Depth and depth-based classification

Pavlo Mozharovskyi

LTCI, Télécom ParisTech, Université Paris Saclay

Machine Learning

Paris, January 16, 2019

Contents

Contents

Linear discriminant analysis

- **Assumptions:** Two classes are normally distributed with the same covariance matrix, *i.e.* $X|Y = j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, $j = 0, 1$ or

$$f_j(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma}_j)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)}, \quad \text{for } j = 0, 1$$

$$\text{and} \quad \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}\,.$$

Linear discriminant analysis

- **Assumptions:** Two classes are normally distributed with the same covariance matrix, *i.e.* $X|Y = j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\,,\ j = 0, 1$ or

$$f_j(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma}_j)}}\, e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)}\,, \quad \text{for } j = 0, 1$$

$$\text{and} \quad \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}\,.$$

- **Plug-in into Bayes:**

$$g(\boldsymbol{x}) = \begin{cases} 1 & \log \frac{\pi_1 f_1(\boldsymbol{x})}{\pi_0 f_0(\boldsymbol{x})} > 0\,, \\ 0 & \text{else}\,. \end{cases}$$

Linear discriminant analysis

- **Assumptions:** Two classes are normally distributed with the same covariance matrix, *i.e.* $X|Y = j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, $j = 0, 1$ or

$$f_j(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma}_j)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)} , \quad \text{for } j = 0, 1$$

$$\text{and} \quad \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma} .$$

- **Plug-in into Bayes:**

$$g(\boldsymbol{x}) = \begin{cases} 1 & \log \frac{\pi_1 f_1(\boldsymbol{x})}{\pi_0 f_0(\boldsymbol{x})} > 0 , \\ 0 & \text{else} . \end{cases}$$

- If $\pi_0 = \pi_1$, this is the same as

$$g(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_0) > (\boldsymbol{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1) , \\ 0 & \text{else} . \end{cases}$$

Linear discriminant analysis

- **Assumptions:** Two classes are normally distributed with the same covariance matrix, *i.e.* $X|Y = j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, $j = 0, 1$ or

$$f_j(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma}_j)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)}, \quad \text{for } j = 0, 1$$

$$\text{and} \quad \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}\,.$$

- **Plug-in into Bayes:**

$$g(\boldsymbol{x}) = \begin{cases} 1 & \log \frac{\pi_1 f_1(\boldsymbol{x})}{\pi_0 f_0(\boldsymbol{x})} > 0\,, \\ 0 & \text{else}\,. \end{cases}$$

- If $\pi_0 = \pi_1$, this is the same as

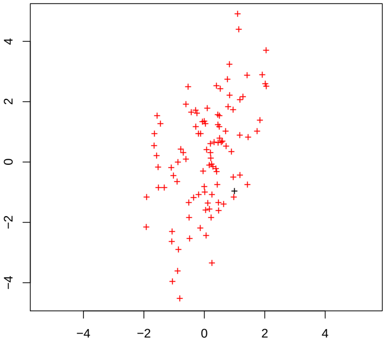
$$g(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_0) > (\boldsymbol{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1)\,, \\ 0 & \text{else}\,. \end{cases}$$

- Equivalently one can write

$$g(\boldsymbol{x}) = \begin{cases} 1 & d^2_{Mah}(\boldsymbol{x}, \boldsymbol{\mu}_0; \boldsymbol{\Sigma}_0) > d^2_{Mah}(\boldsymbol{x}, \boldsymbol{\mu}_1; \boldsymbol{\Sigma}_1)\,, \\ 0 & \text{else}\,. \end{cases}$$

Mahalanobis distance

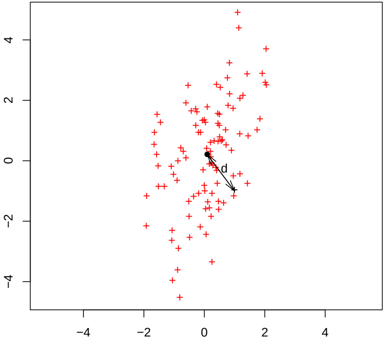
- Regard a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and a point $\boldsymbol{x} \in \mathbb{R}^d$.



- How central (or representative) is $\boldsymbol{x}$ with respect to $\boldsymbol{X}$?

Mahalanobis distance

- Regard a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and a point $\boldsymbol{x} \in \mathbb{R}^d$.



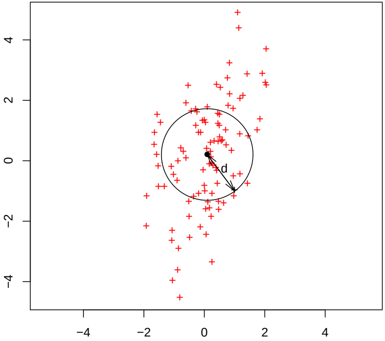
- Euclidean distance from $\boldsymbol{x}$ to $\boldsymbol{\mu}_{\boldsymbol{X}}$:

$$d^2_{Eucl}(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{X}}) = (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})^\top (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}}) \,.$$

- Sample mean: $\boldsymbol{\mu}_{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i$.

Mahalanobis distance

- Regard a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and a point $\boldsymbol{x} \in \mathbb{R}^d$.



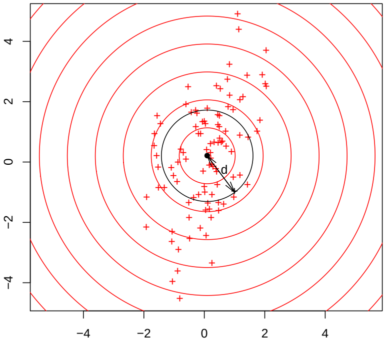
- Euclidean distance from $\boldsymbol{x}$ to $\boldsymbol{\mu}_{\boldsymbol{X}}$:

$$d^2_{Eucl}(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{X}}) = (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})^\top (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}}) .$$

- Sample mean: $\boldsymbol{\mu}_{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i$.

Mahalanobis distance

- Regard a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and a point $\boldsymbol{x} \in \mathbb{R}^d$.



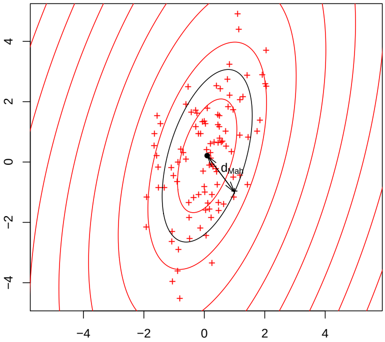
- Euclidean distance from $\boldsymbol{x}$ to $\boldsymbol{\mu}_{\boldsymbol{X}}$:

$$d^2_{Eucl}(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{X}}) = (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})^\top (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}}) \,.$$

- Sample mean: $\boldsymbol{\mu}_{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i$.

Mahalanobis distance

- Regard a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and a point $\boldsymbol{x} \in \mathbb{R}^d$.

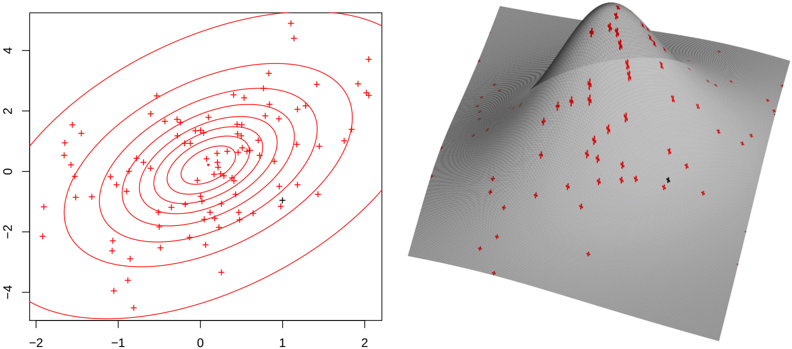


- Mahalanobis distance: $d^2_{Mah}(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{X}}; \boldsymbol{\Sigma}_{\boldsymbol{X}}) = (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})$.

- Sample mean: $\boldsymbol{\mu}_{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i$.
- Sample covariance matrix: $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \frac{1}{n-1}\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{X}})^\top$.

Mahalanobis depth (Mahalanobis, 1936)

- Regard a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and a point $\boldsymbol{x} \in \mathbb{R}^d$.



