Spatial depth for object data

J. Virta

University of Turku

TU Wien, 16th of October 2023

Table of Contents

Ordering in $\mathbb{R}$

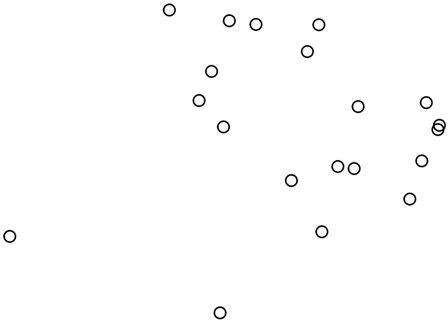
- One of the major differences between $\mathbb{R}$ and $\mathbb{R}^p$, $p \geq 2$, is that the former admits a natural ordering of its elements.
- That is, given a univariate sample $x_1, \ldots, x_n$, we can order it as

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}.$$

- This ordering is invaluable in robust statistics:
  - Extreme observations are seen as outliers.
  - By "trimming" them we obtain outlier-resistant procedures.

Ordering in $\mathbb{R}^2$?

- How to generalize the ordering to bivariate samples?



Depth functions

- Let P be a probability distribution in $\mathbb{R}^p$ and let $\mu \in \mathbb{R}^p$.
- Informally, $\mu \mapsto D(\mu; P)$ is a *statistical depth function* if
  1. $D(\mu; P)$ takes a high value when μ is located close to the center of P.
  2. $D(\mu; P)$ takes a small value when μ is outlying w.r.t. to P.
- [Zuo and Serfling, 2000] formulated a famous set of axioms that any proper depth function should satisfy.
- Depth functions are "unsigned" generalizations of the ordering in $\mathbb{R}$.

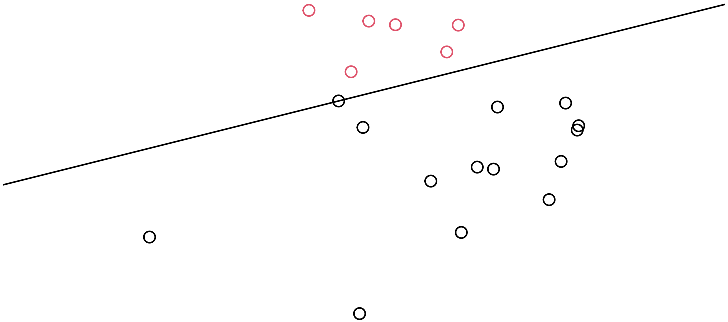
Tukey's halfspace depth

- The most well-known depth function is possibly Tukey's halfspace depth [Tukey, 1975]:

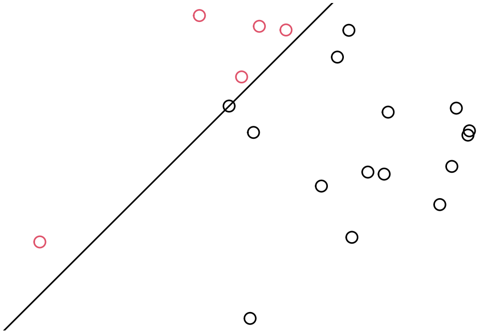
$$D(\mu; P) = \inf_{v \in \mathbb{S}^{p-1}} P(v'X > v'\mu).$$

- Informally, the halfspace depth of μ is the least amount of probability mass one can see when standing at μ and choosing one's orientation appropriately.

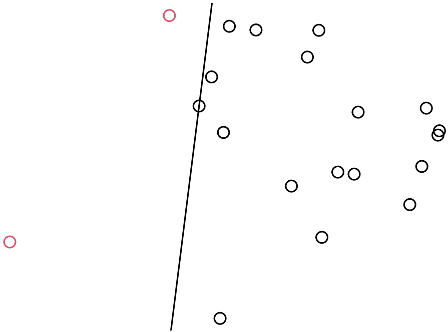
Example



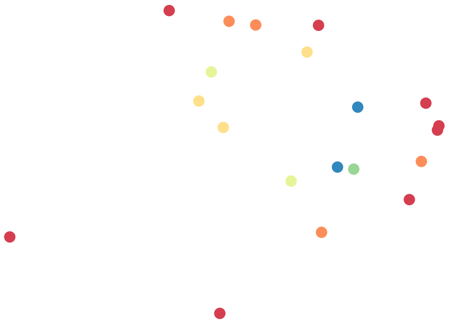
Example



Example



Halfspace depths of all points



Spatial depth

- Depth functions are typically based on geometric ideas.
- Spatial depth [Chaudhuri, 1996, Vardi and Zhang, 2000] is defined as

