolic classification can be used to produce an explicit predictor for the class number of real quadratic fields in our dataset, using the software HeuristicLab [14]. Such a predictor will provide an approximation to the class number formula for our dataset. This approach has been used in theoretical physics, e.g. [13], to develop good approximations for certain physical quantities based on a given set of learning features. In our setting, we will see that reasonable approximations to the class number formula (1.1) can be discovered using simpler learning features. Such approximations, as we will see, are able to shed light on interesting properties of real quadratic fields. In particular, they will allow us to recover some results in Sec. 2 on the parity of the class number.

Using the real quadratic fields of Sec. 4.1, we create a labeled dataset  $\mathcal{D}_{1,2}^{\text{SC}}$  by considering as learning features the number  $n_d$  of ramified primes in  $K_d$  and the ramified primes themselves. Since  $h_d \in \{1, 2\}$ , we have  $n_d \leq 3$  by Lemma 2.6 and write the ramified primes as  $p_1, p_2, p_3$ . By convention, if there is no second (respectively, third) ramified prime, we will set  $p_2 = 0$  (respectively,  $p_3 = 0$ ). Our dataset  $\mathcal{D}_{1,2}^{\text{SC}}$  can be summarized as follows:

$$\mathcal{D}_{1,2}^{\text{SC}} = \{(n_d, p_1, p_2, p_3) \rightarrow h_d\}.$$

As the dataset  $\mathcal{D}_{1,2}^{\text{SC}}$  is quite large, running symbolic classification becomes computationally demanding. We thus sample randomly 50,000 data points from  $\mathcal{D}_{1,2}^{\text{SC}}$  and construct a dataset where fields of class number 1 and 2 are chosen evenly. In this section, we use a training and testing split of 40/60. We will apply our approximate formulas to the whole LMFDB and test their performances.

Our symbolic classifier was built using the following parameters of HeuristicLab. The population size was fixed to 100, the fitness function used was *mean squared error*, the crossover method used was *subtree swapping crossover*, the mutator was *multi symbolic expression tree manipulator* and for *elitism*, we kept one elite at every generation to favor population exploration and avoid exploitation. For the alphabet of functions, we used various mix of the available functions.

Among all the formulas we have generated, the simplest one in terms of length and depth of the model is the following approximation of the class number formula:

$$h_d^{(1,2)} = 0.17257 \sin(1.5703 p_1) + 0.72004 n_d, \quad (4.3)$$

We note that the coefficient 1.5703 of  $p_1$  is approximately equal to  $\pi/2$ . Equation (4.3) yields the following predictor  $\phi(d)$  for  $h_d$ , which has accuracy 98% on the training and testing set, and whose accuracy persists when applied to the entire dataset  $\mathcal{D}_{1,2}^{\text{SC}}$ :

$$\phi(d) = \begin{cases} 1 & \text{if } h_d^{(1,2)} < t, \\ 2 & \text{if } h_d^{(1,2)} \geq t, \end{cases} \quad \text{for } t = 1.5115. \quad (4.4)$$

The value  $t$  appearing in (4.4) serves as the threshold for (4.3). The classification metrics for (4.4) are listed in Fig. 6. Class number 1 fields denote the positive class and Fig. 6 shows in particular that our formula predicts no false negatives, very few false positives, and is thus well-suited to study the relationship between  $p_1, n_d$  and  $h_d$  from the data.

