Using Machine Learning for Number Theory

Seewoo Lee

September 8, 2025. Math254A guest lecture

Goal

Goal of today's talk:

Goal

Goal of today's talk:

- Tell you about the potential usefulness of machine learning in number theory

Goal

Goal of today's talk:

- Tell you about the potential usefulness of machine learning in number theory (but without AI hype)

Goal

Goal of today's talk:

- Tell you about the potential usefulness of machine learning in number theory (but without AI hype)
- Two use cases: Predicting the rank of elliptic curves and Galois groups of number fields

Goal

Goal of today's talk:

- Tell you about the potential usefulness of machine learning in number theory (but without AI hype)
- Two use cases: Predicting the rank of elliptic curves and Galois groups of number fields
- Give you some ideas for your future research

Elliptic Curve

Definition

An **elliptic curve** over a field K is a smooth, projective, algebraic curve of genus one, with a specified point $O \in E(K)$.

More concretely, it can be given by (the projectivization of) a Weierstrass equation of the form

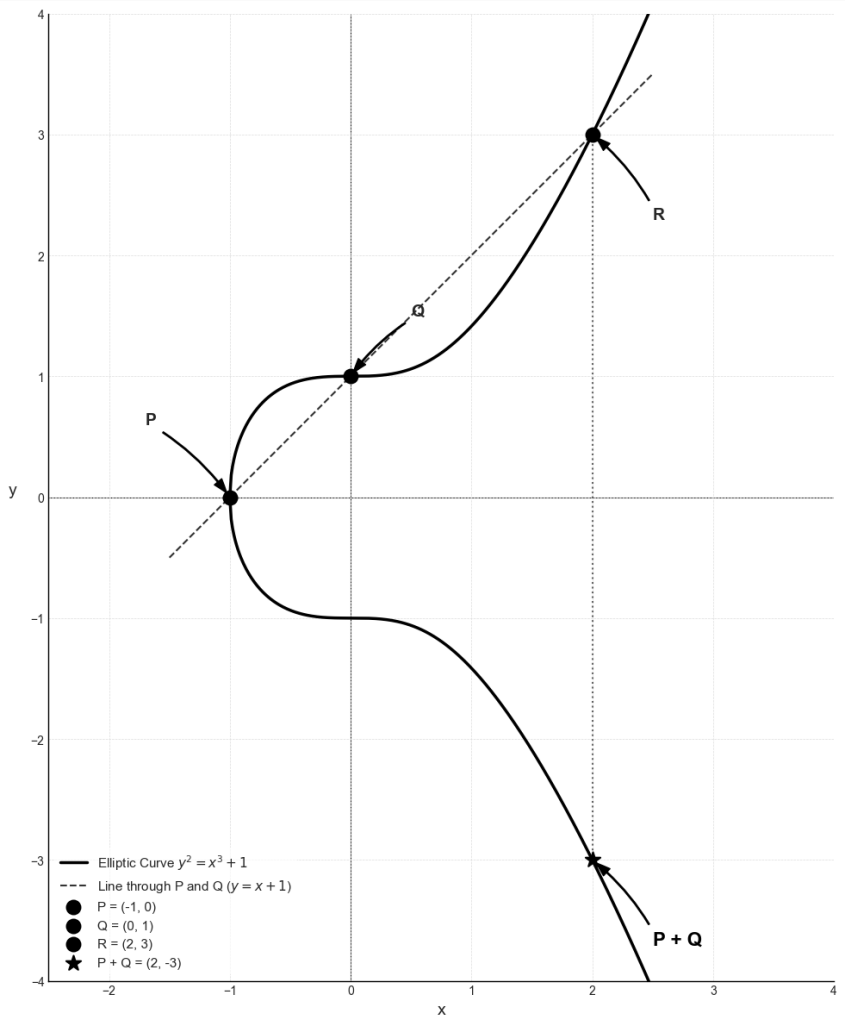
$$y^2 = x^3 + ax + b$$

where $a, b \in K$ and the curve is nonsingular[1].

[1] $\Delta = -16(4a^3 + 27b^2) \neq 0$

Group Structure of $E(K)$

Elliptic curves have a group structure. The point at infinity O serves as the identity element.