- **Mahalanobis depth** of $\boldsymbol{x}$ = a *centrality measure*:

$$D^{Mah(n)}(\boldsymbol{x}|\boldsymbol{X}) = \frac{1}{1 + d^2_{Mah}(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{X}}; \boldsymbol{\Sigma}_{\boldsymbol{X}})} = \frac{1}{1 + (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})^\top \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})}$$

Mahalanobis depth (Mahalanobis, 1936)

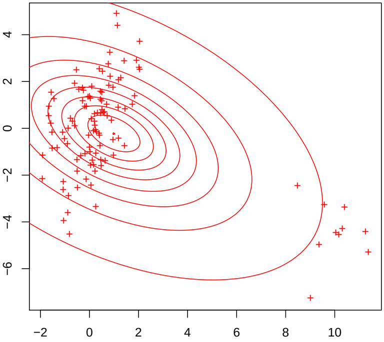
Some **disadvantages**:

- The depth contours are always **ellipses**, or more generally ellipsoids, and thus do not reflect well the underlying data geometry.

Mahalanobis depth (Mahalanobis, 1936)

Some **disadvantages**:

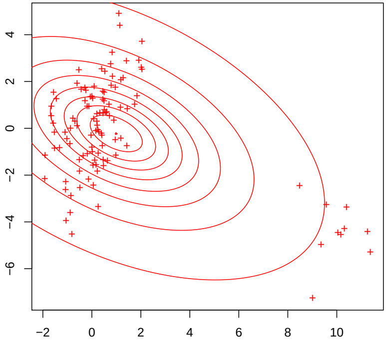
- The depth contours are always **ellipses**, or more generally ellipsoids, and thus do not reflect well the underlying data geometry.
- The traditional estimators for $\boldsymbol{\mu_X}$ and $\boldsymbol{\Sigma_X}$ are **not robust** and thus sensible to the contamination of data with outliers.



Mahalanobis depth (Mahalanobis, 1936)

Some **disadvantages**:

- The depth contours are always **ellipses**, or more generally ellipsoids, and thus do not reflect well the underlying data geometry.
- The traditional estimators for $\boldsymbol{\mu_X}$ and $\boldsymbol{\Sigma_X}$ are **not robust** and thus sensible to the contamination of data with outliers.



- Robust estimates for $\boldsymbol{\mu_X}$ and $\boldsymbol{\Sigma_X}$ can be used, like *minimum covariance determinant* estimator; this is not the topic of the lecture.

Contents

Tukey (halfspace, location) depth (Tukey, 1975)

- **Tukey (1975) — “Mathematics and the picturing of data”**

Tukey (halfspace, location) depth (Tukey, 1975)

- **Tukey (1975) — "Mathematics and the picturing of data"**

- Tukey depth of $\boldsymbol{x} \in \mathbb{R}^d$ with respect to a d-variate random vector X distributed as P is defined as the smallest probability mass of a closed halfspace containing $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}.$$

Tukey (halfspace, location) depth (Tukey, 1975)

- **Tukey (1975) — "Mathematics and the picturing of data"**
- Tukey depth of $\boldsymbol{x} \in \mathbb{R}^d$ with respect to a d-variate random vector X distributed as P is defined as the smallest probability mass of a closed halfspace containing $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

- To obtain the sample version, one substitutes the empirical measure instead of P.

Tukey (halfspace, location) depth (Tukey, 1975)

- **Tukey (1975) — "Mathematics and the picturing of data"**

- Tukey depth of $\boldsymbol{x} \in \mathbb{R}^d$ with respect to a d-variate random vector X distributed as P is defined as the smallest probability mass of a closed halfspace containing $\boldsymbol{x}$:

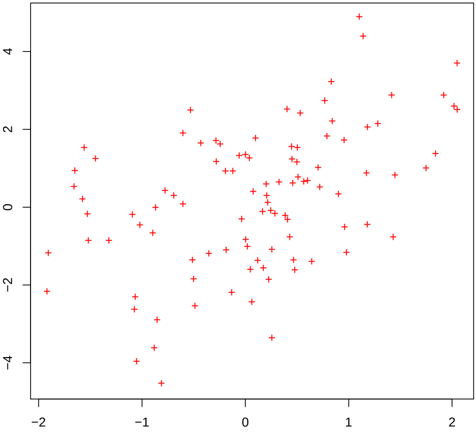
$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

- To obtain the sample version, one substitutes the empirical measure instead of P.

- Thus, Tukey depth of $\boldsymbol{x}$ with respect to a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$ can be formulated as:

$$D^{Tuk(n)}(\boldsymbol{x}|\boldsymbol{X}) = \frac{1}{n} \min_{\boldsymbol{u} \in S^{d-1}} \sharp\{i : \boldsymbol{u}'\boldsymbol{x}_i \geq \boldsymbol{u}'\boldsymbol{x}\}.$$

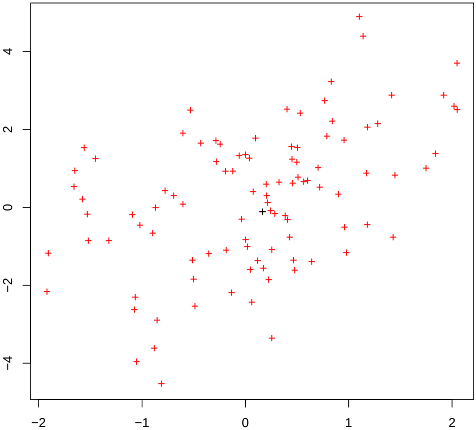
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H)\ :\ H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

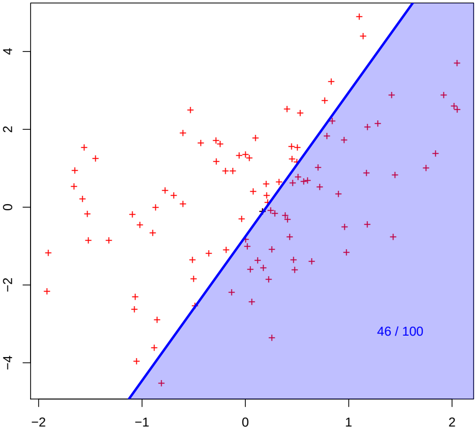
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H)\ :\ H \ \text{is a closed halfspace,}\ \boldsymbol{x} \in H\}\,.$$

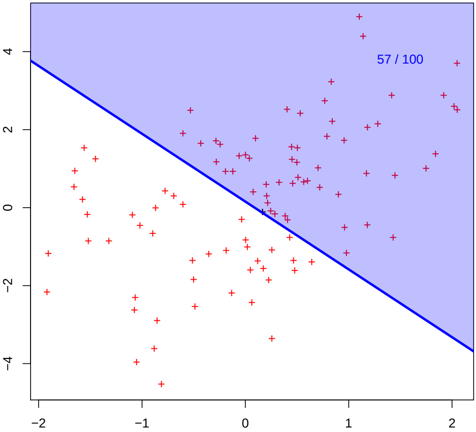
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \;:\; H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

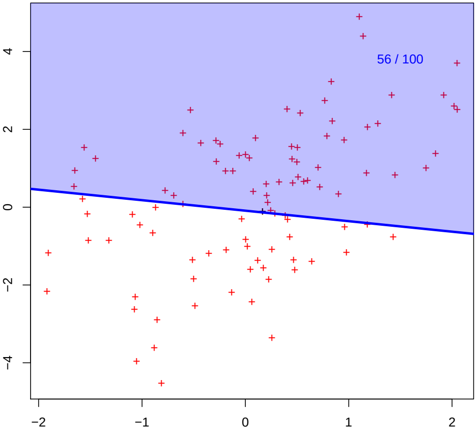
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \; : \; H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

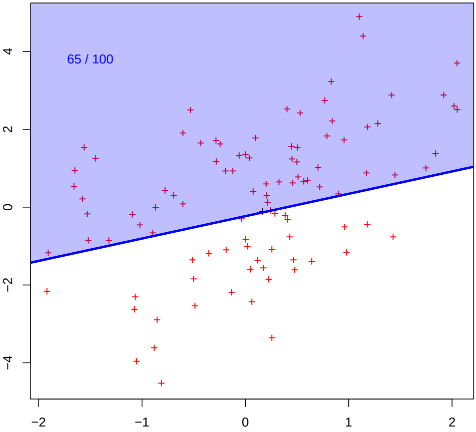
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \; : \; H \; \text{is a closed halfspace,} \; \boldsymbol{x} \in H\}\,.$$

