Machine learning rational L-functions

January 4, 2025 Algebraic and Analytic Aspects of Curves and L-functions

Aly Deines (CCRLJ)
Joint work with Joanna Bieri, Giorgi Butbaia, Edgar Costa, Kyu-Hwan Lee, David Lowry-Duda, Tom Oliver, Tamara Veenstra, and Yidi Qi.

L-functions

L-functions are fundamental objects in number theory

L-functions

L-functions are fundamental objects in number theory

- Generalize the Riemann zeta function.

L-functions

L-functions are fundamental objects in number theory

- Generalize the Riemann zeta function.
- Associated with various objects in number theory.

L-functions

L-functions are fundamental objects in number theory

- Generalize the Riemann zeta function.
- Associated with various objects in number theory.
- Can study families of L-functions at once.

L-functions

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.

L-functions

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.
- d is roughly the number of these Γ-factors
- $\varepsilon \in \{z \in \mathbb{C} : |z| = 1\}$ 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 of them:

www.lmfdb.org: The L-functions and Modular Forms Database

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 of them:

www.lmfdb.org: The L-functions and Modular Forms Database

They contain arithmetic information about their number theoretic sources:

- Class number formula for a number field K:

$$\lim_{s \to 1} (s-1) L(K,s) = \frac{2^{r_1} \cdot (2\pi)^{r_2} \cdot \operatorname{Reg}_K \cdot h_K}{w_K \cdot \sqrt{|D_K|}}$$

- Birch and Swinnerton-Dyer conjecture for an elliptic curve E:

$$L(E,s) \text{ vanishes to order } r := \operatorname{rank} E \text{ and}$$

$$\frac{L^{(r)}(E,1)}{r!} = \frac{\#\mathrm{Sha}(E) \cdot \Omega_E \cdot \operatorname{Reg}_E \cdot \prod_p c_p}{(\#E_{\mathrm{tor}})^2}$$

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?

Need a_n for $n \leq O(N^{d/2})$, for a fixed family of degree d 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

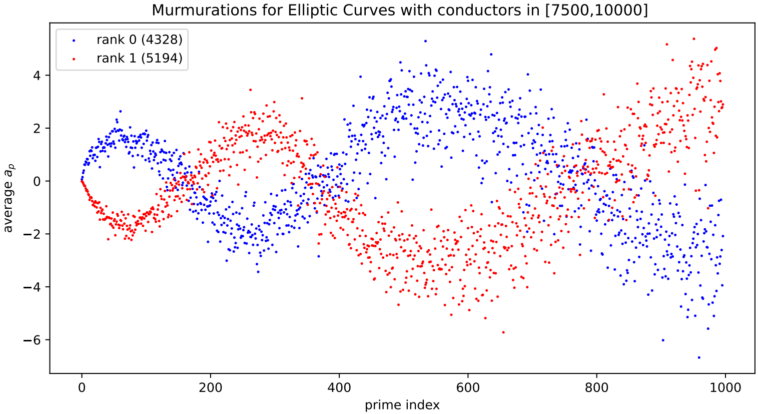
How many a_n does one need to extract this information?

Need a_n for $n \leq O(N^{d/2})$, for a fixed family of degree d L-functions.

Can one do with less?

Several groups have investigated this question with partial success!

Murmurations



Murmurations

"Unexpected, oscillating pattern observed in the statistical analysis of large families of elliptic curves."

Heuristically:

- Pattern in averages of a_p's based on rank.
- Explicit formula based on trace formulas.
- Some success predicting rank using ML.

Similar results with:

- Dirichlet characters
- Modular forms
- Maass forms

Murmurations

"Unexpected, oscillating pattern observed in the statistical analysis of large families of elliptic curves."

Heuristically:

- Pattern in averages of a_p's based on rank.
- Explicit formula based on trace formulas.
- Some success predicting rank using ML.

Similar results with:

- Dirichlet characters
- Modular forms
- Maass forms

Can we learn the order of vanishing on a set of L-functions of differing sources?

What are we even doing?

Motivating quote from David Lowry-Duda's blog:

> *Aside: model success or failure wouldn't say something conclusive about BSD or related conjectures. But in practice, ML can act like a one-sided oracle: if model performance on a particular set of features is very high, this indicates that the arithmetic information is contained within those set of features. If mathematicians don't understand why or how, then at least this can point to a place where we can look for more.*

Old School Cool: Mestre-Nagao Sums

Motivating heuristic: Mestre-Nagao sums

$$S(B) = \frac{1}{\log B} \sum_{p<B} \frac{a_p(E) \log p}{p}$$

- Have been used as a heuristic to predict the rank of elliptic curves
- Further, in "Murmurations of Mestre-Nagao sums", the authors examine an oscillatory behavior in these sums similar to murmurations

Agnostic Murmurations

In our MML 2024 workshop experiments, we saw that L-functions are somewhat agnostic of the source.

The dataset from LMFDB contains:

- 248,359 rational L-functions with root analytic conductor at most 4.
- 186,114 primitive L-functions.
- for each L-function all a_p for primes $p \leq 1000$
- https://zenodo.org/records/14774042

What does the data look like?