Mordell–Weil theorem

Theorem (Mordell–Weil)

The group of rational points $E(\mathbb{Q})$ is a finitely generated abelian group. That is,

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{tors} \oplus \mathbb{Z}^r$$

where $E(\mathbb{Q})_{tors}$ is the torsion subgroup and r is the rank of the elliptic curve.

Torsion is Easy

Theorem (Nagell–Lutz)

Let $(x, y) \in E(\mathbb{Q})$ be a torsion point.

- *If $(x, y) \neq O$, then $x, y \in \mathbb{Z}$.*
- *Either $y = 0$ or y^2 divides the discriminant Δ of the curve.*

Theorem (Mazur)

The torsion subgroup $E(\mathbb{Q})_{tors}$ is isomorphic to one of the following 15 groups:

$$\mathbb{Z}/n \quad (1 \leq n \leq 10, n = 12),$$
$$\mathbb{Z}/2 \times \mathbb{Z}/2m \quad (1 \leq m \leq 4).$$

Rank is Hard

It is not just hard, but it is *very* hard! We have the following list of open problems:

Rank is Hard

It is not just hard, but it is *very* hard! We have the following list of open problems:

- Can you find an elliptic curve over $\mathbb{Q}$ with arbitrarily large rank?
- (Goldfeld's conjecture) How many elliptic curves over $\mathbb{Q}$ have rank 0, 1, 2, . . . ?
- (BSD conjecture, Parity conjecture) How is the rank of an elliptic curve over $\mathbb{Q}$ related to its L-function?

ML?

Question

Can we predict the rank of an elliptic curve over $\mathbb{Q}$ using machine learning?

ML?

Question

Can we predict the rank of an elliptic curve over $\mathbb{Q}$ using machine learning?

If the target is the rank r of an elliptic curve $E/\mathbb{Q}$, then what are the features?

ML?

Question

Can we predict the rank of an elliptic curve over $\mathbb{Q}$ using machine learning?

If the target is the rank r of an elliptic curve $E/\mathbb{Q}$, then what are the features? We may want to predict r from something that is easier to compute.

He–Lee–Oliver

Yanghui He, Kyu-Hwan Lee, and Thomas Oliver[2] used classical ML algorithms (not ChatGPT!) to predict the rank of elliptic curves over $\mathbb{Q}$, using *Frobenius traces* $a_p(E)$ as features:

$$a_p(E) = p + 1 - |E(\mathbb{F}_p)|$$

where $|E(\mathbb{F}_p)|$ is the number of points on the reduced curve modulo p (when E has good reduction at p).

Note that the rank (more precisely, the isogeny class of E) is determined by $a_p(E)$ for *all* p.

[2] He–Lee–Oliver, *Machine learning invariants of arithmetic curves*, 2023

He–Lee–Oliver

Consider rank $0, 1$ curves of conductor ≤ 10000. They used logistic regression:

$$\mathbb{P}[\text{rank}(E) = 1 | \{a_{p_n}\}_{n \leq 300}] \approx \sigma(\mathbf{w} \cdot \mathbf{a} + b)$$

where $\mathbf{a} = (a_2, a_3, a_5, \ldots a_{1987}) \in \mathbb{R}^{300}$,

$$\sigma(x) = \frac{1}{1 + e^{-x}},$$

and $\mathbf{w} \in \mathbb{R}^{300}$, $b \in \mathbb{R}$ are weight and bias to be optimized. It essentially tries to find the a separating hyperplane in $\mathbb{R}^N = \mathbb{R}^{300}$.

Quiz 1

What was the accuracy of the experiment?

1. 0%
2. $(0\%, 50\%]$
3. $(50\%, 75\%]$
4. $(75\%, 90\%]$
5. $(90\%, 95\%]$
6. $(95\%, 100\%)$
7. 100%

Quiz 1

What was the accuracy of the experiment?

1. 0%
2. $(0\%, 50\%]$
3. $(50\%, 75\%]$
4. $(75\%, 90\%]$
5. $(90\%, 95\%]$
6. $(95\%, 100\%)$ (99.1%)
7. 100%

Why?

Why does it work so well?

It is not surprising

The BSD conjecture claims that the number of $\mathbb{F}_p$ points and the rank r of E are related by

$$\prod_{\substack{p \leq x \\ p \nmid \Delta_E}} \frac{|E(\mathbb{F}_p)|}{p} \sim C(\log x)^r$$

for some constant $C > 0$. In other words, the rank and $\{|E(\mathbb{F}_p)|\}_p$ are positively correlated. Since $a_p(E) = p + 1 - |E(\mathbb{F}_p)|$, the rank and $\{a_p\}_p$ are negatively correlated.

