Spatial depth in metric spaces

J. Virta

University of Turku

ICORS 2025
19th May, Stresa

Reference

This talk is based on the following preprint:

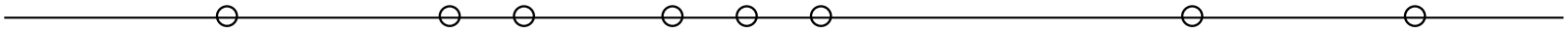
- **Virta, J.**. *Spatial depth for data in metric spaces*. arXiv preprint arXiv:2306.09740 (2023).

These slides are available at the speaker's website
`https://users.utu.fi/jomivi/talks/`

Table of Contents

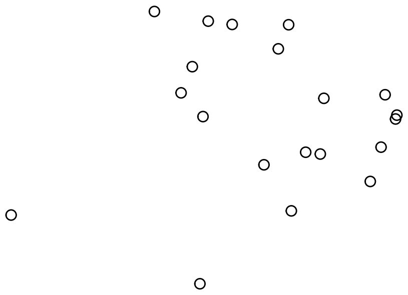
Ordering in $\mathbb{R}$

- Univariate samples of points admit a natural ordering.



Ordering in $\mathbb{R}^2$

- How can we order a bivariate sample?

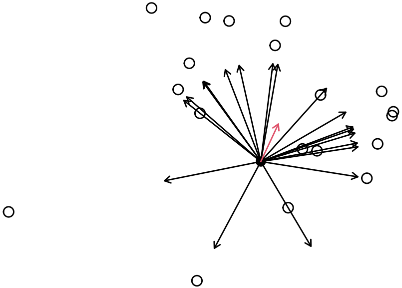


Spatial depth

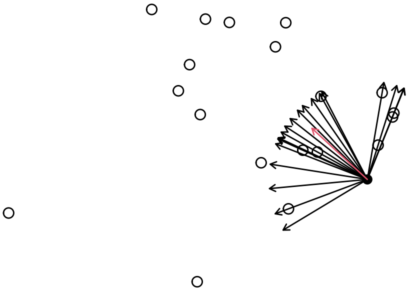
The spatial depth [Chaudhuri, 1996, Vardi and Zhang, 2000] of a point $\mu \in \mathbb{R}^p$ with respect to the distribution of $X \sim P$ is

$$D(\mu; P) := 1 - \left\| \mathrm{E} \left(\frac{X - \mu}{\|X - \mu\|} \right) \right\|.$$

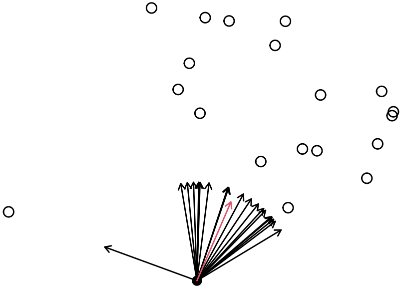
Example



Example



Example



Spatial depths of all points

The points are divided into central and outlying.

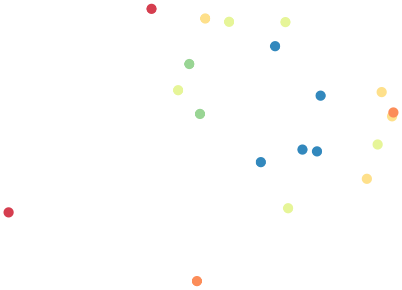


Table of Contents

Object-valued data

- Nowadays, many samples $X_1, \ldots, X_n$ can be seen to consist of objects, such as
  - images,
  - functions,
  - graphs,
  - correlation matrices,
  - ...
- A natural mathematical framework for object-valued data:
  1. A metric space $(\mathcal{X}, d)$.
  2. A random variable $X \sim P$ taking values in $\mathcal{X}$.
  3. Methodology that relies on X only through the metric $d(\cdot, \cdot)$.

Table of Contents

Metric spatial depth

- Let $h : \mathcal{X}^3 \to \mathbb{R}$ be defined as

$$h(x_1, x_2, x_3) := \mathbb{I}(x_3 \notin \{x_1, x_2\}) \frac{d^2(x_1, x_3) + d^2(x_2, x_3) - d^2(x_1, x_2)}{d(x_1, x_3) d(x_2, x_3)},$$

where $\mathbb{I}(\cdot)$ denotes the indicator function.

Metric spatial depth [Virta, 2023]

$$D(\mu; P) := 1 - \frac{1}{2}\mathrm{E}\{h(X_1, X_2, \mu)\},$$

where $X_1, X_2 \sim P$ are independent.

