Novel geometric methods for data analysis focusing on curvature and geodesics in data space

Kei Kobayashi (Keio University)

Boston-Keio-Tsinghua 2023 2023/06/30

Contents

Part I: Data analysis using α and β-metrics

- with Henry P. Wynn (London School of Economics)

Part II: Application of metric cones to graph embedding

- with Daisuke Takehara (Accenture)

Outline of Part I. Data Analysis Using α and β-metrics

Fréchet Mean of a Euclidean Space

Given a sample $x_1, \ldots, x_n$ in a metric space (X, d), the intrinsic mean (Fréchet Mean) is the set of the minimizers of the Fréchet function:

$$\hat{\mu} \in \arg\min_{m \in X} \sum_{i=1}^{n} d(x_i, m)^2$$

For a Euclidean space, the intrinsic mean is unique since the Fréchet function $f(m) = \sum_i \|m - x_i\|^2$ is strictly convex.

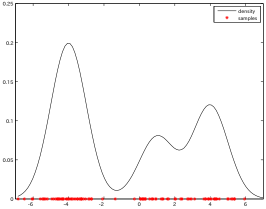
It is easy to see the intrinsic mean is equal to the sample mean $\bar{x}$.

In general, the less (or more negative) curvature, the less the number of (local) minima of the Frechét function.

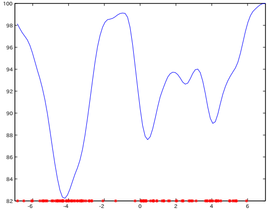
HOWEVER, this "good" behavior of the Frechét function is not always welcome.

Clustering

For a negatively curved space, Fréchet function $f(m) := \sum_i d(m, x_i)^2$ is not necessarily convex and its local minima (called the Karcher means) can be used as the center of each cluster for clustering.



density function and its samples



$f(m)$

$\rightsquigarrow$ Controlling the curvature of the data space should play an important role in data analysis.

Our strategy

- Ordinary data analysis (e.g. classification, regression):

  Data X_i $(i = 1, \ldots, n)$, Metric d

  $\longrightarrow$ Loss function $\hat{f} \in \mathcal{F}$

  (can be selected by cross validation, resampling)

  $$\longrightarrow \hat{\theta} = \arg\min \sum_i \hat{f}(d(X_i, \theta))$$

- Our approach:

  Data X_i $(i = 1, \ldots, n)$, Loss function f

  $\longrightarrow$ Metric $\hat{d} \in \mathcal{D}$

  (can be selected by cross validation, resampling)

  $$\longrightarrow \hat{\theta} = \arg\min \sum_i f(\hat{d}(X_i, \theta))$$

How to set the family $\mathcal{D}$ of metrics?

$\Longrightarrow$ by focusing on their curvature

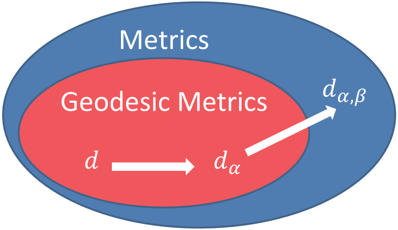
Our policy: keep the problem in geometry
as much as possible.

Two steps of changing metrics

A geodesic metric space is a metric space s.t. the distance between two points is equivalent to the shortest path length connecting them.

We can define a curvature called CAT(k) property for each geodesic metric space.

We change the metric d of the original data space, to a geodesic metric space d_α, and next to a (usually non-geodesic) metric space $d_{\alpha,\beta}$.



The α, β-metric and α, β, γ-mean

We propose a family of metrics:

$$d_{\alpha,\beta}(x, y) = g_\beta(d_\alpha(x, y))$$

and intrinsic means:

$$\hat{\mu}_{\alpha,\beta,\gamma} = \arg\min_{m \in \mathcal{M}} \sum_{i=1}^{n} d_{\alpha,\beta}(x_i, m)^\gamma$$

d_α: a locally transformed geodesic metric ($\alpha \in \mathbb{R}$), defined later
g_β: a concave function corresponding to a specific kind of extrinsic means ($\beta \in (0, \infty]$), defined later
γ: for L_γ-loss ($\gamma \geq 1$)

Data analysis by α, β and γ

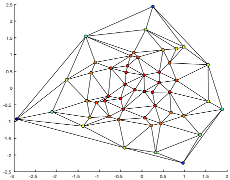
	Euclidean	$d_{\alpha,\beta}$
metrics	$d(x,y) = \|x-y\|$	$d_{\alpha\beta}(x,y) = g_\beta(d_\alpha(x,y))$
intrinsic mean	$\arg\min_{m\in\mathbb{E}^d} \sum \|x_i - m\|^2$	$\arg\min_{m\in\mathcal{M}} \sum g_\beta(d_\alpha(x_i,m))^\gamma$
Fréchet variance	$\min_{m\in\mathbb{E}^d} \frac{1}{n} \sum \|x_i - m\|^2$	$\min_{m\in\mathcal{M}} \frac{1}{n} \sum g_\beta(d_\alpha(x_i,m))^\gamma$
Fréchet function	$f(m) = \sum \|x_i - m\|^2$	$f_{\alpha\beta\gamma}(m) = \sum g_\beta(d_\alpha(x_i,m))^\gamma$

Outline of Part I

Empirical Metric Graphs

We begin from computing empirical graphs, whose vertices are the data points. For example,

1. Complete graph
2. Delaunay graph
3. k-NN graphs
4. ϵ-NN graphs



Delaunay empirical graph

We introduce a metric on the graph by the shortest path length:

$$d(x_0, x_1) := \inf_{\gamma \in \Gamma(x_0, x_1)} \sum_{e_{ij} \in \gamma} d_{ij},$$

where d_{ij} is the length of an edge e_{ij}.

