Machine learning L-functions

Edgar Costa (MIT)

October 28, 2024, Mathematics and Machine Learning Program

Slides available at edgarcosta.org

Joint work with Joanna Biere, Giorgi Butbaia, Alyson Deines, Kyu-Hwan Lee, David Lowry-Duda, Tom Oliver, Tamara Veenstra, and Yidi Qi.

Riemann zeta function: the prototypical L-function

$$\zeta(s = x + iy) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots = \sum_{n=1}^{+\infty} \frac{1}{n^s}$$

$$= \left(1 - \frac{1}{2^s}\right)^{-1} \left(1 - \frac{1}{3^s}\right)^{-1} \cdots = \prod_{p \text{ is prime}} \frac{1}{1 - p^{-s}}$$

Riemann zeta function: the prototypical L-function

$$\zeta(s = x + iy) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots = \sum_{n=1}^{+\infty} \frac{1}{n^s}$$
$$= \left(1 - \frac{1}{2^s}\right)^{-1} \left(1 - \frac{1}{3^s}\right)^{-1} \cdots = \prod_{p \text{ is prime}} \frac{1}{1 - p^{-s}}$$

Used by Chebyshev to study the distribution of primes.

The formula above works for $x > 1$, e.g., $\zeta(2) = \sum_{n\geq 1} \frac{1}{n^2} = \pi^2/6$.

Riemann was the first to consider it as a complex function and showed it has meromorphic continuation to $\mathbb{C}$.

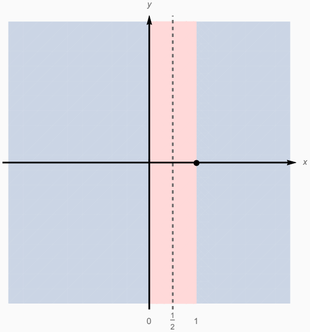
Riemann zeta function functional equation

$$\zeta(s = x + iy) = \sum_{n=1}^{+\infty} \frac{1}{n^s} = \prod_{p \text{ is prime}} \frac{1}{1 - p^{-s}}, \qquad \Re(s) > 1$$

Functional equation relates $s \leftrightarrow 1 - s$

$$\zeta(s) = \Gamma_\zeta(s)\zeta(1-s)$$

Riemann showed $\zeta(s) = 0 \Leftrightarrow \begin{cases} s = -2n \; n \in \mathbb{N} \\ 0 < \Re(s) < 1 \end{cases}$



Riemann hypothesis

$\zeta(s) = 0$ and $0 < \Re(s) < 1 \Longrightarrow \Re(s) = 1/2$

One of the Millennium Prize Problems.

The roots $\zeta(s)$ describe the distribution of the primes.

Riemann zeta function is an L-function

L-functions have certain properties

- Dirichlet series

$$L(s) = \sum_{n \geq 1} a_n n^{-s} \text{ where } a_{nm} = a_n a_m \text{ if } gcd(n,m) = 1$$

  Enough to know a_{p^n} to deduce the rest, where p is a prime number.

Riemann zeta function is an L-function

L-functions have certain properties

- Dirichlet series

$$L(s) = \sum_{n\geq 1} a_n n^{-s} \text{ where } a_{nm} = a_n a_m \text{ if } gcd(n,m) = 1$$

  Enough to know a_{p^n} to deduce the rest, where p is a prime number.
- Functional equation

$$\Lambda(s) := N^{s/2}\Gamma_L(s) \cdot L(s) = \varepsilon\overline{\Lambda}((1+w)-s),$$

where:

- $\Gamma_L(s)$ are defined in terms of Γ-function.
- $\varepsilon \in \{z \in \mathbb{C} : |z| = 1\}$ is the root number (for our examples today $\varepsilon = \pm 1$)
- N is the conductor of $L(s)$,
- $w \in \mathbb{N}$ is the (motivic) weight of $L(s)$.

L-functions: What do they know? Do they know things? Let's find out?

L-functions can arise from many sources, and we have a database about them:
`www.lmfdb.org`: The L-functions and Modular Forms Database

These can be seen as good hash functions for several number theory objects.

L-functions: What do they know? Do they know things? Let's find out?

L-functions can arise from many sources, and we have a database about them:
www.lmfdb.org: The L-functions and Modular Forms Database

These can be seen as good hash functions for several number theory objects.