Looking at Eq. (4.3), we note that only the number  $n_d$  of ramified primes and the first ramified prime  $p_1$  are selected by the genetic algorithm and sufficient to

- ◆ Classification positive class: Class 1
- ◆ F1 score (test): 0.980127159548617
- ◆ F1 score (training): 0.98042349180982824
- ◆ False discovery rate (test): 3.89712122201188 %
- ◆ False discovery rate (training): 3.84012539184953 %
- ◆ False positive rate (test): 3.90783530797932 %
- ◆ False positive rate (training): 3.84879725085911 %
- ◆ Matthews Correlation (test): 0.96097521585687817
- ◆ Matthews Correlation (training): 0.96155538580885913
- ◆ Negative predictive value (test): 100 %
- ◆ Negative predictive value (training): 100 %
- ◆ Positive predictive value (test): 96.1028787779881 %
- ◆ Positive predictive value (training): 96.1598746081505 %
- ◆ True negative rate (test): 96.0921646920207 %
- ◆ True negative rate (training): 96.1512027491409 %
- ◆ True positive rate (test): 100 %
- ◆ True positive rate (training): 100 %

Fig. 6. Summary of the classification metrics for (4.3) used in the binary classification task  $h_d = 1$  versus  $h_d = 2$ .

distinguish class number 1 and 2 real quadratic fields with very high accuracy. It is natural to ask what kind of information we can extract from  $h_d^{(1,2)}$ .

If  $n_d = 1$  (respectively,  $n_d = 3$ ) then it is easy to see that  $\phi(d) = 1$  (respectively,  $\phi(d) = 2$ ) from (4.3), regardless of the value of  $p_1$ . Thus, we recover the (consequence written after) Lemmas 2.2 and 2.6 (both under the additional hypothesis that  $h_d \in \{1, 2\}$ ) and  $\phi$  distinguishes between class numbers 1 and 2 fields with 100% accuracy. In the case  $n_d = 2$ , we have the following conditions imposed on  $p_1$ :

$$h_d^{(1,2)} < t \Leftrightarrow 4n - 2.2724 < p_1 < 4n + 0.27174, \quad n \in \mathbb{Z}_{>0}. \quad (4.5)$$

By analyzing Eq. (4.5), we find that the allowed primes are those  $p_1$  congruent to 3 mod 4 and  $p_1 = 2$ . If, instead, we investigate the  $h_d^{(1,2)} \geq t$  condition, then we find the missing primes  $p_1$  congruent to 1 mod 4. In other words, we have recovered Lemma 2.7. All together it reflects the general fact about parities given in Corollary 2.2.

Though a data point in  $\mathcal{D}_{1,2}^{\text{SC}}$  contains  $p_1, p_2, p_3$ , the formula in (4.3) has only the first ramified prime  $p_1$ , which agrees with Lemma 2.7 (in the case  $n_d = 2$ ). It is interesting to see that our machine learning model has *learned* this fact from data alone.

We will see in Sec. 5.1 that such a classification failed in the case of class number 1 and 3 fields. This gives evidence that one cannot expect to distinguish fields with the same parity using only the simple features  $n_d$  and the list of ramified primes.

## 5. Class Numbers 1 and 3

We now turn to the classification of real quadratic fields of class numbers 1 and 3. We attempt to mimic the structure of Sec. 4, but include new ideas to circumvent the difficulties we encounter. Based on the minimal values of the cost function in Sec. 3.5, we may suspect that classifiers trained on  $\zeta$ -coefficients alone are not as accurate as they are in the case of class numbers 1 and 2. This is borne out in our implementation, in which, trained on these features, neither LightGBM nor symbolic classification succeeded in distinguishing real quadratic fields of class number 1 or 3. Subsequently, we incorporate other features, such as the regulator and partial sums.

### 5.1. *Balancing data and selecting features*

There is a large imbalance between the 25,201 real quadratic fields with class number 3 and the 177,159 with class number 1 included in [11]. To avoid any bias in the construction of our dataset, we will restrict ourselves to the 11,531 real quadratic fields of class number 3 with discriminant  $D \leq 10^6$ . To construct a balanced dataset including the same number of real quadratic fields with class number 1 evenly distributed over discriminant ranges, we sample class number 1 fields randomly from the sets  $\{K_d : i \cdot 10^5 \leq D \leq (i + 1) \cdot 10^5\}$ ,  $i \in \{0, \dots, 9\}$ , as many as class number 3 fields in each interval. For example, there are 1261 fields with class number 3 in

the interval  $1 \leq D \leq 10^5$  and so in this interval, we pick the same amount of class number 1 fields.