Something is very surprising

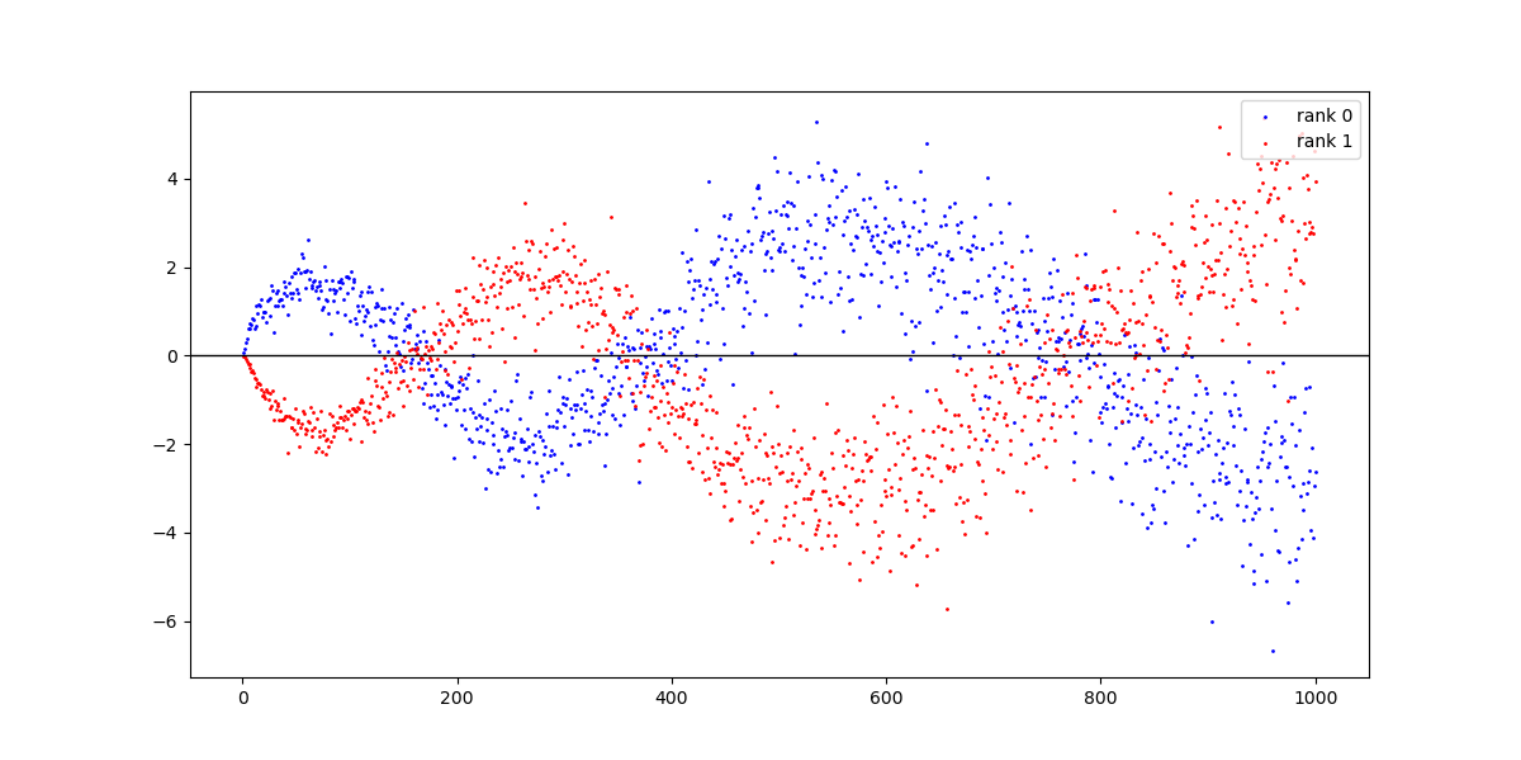
Podznyakov[3] analyzed models and data, e.g. using PCA. He plotted the average of a_p for each p and plotted it as a function in p, and got the following plot:[4]