LMFDB

⌂ → L-functions → Search results

L-function search results

Introduction
Overview Random
Universe Knowledge

L-functions
Rational All

Modular forms
Classical Maass
Hilbert Bianchi

Varieties
Elliptic curves over $\mathbb{Q}$
Elliptic curves over $\mathbb{Q}(\alpha)$
Genus 2 curves over $\mathbb{Q}$
Higher genus families
Abelian varieties over $\mathbb{F}_q$
Belyi maps

Fields
Number fields
p-adic fields

Representations
Dirichlet characters
Artin representations

Groups
Galois groups
Sato-Tate groups
Abstract groups

Database

Refine search

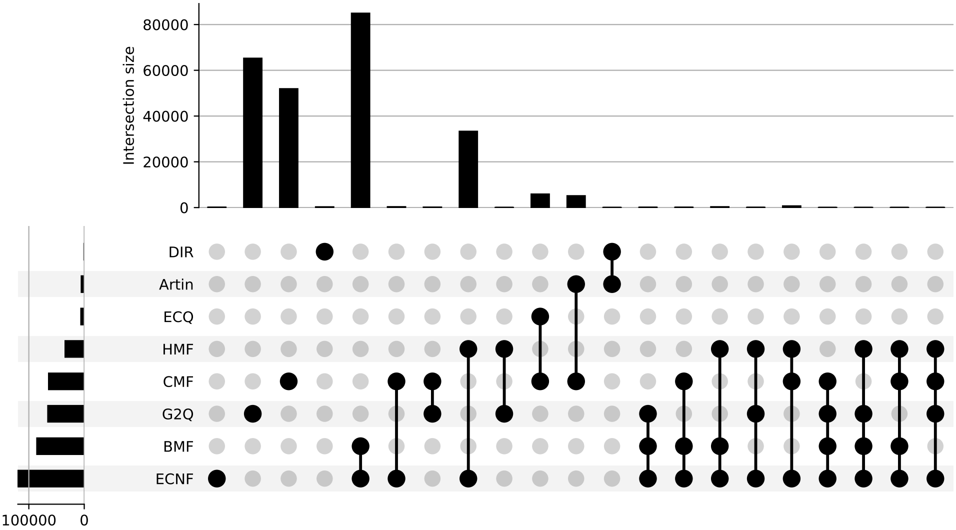
Lowest zero: 9.22237 | Conductor: 37 | Analytic conductor: 0.1-0.3 | Central character: 37.1 | Analytic rank: 2
Degree: 2 | Bad p: include, 2,3 | Root analytic conductor: 0.0-4.0 | Motivic weight: 2 | Spectral label: c1e2
Root angle: 0.5 | Primitive | Origin | Exclude origin | Self-dual
Arithmetic

Search again | Random L-function | Sort order: ▲root analytic conducto | Select: columns to display

Results (1-50 of at least 1000) Next To download results, determine the number of results.

Label	α	A	d	N	χ	μ	ν	w	prim	arith	$\mathbb{Q}$	self-dual	Arg(ϵ)	r	First zero	Origin
1-1-1.1-r0-0-0	0.00464	0.00464	1	1	1.1			0	✓	✓	✓	✓	0	0	14.1347	Artin representation 1.1.1t1.a Artin representation 1.1.1t1.a.a Character $\chi_1(1,\cdot)$ Number field $\mathbb{Q}$
1-5-5.4-r0-0-0	0.0232	0.0232	1	5	5.4	0		0	✓	✓	✓	✓	0	0	6.64845	Character $\chi_5(4,\cdot)$
1-7-7.2-r0-0-0	0.0325	0.0325	1	7	7.2	0		0	✓	✓			−0.0732	0	4.35640	Character $\chi_7(2,\cdot)$
1-7-7.4-r0-0-0	0.0325	0.0325	1	7	7.4	0		0	✓	✓			0.0732	0	6.20123	Character $\chi_7(4,\cdot)$
1-2e3-8.5-r0-0-0	0.0371	0.0371	1	2^3	8.5	0		0	✓	✓	✓	✓	0	0	4.89997	Character $\chi_8(5,\cdot)$
1-3e2-9.4-r0-0-0	0.0417	0.0417	1	3^2	9.4	0		0	✓	✓			0.111	0	5.31957	Character $\chi_9(4,\cdot)$
1-3e2-9.7-r0-0-0	0.0417	0.0417	1	3^2	9.7	0		0	✓	✓			−0.111	0	3.44409	Character $\chi_9(7,\cdot)$
1-11-11.3-r0-0-0	0.0510	0.0510	1	11	11.3	0		0	✓	✓			−0.103	0	3.61004	Character $\chi_{11}(3,\cdot)$
1-11-11.4-r0-0-0	0.0510	0.0510	1	11	11.4	0		0	✓	✓			0.103	0	5.13369	Character $\chi_{11}(4,\cdot)$
1-11-11.5-r0-0-0	0.0510	0.0510	1	11	11.5	0		0	✓	✓			0.142	0	4.62935	Character $\chi_{11}(5,\cdot)$
1-11-11.9-r0-0-0	0.0510	0.0510	1	11	11.9	0		0	✓	✓			−0.142	0	2.69600	Character $\chi_{11}(9,\cdot)$
1-12-12.11-r0-0-0	0.0557	0.0557	1	$2^2\cdot 3$	12.11	0		0	✓	✓	✓	✓	0	0	3.80462	Character $\chi_{12}(11,\cdot)$
1-13-13.10-r0-0-0	0.0603	0.0603	1	13	13.10	0		0	✓	✓			−0.0853	0	3.66097	Character $\chi_{13}(10,\cdot)$

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



Data Normalizations

For each rational L-function $L(s) = \sum_{n\geq 1} a_n n^{-s}$ our dataset includes the 168 a_p for p a primes less than 1000.