Our dataset thus consists of 23,062 fields. For each of them, we compute the vector  $v(d)$  as in (4.1), the discriminant  $D$ , the regulator  $R_d$ , the number  $n_d$  of ramified primes, the ramified primes  $p_1, p_2$  where we set  $p_2 = 0$  in the case that only one prime ramifies, and the partial sums  $S_{\zeta_d}$  and  $S_{\chi_D}$  defined by

$$S_{\zeta_d} := \sum_{n=1}^{1000} \frac{a_n}{n} \quad \text{and} \quad S_{\chi_D} := \sum_{n=1}^{1000} \frac{\chi_D(n)}{n}.$$

The partial sums  $S_{\zeta_d}$  and  $S_{\chi_D}$  are motivated by  $\zeta_d(s)$  and  $L(s, \chi_D)$ , respectively. Our dataset can thus be summarized as

$$\mathcal{D}_{1,3}^{\text{SC}} = \{(v(d), D, R_d, S_{\zeta_d}, S_{\chi_D}, n_d, p_1, p_2) \rightarrow h_d\}.$$

## 5.2. Learning $h_d \in \{1, 3\}$ from $\mathcal{D}_{1,3}^{\text{SC}}$

The problem of classifying real quadratic fields of class numbers 1 and 3 is much more difficult than the problem of classifying those with class numbers 1 and 2. Section 3.5 discussed this problem from a conceptual perspective. In this section, we observe the increased difficulty from a practical standpoint.

In Table 11 we summarize the features and testing accuracy for various experiments using the LightGBM classifier. As in Sec. 4, the data segregation is 70/30, together with a 10-fold cross validation performed on the training set.

There are various observations we can make from Table 11. Row 4 shows that the regulator and the prime index coefficients are strong predictors for the class number. When we add the discriminant in row 6, we obtain an almost perfect classification. Also look at the following rows where high accuracies are attained with  $S_{\zeta_d}$  and

Table 11. Performance of LightGBM on various combinations of learning features for the binary classification task of  $h_d = 1$  versus  $h_d = 3$ .

Row number	Features	Testing accuracy
1	$(a_p)_{p \leq 1000}$	53.34%
2	$(a_p)_{p \leq 1000}, n_d, p_i$	54.87%
3	$(a_p)_{p \leq 1000}, S_{\zeta_d}$	52.87%
4	$(a_p)_{p \leq 1000}, R_d$	91.21%
5	$(a_p)_{p \leq 1000}, D$	53.75%
6	$(a_p)_{p \leq 1000}, D, R_d$	99.38%
7	$D, R_d, S_{\zeta_d}$	99.84%
8	$(a_p)_{p \leq 1000}, S_{\chi_D}$	53.45%
9	$D, R_d, S_{\chi_D}$	99.93%
10	$(a_{p_i})_{i=1}^{10}, D, R_d$	99.90%
11	$(a_{p_i})_{i=1}^5, D, R_d$	99.54%
12	$(a_{p_i})_{i=1}^3, D, R_d$	99.55%
13	$(a_{p_i})_{i=1}^2, D, R_d$	97.51%
14	$a_{p_1}, D, R_d$	88.46%

Table 12. Detected ramified primes in the first 1000 coefficients of the function  $\zeta_d(s)$  for real quadratic fields of class numbers 1 and 3 in dataset  $\mathcal{D}_{1,3}^{\text{SC}}$ .

$h_d$	# of detected primes	# of fields with this property
1	0	4250
	1	6904
	2	386
3	0	4284
	1	6858
	2	389

Table 13. Number of ramified primes in fields  $K_d$  of class number  $h_d \in \{1, 3\}$  from dataset  $\mathcal{D}_{1,3}^{\text{SC}}$  in terms of their class number.