[3] was an undergraduate student supervised by K.-H. Lee

[4] Conductor in $[7500, 10000]$, rank $\in \{0, 1\}$

Something is very surprising

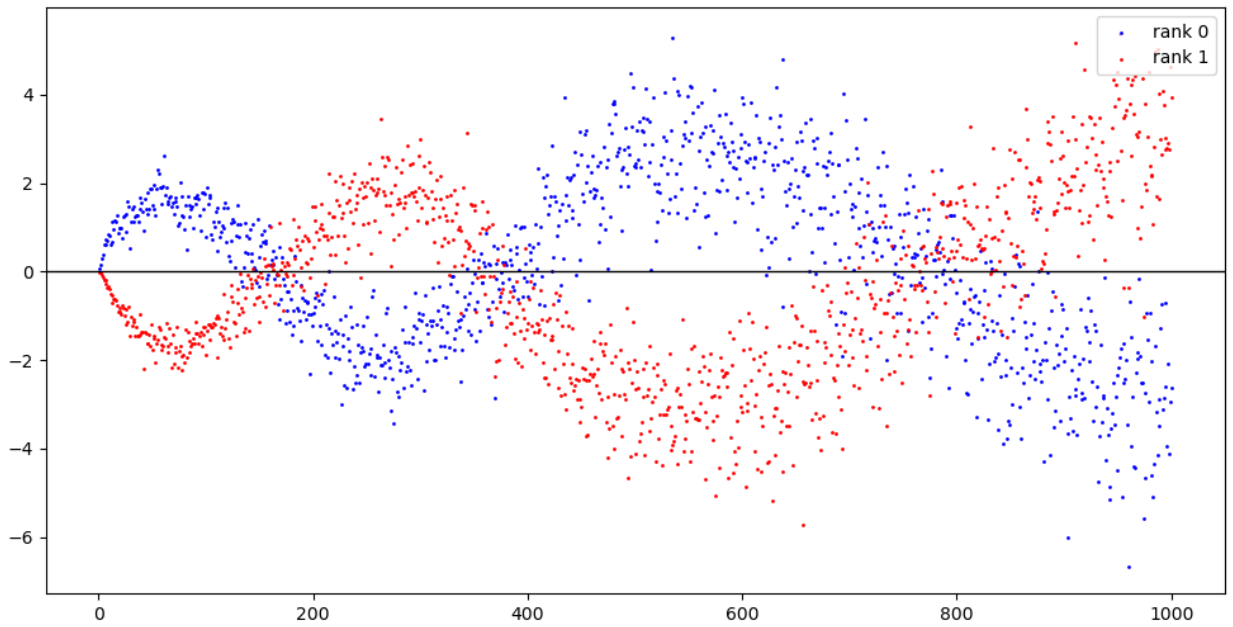
Podznyakov[3] analyzed models and data, e.g. using PCA. He plotted the average of a_p for each p and plotted it as a function in p, and got the following plot:[4]



[3] was an undergraduate student supervised by K.-H. Lee

[4] Conductor in $[7500, 10000]$, rank $\in \{0, 1\}$

Something is very surprising



This IS surprising, since

1. We only expected a negative correlation between rank and a_p's, but this shows an oscillating pattern.
2. The equation for the "limit curve" is still unknown.

Murmuration

It is now called *murmuration*, and has become an active area of study. The oscillating pattern is observed for other "families" of L-functions, and the "limit curve" (*murmuration density*) is now known for

- Modular forms
- Maass forms
- Dirichlet characters
- Hecke characters of imaginary quadratic fields
- Elliptic curves, but in a different setup

In this case, ML **"motivated"** mathematicians to find a new phenomenon.

Murmuration

It is now called *murmuration*, and has become an active area of study. The oscillating pattern is observed for other "families" of L-functions, and the "limit curve" (*murmuration density*) is now known for

- Modular forms
- Maass forms
- Dirichlet characters
- Hecke characters of imaginary quadratic fields
- Elliptic curves, but in a different setup

In this case, ML **"motivated"** mathematicians to find a new phenomenon. **However, it is incorrect to say that ML "found" new mathematics.**

Number fields

Definition

A **number field** is a finite extension K of $\mathbb{Q}$.

Number fields

Definition

A **number field** is a finite extension K of $\mathbb{Q}$.