The α-Metric: Empirical Graph Case

α-metric for an empirical graph is defined by the shortest path length with powered edge lengths:

$$d_\alpha(x_0, x_1) := \inf_{\gamma \in \Gamma(x_0, x_1)} \sum_{e_{ij} \in \gamma} d_{ij}^{1-\alpha}.$$

This can be seen as an empirical approximation of

$$\tilde{d}_\alpha(x_0, x_1) := \inf_{\gamma \in \Gamma(x_0, x_1)} \int_0^1 f^{\alpha p}(z(t)) d|\gamma(t)|.$$

where p is the dimension of the original data space.
Here we use a fact, under some regularity conditions, $d_{ij}^{-1/p}$ is an unbiased estimator of the local density and

$$(d_{ij}^{-1/p})^{\alpha p} d_{ij} = d_{ij}^{1-\alpha}.$$

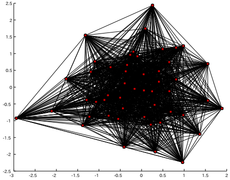
Some facts about the α-metric

- Stochastic convergence of the empirical version d_α to a continuous version $\tilde{d}_\alpha$ has been proved by Hwang, et al.(2016) under some condition.

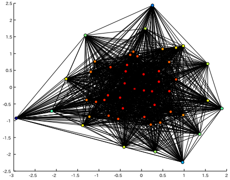
- Geodesic subgraphs computed using α-metric are a special case of "Pathfinder networks" used mainly in areas of Psychology.

Ex: Geodesic subgraphs (starting from the complete graph)

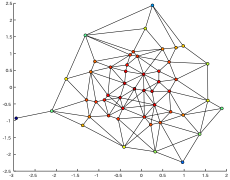
A geodesic subgraph is computed by removing edges that cannot be used in any geodesics.



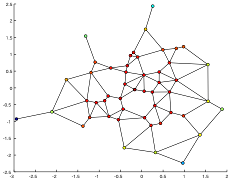
(a) $\alpha = 1$



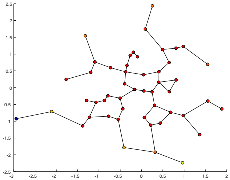
(b) $\alpha = 0$



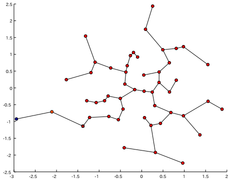
(c) $\alpha = -0.3$



(d) $\alpha = -1$



(e) $\alpha = -5$



(f) $\alpha = -30$

α and CAT(k) property

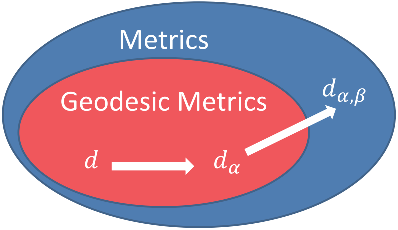
In K. and Wynn (2020), we proved the following facts under some conditions:

- The edge set of a geodesic subgraph of an α-empirical graph becomes smaller as α decreases.
- Each geodesic subgraph becomes a tree (minimum spanning tree) for sufficiently small α.
- The curvature of a geodesic subgraph decreases as α decreases in the sense of CAT(k) property and finally becomes CAT(0).

Outline of Part I

Further metric transformation by β

We next introduce a parameter β to change a geodesic metric d_α to a (not necessarily geodesic) metric $d_{\alpha,\beta}$.



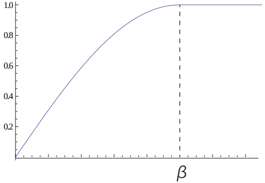
β-Metric

Let (X, d) be a geodesic metric space.
For $\beta > 0$, transform the metric d as

$$d_\beta(x_0, x_1) = g_\beta(d(x_0, x_1))$$

where

$$g_\beta(z) = \begin{cases} \sin(\frac{\pi z}{2\beta}), \ \text{for} \ \ 0 \le z \le \beta, \\ 1, \ \ \text{for} \ \ z > \beta. \end{cases}$$



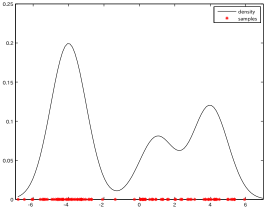
For $\beta = \infty$, $d_\beta = d$.

d_β satisfies the triangle inequality and becomes a metric but not necessarily a geodesic metric.

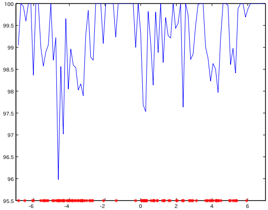
d_β-mean: $\hat{\mu}_\beta = \arg\min_{m \in X} \sum_i g_\beta(d(x_i, m))^2$.

β and clustering

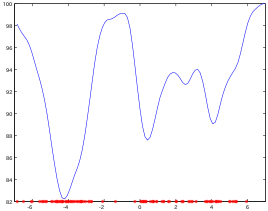
$f(m) = \sum_i g_\beta(|x_i - m|)^2$ with various β:



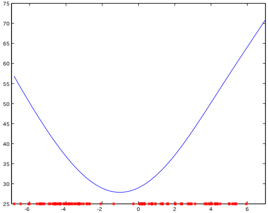
(a) Density function



(b) $\beta = 0.1$



(c) $\beta = 1$



(d) $\beta = 10$

Can β-metric be explained by curvature?

d_β-mean: $$\hat{\mu}_\beta = \arg\min_{m \in X} \sum_i g_\beta(d(x_i, m))^2.$$

Fréchet function on the d_β metric space behaves similar to positively curved space.

But the curvature (i.e. CAT(k) property) can not be defined for non geodesic metric spaces.

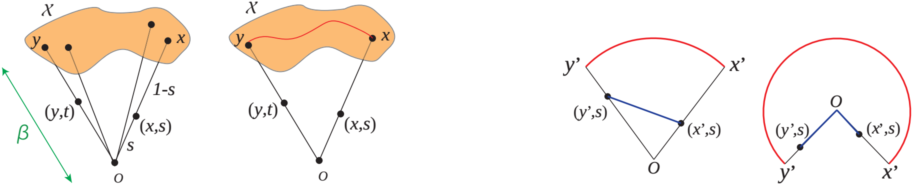
The d_β-mean can be redefined as an **extrinsic mean** when the data space is embedded in a geodesic metric space called a "metric cone".