Class number $h_d$	$n_d = 1$	$n_d = 2$
1	4258	7282
3	4288	7243

$S_{\chi_D}$ . Of course, this could be expected from the class number formula (1.1), and we will generate some approximate formulas to  $h_d$  in Sec. 5.3. As in Sec. 4, rows 10–14 show the importance of the small prime index coefficients.

Row 2 suggests that the values  $n_d$  and the number of ramified primes detected in the sequence of prime index  $\zeta$ -coefficients may be uniformly distributed over our dataset of class numbers 1 and 3. Indeed, Tables 12 and 13 show this phenomenon. Consequently,  $n_d$  cannot be taken as a meaningful feature for classification. This situation is in stark contrast with the case of class numbers 1 and 2.

### 5.3. Learning $h_d \in \{1, 3\}$ with symbolic classification

Using the same parameters as in Sec. 4.2, we will explore two formulas yielded by symbolic classification. The classification metrics are listed in Fig. 7. Unlike in Sec. 4.2, the formulas are not tested on the entire LMFDB but only on our dataset.

The first formula obtained is closely related to the class numbers formula (1.1):

$$h_d^{(1,3)} = \frac{1}{2} \frac{\sqrt{D} S_{\chi_D}}{R_d}. \tag{5.1}$$

The following predictor attains 100% accuracy on both the training and testing data:

$$\phi(d) = \begin{cases} 1 & \text{if } h_d^{(1,3)} < t, \\ 3 & \text{if } h_d^{(1,3)} \geq t, \end{cases} \quad \text{for } t = 1.963. \tag{5.2}$$

This result shows the capability of symbolic classification to discover an effective formula from a dataset. On the other hand, [2, Proposition 5] implies that  $S_{\chi_D}$  may be well approximated by the sum  $\chi_D(2)/2 + \chi_D(3)/3 + \chi_D(5)/5$ . Consequently,

- Classification positive class: Class 1
- F1 score (test): 1
- F1 score (training): 1
- False discovery rate (test): 0 %
- False discovery rate (training): 0 %
- False positive rate (test): 0 %
- False positive rate (training): 0 %
- Matthews Correlation (test): 1
- Matthews Correlation (training): 1
- Negative predictive value (test): 100 %
- Negative predictive value (training): 100 %
- Positive predictive value (test): 100 %
- Positive predictive value (training): 100 %
- True negative rate (test): 100 %
- True negative rate (training): 100 %
- True positive rate (test): 100 %
- True positive rate (training): 100 %

Equation (5.1)

- Classification positive class: Class 1
- F1 score (test): 0.99773091297383876
- F1 score (training): 0.99814195466369382
- False discovery rate (test): 0.106908993719097 %
- False discovery rate (training): 0.123946455131383 %
- False positive rate (test): 0.106737825216811 %
- False positive rate (training): 0.12388503468781 %
- Matthews Correlation (test): 0.99546832956286935
- Matthews Correlation (training): 0.9962855948744499
- Negative predictive value (test): 99.6539331824837 %
- Negative predictive value (training): 99.7525365008661 %
- Positive predictive value (test): 99.8930910062809 %
- Positive predictive value (training): 99.8760535448686 %
- True negative rate (test): 99.8932621747832 %
- True negative rate (training): 99.8761149653122 %
- True positive rate (test): 99.6533795493934 %
- True positive rate (training): 99.7524139638524 %

Equation (5.3)

Fig. 7. Classification metrics of the formulas (5.1) and (5.3).

the analytic class number formula suggests that  $h_d$  is determined by  $D$ ,  $R_d$ , and the coefficients  $a_2$ ,  $a_3$  and  $a_5$ . Given these inputs, symbolic classification yields the following unexpected formula:

$$\tilde{h}_d^{(1,3)} = \frac{1.8858\sqrt{D}}{R_d \exp(-0.5468a_2 - 0.2718a_3) \cos(\sin(-0.2556a_3)) \times \cos(-0.1962a_5) \cos^4(-0.1952a_5)}. \quad (5.3)$$