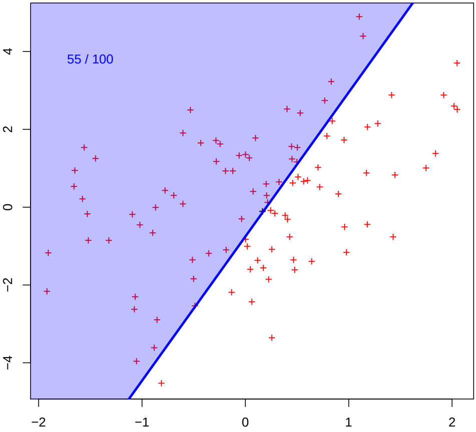
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \;:\; H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

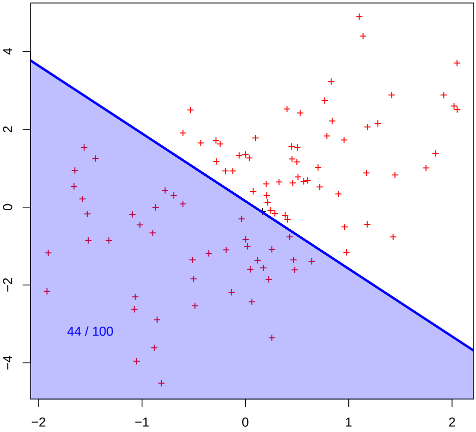
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \; : \; H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

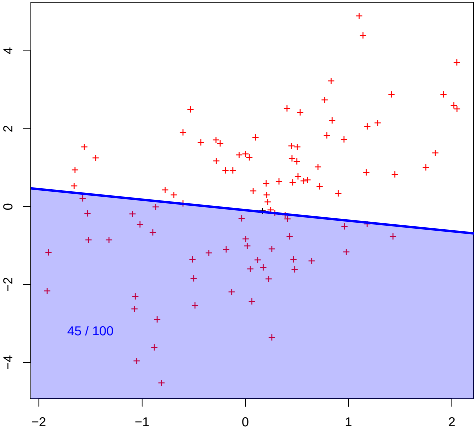
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

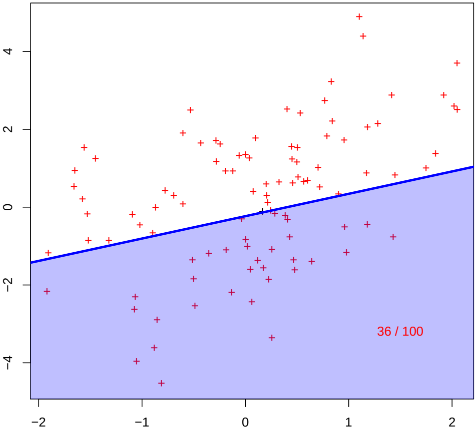
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

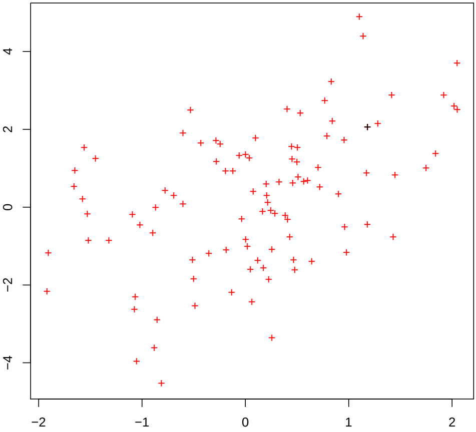
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H)\ :\ H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

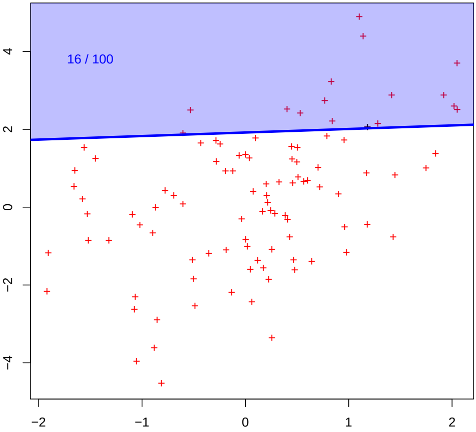
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \, : \, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}.$$

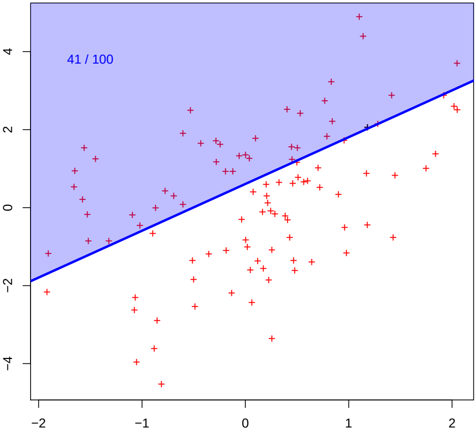
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \; : \; H \text{ is a closed halfspace, } \boldsymbol{x} \in H\} \, .$$

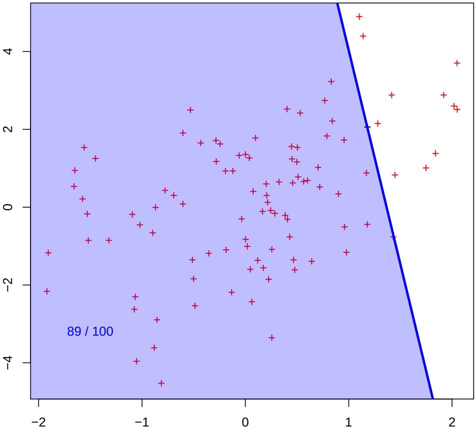
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \;:\; H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

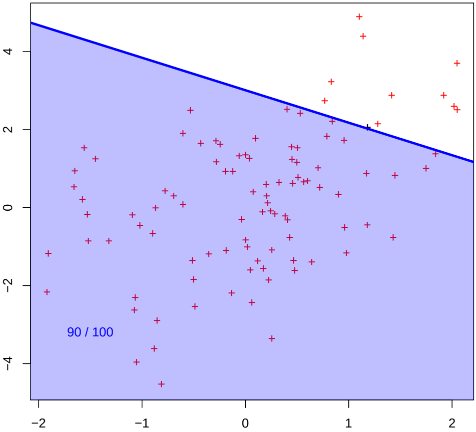
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

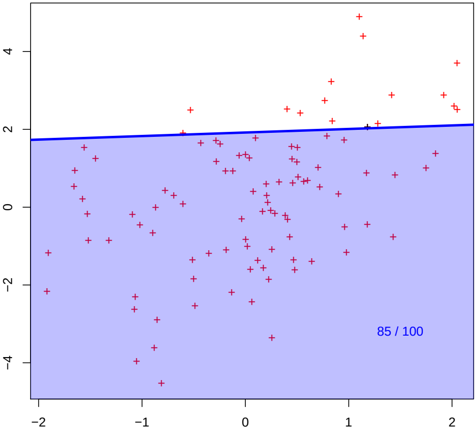
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H)\ :\ H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

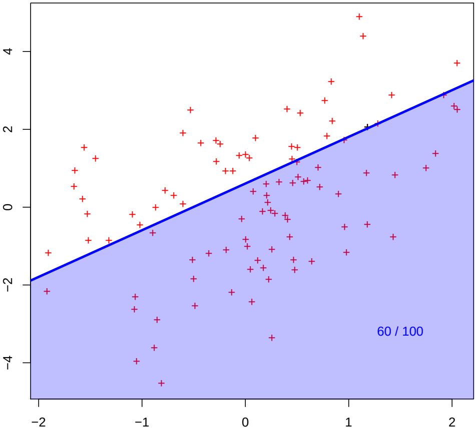
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

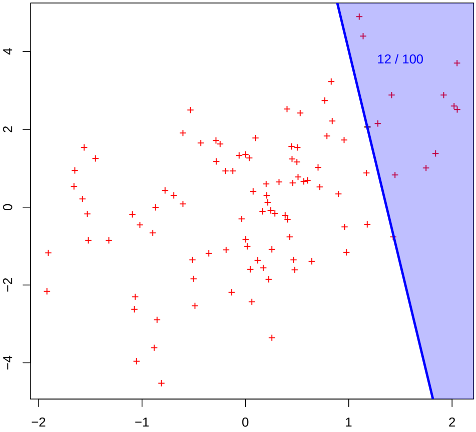
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