Theoretical properties

Property 1

When $(\mathcal{X}, d)$ is an Euclidean space, then the metric spatial depth reduces to (a one-to-one transformation of) the classical spatial depth.

Theoretical properties

Property 2

$D(\mu; P)$ is finite for all $\mu \in \mathcal{X}$ and all P.

Property 3

The influence function of D satisfies

$$\sup_{z \in \mathcal{X}} |IF(z; D, \mu, P)| \leq 4.$$

Theoretical properties

Let $L[x, y, z]$ denote the event that

$$d(x, z) = d(x, y) + d(y, z),$$

i.e., that the point y is along the way from x to z

Property 4

Assume that P has no atoms. Then,

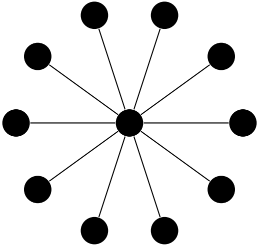
(i) $D(\mu; P) \in [0, 2]$.

(ii) $D(\mu; P) = 0$ if and only if $P(L[X_1, X_2, \mu] \cup L[X_2, X_1, \mu]) = 1$.

(iii) $D(\mu; P) = 2$ if and only if $P(L[X_1, \mu, X_2]) = 1$.

(iv) If $(\mathcal{X}, d)$ is a Hilbert space, then $D(\mu; P) \leq 1$.

Railroad metric



Theoretical properties

Property 5

Let μ_n be a divergent sequence in $\mathcal{X}$. Then $D(\mu_n; P) \to 0$ as $n \to \infty$.

Theoretical properties

Property 6

Assume that P has no atoms. Then $D(\mu; P)$ is continuous in both μ and P.

Table of Contents

Non-convex depth regions

- It is not necessary to use the Euclidean metric with data in $\mathbb{R}^p$ and we can instead use metrics that promote non-linearity

Depth regions

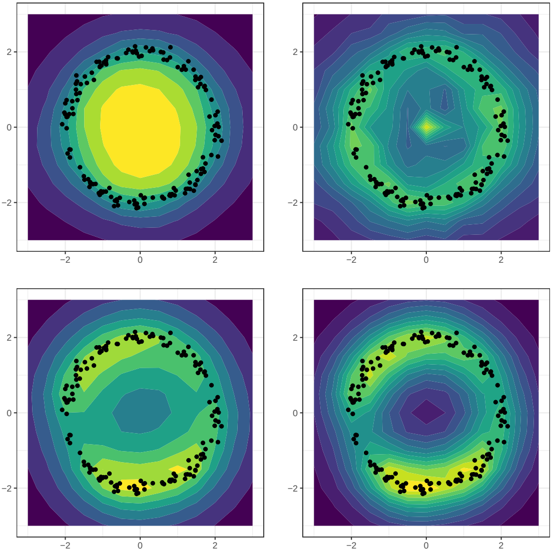


Table of Contents

Literature

- A review of the current state of object-valued data analysis can be found in:

  Dubey Paromita, Yaqing Chen, and Hans-Georg Müller. *Metric statistics: Exploration and inference for random objects with distance profiles.* Annals of Statistics 52.2 (2024): 757-792.

Thank you for your attention!

References I

Chaudhuri, P. (1996).
On a geometric notion of quantiles for multivariate data.
Journal of the American Statistical Association, 91(434):862–872.

Li, J., Cuesta-Albertos, J. A., and Liu, R. Y. (2012).
DD-classifier: Nonparametric classification procedure based on DD-plot.
Journal of the American Statistical Association, 107(498):737–753.

Vardi, Y. and Zhang, C.-H. (2000).
The multivariate L_1-median and associated data depth.
Proceedings of the National Academy of Sciences, 97(4):1423–1426.

References II

Virta, J. (2023).
Spatial depth for data in metric spaces.
arXiv preprint arXiv:2306.09740.

Xiao, H., Rasul, K., and Vollgraf, R. (2017).
Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms.
arXiv preprint arXiv:1708.07747.

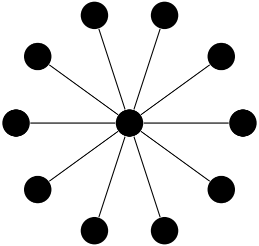
Table of Contents

Example I

- Assume that P puts equal mass $1/n$ to each of the fixed points $z_1, \ldots, z_n \in \mathcal{X}$.
- Question: Under what metric is the depth $D(z_1; P)$ maximal?