$\mathbb{Q}, \mathbb{Q}(i), \mathbb{Q}(\sqrt{2}), \mathbb{Q}(\sqrt[3]{3}), \mathbb{Q}[x]/(x^3 - 3x - 1), \ldots$

Galois group

Question

How do we compute the Galois group $\mathrm{Gal}(K/\mathbb{Q})$ of a number field K?

It is not an easy problem.

Quiz 2-1

What is the Galois group of

$$K = \mathbb{Q}[x]/(x^4 + 2x^2 + 4)$$

1. $\mathbb{Z}/4$
2. $(\mathbb{Z}/2)^2$

Quiz 2-1

What is the Galois group of

$$K = \mathbb{Q}[x]/(x^4 + 2x^2 + 4) \simeq \mathbb{Q}(\sqrt{2}, \sqrt{-3})$$

1. $\mathbb{Z}/4$
2. $(\mathbb{Z}/2)^2$

Quiz 2-2

What is the Galois group of

$$K = \mathbb{Q}[x]/(x^8 - x^7 + x^5 - x^4 + x^3 - x + 1)$$

1. $\mathbb{Z}/8$
2. $\mathbb{Z}/4 \times \mathbb{Z}/2$
3. $(\mathbb{Z}/2)^3$
4. D_4
5. Q_8

Quiz 2-2

What is the Galois group of

$$K = \mathbb{Q}[x]/(x^8 - x^7 + x^5 - x^4 + x^3 - x + 1) \simeq \mathbb{Q}(\zeta_{15})$$

1. $\mathbb{Z}/8$
2. $\mathbb{Z}/4 \times \mathbb{Z}/2 \simeq (\mathbb{Z}/15)^\times$
3. $(\mathbb{Z}/2)^3$
4. D_4
5. Q_8

ML?

Question

Can we use ML to predict Galois group?

Dedekind zeta function

The analogue of $a_p(E)$ for number fields is the *Dedekind zeta coefficients*.

Definition

For a number field K, the **Dedekind zeta function** $\zeta_K(s)$ is defined as

$$\zeta_K(s) = \sum_{\mathfrak{a} \subset \mathcal{O}_K} \frac{1}{(N\mathfrak{a})^s}$$

where the sum is over all nonzero ideals $\mathfrak{a}$ of the ring of integers $\mathcal{O}_K$ of K, and $N\mathfrak{a} = |\mathcal{O}_K/\mathfrak{a}|$ is the norm of $\mathfrak{a}$.

Dedekind zeta function

It can be rewritten as

$$\zeta_K(s) = \sum_{n \geq 1} \frac{a_n(K)}{n^s}$$

for $a_n(K) = |\{\mathfrak{a} \subset \mathcal{O}_K : N\mathfrak{a} = n\}|$. These will be used as features.

Examples

$$\zeta_{\mathbb{Q}}(s) = \sum_{n \geq 1} \frac{1}{n^s} = \zeta(s)$$

$$\begin{aligned}\zeta_{\mathbb{Q}(i)}(s) &= \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{4^s} + \frac{2}{5^s} + \frac{1}{8^s} + \frac{1}{9^s} + \cdots \\ &= \zeta(s)\left(\frac{1}{1^s} - \frac{1}{3^s} + \frac{1}{5^s} - \frac{1}{7^s} + \cdots\right)\end{aligned}$$

Setup

We follow the setup of He–Lee–Oliver[5]

- Fix degree $\in \{4, 6, 8, 9, 10\}$.
- Input (feature): $\{a_n(K)\}_{n \leq N}$, let's say $N = 1000$.
- Output (target): Galois group
- Model: Decision Tree (= bunch of if-else statements)
- **We'll only focus on Galois (normal) extensions.**

[5] He–Lee–Oliver, *Machine Learning Number Fields*, 2022

Quiz 3-1

Let K be a degree 9 Galois extension of $\mathbb{Q}$. What are the possible groups that appear as $\mathrm{Gal}(K/\mathbb{Q})$?

Quiz 3-1

Let K be a degree 9 Galois extension of $\mathbb{Q}$. What are the possible groups that appear as $\mathrm{Gal}(K/\mathbb{Q})$?

$$\mathbb{Z}/9, \quad (\mathbb{Z}/3)^2$$

Quiz 3-2

There are 1266 Galois nonic fields in LMFDB, 22% of them are $\mathbb{Z}/9$ and 78% of them are $(\mathbb{Z}/3)^2$. Split them randomly into train (80%) and test (20%) set. What was the accuracy of the decision tree model?