This is rather unlike anything from the classical theory, and offers limited mathematical insight to the problem. Nevertheless, the following predictor based on Eq. (5.3) attains around 99.8% accuracy both on the training and test set:

$$\phi(d) = \begin{cases} 1 & \text{if } \tilde{h}_d^{(1,3)} < t, \\ 3 & \text{if } \tilde{h}_d^{(1,3)} \geq t, \end{cases} \quad \text{for } t = 15.97. \quad (5.4)$$

## Acknowledgments

MA is supported by Microsoft Research NE, YHH is indebted to STFC UK, for grant ST/J00037X/2, KHL is partially supported by a grant from the Simons Foundation (#712100), and TO acknowledges support from the EPSRC through research grant EP/S032460/1. MA is also grateful to Vlad Serban, Maryna Viazovska, Mounir Boukadoum and Henry Cohn for their valuable comments, discussions and teachings. We are grateful to the anonymous referee, whose insight has had a significant impact on the clarity of this paper.

## Appendix A

### A.1. Dimensionality reduction

In Sec. 3, we explored how class number data may be separated using the *bubble* algorithm. In this appendix, we visualize how PCA clusters our data with respect to the axes of maximal variance. To perform this dimensionality reduction, we compress the features  $(a_n)_{n=1}^{1000}$  to a 3-dimensional space.

The case of class numbers 1 and 2 real quadratic fields is depicted in Fig. A.1, where we observe that there are forbidden zones for class number 1 fields. It would be interesting to study explicitly these constraints. In addition, we can see that the two classes are not entirely separated or mixed in this 3-dimensional representation of the data. This shows that our data responds moderately to linear methods and points towards the fact that nonlinear algorithms, such as LightGBM and CatBoost, can perform better. Furthermore, looking at Fig. A.2, we see that the prime index coefficients alone are not separable through PCA. Note that this does not imply that nonlinear separability is unfeasible as we have successfully distinguished these two classes in Sec. 4 using nonlinear algorithms.

In the case of class numbers 1 and 3 fields in Figs. A.3–A.5, we see that the data is much less separable from the point of view of PCA. In contrast to Fig. A.1, Fig. A.3 shows that the coefficients of class numbers 1 and 3 fields are completely mixed, which can explain the greater difficulty in classifying such fields, and there is no noticeable difference between Figs. A.4 and A.2. Figure A.5 shows the result when the first 10 prime coefficients are combined with other features  $R_d$  and  $D$ . Recall that, with all these features combined, we could obtain a high accuracy as indicated in Table 11.

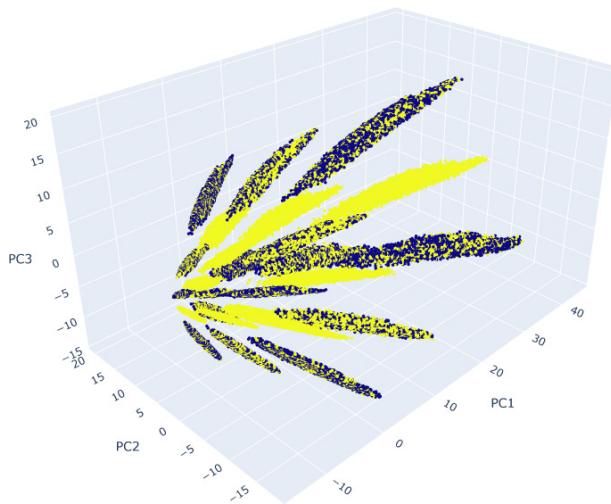


Fig. A.1. (Color online) Dimensionality reduction using PCA for  $(a_n)_{n=1}^{1000}$  in the case of real quadratic fields of class number  $h_d \in \{1, 2\}$  (1: blue; 2: yellow).