They contain a lot of arithmetic information about their sources.

- Class number formula for a number field K:
$$\lim_{s\to 1}(s-1)L(K,s) = \frac{2^{r_1}\cdot(2\pi)^{r_2}\cdot \mathrm{Reg}_K \cdot h_K}{w_K\cdot\sqrt{|D_K|}}$$
- Birch and Swinnerton-Dyer conjecture for an elliptic curve E:
$$L(E,s) \text{ vanishes to order } r := \operatorname{rank} E \text{ and}$$
$$\frac{L^{(r)}(E,1)}{r!} = \frac{\#\mathrm{Sha}(E)\cdot\Omega_E\cdot\mathrm{Reg}_E\cdot\prod_p c_p}{(\#E_{\mathrm{tor}})^2}$$

L-functions: What do they know? Do they know things? Let's find out?

Can we harvest this arithmetic information about their sources from an approximation?

$$L(s) = \sum_{n \geq 1} a_n n^{-s}$$

Question

How many a_n does one need to extract this information?

L-functions: What do they know? Do they know things? Let's find out?

Can we harvest this arithmetic information about their sources from an approximation?

$$L(s) = \sum_{n \geq 1} a_n n^{-s}$$

Question

How many a_n does one need to extract this information?

It is sufficient a_n for $n \leq O(\sqrt{N})$, for a fixed family of L-functions.

L-functions: What do they know? Do they know things? Let's find out?

Can we harvest this arithmetic information about their sources from an approximation?

$$L(s) = \sum_{n \geq 1} a_n n^{-s}$$

Question

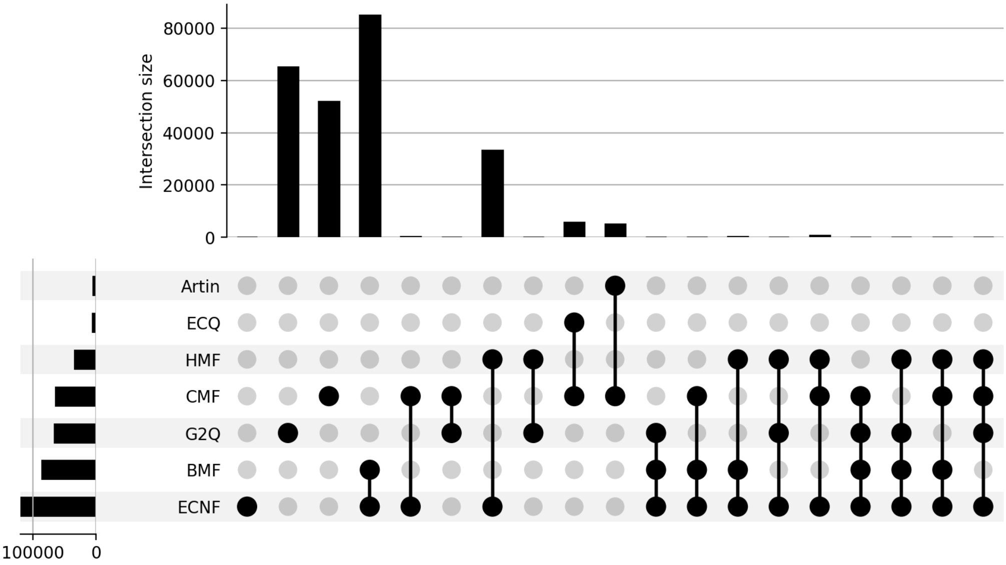
How many a_n does one need to extract this information?

It is sufficient a_n for $n \leq O(\sqrt{N})$, for a fixed family of L-functions.