1. 0%
2. $(0\%, 50\%]$
3. $(50\%, 75\%]$
4. $(75\%, 90\%]$
5. $(90\%, 95\%]$
6. $(95\%, 100\%)$
7. 100%

Quiz 3-2

There are 1266 Galois nonic fields in LMFDB, 22% of them are $\mathbb{Z}/9$ and 78% of them are $(\mathbb{Z}/3)^2$. Split them randomly into train (80%) and test (20%) set. What was the accuracy of the decision tree model?

1. 0%
2. $(0\%, 50\%]$
3. $(50\%, 75\%]$
4. $(75\%, 90\%]$
5. $(90\%, 95\%]$
6. $(95\%, 100\%)$
7. 100%

Why? Let's look inside the tree.

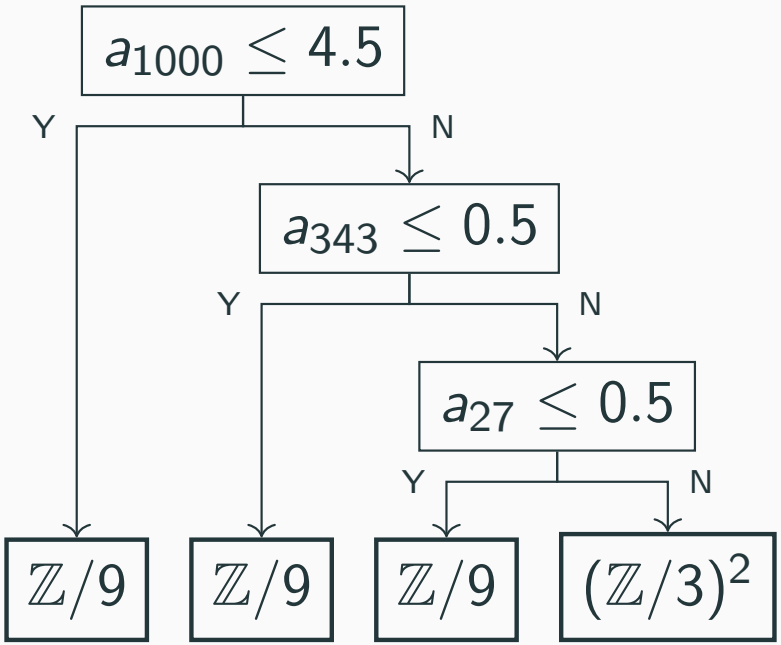
Decision trees are great since they are often easy to interpret.

Here's the tree achieving 100% accuracy:

Why? Let's look inside the tree.

Decision trees are great since they are often easy to interpret.

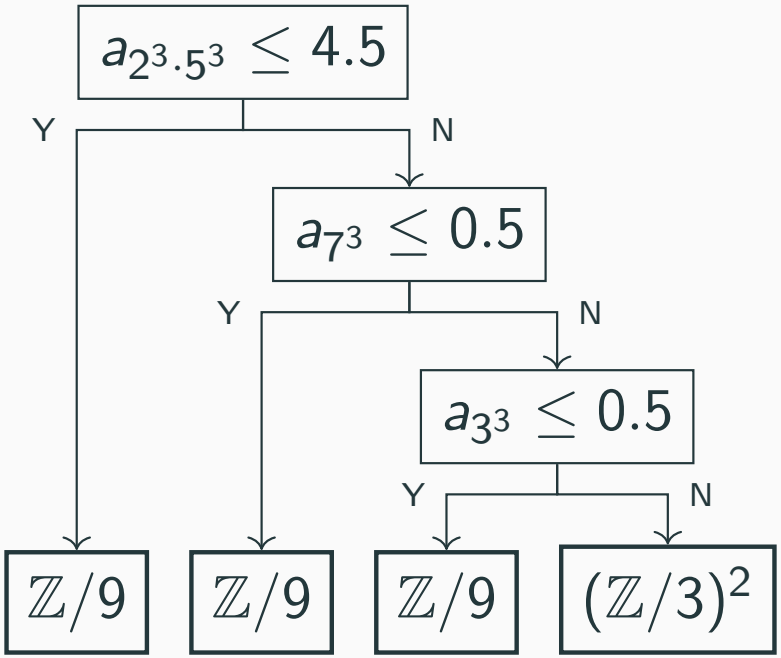
Here's the tree achieving 100% accuracy:



Why? Let's look inside the tree.