We use two normalizations:

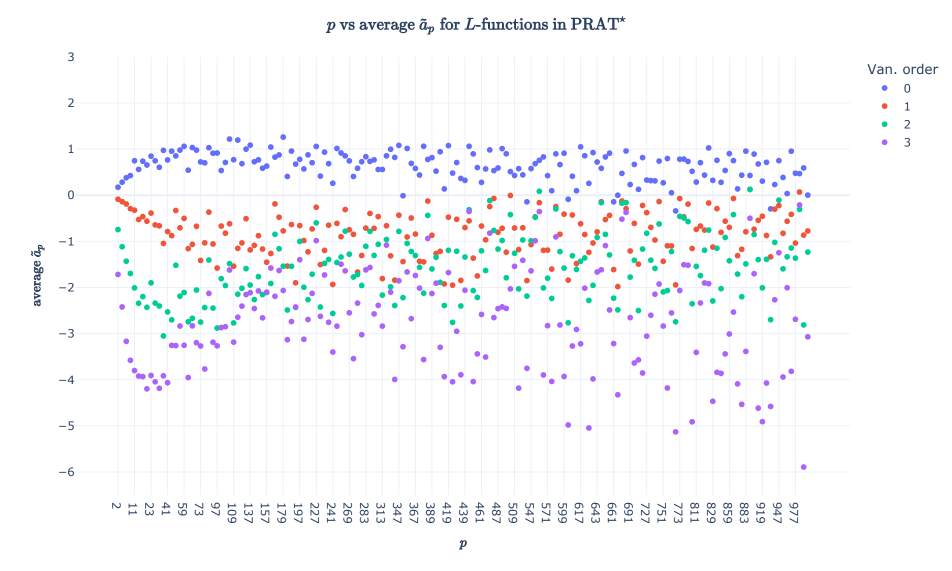
Murmurations:

$$\widetilde{a_p} = \frac{a_p}{p^{(w-1)/2}}$$

Machine Learning:

$$\overline{a_p} = \frac{a_p}{dp^{w/2}}$$

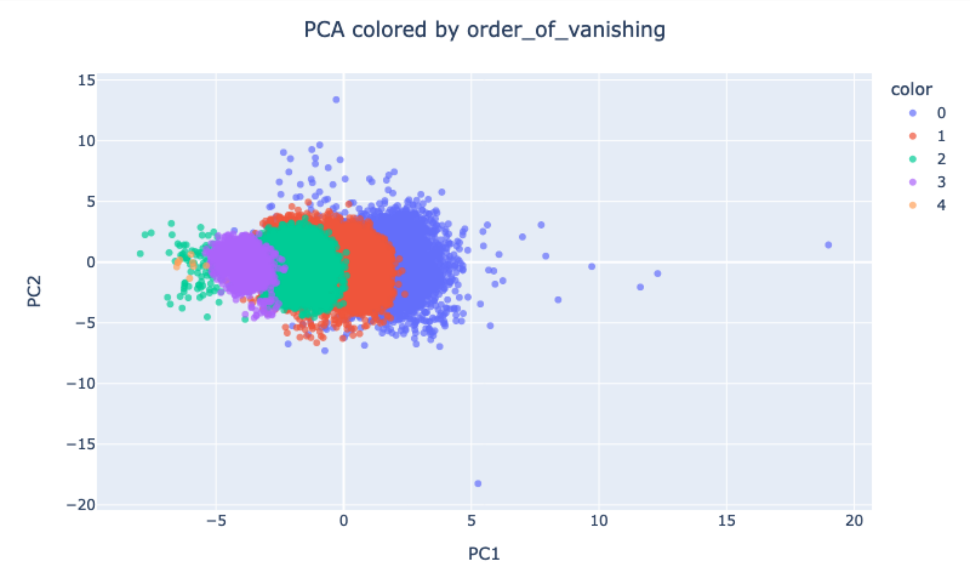
Agnostic Murmurations



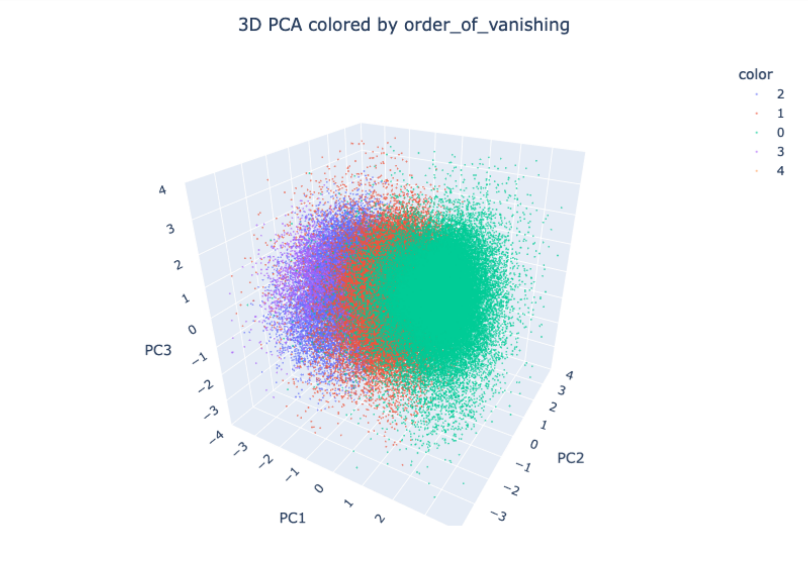
What ML algorithms to use?