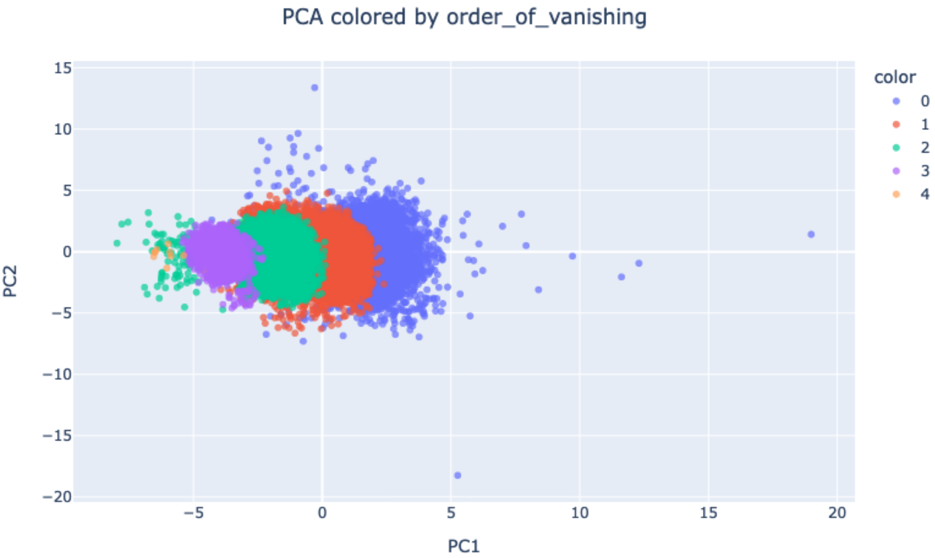
Can one do with less?
Several groups have investigated this question with partial success!

In our first experiment, we set out to investigate this question agnostic of the source, with a focus on the order of vanishing, for rational L-functions, i.e., $a_n \in \mathbb{Q}$.

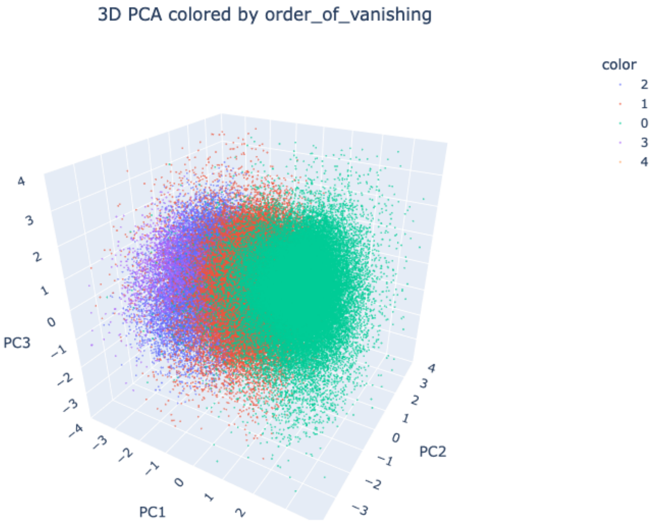
We looked at about 250k rational L-functions of small arithmetic complexity



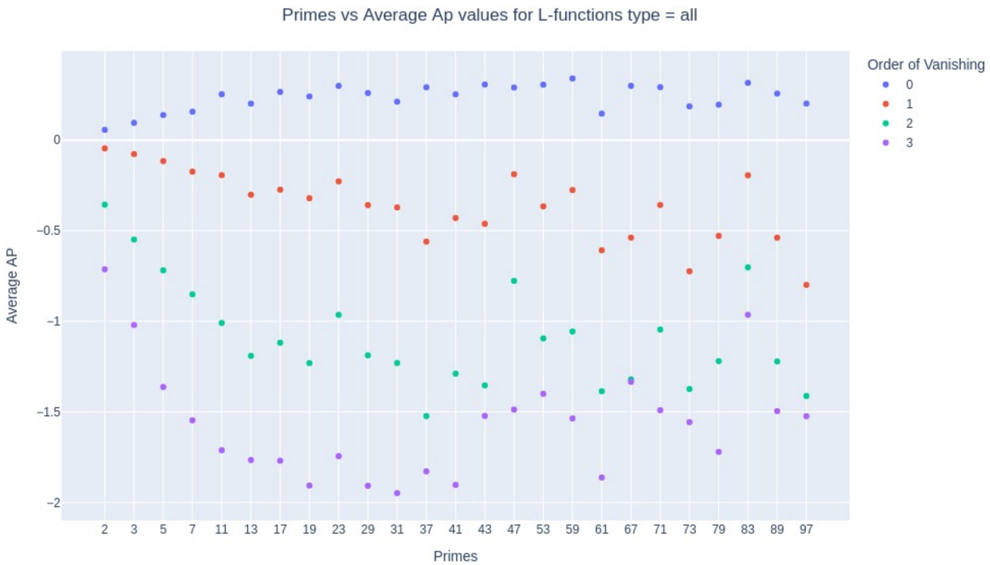
We did some principal component analysis