Metric Cone

$\mathcal{X}$: a geodesic metric space
A metric cone $\tilde{\mathcal{X}}_\beta$ with $\beta > 0$ is a (truncated) cone $\mathcal{X} \times [0,1]/\mathcal{X} \times \{0\}$ with a metric

$$\tilde{d}_\beta((x,s),(y,t)) = \sqrt{t^2 + s^2 - 2ts\cos(\pi \min(d_{\mathcal{X}}(x,y)/\beta, 1))}$$

for any $(x,s),(y,t) \in \tilde{\mathcal{X}}_\beta$.

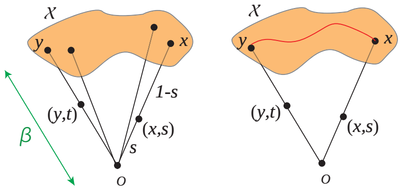


The d_β-mean can be redefined as an extrinsic mean when the data space is embedded in a "metric cone".

Extrinsic Mean in Metric Cone

Merits of embedding in a metric cone:

- Curvature of embedding space can be tuned by β in the sense of CAT(k) property (K. and Wynn(2020)) and Riemannian sense (Takehara and K.(2021)).
- The embedding space is only 1-dimensional higher than the original data space.
- Every geodesic metric space can be embedded. (Remark this is not true for embedding in a Euclidean space, e.g. d_α metric spaces.)



The α, β, γ-mean: Summary

We proposed a class of the intrinsic means:

$$\hat{\mu}_{\alpha,\beta,\gamma} = \arg\min_{m \in \mathcal{M}} \sum_i g_\beta(d_\alpha(x_i, m))^\gamma$$

and corresponding Frechét variances:

$$V_{\alpha,\beta,\gamma} = \min_{m \in \mathcal{M}} \frac{1}{n} \sum_i g_\beta(d_\alpha(x_i, m))^\gamma.$$

The Frechét function can be used for clustering.

Outline of Part I

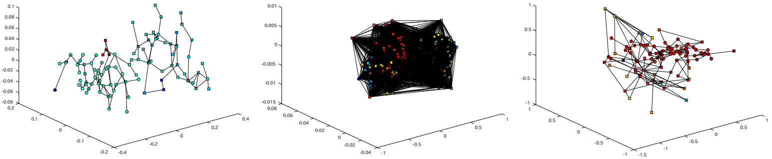
Application: Clustering

Data: five kinds of data from UCI Machine Learning Repository (iris, wine, ionosphere, breast cancer, yeast)
- The clustering error by k-means method decreases significantly by selecting an adequate value.
- $\alpha \in \{-5.0, -4.8, \ldots, 0.8, 1\}$ and $\beta \in \{2^{-3}, 2^{-2}, \ldots, 2^{6}, \infty\}$.

	k-means with $d_{\alpha,\beta}$			Euclid
data set	$\hat{\alpha}$	$\hat{\beta}$	r^*	r
(i) iris	-4.4	0.125	0.0333	0.1067
(ii) wine	0.8	8	0.2753	0.2978
(iii) ionosphere	-5.0	16	0.0798	0.2877
(iv) cancer	0.8	16	0.0914	0.1459
(v) yeast	-0.6	2	0.4447	0.4515

Application: Clustering

- The structure of the "optimal" geodesic graphs differs depending on the data:



(a) iris ($\hat{\alpha} = -5.0, \hat{\beta} = 0.125$) (b) wine ($\hat{\alpha} = 0.8, \hat{\beta} = 4$) (c) ionosphere ($\hat{\alpha} = -3.2, \hat{\beta} = 8$)

Figure 1: The geodesic graph of each data set with an optimum value of α and β for a randomly selected 100 sub-samples.

Example: Rainfall Data (1)

Data :

Daily precipitation data of 9 regions in UK since 1931
(UK Met office Hadley Centre observation data)

Target:

Check if the variance throughout each year changed

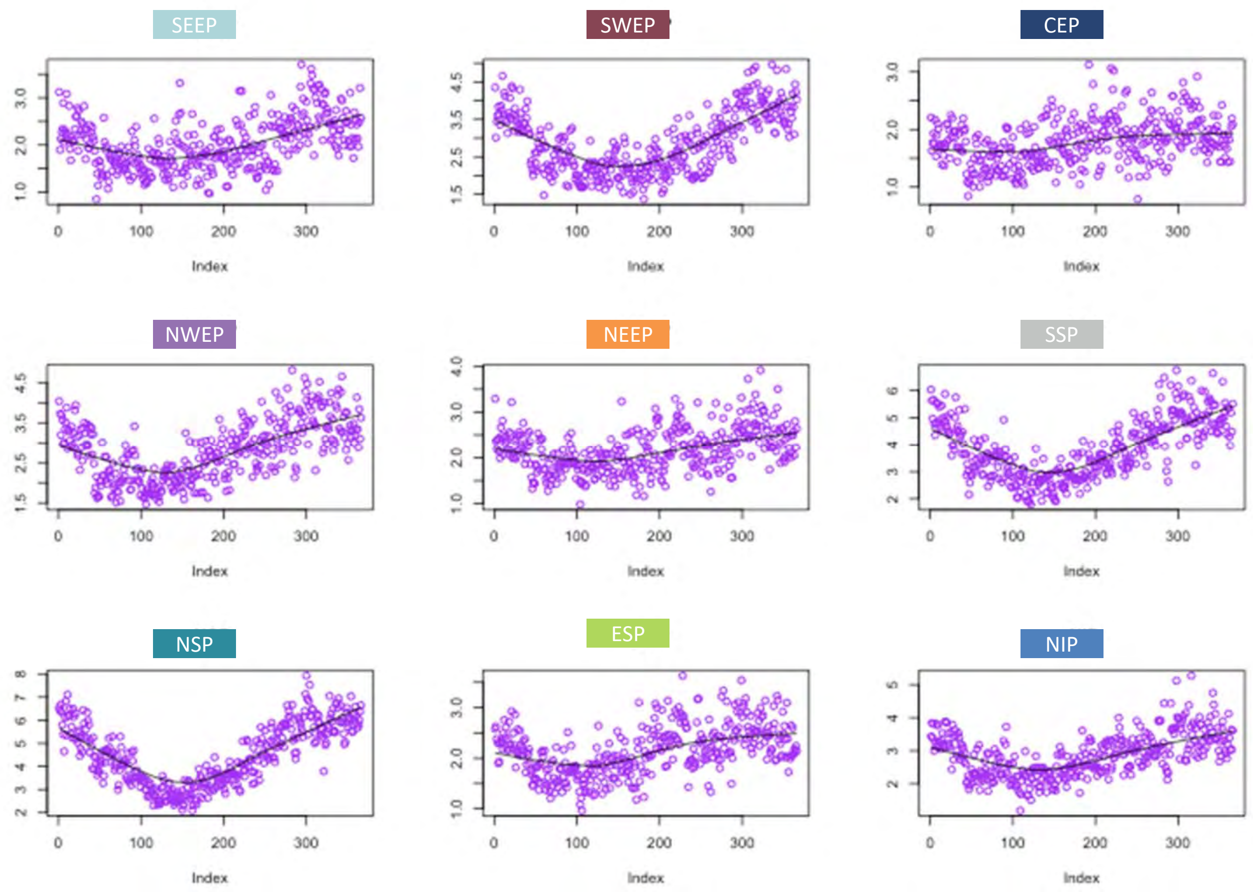
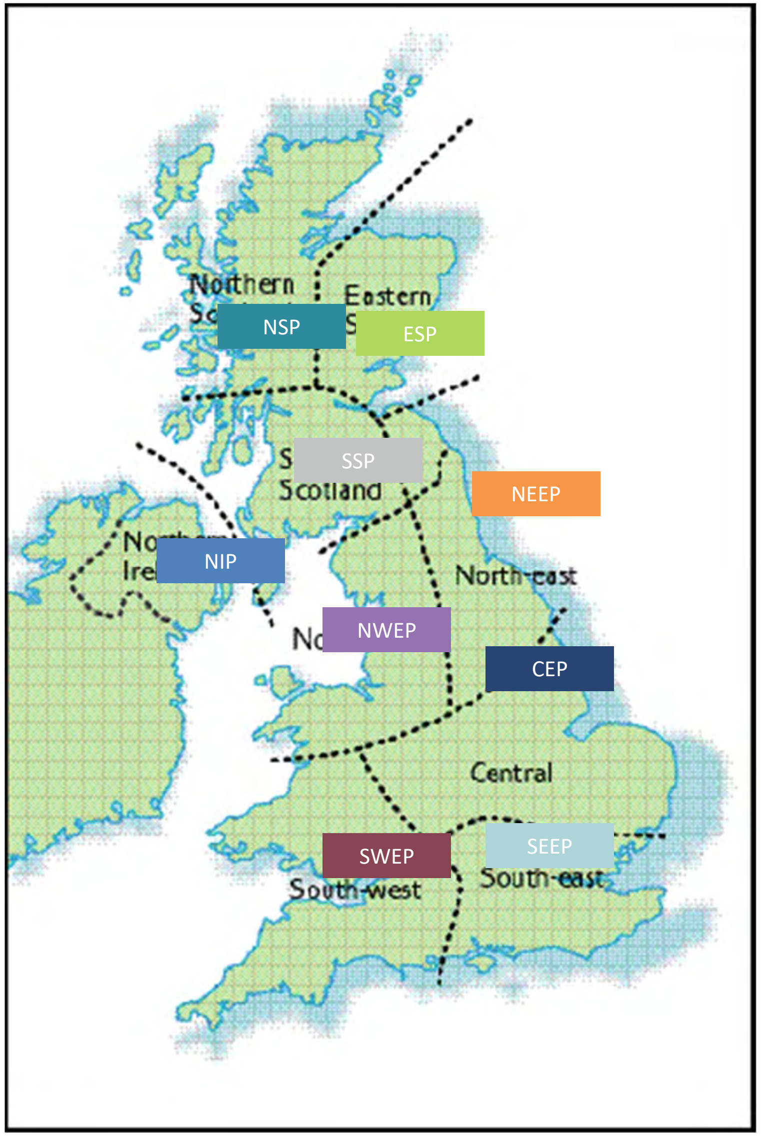
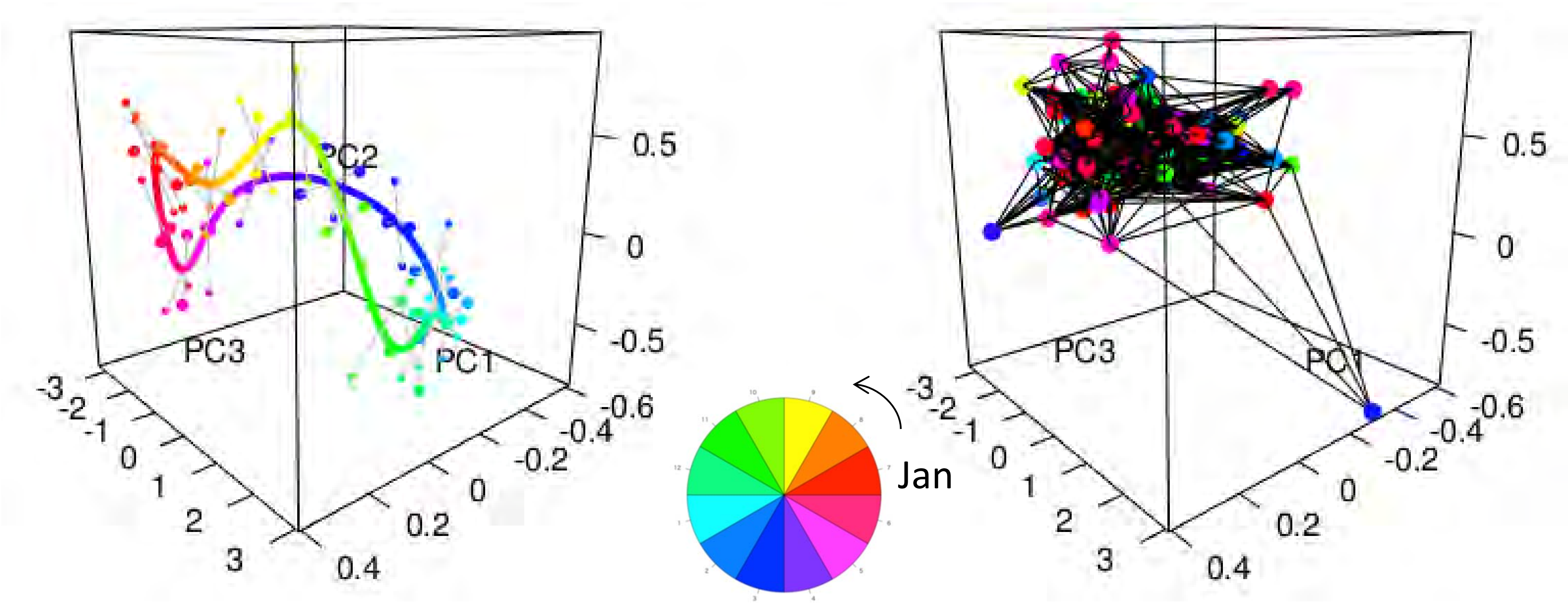