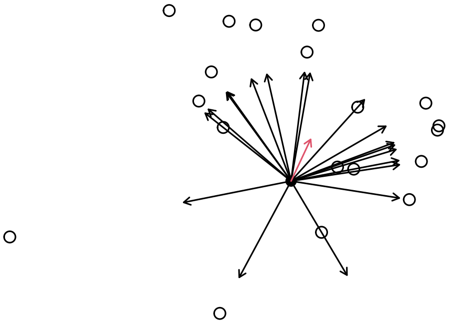
$$D(\mu; P) := 1 - \left\| \mathrm{E} \left(\frac{X - \mu}{\|X - \mu\|} \right) \right\|^2,$$

- When P is an empirical distribution of a sample, the quantity

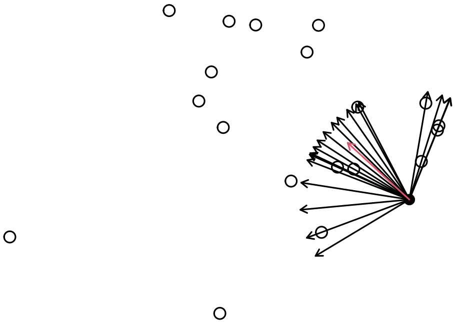
$$\left\| \mathrm{E} \left(\frac{X - \mu}{\|X - \mu\|} \right) \right\|$$

is the length of the average of unit-length arrows drawn from μ towards the sample points.

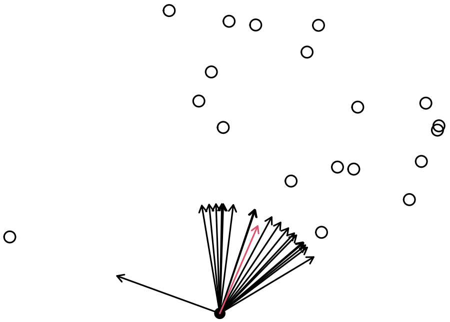
Example



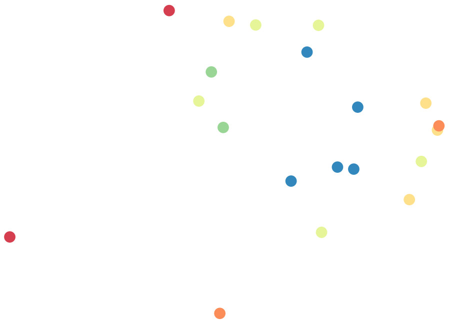
Example



Example



Spatial depths of all points



Other depths

- Both the halfspace and spatial depth can be shown to be robust (extremely outlying observations do not render them useless).
- This is true also for most developed depth functions:
  - Projection depth [Liu, 1992]
  - Simplicial depth [Liu, 1990]
  - Oja depth [Oja, 1983]
  - Lens depth [Liu and Modarres, 2011]
  - ...

Using depths

- The ordering offered by depths leads to natural generalizations of univariate trimmed estimators.
- The point $\mu_0 \in \mathbb{R}^p$ with the largest depth can act as a measure of location.
  - In $\mathbb{R}$, the deepest point equals the median for most depth functions (such as for both the halfspace and spatial depth).
- Outlier detection.
- Depth-depth plots [Li et al., 2012] can be used for classification.

Table of Contents

Object data

- In many contemporary applications, the sample $X_1, \dots, X_n$ does not consist of points in Euclidean space but, rather, of more general *objects*, such as images, functions, graphs...
- Computing descriptive statistics for a sample of objects is tricky but often the difference between two objects is easy to quantify.

Analysing object data

- A natural mathematical framework for object data is as follows.
  - A metric space $(\mathcal{X}, d)$.
  - A distribution P (possibly a sample) taking values in $\mathcal{X}$
  - Methodology that relies on $X \sim P$ only through the metric $d(\cdot, \cdot)$.
- **Example 1**: the Frechét mean of P is

$$\mathrm{argmin}_{\mu \in \mathcal{X}} \mathrm{E}\{d^2(X, \mu)\}$$

and the minimal value of the objective function is the Frechét variance of P.

Analysing object data

- A simple way of developing object data methods is to
  1. Formulate a Euclidean method as a function of "$\|X - Y\|$" only.
  2. Replace $\|X - Y\|$ with $d(X, Y)$.
- **Example 2**: Applying the above to PCA results into multidimensional scaling (MDS).