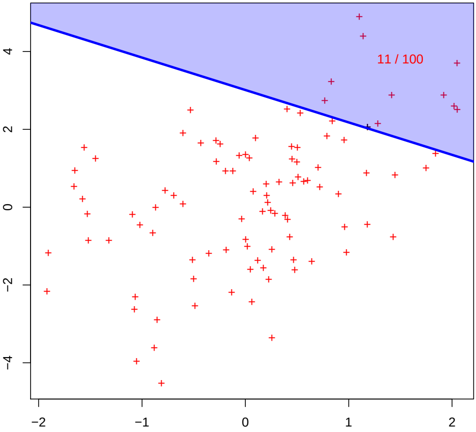
Tukey (halfspace, location) depth (Tukey, 1975)



- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H)\ :\ H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

Tukey (halfspace, location) depth

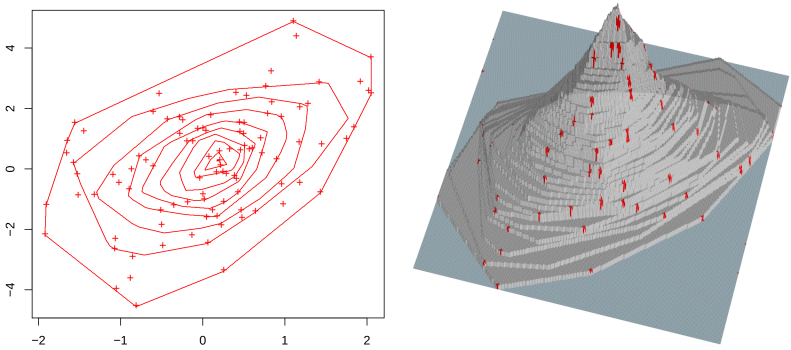


- **Tueky depth** of $\boldsymbol{x}$:

$$D^{Tuk}(\boldsymbol{x}|X) = \inf\{P(H) \,:\, H \text{ is a closed halfspace, } \boldsymbol{x} \in H\}\,.$$

Tukey (halfspace, location) depth (Tukey, 1975)

- Depth function for a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$.



- **Tueky depth** of $\boldsymbol{x}$:

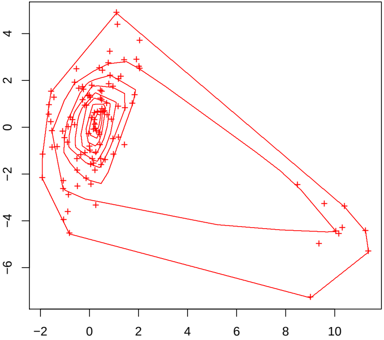
$$D^{Tuk(n)}(\boldsymbol{x}|\boldsymbol{X}) = \frac{1}{n} \min_{\boldsymbol{u} \in S^{d-1}} \sharp\{i : \boldsymbol{u}'\boldsymbol{x}_i \geq \boldsymbol{u}'\boldsymbol{x}\}.$$

Tukey (halfspace, location) depth (Tukey, 1975)

- The depth contours adapt to the observations and thus reflect the underlying data geometry.

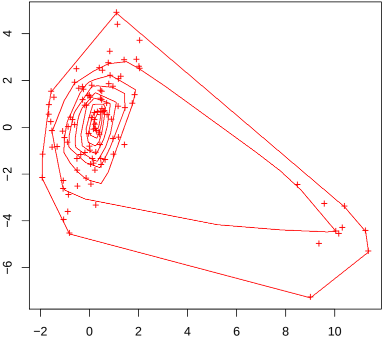
Tukey (halfspace, location) depth (Tukey, 1975)

- The depth contours adapt to the observations and thus reflect the underlying data geometry.
- The depth is robust with respect to outliers and ignores them for deeper contours.



Tukey (halfspace, location) depth (Tukey, 1975)

- The depth contours adapt to the observations and thus reflect the underlying data geometry.
- The depth is robust with respect to outliers and ignores them for deeper contours.



- The depth is entirely non-parametric since is based on the information about halfspaces only.

Contents

Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

Koshevoy & Mosler (1997) and Mosler (2002) define the **zonoid trimmed region** as:

$$\begin{aligned} D_\alpha^{zon}(P) &= \Big\{ \int_{\mathbb{R}^d} \mathbf{x} g(\mathbf{x}) dP(\mathbf{x}) : \\ & \qquad g : \mathbb{R}^d \mapsto \left[0, \frac{1}{\alpha}\right] \text{ measurable and } \int_{\mathbb{R}^d} g(\mathbf{x}) dP(\mathbf{x}) = 1 \Big\} \end{aligned}$$

for $\alpha \in (0, 1]$,

Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

Koshevoy & Mosler (1997) and Mosler (2002) define the **zonoid trimmed region** as:

$$
\begin{aligned}
D_\alpha^{zon}(P) &= \Big\{ \int_{\mathbb{R}^d} \mathbf{x} g(\mathbf{x}) dP(\mathbf{x}) : \\
&\qquad g : \mathbb{R}^d \mapsto \left[0, \frac{1}{\alpha}\right] \text{ measurable and } \int_{\mathbb{R}^d} g(\mathbf{x}) dP(\mathbf{x}) = 1 \Big\}
\end{aligned}
$$

for $\alpha \in (0, 1]$, and

$$
D_0^{zon}(P) = \mathrm{cl}\Big(\bigcup_{\alpha \in (0,1]} D_\alpha(P) \Big)
$$

for $\alpha = 0$, where "cl" denotes closure.

Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

Koshevoy & Mosler (1997) and Mosler (2002) define the **zonoid trimmed region** as:

$$
\begin{aligned}
D_\alpha^{zon}(P) &= \Big\{ \int_{\mathbb{R}^d} \mathbf{x} g(\mathbf{x}) dP(\mathbf{x}) : \\
&\qquad g : \mathbb{R}^d \mapsto \left[0, \frac{1}{\alpha}\right] \text{ measurable and } \int_{\mathbb{R}^d} g(\mathbf{x}) dP(\mathbf{x}) = 1 \Big\}
\end{aligned}
$$

for $\alpha \in (0, 1]$, and

$$
D_0^{zon}(P) = \mathrm{cl}\Big(\bigcup_{\alpha \in (0,1]} D_\alpha(P) \Big)
$$

for $\alpha = 0$, where "cl" denotes closure.

Zonoid depth is then defined as:

$$
D^{zon}(\boldsymbol{x}|X) = \begin{cases} \sup\{\alpha : \boldsymbol{x} \in D_\alpha^{zon}(X)\} & \text{if } \boldsymbol{x} \in \mathrm{conv}\big(\mathrm{supp}(X)\big), \\ 0 & \text{otherwise}. \end{cases}
$$

Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

Zoonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

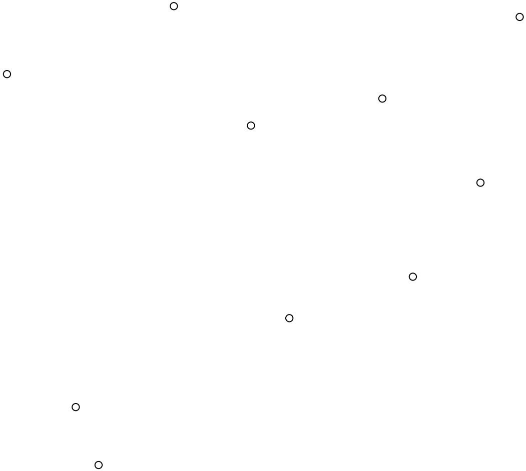
For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

$$D_{\alpha}^{zon(n)}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$

Zoonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

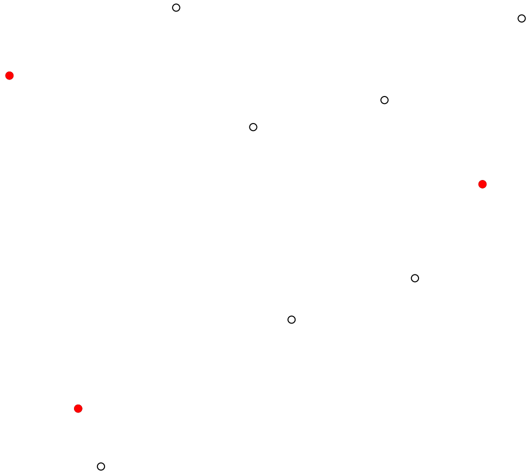
$$D_{\alpha}^{zon(n)}(\boldsymbol{X}) = \mathrm{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zoonid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