Fig: precipitation plot for 85 years



Example: Rainfall Data (2)



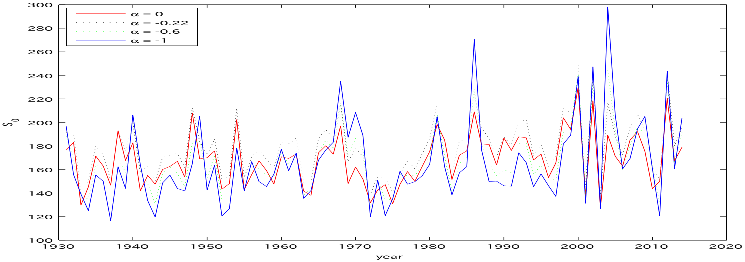
Annual cycle structure of the average of the 85 years

The geodesic graph of 1986, $\alpha = -0.6$

- Precipitation data has some annual cyclic structure.
- The geodesic subgraph and the corresponding generalized variance are expected to reflect such geometric structure.

Application: Rainfall Data

Time series of "variance" $s_0^2 := \{\min_i \sum_j d_\alpha(x_i, x_j)^2\}^{1/(1-\alpha)}$ are plotted for $\alpha = 0$ (red solid line), -0.22 (black dashed line) and -1 (blue solid line).



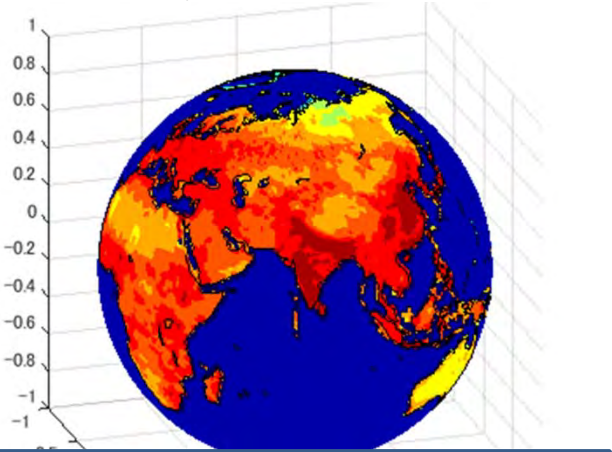
This generalized "variance" is expected to detect change of another type of volatility incorporating spacio-temporal geometrical structure of the precipitation data.

Example: Clustering of Populations on the earth

X_i: a person on the earth

y : a point on the earth

$$f_\beta(y) = \sum_i g_\beta(\|X_i - y\|)^2$$



population density

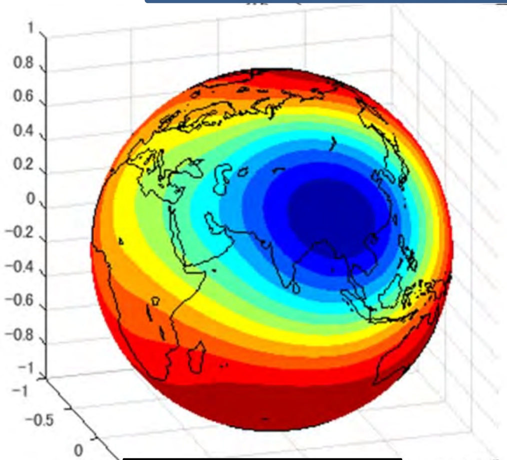
$\beta = 1$: Embedding into a Euclidean space

↓ Let β smaller

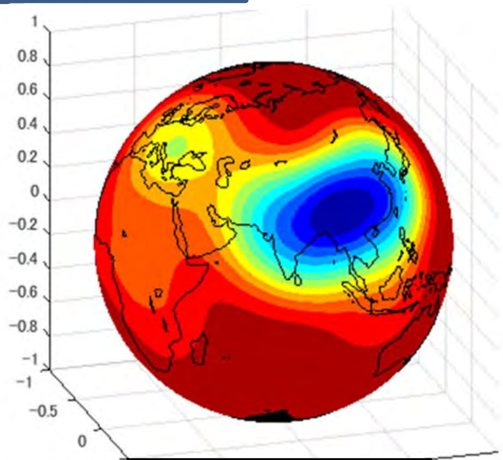
Curvature of the embedding metric cone becomes more positive