Metric halfspace depth

- [Dai and Lopez-Pintado, 2022] formulated metric halfspace depth as

$$D(\mu; P) = \inf_{\substack{z_1, z_2 \in \mathcal{X} \\ d(\mu, z_1) \leq d(\mu, z_2)}} P(d(X, z_1) \leq d(X, z_2)).$$

- Computing the deepest point μ_0 is highly non-trivial.

Table of Contents

Metric spatial depth

- We 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.

- The *metric spatial depth* [Virta, 2023] is then

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

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

Theoretical properties

Property I

When $(\mathcal{X}, d)$ is an Euclidean space, then the metric spatial depth equals the classical spatial depth.

Theoretical properties

Property II

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

Property III

The influence function of D satisfies

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

Theoretical properties

$L[x_1, x_2, x_3]$ denotes the event that

$$d(x_1, x_3) = d(x_1, x_2) + d(x_2, x_3),$$

i.e., that the three points $x_1, x_2, x_3 \in \mathcal{X}$ fall in a line (in the sense of the metric d) such that x_2 is in the middle of the other two.

Property IV

$D(\mu; P)$ takes values in the interval $[0, 2]$. Additionally,

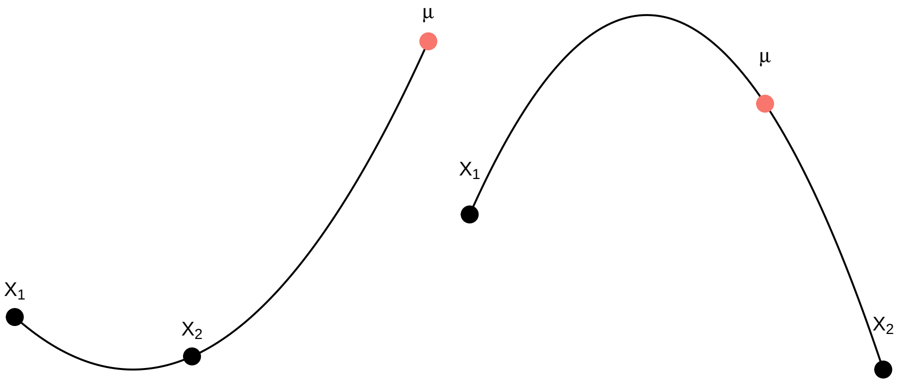
(i) $D(\mu; P) = 0$ if and only if

$$P(\{\mu\}) = 0 \quad \text{and} \quad P(L[X_1, X_2, \mu] \cup L[X_2, X_1, \mu]) = 1.$$

(ii) $D(\mu; P) = 2$ if and only if

$$P(\{\mu\}) = 0 \quad \text{and} \quad P(L[X_1, \mu, X_2]) = 1.$$

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


μ
X1
X2
μ
X1
X2

Theoretical properties

Property V

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

Theoretical properties

Property VI

Assume that $P(\{\mu\}) = 0$ for all $\mu \in \mathcal{X}$. Then $D(\mu; P)$ is continuous in both μ and P.

Remarks

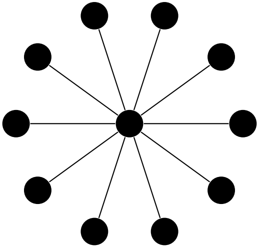
- Metric spatial depth also inherits the invariance properties of the used metric.
- Interestingly, the previous results use all four axioms of a metric.

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?

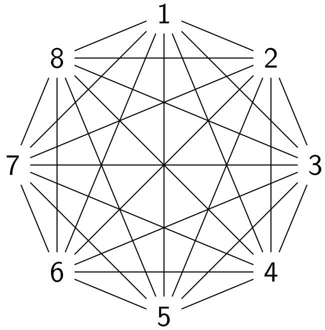
Example I

- 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.

Sample metric spatial depth

- Let P_n be the empirical distribution of $X_1, \ldots, X_n \sim P$. The sample metric spatial depth is

$$D(\mu; P_n) = 1 - \frac{1}{2n^2} \sum_{i,j=1}^{n} h(X_i, X_j, \mu).$$

- Standard U-statistic theory says that, for a fixed $\mu \in \mathcal{X}$, we have,

$$D(\mu; P_n) = D(\mu; P) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Table of Contents

Non-convex depth regions

- The freedom to choose the metric d in the formulation of the metric spatial depth means that, in the case of Euclidean data, we can use metrics that promote non-linearity:
    1. Kernel trick [Chen et al., 2008]
    2. ISOMAP.