- Principal Component Analysis (cite ??)
- Neural Nets (cite ??)
- Linear Discriminant Analysis

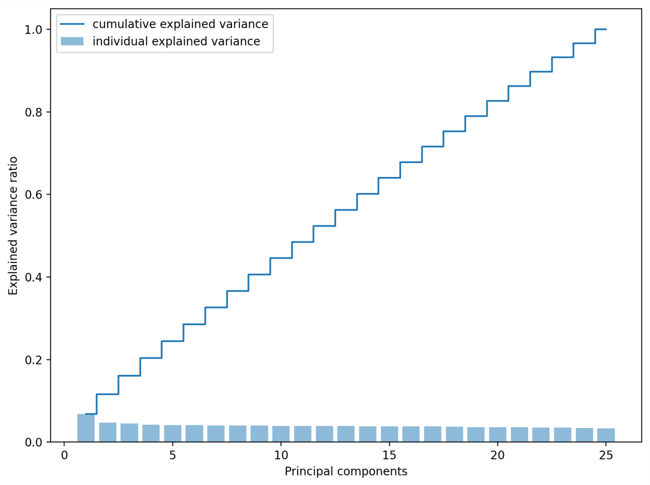
We did some principal component analysis



We did some principal component analysis

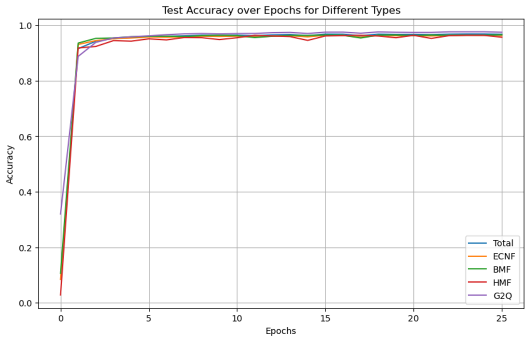


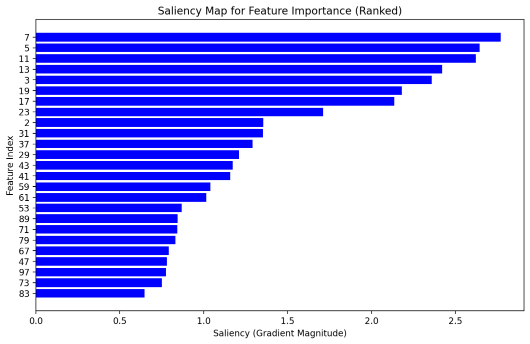
We did some principal component analysis



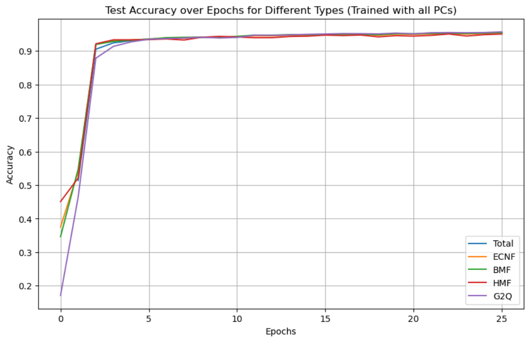
Cumulative explained variance for PCA

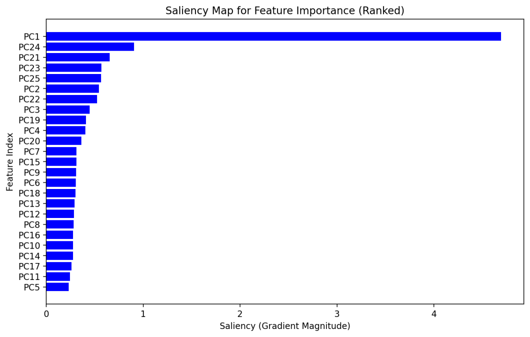
Neural Nets: Training order of vanishing via a_p's