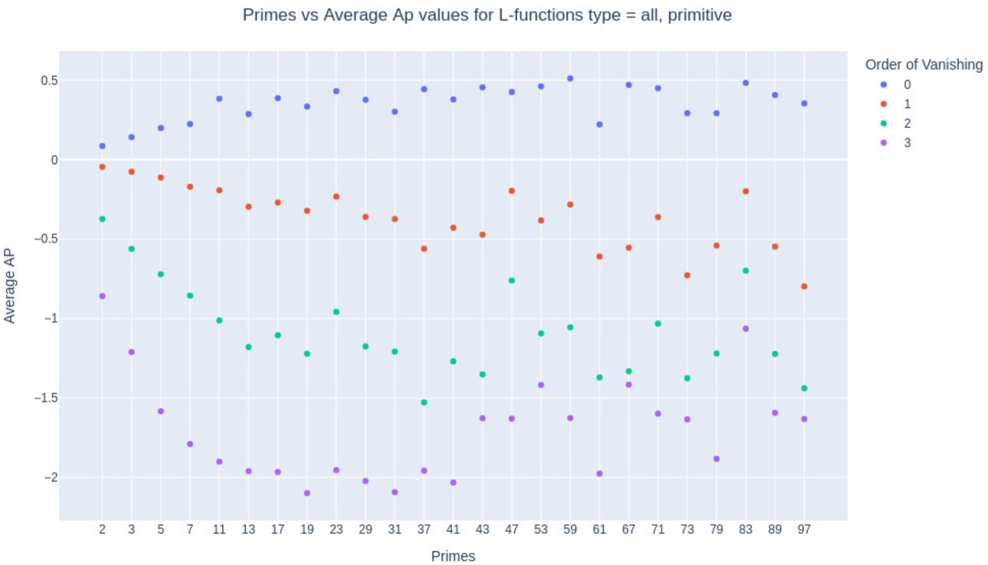
We did some principal component analysis