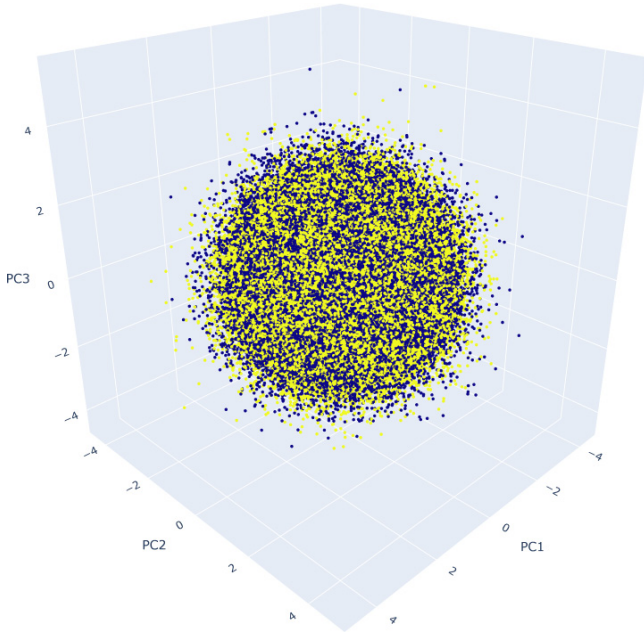


Fig. A.2. (Color online) Dimensionality reduction using PCA for  $(a_p)_{\substack{p \leq 1000 \\ p: \text{prime}}}$  in the case of real quadratic fields of class number  $h_d \in \{1, 2\}$  (1: blue; 2: yellow).

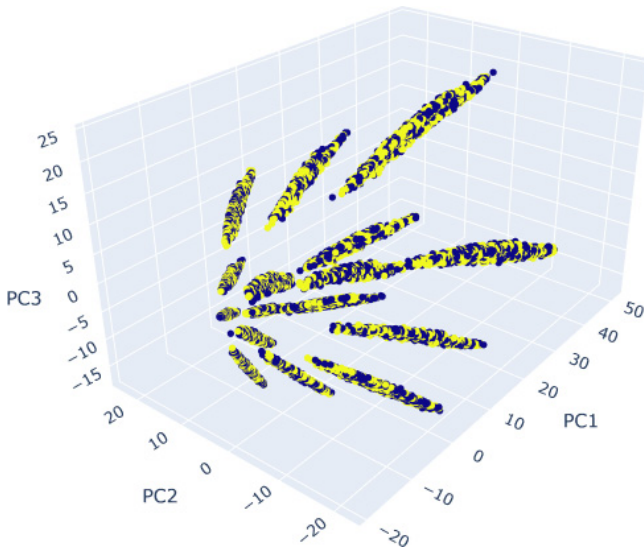


Fig. A.3. (Color online) Dimensionality reduction using PCA for  $(a_n)_{n=1}^{1000}$  in the case of real quadratic fields of class number  $h_d \in \{1, 3\}$  (1: blue; 3: yellow).