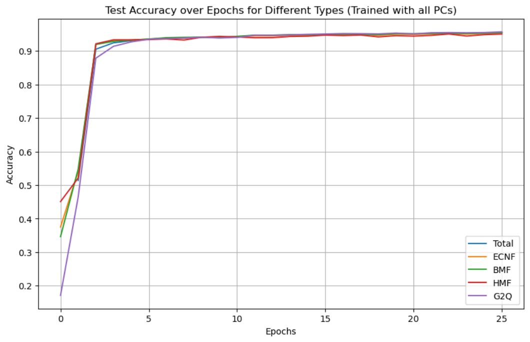


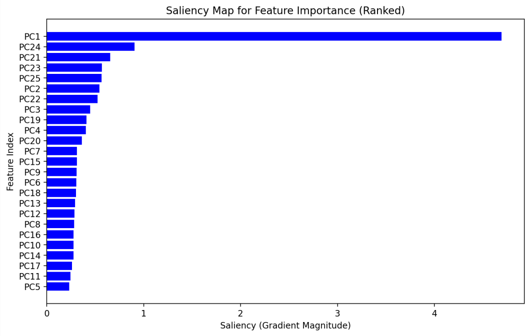
Neural Nets: Training order of vanishing via PCA





Neural Nets: Training order of vanishing via PCA



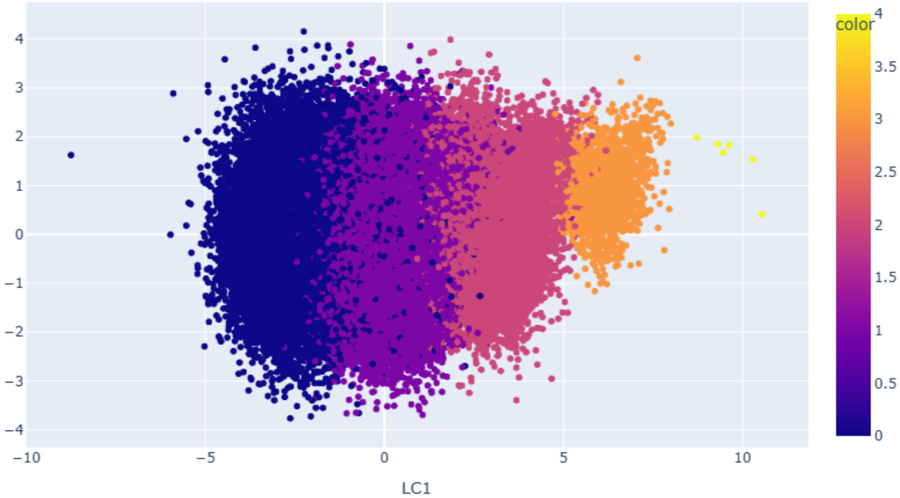


Indeed, training just with the first principle component retains much accuracy.

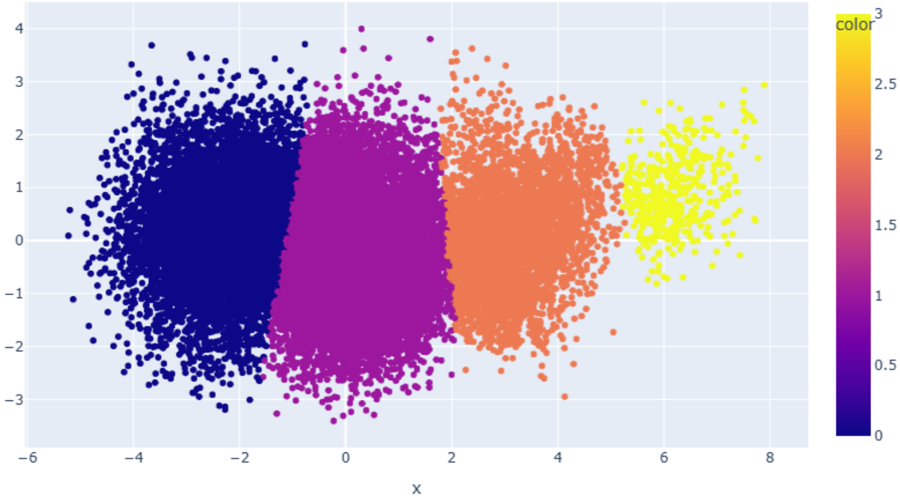
Neural Nets: Even just two components does a lot!

Data(sub)set	PCA 2 comps. accuracy	a_p accuracy
ECNF	0.9122	0.9537
BMF	0.9148	0.9548
HMF	0.9054	0.9504
G2Q	0.9113	0.9571