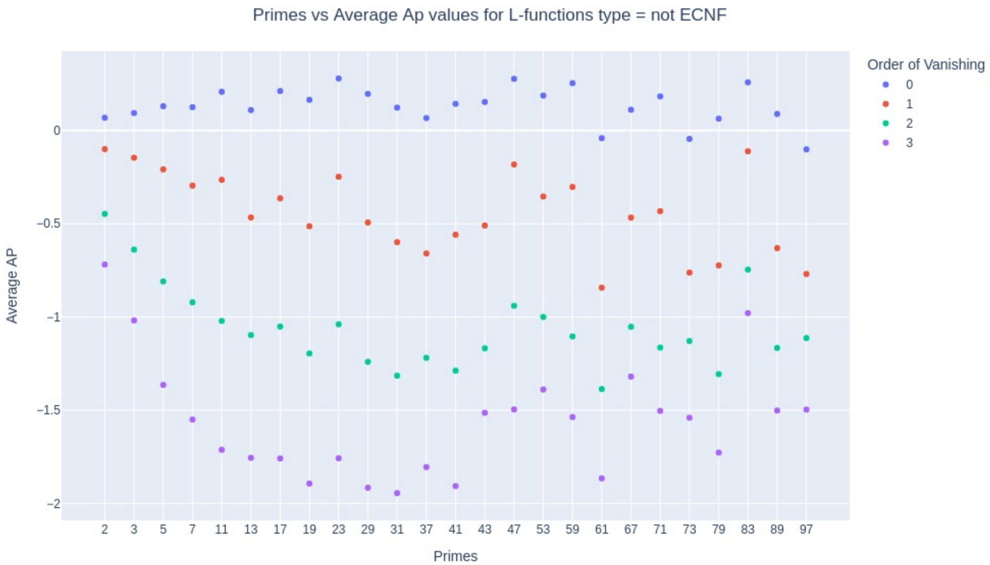
Looked at averaged a_p



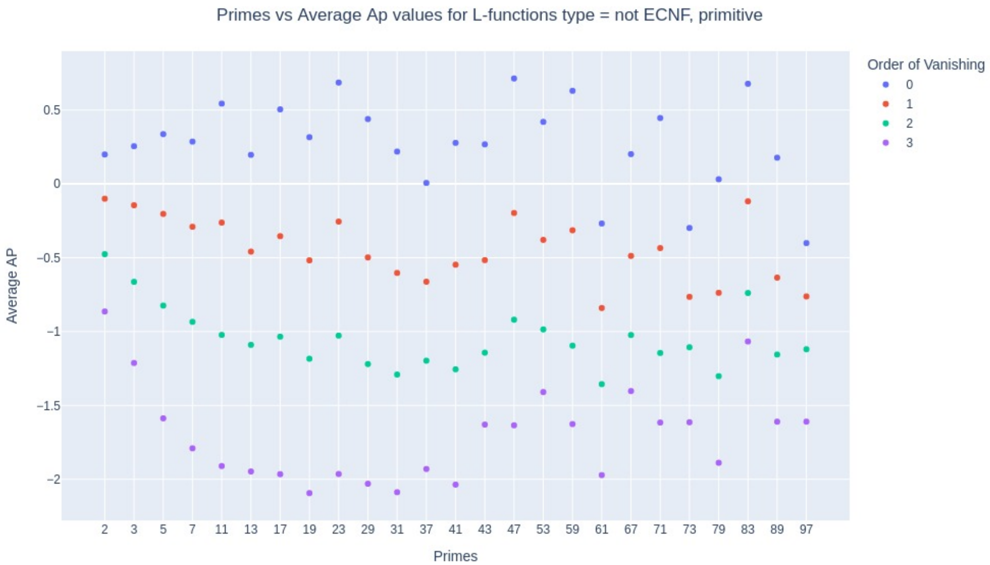
🧐 Looked at averaged a_p, restricted to primitive L-functions



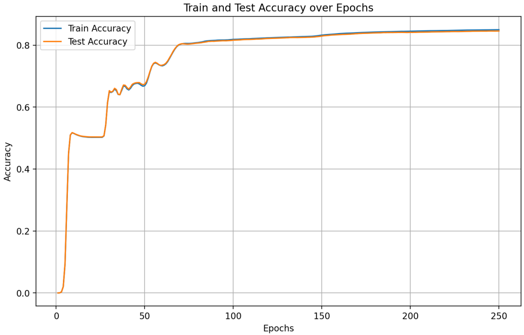
🤨 Looked at averaged a_p, excluding the largest source