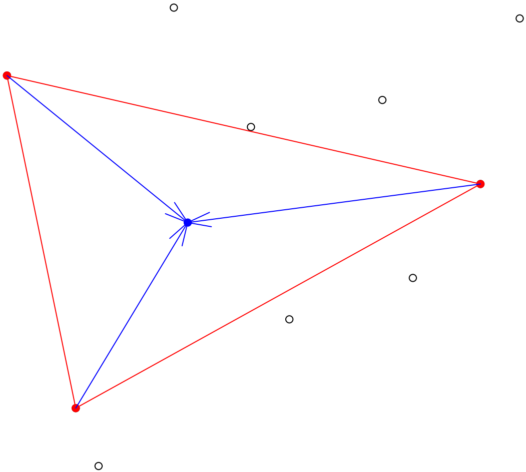
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \operatorname{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \right\} :$$



Zoonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

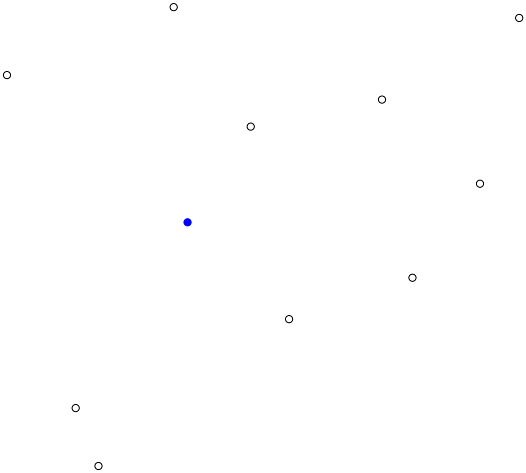
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \text{conv}\Big\{\frac{1}{\alpha n}\sum_{j=1}^{k}\mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N\Big\} :$$



Zoonid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

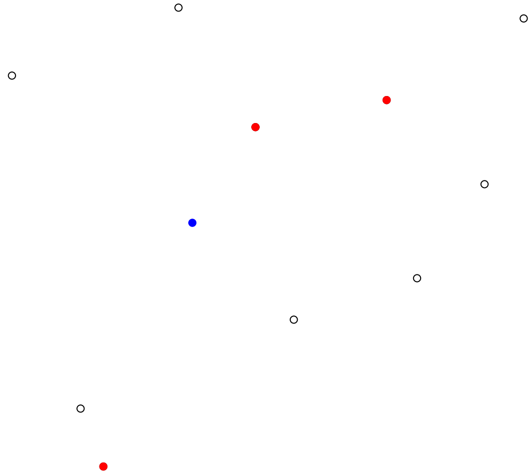
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \right\} :$$



Zoonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

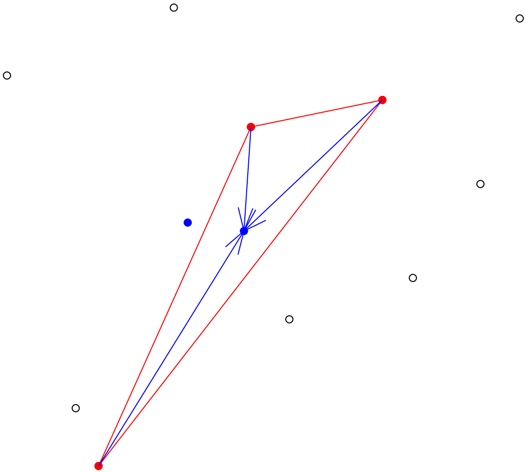
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \operatorname{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big)\mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

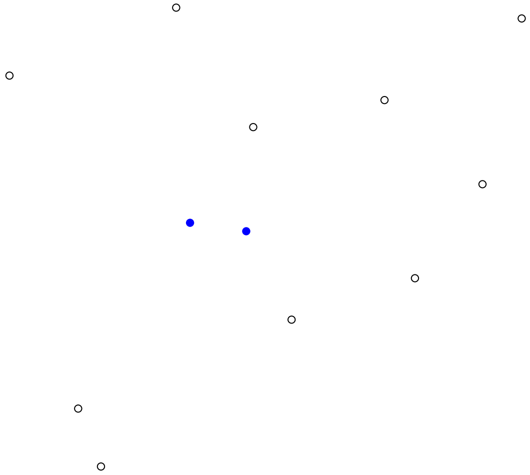
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \right\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

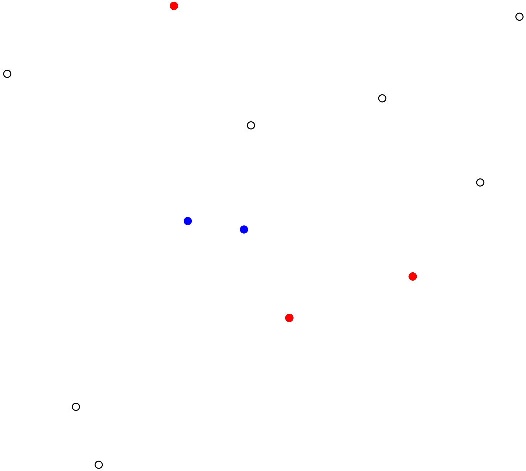
$$D_{\alpha}^{zon(n)}(\boldsymbol{X}) = \operatorname{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \right\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

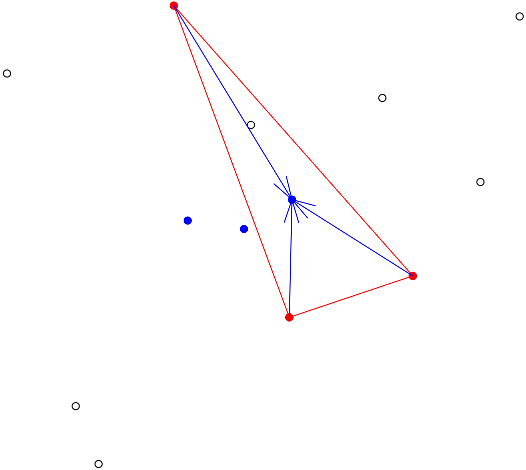
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \mathrm{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big) \mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \Big\} :$$



Zoonid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

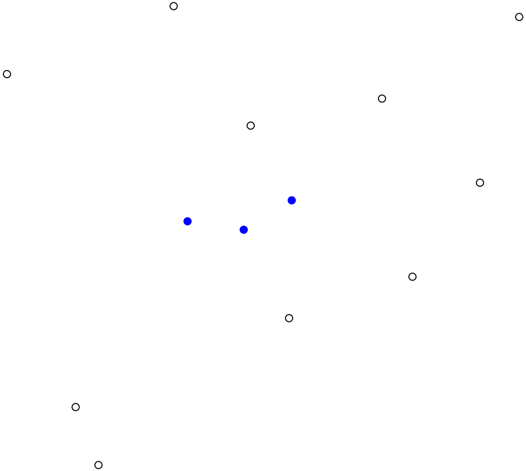
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

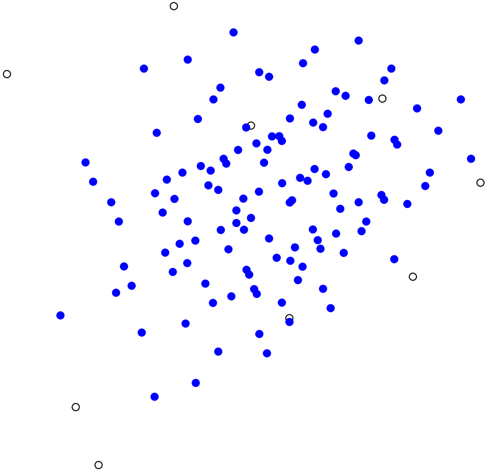
$$D_{\alpha}^{zon(n)}(\boldsymbol{X}) = \text{conv}\left\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \right\} :$$



Zonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

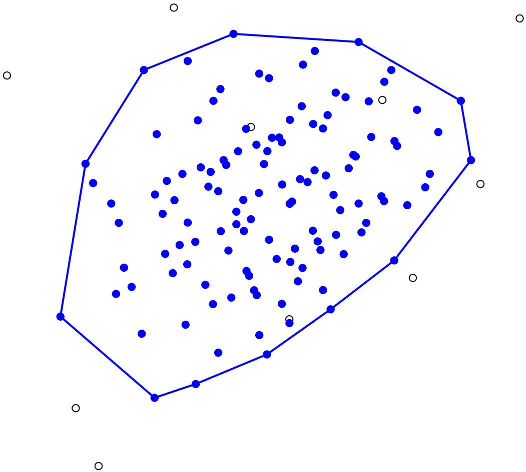
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \operatorname{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big) \mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \Big\} :$$



Zoonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

For $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, \ldots, n-1$, $N = \{1, \ldots, n\}$

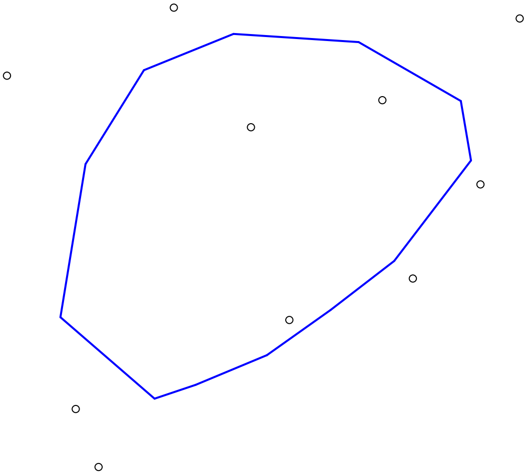
$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \mathrm{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \Big(1 - \frac{k}{\alpha n}\Big) \mathbf{x}_{i_{k+1}} : \{i_1, \ldots, i_{k+1}\} \subset N \Big\} :$$



Zoonoid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

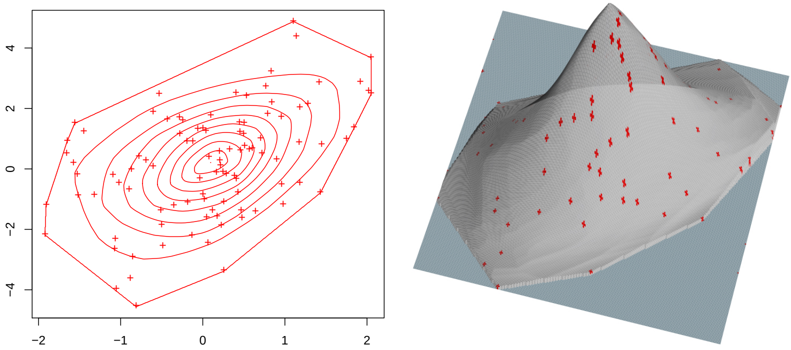
For $\boldsymbol{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, $\alpha \in [\frac{k}{n}, \frac{k+1}{n}]$, $k = 1, ..., n-1$, $N = \{1, ..., n\}$

$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \text{conv}\Big\{ \frac{1}{\alpha n} \sum_{j=1}^{k} \mathbf{x}_{i_j} + \left(1 - \frac{k}{\alpha n}\right) \mathbf{x}_{i_{k+1}} : \{i_1, ..., i_{k+1}\} \subset N \Big\} :$$



Zoonid depth (Koshevoy & Mosler, 1997; Mosler, 2002)

- Depth function for a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$.



- **Zonoid depth** of $\boldsymbol{x}$:

$$D^{zon(n)}(\boldsymbol{x}|\boldsymbol{X}) = \begin{cases} \sup\{\alpha : \boldsymbol{x} \in D_{\alpha}^{zon(n)}(\boldsymbol{X})\} & \text{if } \boldsymbol{x} \in \mathrm{conv}(\boldsymbol{X}), \\ 0 & \text{otherwise}. \end{cases}$$

Zonoid depth (Dyckerhoff, Koshevoy, Mosler, 1996):

- Zonoid depth can be defined as:

$$D^{zon(n)}(\boldsymbol{x}|\boldsymbol{X}) = \sup\{\alpha \in [0,1] : \boldsymbol{x} \in D_\alpha^{zon(n)}(\boldsymbol{X})\},$$

where sup of $\emptyset$ is defined to be 0,

Zonoid depth (Dyckerhoff, Koshevoy, Mosler, 1996):

- Zonoid depth can be defined as:

$$D^{zon(n)}(\boldsymbol{x}|\boldsymbol{X}) = \sup\{\alpha \in [0,1] : \boldsymbol{x} \in D_\alpha^{zon(n)}(\boldsymbol{X})\},$$

where sup of $\emptyset$ is defined to be 0, with:

$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \left\{\sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \ : \ \sum_{i=1}^{n} \lambda_i = 1, 0 \leq \lambda_i, \alpha\lambda_i \leq \frac{1}{n} \text{ for all } i\right\}.$$

Zonoid depth (Dyckerhoff, Koshevoy, Mosler, 1996):

- Zonoid depth can be defined as:

$$D^{zon(n)}(\boldsymbol{x}|\boldsymbol{X}) = \sup\{\alpha \in [0,1] : \boldsymbol{x} \in D_\alpha^{zon(n)}(\boldsymbol{X})\},$$

where sup of $\emptyset$ is defined to be 0, with:

$$D_\alpha^{zon(n)}(\boldsymbol{X}) = \left\{\sum_{i=1}^n \lambda_i \boldsymbol{x}_i \,:\, \sum_{i=1}^n \lambda_i = 1, 0 \le \lambda_i, \alpha\lambda_i \le \frac{1}{n} \text{ for all } i\right\}.$$

- $D^{zon(n)}(\boldsymbol{x}|\boldsymbol{X}) = \frac{1}{n\gamma^*}$ with γ^* being an optimizer of:

$$\begin{aligned} \min \gamma \quad \text{s. t.} \quad \boldsymbol{X}^T\boldsymbol{\lambda} &= \boldsymbol{x}, \\ \boldsymbol{\lambda}^T\mathbf{1}_n &= 1, \\ \gamma\mathbf{1}_n - \boldsymbol{\lambda} &\ge \mathbf{0}_n, \\ \boldsymbol{\lambda} &\ge \mathbf{0}_n. \end{aligned}$$

(Notation: $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^T$, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)^T$.)

Contents

Projection depth (Zuo & Serfling, 2000)

- According to Zuo & Serfling (2000), **projection depth** is defined as:

$$D^{prj}(\boldsymbol{x}|X) = \frac{1}{1 + O^{prj}(\boldsymbol{x}|X)} ,$$

Projection depth (Zuo & Serfling, 2000)

- According to Zuo & Serfling (2000), **projection depth** is defined as:

$$D^{prj}(\boldsymbol{x}|X) = \frac{1}{1 + O^{prj}(\boldsymbol{x}|X)},$$

- where

$$O^{prj}(\boldsymbol{x}|X) = \sup_{\boldsymbol{u} \in S^{d-1}} \frac{|\boldsymbol{x}^T \boldsymbol{u} - \mathrm{med}(X^T \boldsymbol{u})|}{\mathrm{MAD}(X^T \boldsymbol{u})}$$

is the **projected outlyingness** (Stahel, 1981; Donoho, 1982),

Projection depth (Zuo & Serfling, 2000)

- According to Zuo & Serfling (2000), **projection depth** is defined as:

$$D^{prj}(\boldsymbol{x}|X) = \frac{1}{1 + O^{prj}(\boldsymbol{x}|X)} \,,$$

- where

$$O^{prj}(\boldsymbol{x}|X) = \sup_{\boldsymbol{u} \in S^{d-1}} \frac{|\boldsymbol{x}^T \boldsymbol{u} - \mathrm{med}(X^T \boldsymbol{u})|}{\mathrm{MAD}(X^T \boldsymbol{u})}$$

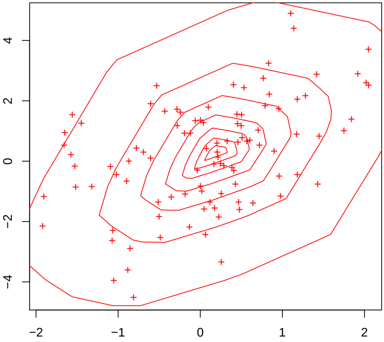
  is the **projected outlyingness** (Stahel, 1981; Donoho, 1982),

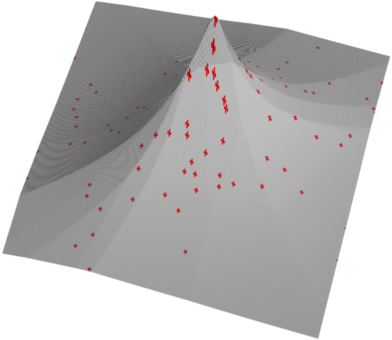
- $\mathrm{med}(Y)$ is the *univariate median*,

- $\mathrm{MAD}(Y) = \mathrm{med}\big(\big|Y - \mathrm{med}(Y)\big|\big)$ is the *median absolute deviation from the median*.