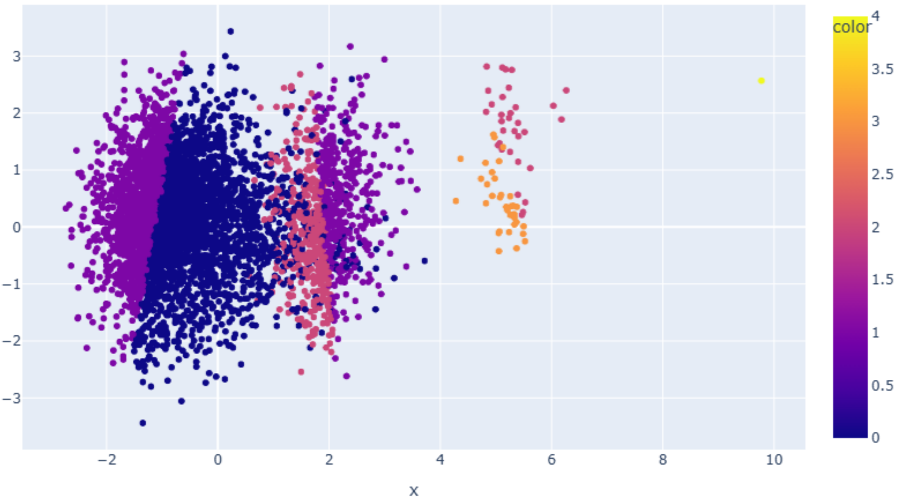
We looked at LDA



We looked at LDA - correctly classified



We looked at LDA - incorrectly classified



LDA on individual types

Dataset	Training obs.	Validation obs.	Accuracy	Explained Variance	Counts	
PRAT*	140 924	35 232	0.959	0.982	0	53 344
					1	90 327
					2	29 648
					3	2 837
BMF	65 442	16 361	0.958	0.979	0	28 280
					1	44 773
					2	8 724
					3	26
ECNF	90 791	22 698	0.956	0.983	0	42 558
					1	61 243
					2	9 661
					3	27
G2Q	50 224	12 556	0.971	0.997	0	10 827
					1	29 155
					2	19 988
					3	2 810
HMF	25 571	6 393	0.963	0.988	0	14 443
					1	16 582
					2	938
					3	1

Table 1: LDA results for predicting vanishing order in PRAT* and various subsets.

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.
  - First principle component strongly contributes to training accuracy.
  - Neural Nets perform surprisingly well on the the PCA components.
  - Linear discriminant analysis gives good predictors for the order of vanishing.
- 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.
  - First principle component strongly contributes to training accuracy.
  - Neural Nets perform surprisingly well on the the PCA components.
  - Linear discriminant analysis gives good predictors for the order of vanishing.
- The data set is quite skewed, so all this should be taken with a grain of salt.

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

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.
  - First principle component strongly contributes to training accuracy.
  - Neural Nets perform surprisingly well on the the PCA components.
  - Linear discriminant analysis gives good predictors for the order of vanishing.
- The data set is quite skewed, so all this should be taken with a grain of salt.

How do these tools perform for non-rational L-functions? See the next talk!

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.
  - First principle component strongly contributes to training accuracy.
  - Neural Nets perform surprisingly well on the the PCA components.
  - Linear discriminant analysis gives good predictors for the order of vanishing.
- The data set is quite skewed, so all this should be taken with a grain of salt.

How do these tools perform for non-rational L-functions? See the next talk!

How well does one type of L-function learn on another?

Agnostic Murmurations: Transfer Learning

How well does one type of *L*-function generalize?

What happens when we train on one type, but test on another?

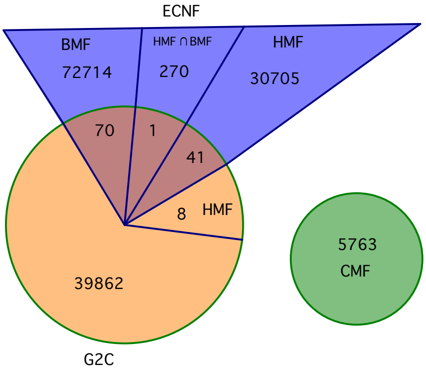
Now we restrict to:

- Primitive
- Order of vanishing 0 **and** 1
- Motivic weight 1

Which gives:

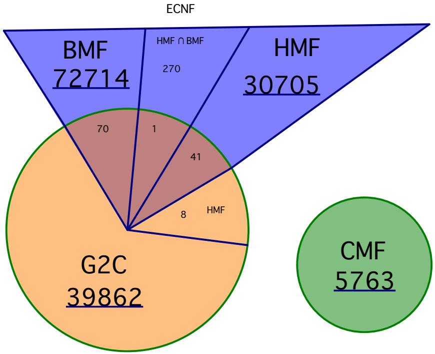
- All CMF's are ECQ's
- All CMF's have degree 2
- Everything else has degree 4

We looked at about 150k rational L-functions of small arithmetic complexity



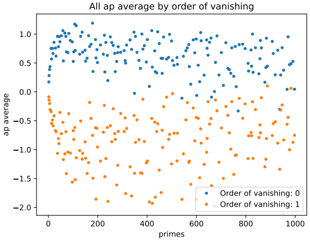
Here's the ~~weird~~ interesting L-function in the intersection ECNF, HMF, BMF, and G2C: `https://beta.lmfdb.org/L/4/2e13/1.1/c1e2/0/0`

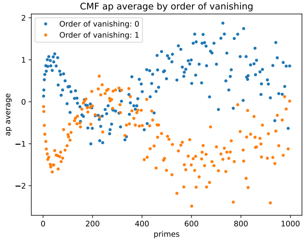
We looked at about 150k rational L-functions of small arithmetic complexity

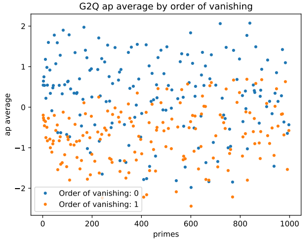


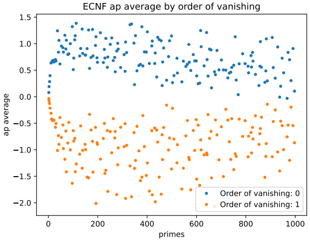
Our training and testing sets come from the four disjoint sets, that we'll refer to as BMF, HMF, G2C, and CMF.