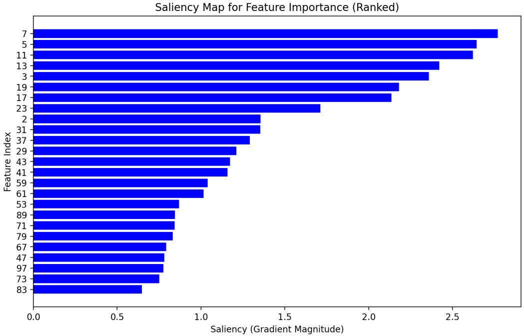


🫢 Looked at averaged a_p, excluding the largest source and primitive

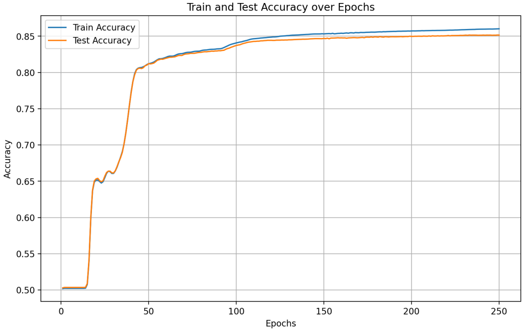


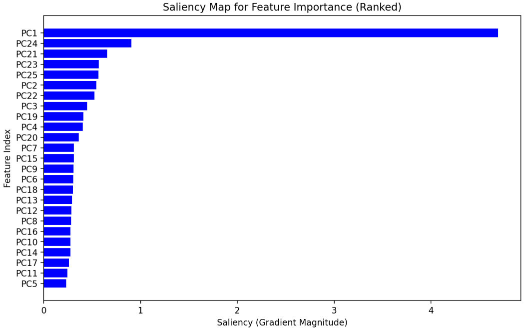
Training order of vanishing via a_p's



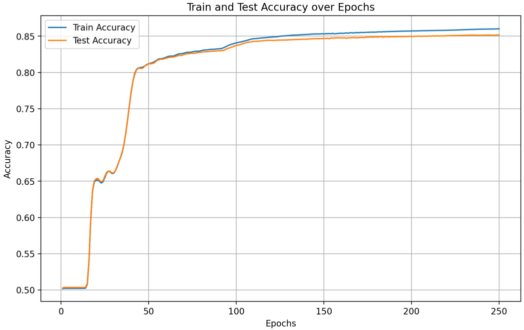


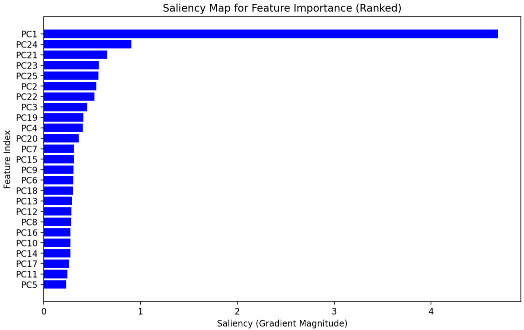
🧐 Training order of vanishing via PCA





🧐 Training order of vanishing via PCA





Indeed, training just with the first principle component retains much accuracy.

Upshot

- Rational L-functions as a dataset seem to be agnostic to their source, when normalized accordingly.
- Techniques employed for specific classes of L-function should generalize.
    - Linear discriminant analysis gives a good predictors for the order of vanishing.
    - First principle component strongly contributes to training accuracy.
- The data set is quite skewed, so all this should be taken with a grain of salt.

Upshot

- Rational L-functions as a dataset seem to be agnostic to their source, when normalized accordingly.
- Techniques employed for specific classes of L-function should generalize.
  - Linear discriminant analysis gives a good predictors for the order of vanishing.
  - First principle component strongly contributes to training accuracy.
- The data set is quite skewed, so all this should be taken with a grain of salt.

Can we put this theory to the test?

How do these tools perform for non-rational L-functions?

L-functions associated to Maass forms

Maass forms are very similar to classical modular forms.

	Classical modular form	Maass form
Domain	$\{z \in \mathbb{C} : \Im(z) > 0\}$	
Symmetry group	$\subset \mathrm{GL}_2(\mathbb{Z})$	$\subset \mathrm{GL}_2(\mathbb{R})$
eigenfunction for Δ		✅ $\Delta f = \lambda f$
Fourier expansion	$\sum_{n \geq 1} a_n e^{2\pi i n z}$	$\sum_{n \geq 1} a_n \phi_{n,s,\lambda}(z)$
a_n	algebraic	transcendental in general
$L(f, s)$	$\sum_{n \geq 1} a_n n^{-s}$	
Difficulty to compute L	🥱	😩

L-functions associated to Maass forms

Maass forms are very similar to classical modular forms.

	Classical modular form	Maass form
Domain	$\{z \in \mathbb{C} : \Im(z) > 0\}$	
Symmetry group	$\subset \mathrm{GL}_2(\mathbb{Z})$	$\subset \mathrm{GL}_2(\mathbb{R})$
eigenfunction for Δ		✅ $\Delta f = \lambda f$
Fourier expansion	$\sum_{n\geq 1} a_n e^{2\pi i n z}$	$\sum_{n\geq 1} a_n \phi_{n,s,\lambda}(z)$
a_n	algebraic	transcendental in general
$L(f,s)$	$\sum_{n\geq 1} a_n n^{-s}$	
Difficulty to compute L	🥱	😩

`www.lmfdb.org` recently added about 35k of these to its database.

Unfortunately, for about half of them, the data is incomplete 😒.

Maass forms: the missing data, the Fricke sign

$$f(z) = w_N f(-1/Nz), \quad w_N = \pm 1 = \prod_{p|N} w_p \text{ where } w_p \in \{\pm 1\}$$

and N is the Maass form's level (or conductor).

Furthermore, $a_p = -w_p/\sqrt{p}$, thus if w_p is unknown, then a_p is also unknown.

Maass forms: the missing data, the Fricke sign

$$f(z) = w_N f(-1/Nz), \quad w_N = \pm 1 = \prod_{p|N} w_p \text{ where } w_p \in \{\pm 1\}$$

and N is the Maass form's level (or conductor).