Depth regions

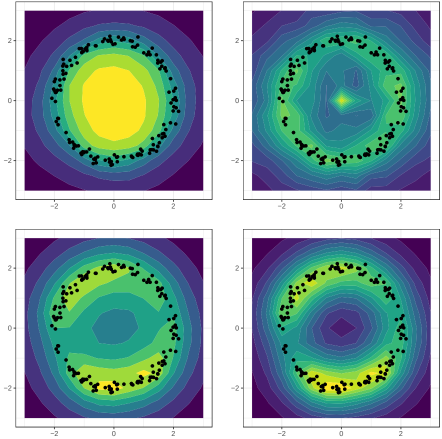


Figure: Used "metrics" from left to right, top to bottom: Euclidean, ISOMAP, Rational quadratic kernel, Gaussian kernel.

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

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

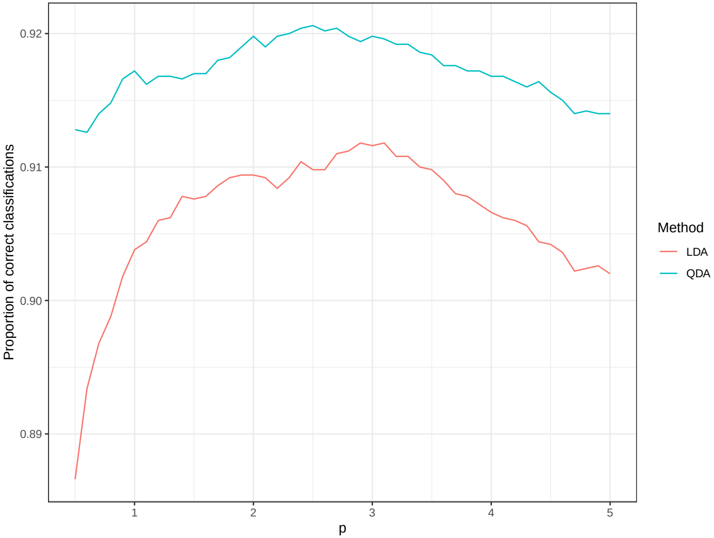
Data example

- 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.

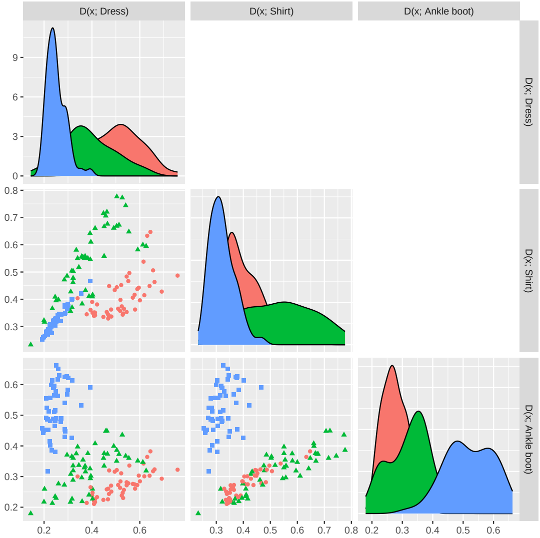
Specifications

- 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.

Final remarks

- Is the point μ_0 with the largest depth ("spatial median") unique under some natural conditions?
- In the Euclidean case, the spatial median is equivalent to the Frechét median. Does the same hold for some object spaces?
- How to find the spatial median in practice?

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.

Chen, Y., Dang, X., Peng, H., and Bart, H. L. (2008).
Outlier detection with the kernelized spatial depth function.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2):288–305.

Dai, X. and Lopez-Pintado, S. (2022).
Tukey's depth for object data.
Journal of the American Statistical Association, pages 1–13.

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.

References II

Liu, R. Y. (1990).
On a notion of data depth based on random simplices.
Annals of Statistics, pages 405–414.

Liu, R. Y. (1992).
Data depth and multivariate rank tests.
L1-Statistical Analysis and Related Methods, pages 279–294.

Liu, Z. and Modarres, R. (2011).
Lens data depth and median.
Journal of Nonparametric Statistics, 23(4):1063–1074.

Oja, H. (1983).
Descriptive statistics for multivariate distributions.
Statistics & Probability Letters, 1(6):327–332.

References III

Tukey, J. W. (1975).
Mathematics and the picturing of data.
In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, volume 2, pages 523–531.

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.

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.

References IV

Zuo, Y. and Serfling, R. (2000).
General notions of statistical depth function.
Annals of Statistics, pages 461–482.