↓

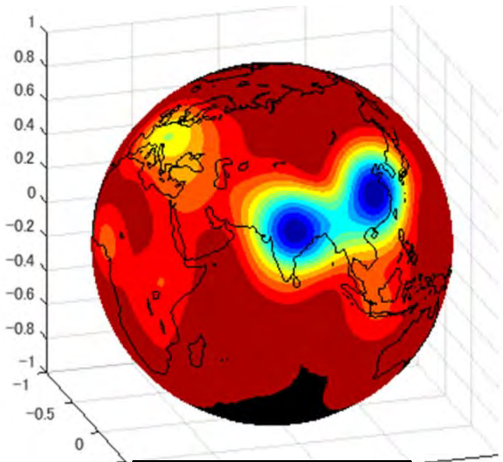
Local minima of f_β increases and are applicable to clustering



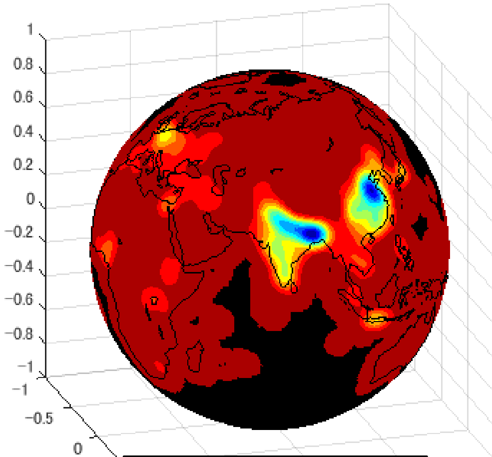
$\beta = 1$



$\beta = 0.3$



$\beta = 0.1$



$\beta = 0.01$

References for Part I

- Hwang, Sung Jin, Steven B. Damelin, and Alfred O. Hero III. (2016), "Shortest Path through Random Points." The Annals of Applied Probability: An Official Journal of the Institute of Mathematical Statistics 26 (5): 2791–2823.
- Kobayashi, K. and H. Wynn, (2020), Empirical geodesic graphs and CAT(k) metrics for data analysis, *Statistics and Computing*, 30(1), 1-18.
- Mckenzie, D. and Damelin, S. (2019). Power weighted shortest paths for clustering Euclidean data. arXiv preprint arXiv:1905.13345.

Part II
Application of metric cones to graph embedding

Kei Kobayashi (Keio University)
collaboration with **Daisuke Takehara (Accenture)**

Boston-Keio-Tsinghua2023 2023/06/30

Original problem setting

True model (unknown): A finite directed acyclic graph $G = (E, V)$ whose vertex corresponds to a data.

Input:

- an undirected finite graph $\bar{G} = (\bar{E}, V)$ without weights
- a metric space M with a "height" coordinate $h_M : M \to \mathbb{R}_{\geq 0}$

Output:

- a "representation of hierarchy": i.e. an embedding $\varphi : V \to M$ s.t. $h_M(\varphi(V))$ is consistent to the directions of E.

The problem does not make sense without further conditions.

Here we assume G is a rooted tree or approximately a rooted tree.

Example data and task

- Social Networks
  Input: Tweets of a buzz topic (w/o time stamps)
  Output: Estimated time order (and time intervals) of the tweets
- Chemical/Biological reactions
  Input: List of substance pairs whose interactions have been experimentally observed
  Output: Estimated direction and timing of each interaction

E.g., the Poincaré embeddings are applied to extract hierarchical structures from biological cell data (Kilmovskaia et al., 2020).

Based method: Poincaré embedding

- The Poincaré embedding (Nickel and Kiela, 2017) is a method for learning hierarchical representations of symbolic data by embedding them into a hyperbolic space (Poincaré ball).
- The Poincaré embeddings were reported "outperforming Euclidean embeddings significantly on data with latent hierarchies, both in terms of representation capacity and in terms of generalization ability."

Poincaré ball

The Poincaré embeddings embed data points(= vertices) to a **Poincaré ball**.

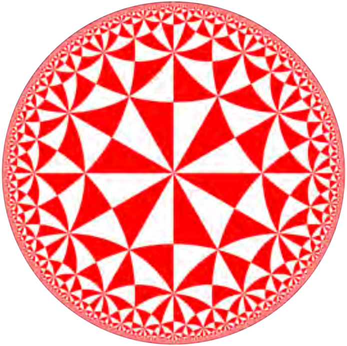
The Poincaré ball is a ball $B^d := \{\boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| < 1\}$ with a metric

$$g_{\boldsymbol{x}} := \left(\frac{2}{1 - \|\boldsymbol{x}\|^2}\right)^2 g^E.$$

The corresponding distance function becomes

$$d(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arccos}\left(1 + 2\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{(1 - \|\boldsymbol{u}\|^2)(1 - \|\boldsymbol{v}\|^2)}\right).$$

The Poincaré ball is a hyperbolic space and has a negative constant curvature.



M.C. Escher's hyperbolic tiling from Wikipedia

Poincaré embedding 1

The Poincaré embeddings learn the embedding of an undirected graph G by maximizing the following objective function:

$$L(\varphi) = \sum_{(u,v)\in E} \log \frac{\exp\left(-d(\varphi(u), \varphi(v))\right)}{\sum_{v'\in N(u)} \exp\left(-d(\varphi(u), \varphi(v'))\right)}$$

where $N(u) := \{v' \in V | (u, v') \notin E\}$ and d denotes the distance function on a Poincaré ball.

Poincaré embedding 2

The maximization of the objective function is done by the stochastic gradient descent on Riemannian manifolds (Riemannian SGD).