Projection depth (Zuo & Serfling, 2000)

- Depth function for a data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \subset \mathbb{R}^d$.





- **Projection depth** of $\boldsymbol{x}$:

$$D^{prj(n)}(\boldsymbol{x}|\boldsymbol{X}) = \inf_{\boldsymbol{u} \in S^{d-1}} \frac{1}{1 + \frac{|\boldsymbol{x}^T \boldsymbol{u} - \text{med}(\boldsymbol{X}\boldsymbol{u})|}{\text{MAD}(\boldsymbol{X}\boldsymbol{u})}} .$$

Contents

Contents

Statistical data depth (Mosler, 2013)

A **data depth** measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to [0, 1], (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

Statistical data depth (Mosler, 2013)

A **data depth** measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to [0, 1], (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

that is:

D1 **translation invariant:** $D(\boldsymbol{x} + b|X + b) = D(\boldsymbol{x}|X)$ for any $b \in \mathbb{R}^d$;

Statistical data depth (Mosler, 2013)

A **data depth** measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to [0,1], (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

that is:

D1 **translation invariant:** $D(\boldsymbol{x} + b|X + b) = D(\boldsymbol{x}|X)$ for any $b \in \mathbb{R}^d$;

D2 **linear invariant:** $D(A\boldsymbol{x}|AX) = D(\boldsymbol{x}|X)$ for any $d \times d$ non-singular matrix A;

Statistical data depth (Mosler, 2013)

A **data depth** measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to [0,1], (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

that is:

D1 **translation invariant:** $D(\boldsymbol{x} + b|X + b) = D(\boldsymbol{x}|X)$ for any $b \in \mathbb{R}^d$;

D2 **linear invariant:** $D(A\boldsymbol{x}|AX) = D(\boldsymbol{x}|X)$ for any $d \times d$ non-singular matrix A;

D3 **vanishing at infinity:** $\lim_{||\boldsymbol{x}||\to\infty} D(\boldsymbol{x}|X) = 0$;

Statistical data depth (Mosler, 2013)

A **data depth** measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to [0,1], (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

that is:

D1 **translation invariant:** $D(\boldsymbol{x} + b|X + b) = D(\boldsymbol{x}|X)$ for any $b \in \mathbb{R}^d$;

D2 **linear invariant:** $D(A\boldsymbol{x}|AX) = D(\boldsymbol{x}|X)$ for any $d \times d$ non-singular matrix A;

D3 **vanishing at infinity:** $\lim_{||\boldsymbol{x}|| \to \infty} D(\boldsymbol{x}|X) = 0$;

D4 **monotone on rays:** for any $\boldsymbol{x}^* \in \arg\max_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}|X)$, any $\boldsymbol{x} \in \mathbb{R}^d$, and any $0 \leq \alpha \leq 1$ it holds: $D(\boldsymbol{x}|X) \leq D(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)|X)$;

Statistical data depth (Mosler, 2013)

A **data depth** measures how "close" a given point is located to the "center" of a distribution. For $\boldsymbol{x} \in \mathbb{R}^d$ and a d-variate random vector X distributed as $P \in \mathcal{P}$, a data depth is a function

$$D : \mathbb{R}^d \times \mathcal{P} \to [0,1], (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}|P)$$

that is:

D1 **translation invariant:** $D(\boldsymbol{x} + b|X + b) = D(\boldsymbol{x}|X)$ for any $b \in \mathbb{R}^d$;

D2 **linear invariant:** $D(A\boldsymbol{x}|AX) = D(\boldsymbol{x}|X)$ for any $d \times d$ non-singular matrix A;

D3 **vanishing at infinity:** $\lim_{||\boldsymbol{x}|| \to \infty} D(\boldsymbol{x}|X) = 0$;

D4 **monotone on rays:** for any $\boldsymbol{x}^* \in \arg\max_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}|X)$, any $\boldsymbol{x} \in \mathbb{R}^d$, and any $0 \leq \alpha \leq 1$ it holds: $D(\boldsymbol{x}|X) \leq D(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)|X)$;

D5 **upper semicontinuous in $\boldsymbol{x}$:** the upper-level sets $D_\alpha(X) = \{\boldsymbol{x} \in \mathbb{R}^d : D(\boldsymbol{x}|X) \geq \alpha\}$ are closed for all α.

Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.

Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.
- **D4** implies star-shaped upper-level sets of D.

Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.

- **D4** implies star-shaped upper-level sets of D.

  One can strengthen to:

  - **D4con**: $D(\cdot|X)$ is a **quasiconcave** function, *i.e.* the upper-level sets $D_\alpha(X)$ are convex for all α.

Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.

- **D4** implies star-shaped upper-level sets of D.

  One can strengthen to:

  - **D4con**: $D(\cdot|X)$ is a **quasiconcave** function, *i.e.* the upper-level sets $D_\alpha(X)$ are convex for all α.

- **D1** and **D2** define **affine invariante depth**.

Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.

- **D4** implies star-shaped upper-level sets of D.

  One can strengthen to:

  - **D4con**: $D(\cdot|X)$ is a **quasiconcave** function, *i.e.* the upper-level sets $D_\alpha(X)$ are convex for all α.

- **D1** and **D2** define **affine invariante depth**.

  One can also weaken to:

  - **D2iso**: $D(A\mathbf{x}|AX) = D(\mathbf{x}|X)$ for every isometric linear A to define **orthogonal invariant depth**;

Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.

- **D4** implies star-shaped upper-level sets of D.

  One can strengthen to:

  - **D4con**: $D(\cdot|X)$ is a **quasiconcave** function, *i.e.* the upper-level sets $D_\alpha(X)$ are convex for all α.

- **D1** and **D2** define **affine invariante depth**.

  One can also weaken to:

  - **D2iso**: $D(A\mathbf{x}|AX) = D(\mathbf{x}|X)$ for every isometric linear A to define **orthogonal invariant depth**;

  - **D2sca**: $D(\lambda\mathbf{x}|\lambda X) = D(\mathbf{x}|X)$ for any $\lambda > 0$ to define **scale invariant depth**.

Statistical data depth (Mosler, 2013)

Some remarks:

- **D1 – D5** implies due to Dyckerhoff (2004):
  If X is centrally symmetric about $\boldsymbol{x}^*$, then any depth function satisfying D1 – D5 $D(\cdot|X)$ is maximal at $\boldsymbol{x}^*$.

- **D4** implies star-shaped upper-level sets of D.

  One can strengthen to:

  - **D4con**: $D(\cdot|X)$ is a **quasiconcave** function, *i.e.* the upper-level sets $D_\alpha(X)$ are convex for all α.

- **D1** and **D2** define **affine invariante depth**.

  One can also weaken to:

  - **D2iso**: $D(A\mathbf{x}|AX) = D(\mathbf{x}|X)$ for every isometric linear A to define **orthogonal invariant depth**;

  - **D2sca**: $D(\lambda\mathbf{x}|\lambda X) = D(\mathbf{x}|X)$ for any $\lambda > 0$ to define **scale invariant depth**.

- **Mahalanobis**, **Tukey**, **zonoid** and **projection** depths **satisfy all D1 – D5** properties.

Contents

Central regions

- For given P and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.

Central regions

- For given P and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.
- The innermost region arises at some depth $\alpha_{\text{max}} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of **deepest points**.

Central regions

- For given P and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.
- The innermost region arises at some depth $\alpha_{\text{max}} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of **deepest points**.
- Central regions describe distribution w.r.t. location, dispersion, and shape.