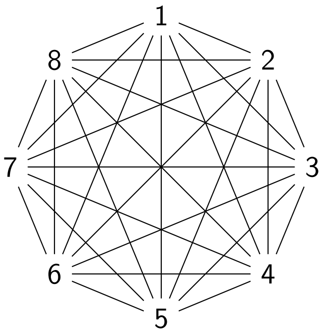
Example I

- The railway metric uniquely gives $D(z_1; P) = 1 + \left(1 - \frac{1}{n}\right)\left(1 - \frac{3}{n}\right)$.



Example II

- Equip the finite set $\mathcal{X} = \{1, \ldots, n\}$ with the discrete metric $d(i,j) = 1 - \mathbb{I}(i = j)$.



- For which probability distribution $P = (p_1, \ldots, p_n)$ is $D(1; P)$ maximized/minimized?

Example II

(i) $D(1; P) = 1$ if and only if $p_1 = 1$.

(ii) $D(1; P) = 0$ if and only if exactly one of $p_2, \ldots, p_n$ equals 1.

Table of Contents

Depth-depth plot

- Let P_{n1} and P_{n2} be the empirical distributions of two samples corresponding to different groups.
- In DD-classification [Li et al., 2012], we compute the depth vectors

$$z_i := (D(x_i; P_{n1}), D(x_i; P_{n2})).$$

- A test point $x \in \mathcal{X}$ is then classified based on the vector

$$z := (D(x; P_{n1}), D(x; P_{n2})).$$

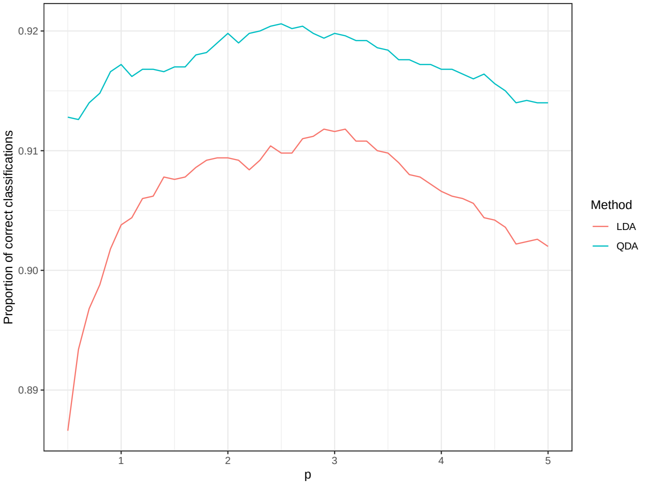
Data example

- We applied DD-classification to a subsample of the FashionMNIST data set [Xiao et al., 2017] consisting of images of dresses, shirts and ankle boots.

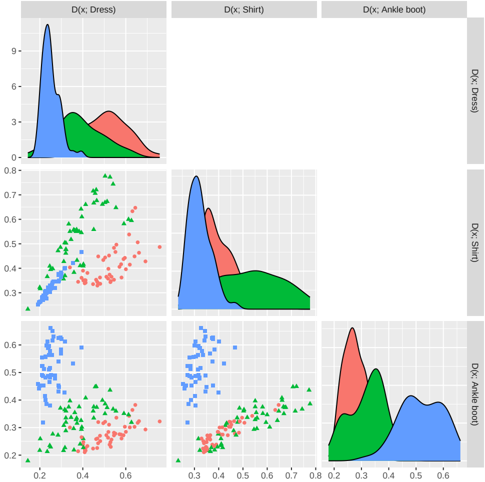
Specifications

- We randomly drew a training sample of $n = 150$ images and a test sample of $n_0 = 50$ images and used DD-classifier to predict the labels of the test images.
- We used metric spatial depth with the L_p-distance with $p = 0.5, 0.6, \ldots, 5$ as the metric.
- We considered both LDA and QDA.
- The experiment was repeated a total of 100 times.

Results


0.92
0.91
0.90
0.89
Proportion of correct classifications
1
2
3
4
5
p
Method
LDA
QDA

DD-plot


D(x; Dress)
D(x; Shirt)
D(x; Ankle boot)
D(x; Dress)
D(x; Shirt)
D(x; Ankle boot)
0
3
6
9
0.3
0.4
0.5
0.6
0.7
0.8
0.2
0.3
0.4
0.5
0.6
0.2
0.4
0.6
0.3
0.4
0.5
0.6
0.7
0.8
0.2
0.3
0.4
0.5
0.6

Interpretation

- Both LDA and QDA reach their maximal performance at super-Euclidean geometry.
- QDA is uniformly superior to LDA.