Applications of Metric Space Magnitude to Machine Learning

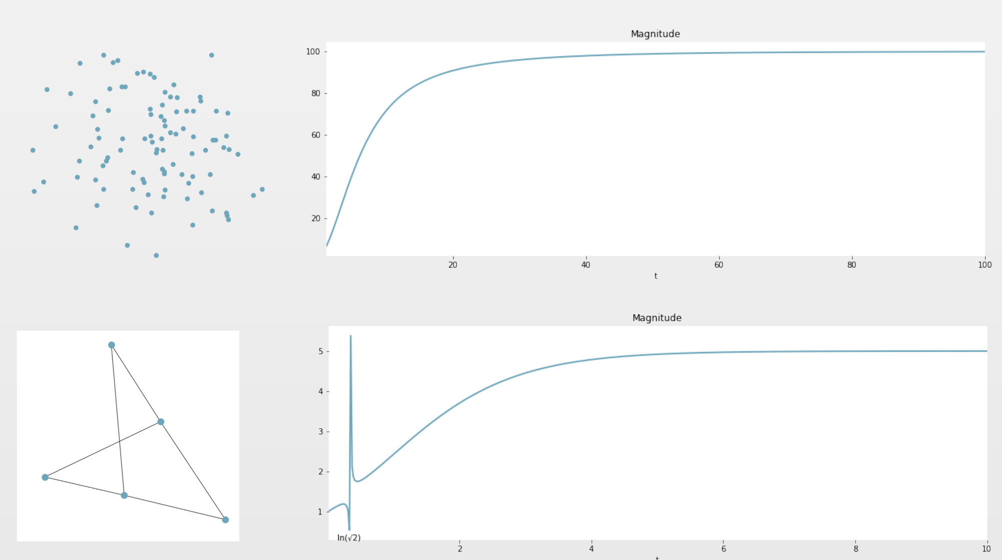
Introduction

Given a finite metric space X, denote the similarity matrix by $\zeta_X(x, y) := e^{-d(x,y)}$

The magnitude of X is defined to be

$$|X| := \sum_{x,y} \zeta_X^{-1}(x, y)$$

For $t \in (0, \infty)$ define the metric space tX to have distance given by the distance of X scaled by t. We then study $|tX|$



Behavior of points in $\mathbb{R}^n$

The magnitude of a (finite) metric space is thought of as the number of "effective points" in X. For any finite metric space, it is known that $\lim_{t\to\infty} |tX| = \#X$ It may not be the case that $|tX|$ is well-defined for all t. However, we have the following.

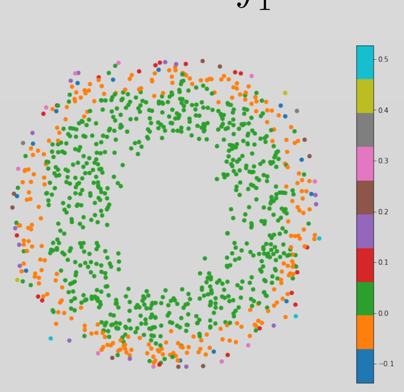
Theorem [Leinster & Meckes] If $X \subset \mathbb{R}^n$ finite, then $|tX|$ is defined for all $t \in (0, \infty)$

This allows us to define the *power of a point* $x \in X$ as

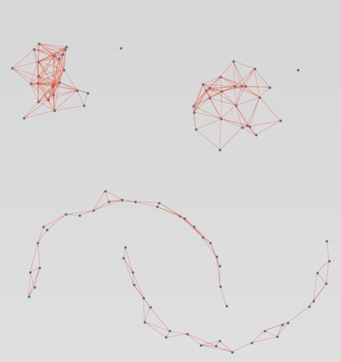
$$p_x := \int_0^\infty e^{-t} \sum_{y \in X} \zeta_{tX}^{-1}(x, y) dt$$

Similarly, we can define the *ϵ-power of an edge* between points $x, y \in X$ to be

$$E_\epsilon(x, y) := \int_1^\infty \frac{I(|\zeta_{tX}^{-1}(x, y)| > \epsilon)}{t^2} dt$$



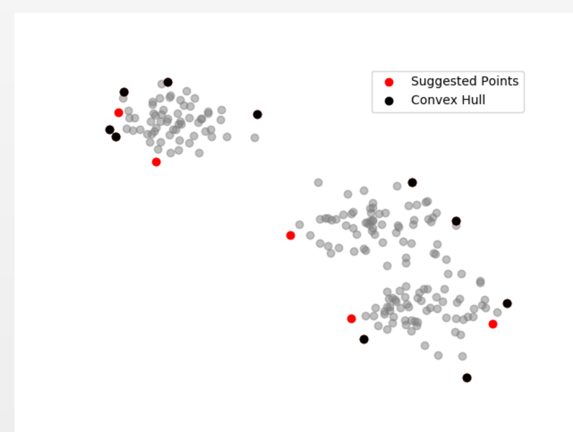
The power of points plotted for a noisy circle.



Edge between two points if power of edge is greater than a fixed constant.

Experiments

Convex Hulls



By ranking the points in X by their power we generate candidates for the convex hull of X. Each data set in these trials contains 200 points.