Decision trees are great since they are often easy to interpret.

Here's the tree achieving 100% accuracy:



Tree's logic

All a_n's are integers, so $a_n \leq 0.5$ is equivalent to $a_n = 0$. From the two nodes below, it is natural to conjecture that

Conjecture

Let $K/\mathbb{Q}$ be a nonic Galois extension. If $a_{p^3}(K) = 0$ for some prime p, then $\mathrm{Gal}(K/\mathbb{Q}) \simeq \mathbb{Z}/9$.

Tree's logic

All a_n's are integers, so $a_n \leq 0.5$ is equivalent to $a_n = 0$. From the two nodes below, it is natural to conjecture that[6]

Theorem (Lee[2])

Let $K/\mathbb{Q}$ be a nonic Galois extension. Then $a_{p^3}(K) = 0$ for some prime p if and only if $\mathrm{Gal}(K/\mathbb{Q}) \simeq \mathbb{Z}/9$.

[6] Lee–Lee, *Machines Learn Number fields, But How? The Case of Galois Groups*, 2025

Tree's logic

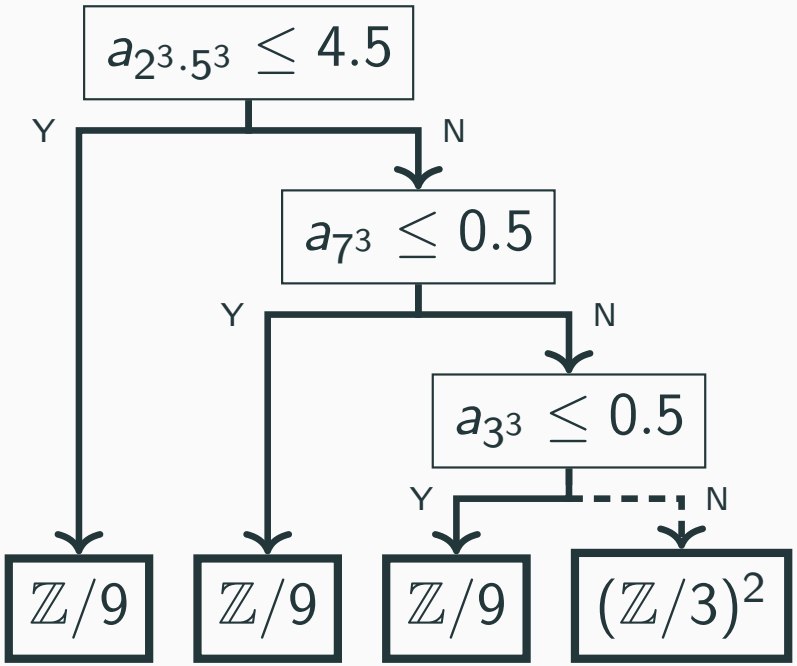
All a_n's are integers, so $a_n \leq 0.5$ is equivalent to $a_n = 0$. From the two nodes below, it is natural to conjecture that[7]

Theorem (Lee[2])

Let ℓ be a prime and $K/\mathbb{Q}$ be a degree ℓ^2 Galois extension. Then $a_{p^\ell}(K) = 0$ for some prime p if and only if $\mathrm{Gal}(K/\mathbb{Q}) \simeq \mathbb{Z}/\ell^2$.

[7] Lee–Lee, *Machines Learn Number fields, But How? The Case of Galois Groups*, 2025

Provable prediction



The bold paths *always gives a correct prediction*!

Proof

Proof

Exercise (for the end of the semester)! Use Euler factorization of Dedekind zeta function, decomposition/inertia group, etc.

Other degrees

We did it for degrees 4, 6, 8, 9, 10. Nonabelian ones are more interesting. For example, for degree 8, we have

Other degrees

We did it for degrees 4, 6, 8, 9, 10. Nonabelian ones are more interesting. For example, for degree 8, we have

Theorem (Lee[2])

Let $K/\mathbb{Q}$ be an octic Galois extension.