Furthermore, $a_p = -w_p/\sqrt{p}$, thus if w_p is unknown, then a_p is also unknown.

Knowing w_N and the symmetry type of f ($\in \{\pm 1\}$) determines the root number ε in

$$\Lambda(s) := N^{s/2}\Gamma_L(s) \cdot L(s) = \varepsilon\overline{\Lambda}(1-s).$$

Theoretically, one could compute enough other a_n for $n \leq O(\sqrt{N})$ to deduce w_p from the functional equation.

Maass forms: the missing data, the Fricke sign

$$f(z) = w_N f(-1/Nz), \quad w_N = \pm 1 = \prod_{p|N} w_p \text{ where } w_p \in \{\pm 1\}$$

and N is the Maass form's level (or conductor).

Furthermore, $a_p = -w_p/\sqrt{p}$, thus if w_p is unknown, then a_p is also unknown.

Knowing w_N and the symmetry type of f ($\in \{\pm 1\}$) determines the root number ε in

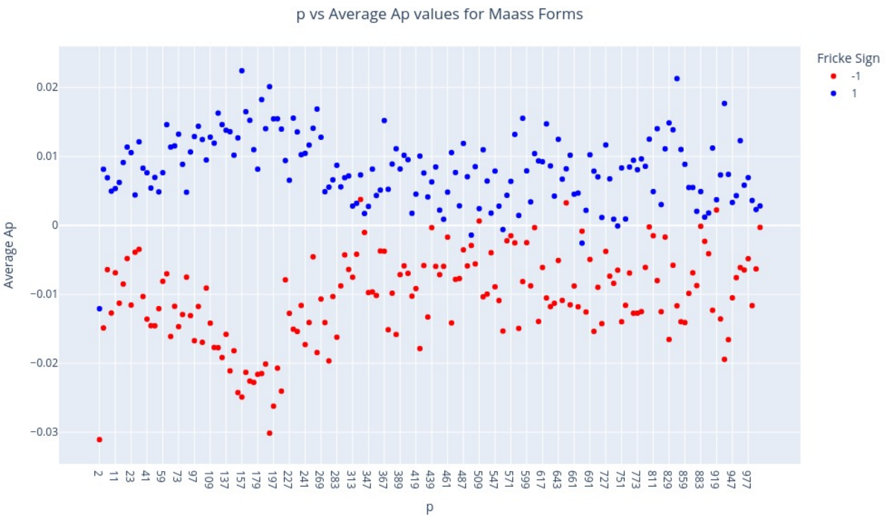
$$\Lambda(s) := N^{s/2}\Gamma_L(s) \cdot L(s) = \varepsilon\overline{\Lambda}(1-s).$$

Theoretically, one could compute enough other a_n for $n \leq O(\sqrt{N})$ to deduce w_p from the functional equation.

This is impractical; instead, we would like to guess w_N (or w_p).

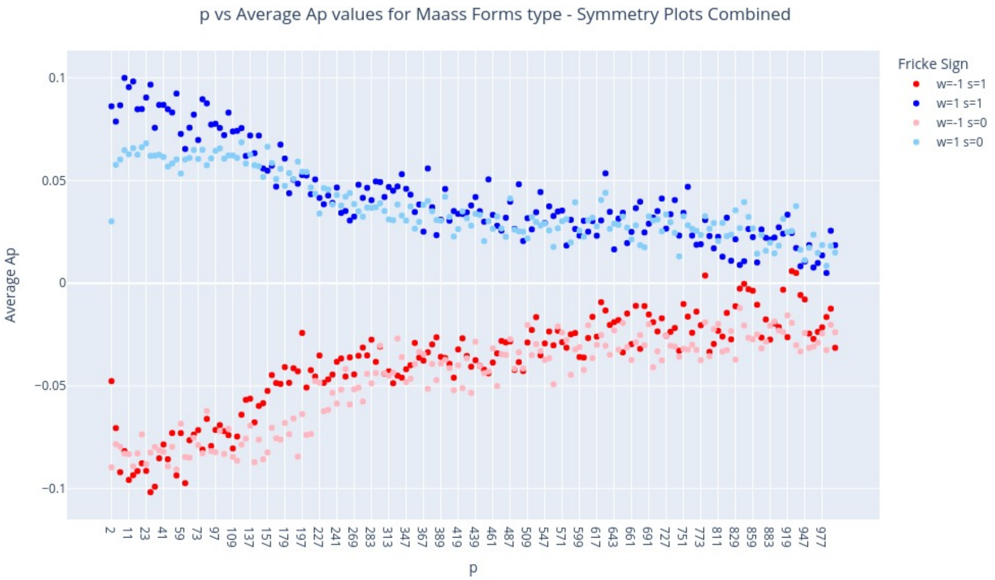
Can we predict it some other way?