Euclidean SGD updates:

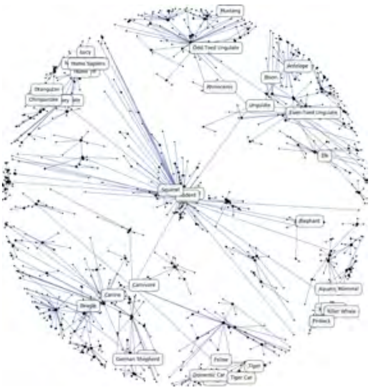
$$\boldsymbol{u} \leftarrow \boldsymbol{u} - \eta \nabla_{\boldsymbol{u}} L(\boldsymbol{u}),$$

Riemannian SGD updates:

$$\boldsymbol{u} \leftarrow \exp_{\boldsymbol{u}}(-\eta \nabla^R_{\boldsymbol{u}} L(\boldsymbol{u})).$$

Here $\eta = \eta_t > 0$ is a learning rate.
With the metric matrix of the embedding space (now, the Poincare ball) as g_u at $\boldsymbol{u}$, the gradient on the Riemannian manifold $\nabla^R_{\boldsymbol{u}} L(\boldsymbol{u})$ is a scaled Euclidean gradient:

$$\nabla^R_{\boldsymbol{u}} L(\boldsymbol{u}) = g_u^{-1} \nabla_{\boldsymbol{u}} L(\boldsymbol{u}).$$



(a) Intermediate embedding after 20 epochs

(b) Embedding after convergence

Learned embedding (cited from Nickel and Kiela(2017))

Pros and Cons of Poincaré embedding

Pros

- In experiments, embedding in a Poincaré ball works better than a Euclidean Space.
- The negative curvature of Poincaré balls suits the graph embedding problem, especially for tree-like networks.

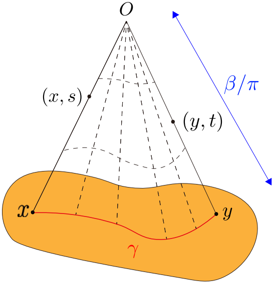
Cons

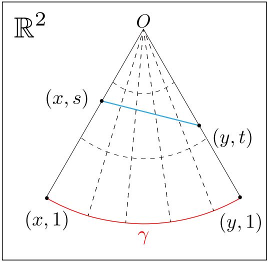
- It requires much more computational costs compared to the Euclidean embedding.
- The model does not have much flexibility thus we cannot select a model suitable to each graph structure.
- Especially the curvature of the Poincare ball is fixed and cannot be controlled.
- The computed hierarchical relation is not invariant under isometric transformations (e.g. the Möbius transformation).

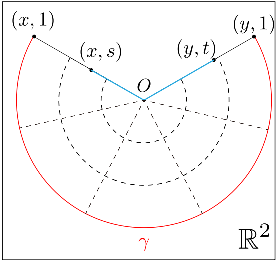
Our proposed method: Cone embedding 1

In Takehara and K. (2023), we proposed the **Cone embedding**, a method embedding a graph into **a metric cone**.

We first use another embedding algorithm to embed a graph into a Riemannian space (called "original embedding space"), and next optimize the height parameters.







Our proposed method: Cone embedding 2

The distance function on a metric cone is

$$\bar{d}_\beta((x,s),(y,t)) := \beta\sqrt{t^2+s^2-2ts\cos\left(\pi\min\left(d_Z(x,y)/\beta,1\right)\right)}.$$

The corresponding metric matrix at non-apex points is

$$\bar{g}(x,r) = \begin{pmatrix} r^2\pi^2 g(x) & 0 \\ 0 & \beta^2 \end{pmatrix}.$$

Thus we use the Riemannian SGD updates with this metric:

$$\boldsymbol{u} \leftarrow \exp_{\boldsymbol{u}}(-\eta\nabla^R_{\boldsymbol{u}}L(\boldsymbol{u})).$$

$$\nabla^R_{\boldsymbol{u}}L(\boldsymbol{u}) = \bar{g}_u^{-1}\nabla_{\boldsymbol{u}}L(\boldsymbol{u}).$$

Uniformly variable curvatures

The Ricci curvatures $\tilde{R}_{\alpha\gamma}$ and the scalar curvature $\tilde{R}$ at (x, s) become

$$\tilde{R}_{\alpha\gamma} = R_{\alpha\gamma} - \pi^2(n-1)\beta^{-2}\bar{g}_{\alpha\gamma},$$
$$\tilde{R}_{\alpha 0} = \tilde{R}_{0\gamma} = \tilde{R}_{00} = 0, \tilde{R} = \{\pi^{-2}R - n(n-1)\beta^{-2}\}s^{-2}$$

- The scalar curvature and the Ricci curvatures become more negative than (a constant times of) the corresponding original curvatures for $\beta < \infty$ and $n \geq 2$.
- β changes the curvatures uniformly thus we can control the curvatures by tuning β.
- The closer to the apex, i.e. the smaller the value of s, the greater the change of the scalar curvature.

When the original space is a geodesic metric space, similar results in the sense of CAT(k) property were proved in K. and Wynn(2020).

Identifiablity 1

The Cone embedding has identifiability of hierarchical values from the learnt metrics that the Poincaré embedding does not satisfy.

Assume the original embedding space Z is a Riemannian manifold, and let X be the metric cone of Z with a parameter $\beta > 0$.

We assume that each data point $z_i \in Z$ $(i = 1, \ldots, n)$ has its specific "height" $t_i \in [0, 1]$ in the metric cone X.

Our proposed method embeds data points into a metric cone based on the estimated distances $\tilde{d}_\beta(x_i, x_j)$ $(i, j = 1, \ldots, n)$ and tries to compute the heights $t_1, \ldots, t_n$ as a measure of the hierarchy level.

However, it is not evident if these heights are identifiable only from the information of the original data points in Z and the distances $\tilde{d}_\beta(x_i, x_j)$ $(i, j = 1, \ldots, n)$ in the metric cone.