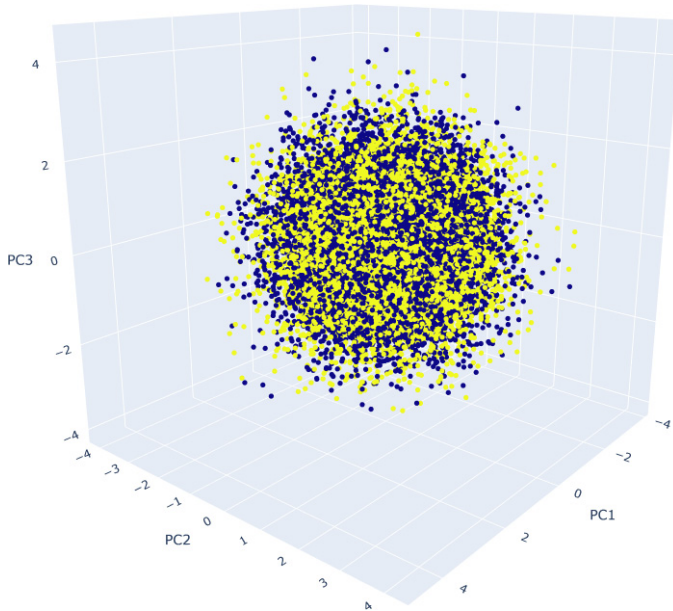


Fig. A.4. (Color online) Dimensionality reduction using PCA for  $(a_p)_{\substack{p \leq 1000 \\ p: \text{prime}}}$  in the case of real quadratic fields of class number  $h_d \in \{1, 3\}$  (1: blue; 3: yellow).

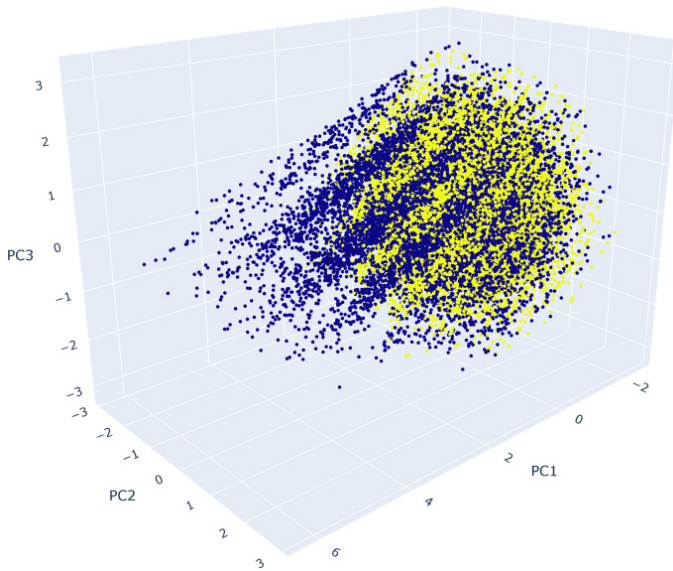


Fig. A.5. (Color online) Dimensionality reduction using PCA for  $D, R_d$  and the first 10  $a_p$ -coefficients in the case of real quadratic fields of class number  $h_d \in \{1, 3\}$  (1: blue; 3: yellow).

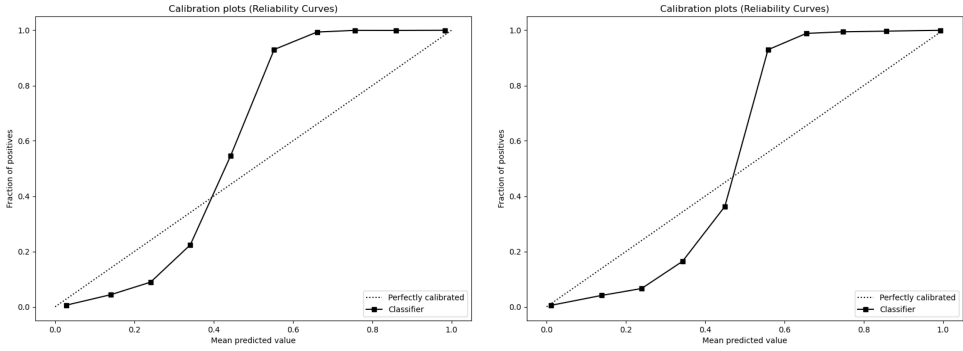


Fig. A.6. Calibration curves of the LightGBM model (left) and CatBoost model (right) on a 70/30 split of the dataset for the binary classification task of quadratic fields with class number  $h_d = 1$  versus  $h_d = 2$ .