😀 Averaged a_p separated by w_N = Fricke sign

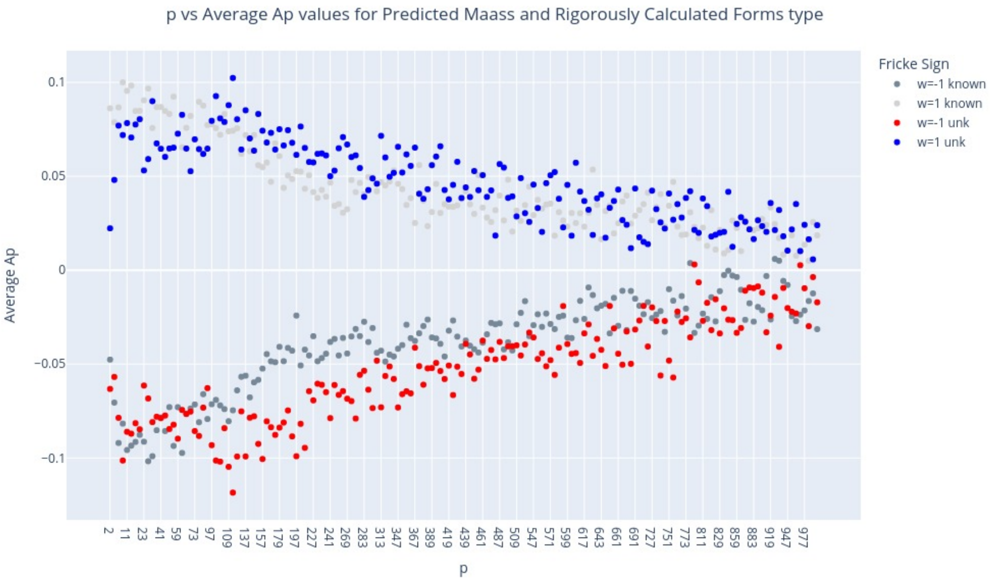


Linear Discriminant Analysis is a good candidate for a predictor.

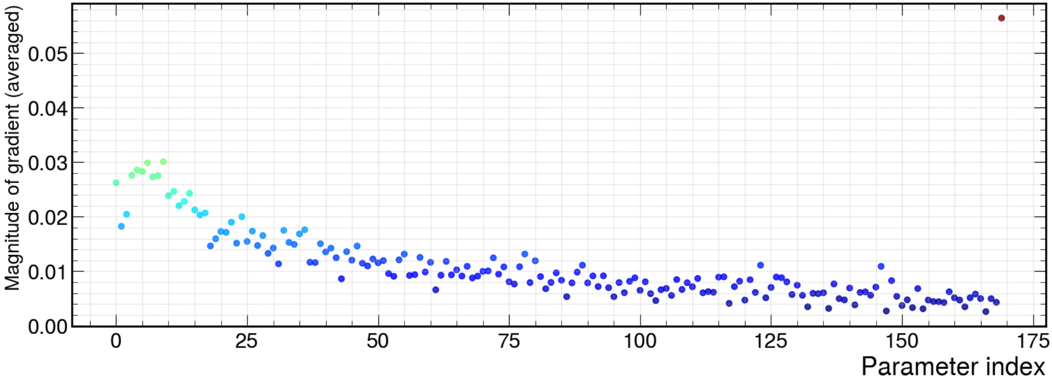
😁 Averaged $(-1)^s a_p$ separated by w_N = Fricke sign and symmetry type



🤯 Averaged $-a_p$ for odd forms, separated by (rigorous/LDA predicted) w_N



Neural networks approach: Earlier a_p and the eigenvalue λ play a bigger role



We also observed that simpler neural networks performed better.