Identifiablity 2

Theorem 1 (Theorem 1 of Takehara and K. (2023))

(a) *Let* $n \geq 3$ *and assume that* $z_1, \ldots, z_n$ *are not all aligned on a geodesic in* Z*. Then, the heights* $t_1, \ldots, t_n$ *are "identifiable" up to at most four candidates.*

(b) *Let* $n \geq 4$ *and assume* $z_1, \ldots, z_n$ *and* $t_1, \ldots, t_n$ *take "general" positions and heights, respectively. Then, the heights* $t_1, \ldots, t_n$ *are identifiable uniquely.*

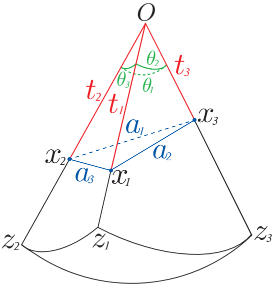
(c) *If* $d_Z(z_i, z_j) \geq \beta/2$ *for all* $i, j = 1, \ldots, n,\ i \neq j$*, then the heights* $t_1, \ldots, t_n$ *are identifiable uniquely.*

See Takehara and K. (2023) for the definition of "ientifiable" and "general".

Identifiablity 3

Comments on the proof)

- For $n = 3$, the problem becomes an elementary geometrical problem.
- We used algebraic approach and evaluate the number of solutions of a quadratic polynomial system using the Gröbner basis.
- For $n \geq 4$, the system becomes overdetermined.



$$t_2^2 + t_3^2 - 2t_2 t_3 \cos\theta_1 = a_1^2,$$
$$t_3^2 + t_1^2 - 2t_3 t_1 \cos\theta_2 = a_2^2,$$
$$t_1^2 + t_2^2 - 2t_1 t_2 \cos\theta_3 = a_3^2.$$

Experiments

We evaluate the proposed method in two experiments:

- E1 embedding graphs (coauthors' network of academic papers),
- E2 embedding taxonomies (WordNet).

For every vertex u, we rank $\{\tilde{d}_\beta(u, v) \mid v \in V\}$ and an index MR(mean ranking) is computed as the difference from the ranking of u and the mean ranking of vs adjacent to u.
Mean average precisions(MAP) are also compared[1].

In E2, the following score is used instead of $\tilde{d}_\beta(u, v)$:

$$\text{score}(\text{is-a}((\boldsymbol{u}, s), (\boldsymbol{v}, t))) = -(1 + \alpha(s - t))d(\boldsymbol{u}, \boldsymbol{v})$$

where $\alpha = 10^3$.

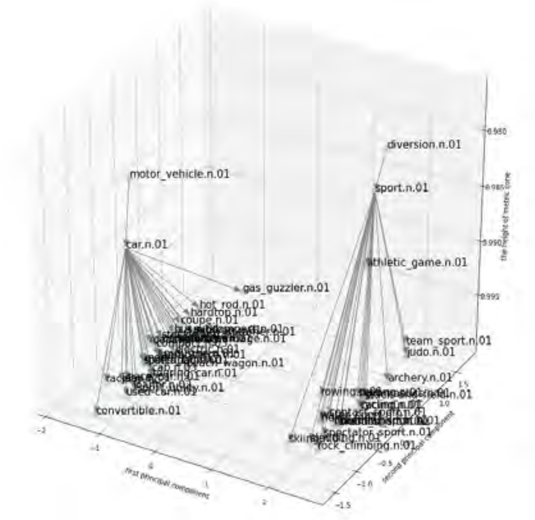
[1] Average precision is similar to AUC (AUROC) but using a plot of precision (not specificity) as a function of sensibility. MAP is the mean of APs over u.

Experimental results 1

Table 1: Embedding accuracy for WordNet

		dimension			
		10	20	50	100
Euclidean	MR	1471.70	232.88	2.51	1.82
	MAP	0.070	0.122	0.838	0.899
Poincare	MR	**19.94**	**19.62**	19.47	19.36
	MAP	**0.528**	**0.534**	0.537	0.538
Our Model (Metric Cone)	MR	1401.28	209.11	**2.30**	**1.79**
	MAP	0.052	0.126	**0.853**	**0.902**

Figure: Visualization of WordNet embedding using metric cone



Experimental results 3

Table 2: Results of GrQc embedding into low-dimensional space

Model	eval.	2	3	4	5	6	7	8	9
Euclid.	MR	88.99	37.17	17.15	9.42	5.78	4.27	3.42	3.18
	MAP	0.375	0.488	0.600	0.719	0.842	0.929	0.983	0.998
Metric Cone	MR	72.35	26.39	14.50	8.65	5.50	4.16	3.40	3.18
	MAP	0.450	0.551	0.614	0.726	0.851	0.935	0.986	0.998

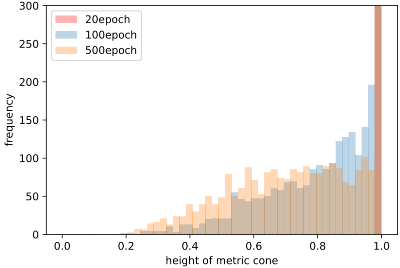


Figure 1: Changes in the distribution of the heights of data points

Summary: Merits of cone embeddings

- We provide an indicator of hierarchical information that is both geometrically and intuitively natural to interpret.
- The model can inherit the merits of the original embedding method. Thus it can suit each graph structure more flexibly.
- Computation for cone embeddings is light since it optimizes only one coordinate variable per each data point.
- The curvatures of the embedding space can be controlled by parameter β.
- The hierarchical relation is determined uniquely by the distances between the data points under mild conditions.

Bibliography

- Klimovskaia, A., Lopez-Paz, D., Bottou, L., and Nickel, M., Poincaré maps for analyzing complex hierarchies in single-cell data. *Nature Communications*, 11(1):1–9, 2020.
- Kobayashi, K. and Wynn, H. P., Empirical geodesic graphs and cat (k) metrics for data analysis. *Statistics and Computing*, 30(1):1–18, 2020.
- Nickel, M. and Kiela, D., Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*, pp. 6338–6347, 2017.
- Takehara, D. and K., Representing hierarchical structured data using cone embedding, *Mathematics*, 11(10), 2294, 2023.