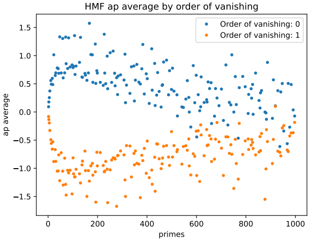
Non-agnostic murmurations?

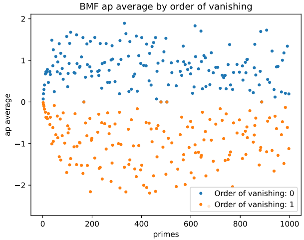




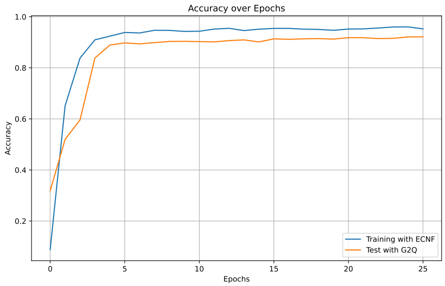


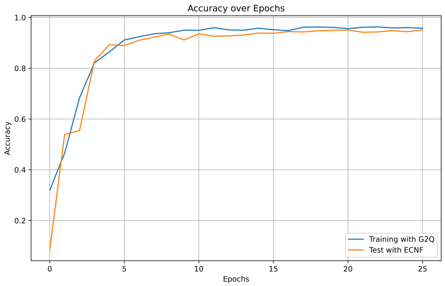






Neural Nets: Transfer learning

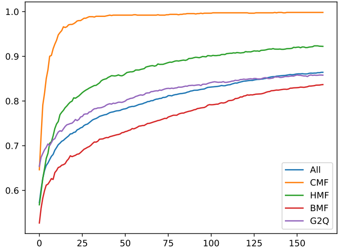




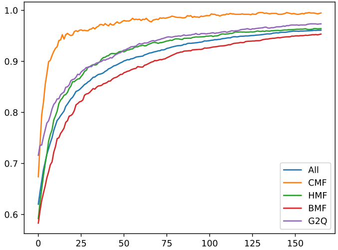
LDA Transfer learning on different sets of L-functions

How many coefficients do we actually need?

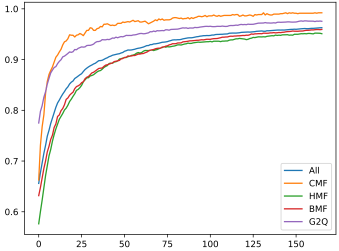
- CMF



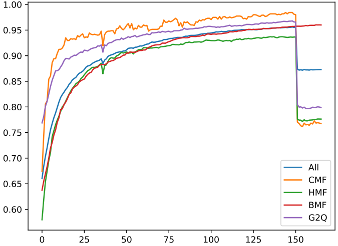
- HMF



- G2C



- BMF



Individual Primes

- BMF's do poorly - why?
- What primes do the two dips occur at? 167 and 887
- In our original dataset, we were missing two labels - just labels, not L-functions!

Individual Primes

- BMF's do poorly - why?
- What primes do the two dips occur at? 167 and 887
- In our original dataset, we were missing two labels - just labels, not L-functions!

Two L-functions were labeled as ECNF when they should have been labeled as **both** ECNF and BMF:

- L-function 4-643e2-1.1-c1e2-0-0
- L-function 4-1879e2-1.1-c1e2-0-0

Sparse columns in BMFs

There are four sparse columns in the BMF dataset (and none in the others):

- 167 - one nonzero at 4-643e2-1.1-c1e2-0-0
- 479 - one nonzero at 4-643e2-1.1-c1e2-0-0
- 503 - all zeros
- 887 - one nonzero 4-1879e2-1.1-c1e2-0-0

Note: in the old BMF dataset, the columns for the primes 167, 479, 503, and 887 were all zeros. No other subset has any columns that are all zeros!

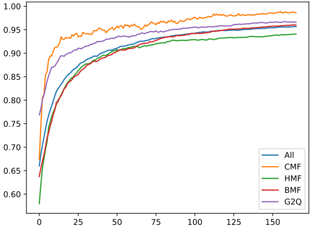
Take out the sparse primes:

Removing a single prime column

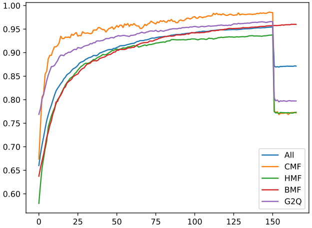
Take out the weird L-functions

Take out the weird L-functions:

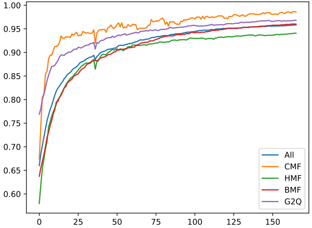
- Neither in training set



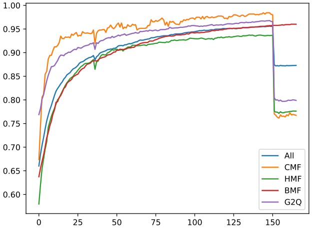
- Only the second in training set



- Only the first in training set



- Both used in training set



L-functions from elliptic curves over number fields

All our L-functions are rational of the form $L = \sum_{n \geq 1} a_n n^{-s}$, however, we can say more about those coming from ECNFs.