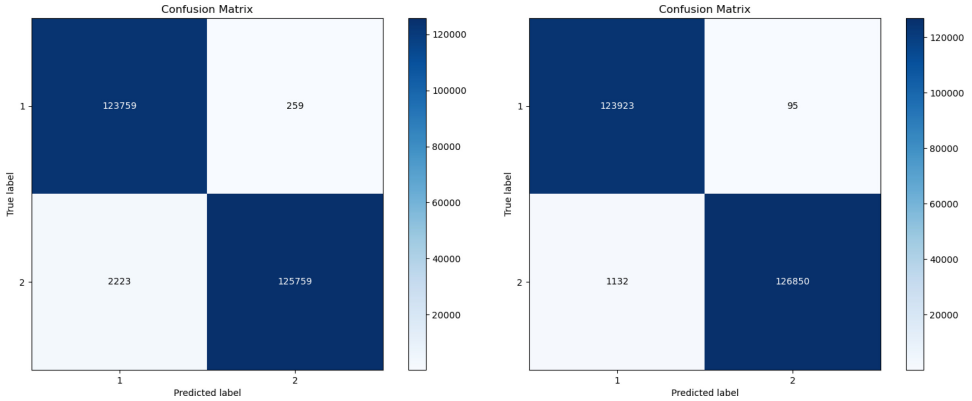


Fig. A.7. Confusion matrix of the LightGBM model (left) and CatBoost model (right) on a 70/30 split of the dataset for the binary classification task of quadratic fields with class number  $h_d = 1$  versus  $h_d = 2$ .

### A.2. Supplementary material for Sec. 4.1

In Sec. 4.1, we focused largely on the permutation feature importance. In Figs. A.6 and A.7, we record the calibration curves and confusion matrices for 70/30 splits.

### References

- [1] M. Amir (2022), <https://github.com/malik-amir/Machine-Learning-of-the-Tate-Shafarevich-group>.
- [2] W. Duke, Extreme values of Artin L-functions and class numbers, *Compos. Math.* **136**(1) (2003) 103–115.
- [3] Y.-H. He, K.-H. Lee and T. Oliver, Machine-learning the Sato–Tate conjecture, *J. Symb. Comput.* **111** (2022) 61–72.
- [4] Y.-H. He, K.-H. Lee and T. Oliver, Machine-learning number fields, *Math. Comput. Geom. Data* **2** (2022) 49–66.

- [5] Y.-H. He, K.-H. Lee and T. Oliver, Machine learning the invariants of arithmetic curves, *J. Symb. Comput.* **115** (2023) 478–491.
- [6] Y.-H. He, K.-H. Lee, T. Oliver and A. Pozdnyakov, Murmurations of elliptic curves, preprint (2022), arXiv:2204.10140.
- [7] G. J. Janusz, *Algebraic Number Fields*, Pure and Applied Mathematics, Vol. 55 (Academic Press, New York, 1973).
- [8] M. F. Kornš, Genetic programming symbolic classification: A study, *Genetic Programming Theory and Practice XV* (Springer, 2018), pp. 39–54.
- [9] E. Kowalski, Elementary theory of L-functions, II, *An Introduction to the Langlands Program (Jerusalem, 2001)* (Birkhäuser Boston, Boston, MA, 2003), pp. 21–37.
- [10] F. Lemmermeyer, *Reciprocity Laws: From Euler to Eisenstein*, Springer Monographs in Mathematics (Springer-Verlag, Berlin, 2000).
- [11] The LMFDB Collaboration, The L-functions and modular forms database (2022), <http://www.lmfdb.org>.
- [12] A. Płońska and P. Płoński, MLJAR: State-of-the-art automated machine learning framework for tabular data, Version 0.10.3, MLJAR (2021), <https://github.com/mljar/mljar-supervised>.
- [13] S. Udrescu and M. Tegmark, AI Feynman: A physics-inspired method for symbolic regression, *Sci. Adv.* **6**(16) (2020) eaay2631, arXiv:1905.11481.
- [14] S. Wagner *et al.*, Architecture and design of the heuristiclab optimization environment, *Advanced Methods and Applications in Computational Intelligence*, Topics in Intelligent Engineering and Informatics Series (Springer, 2014), pp. 197–261.