Dimension	Trials	Average Number of Points in Convex Hull	Average Number of Points Suggested By Point Power	Precision at 100% Recall of Convex Hull
2	20	10.6	12.85	0.8469
3	20	31.2	58.25	0.5511
4	20	65.2	115.1	0.5687
5	20	102.8	152.75	0.6748

Clustering

The ϵ-power of an edge can be seen as a similarity measure between points. We can convert this to a distance measure by

$$D_\epsilon(x, y) := -\ln(E_\epsilon(x, y))$$

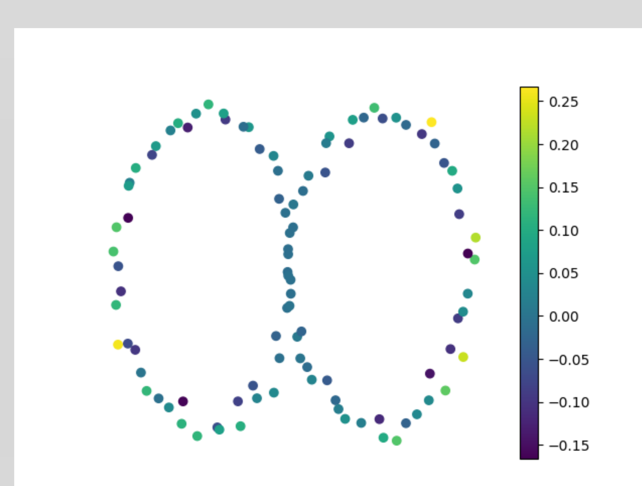
Once we have this, we can formulate clustering methods in a number of ways. We show here a clustering based on the HDBSCAN algorithm.

Persistent Homology

By using the point power of a dataset X we can formulate a downsampling technique as follows. If we wish to downsample X to 60% resolution, choose the points whose power is in the 60th percentile across all the point powers in X.

We compare this method to a random downsampling technique and compare how well the persistent homology approximates that of the full resolution data set. To do this, the bottleneck distance between the persistence diagram of the downsampled data set and the persistence diagram of the full resolution data set is computed.

For each downsample rate, 10 trials were performed and the bottleneck distance shown is the average across trials.



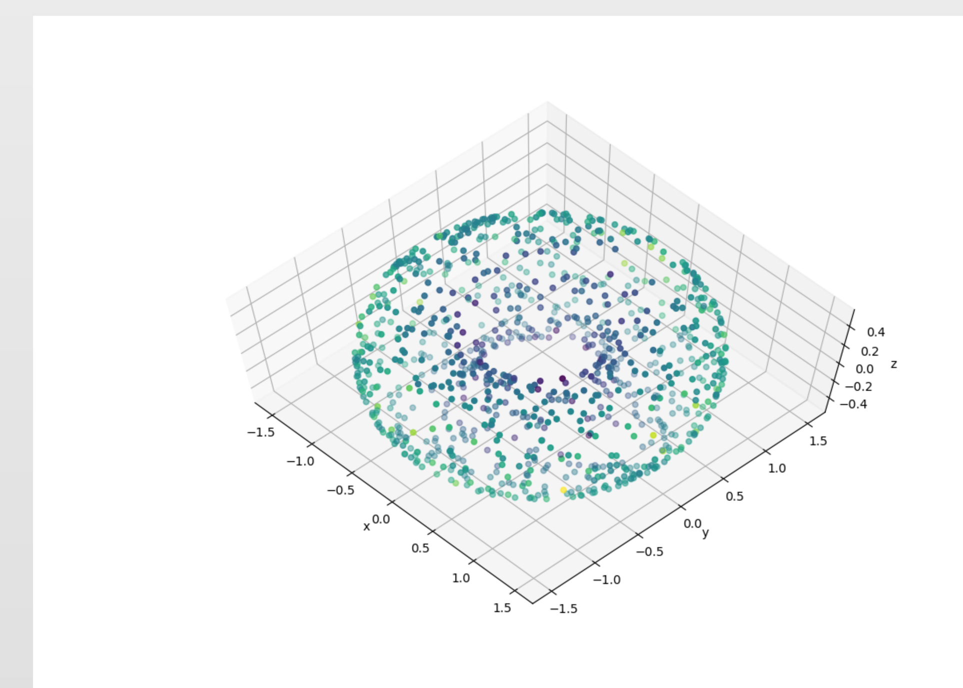
Downsample rate (%)	Bottleneck distance power	Bottleneck distance random sample
10	0.693	0.693
20	0.693	0.693
30	0.693	0.670
40	0.670	0.475
50	0.236	0.382
60	0.208	0.284
70	0.137	0.318
80	0.078	0.115
90	0.047	0.036

Conclusions

The magnitude of a metric space has a slogan of giving the "number of effective points" in the space. We give definitions of the power of a point, and power of an edge that can give more formalism to this slogan.

The power of a point allows us to order the points in X defined by an intrinsic measure. This leads to a number of interesting applications, including

- Convex hull computation
- Clustering methods
- Persistent homology computation
- Choosing candidates for active learning
- Supervised machine learning



Torus colored by point power

For further information

Please contact *ebunch@amfam.com*
https://www.ai-ml-amfam.com/