1. *If $a_{p^4}(K) = 0$, then $\mathrm{Gal}(K/\mathbb{Q})$ is a C_8-extension (hence abelian).*
2. *For $p \equiv 1 \pmod 4$, if $a_{p^4}(K) = 1$ or $a_{p^2}(K) = 1$, then $\mathrm{Gal}(K/\mathbb{Q})$ is abelian.*
3. *If $p \equiv 3 \pmod 4$, $a_{p^4}(K) = 1$, and $a_{p^2}(K) > 0$, then $\mathrm{Gal}(K/\mathbb{Q})$ is nonabelian.*

Other degrees

We did it for degrees 4, 6, 8, 9, 10. Nonabelian ones are more interesting. For example, for degree 8, we have

Theorem (Lee[2])

Let $K/\mathbb{Q}$ be an octic Galois extension.

1. *If $a_{p^4}(K) = 0$, then $\mathrm{Gal}(K/\mathbb{Q})$ is a C_8-extension (hence abelian).*
2. *For $p \equiv 1 \pmod 4$, if $a_{p^4}(K) = 1$ or $a_{p^2}(K) = 1$, then $\mathrm{Gal}(K/\mathbb{Q})$ is abelian.*
3. *If $p \equiv 3 \pmod 4$, $a_{p^4}(K) = 1$, and $a_{p^2}(K) > 0$, then $\mathrm{Gal}(K/\mathbb{Q})$ is nonabelian.*

And the decision tree uses this as a part of its prediction logic!

Review Quiz

$$K = \mathbb{Q}[x]/(x^8 - x^7 + x^5 - x^4 + x^3 - x + 1) = \mathbb{Q}(\zeta_{15})$$

One can compute (using 'zeta_coefficients()' in SageMath) that $a_{5^2} = a_{5^4} = 1$, hence $K/\mathbb{Q}$ is abelian.

Review Quiz

$$K = \mathbb{Q}[x]/(x^8 - x^7 + x^5 - x^4 + x^3 - x + 1) = \mathbb{Q}(\zeta_{15})$$

One can compute (using 'zeta_coefficients()' in SageMath) that $a_{5^2} = a_{5^4} = 1$, hence $K/\mathbb{Q}$ is abelian. In fact, this also shows $\mathrm{Gal}(K/\mathbb{Q}) \simeq \mathbb{Z}/4 \times \mathbb{Z}/2$, since $a_{p^2} = a_{p^4} = 1$ can only happen when p is totally ramified in K.

General workflow

1. Problem setup: Define target (hard to compute) and feature (easy to compute) of ML model.
2. Experiments: Start from small and simpler algorithms. If they don't work well, try larger and more complex models.
3. Interpret: If the model works much better than you expected, there should be something. Analyze the models.
4. Math: Make a conjecture from your observation, and try to prove it.

General workflow

Ideal scenarios:

- ML model works so well and is easy to interpret.
- Find something new and prove a (well-known) conjecture.
- Design a new algorithm to compute the target that is more efficient than existing algorithms.
- Find rare examples (e.g. elliptic curve of rank ≥ 30?). Probably use Reinforcement Learning.

If model works so bad

If model works so bad

...then it is also good! If you tried several models and still get a poor performance *close to random guess*, then it suggests an *equidistribution* property of the target with respect to the feature.

Unit rank of quadratic fields

For example, when you try to predict the rank of the unit group $U_K = \mathcal{O}_K^\times$ of a quadratic field $K = \mathbb{Q}(\sqrt{d})$ from the Dedekind zeta coefficients $a_n(K)$, you may get about 50% accuracy *for any model*.

Unit rank of quadratic fields

For example, when you try to predict the rank of the unit group $U_K = \mathcal{O}_K^\times$ of a quadratic field $K = \mathbb{Q}(\sqrt{d})$ from the Dedekind zeta coefficients $a_n(K)$, you may get about 50% accuracy *for any model*. This suggests that the distribution

$$\mathbb{P}[a_n(K) = a \mid \mathrm{rank}(U_K) = r]$$

only depends on n and $a \in \{0, 1, 2\}$, not on $r \in \{0, 1\}$.

Possible projects

- Read the works where machine learning is used to study number theoretic objects[8]. If they only report performance but no interpretation, try to interpret the model.
- Think about your favorite number theoretic object. Can you learn any invariants of it from other invariants, using machine learning? Does LMFDB have data on it?
- There are some possible follow-up works for Lee[2], e.g. non-Galois extensions (both with or without ML).

[8] https://seewoo5.github.io/awesome-ai-for-math/