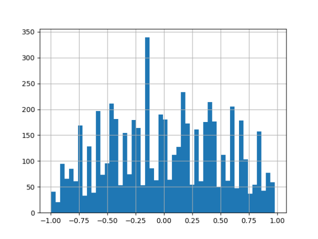
Let E/K be an elliptic curve defined over a number field K with ring of integers $\mathcal{O}_K$, then

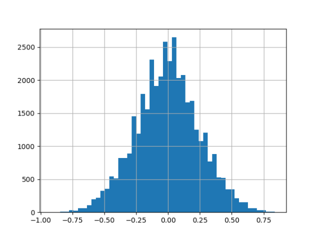
$$L(E/K, s) = \sum_{\mathfrak{n} < \mathcal{O}_K} a_{\mathfrak{n}} N_{K/\mathbb{Q}}(\mathfrak{n})^{-s}$$

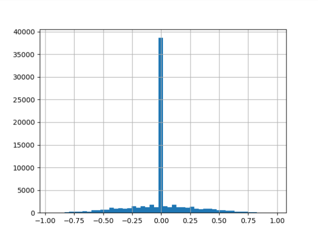
For quadratic fields, a_p depends on how p splits in K:

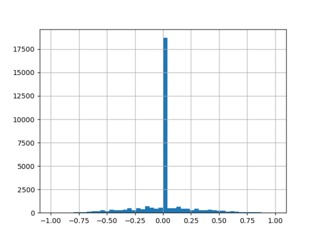
- $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$, then $a_p = a_{\mathfrak{p}_1} + a_{\mathfrak{p}_2}$
- $p\mathcal{O}_K = \mathfrak{p}$, then $a_p = 0$
- $p\mathcal{O}_K = \mathfrak{p}^2$, then $a_p = a_{\mathfrak{p}}$

You see the zeros in the distribution of a_{997}'s

- CMF 

- G2Q 

- BMF 

- HMF 

BMF Number Fields

LMFDB Number Field	Count	167	479	503	887
2.0.4.1	40275	Inert	Inert	Inert	Inert
2.0.3.1	42808	Inert	Inert	Inert	Inert
2.0.8.1	36907	Inert	Inert	Inert	Inert
2.0.7.1	28322	Inert	Inert	Inert	Inert
2.0.11.1	30608	Inert	Inert	Inert	Inert
2.0.643.1	1	Split	Split	Split	Inert
2.0.1879.1	1	Inert	Inert	Inert	Split

These are the smallest primes that are inert for all number fields in this set.

So many questions!

- Why is one of the originally mis-labled L-functions weird and the other only slightly weird?
- Why is there a slight dip at 167 but no dip at 479?
- Why does the one value at 887 tank the training set?
- Is there anything weird about 4-1879e2-1.1-c1e2-0-0?
- Is this all an artifact of LDA?
- What would happen with a more complete/larger BMF dataset?

SVM


Transfer Learning L-functions
Testing Set
CMF
HMF
BMF
G2Q
Training Set
CMF
HMF
BMF
G2Q
0.999133
0.8363459
0.737537
0.7362348
0.993062
0.9653151
0.9390085
0.9613696
0.986123
0.9474027
0.9611497
0.9686442
0.993062
0.9579873
0.9586055
0.9769221
0.95
0.9
0.85
0.8
0.75

LDA and outliers

- There is no weirdness for other types of machine learning such as linear support vector machines (good) or decision trees (bad)
- LDA is fragile against outliers, but this is still striking!
- Transfer learning does remarkably well on the different rational L-functions

Future Directions

- Raise analytic rank/root analytic conductor
- Could we just get a few more BMFs?
- Can we learn the origin of an L-function?
- Can we predict in advance what the outliers are?
- Is LDA fragile against incorrect values?
- Study in the context of Mestre-Nagao sums

References

- Machine learning rational L-functions J. Bieri, G. Butbaia, E. Costa, A. Deines, K. Lee, D. Lowry-Duda, T. Oliver, T. Veenstra, and Y. Qi. arxiv.org:2502.10360
- Murmurations of Mestre-Nagao sums Z. Bujanović, Matija Kazalicki, and L. Novak. Special Issue on DANGER (Data, Numbers, and Geometry) pp. 51-61 (2024)
- Coda and data available on Zenoda: doi.org/10.5281/zenodo.14774042
- Ranks of elliptic curves and deep neural networks M. Kazalicki, D. Vlah, arXiv:2207.06699
- Murmurations of elliptic curves, Y.-H. He, K.-H. Lee, T. Oliver, and A. Pozdnyakov, Experimental Mathematics (2024), 1–13.
- The L-functions and modular forms database, available at `https://www.lmfdb.org`,

Thank you!

Special thanks to the Mathematics and Machine Learning Program at Harvard University's Center of Mathematical Sciences and Applications where this research project started!

Any questions?

Teaser:

Can you transfer learn on non-rational L-functions?

Teaser:

Can you transfer learn on non-rational L-functions?

Teaser:

Can you transfer learn on non-rational L-functions?

Maybe a_n's are better? See next week!