Central regions

- For given P and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.
- The innermost region arises at some depth $\alpha_{\max} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of **deepest points**.
- Central regions describe distribution w.r.t. location, dispersion, and shape.
- Properties of central regions, for any α:
  - Due to **D1** and **D2** $D_\alpha(X)$ is **affine equivariant**: $D_\alpha(AX + b) = AD_\alpha(X) + b$ for any $d \times d$ non-singular matrix A and any $b \in \mathbb{R}^d$;

Central regions

- For given P and $\alpha \in [0,1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.
- The innermost region arises at some depth $\alpha_{\max} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of **deepest points**.
- Central regions describe distribution w.r.t. location, dispersion, and shape.
- Properties of central regions, for any α:
  - Due to **D1** and **D2** $D_\alpha(X)$ is **affine equivariant**: $D_\alpha(AX + b) = AD_\alpha(X) + b$ for any $d \times d$ non-singular matrix A and any $b \in \mathbb{R}^d$;
  - Due to **D3** $D_\alpha(X)$ is bounded;

Central regions

- For given P and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.
- The innermost region arises at some depth $\alpha_{\max} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of **deepest points**.
- Central regions describe distribution w.r.t. location, dispersion, and shape.
- Properties of central regions, for any α:
  - Due to **D1** and **D2** $D_\alpha(X)$ is **affine equivariant**: $D_\alpha(AX + b) = AD_\alpha(X) + b$ for any $d \times d$ non-singular matrix A and any $b \in \mathbb{R}^d$;
  - Due to **D3** $D_\alpha(X)$ is bounded;
  - Due to **D4** $D_\alpha(X)$-s are nested: if $\alpha \geq \beta$, then $D_\alpha(X) \subseteq D_\beta(X)$, and star-shaped;

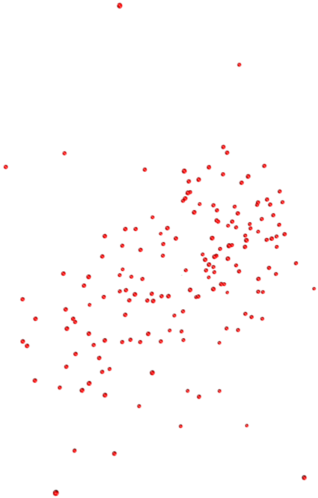
Central regions

- For given P and $\alpha \in [0, 1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.
- The innermost region arises at some depth $\alpha_{\max} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of **deepest points**.
- Central regions describe distribution w.r.t. location, dispersion, and shape.
- Properties of central regions, for any α:
  - Due to **D1** and **D2** $D_\alpha(X)$ is **affine equivariant**:
    $D_\alpha(AX + b) = AD_\alpha(X) + b$ for any $d \times d$ non-singular matrix A and any $b \in \mathbb{R}^d$;
  - Due to **D3** $D_\alpha(X)$ is bounded;
  - Due to **D4** $D_\alpha(X)$-s are nested:
    if $\alpha \geq \beta$, then $D_\alpha(X) \subseteq D_\beta(X)$, and star-shaped;
    due to **D4con** $D_\alpha(X)$ is in addition convex;

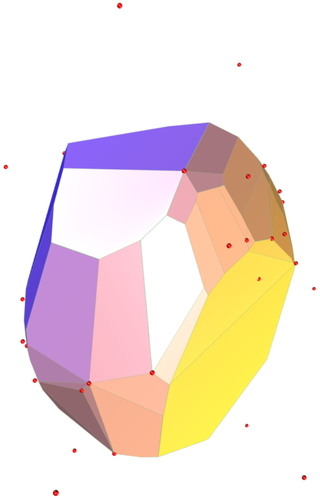
Central regions

- For given P and $\alpha \in [0,1]$, the level sets $D_\alpha(P)$ form a family of **depth-trimmed** of **central regions**.
- The innermost region arises at some depth $\alpha_{\mathsf{max}} \leq 1$, which depends on the depth notion D and distribution P. Then $D_\alpha(X)$ is the set of **deepest points**.
- Central regions describe distribution w.r.t. location, dispersion, and shape.
- Properties of central regions, for any α:
  - Due to **D1** and **D2** $D_\alpha(X)$ is **affine equivariant**: $D_\alpha(AX+b) = AD_\alpha(X)+b$ for any $d \times d$ non-singular matrix A and any $b \in \mathbb{R}^d$;
  - Due to **D3** $D_\alpha(X)$ is bounded;
  - Due to **D4** $D_\alpha(X)$-s are nested: if $\alpha \geq \beta$, then $D_\alpha(X) \subseteq D_\beta(X)$, and star-shaped; due to **D4con** $D_\alpha(X)$ is in addition convex;
  - Due to **D5** $D_\alpha(X)$ is closed.

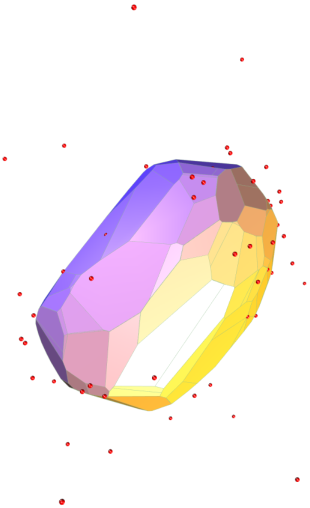
Tukey (halfspace, location) depth region



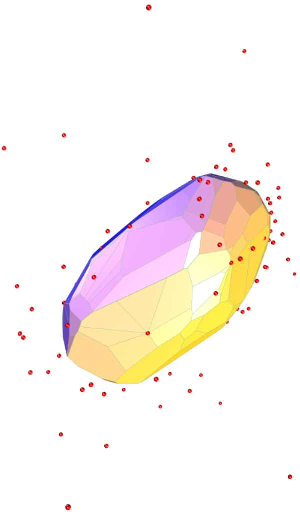
Tukey (halfspace, location) depth region: $\alpha = 2/161$



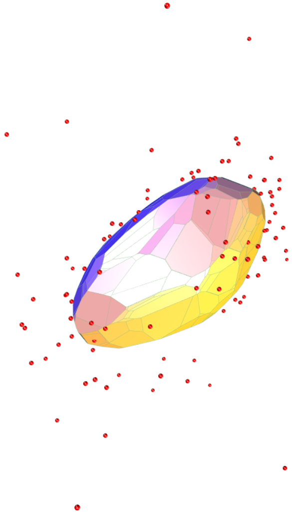
Tukey (halfspace, location) depth region: $\alpha = 5/161$



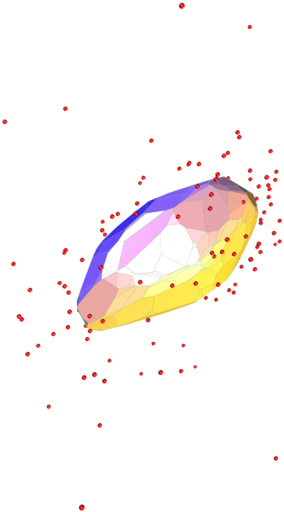
Tukey (halfspace, location) depth region: $\alpha = 9/161$



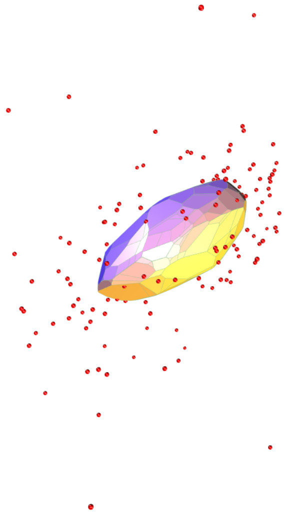
Tukey (halfspace, location) depth region: $\alpha = 13/161$



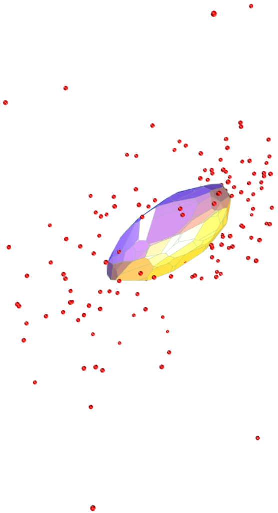
Tukey (halfspace, location) depth region: $\alpha = 17/161$



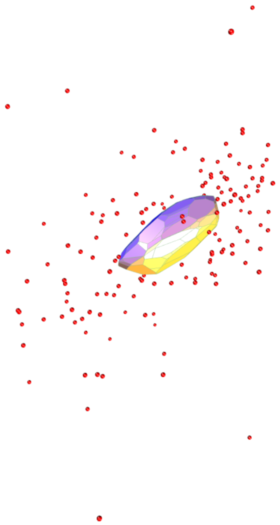
Tukey (halfspace, location) depth region: $\alpha = 25/161$



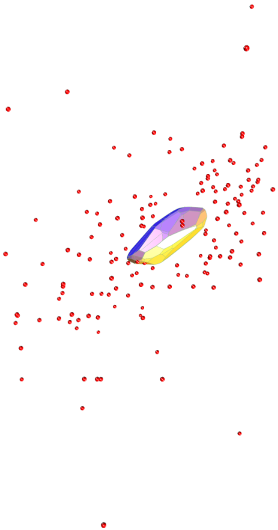
Tukey (halfspace, location) depth region: $\alpha = 33/161$



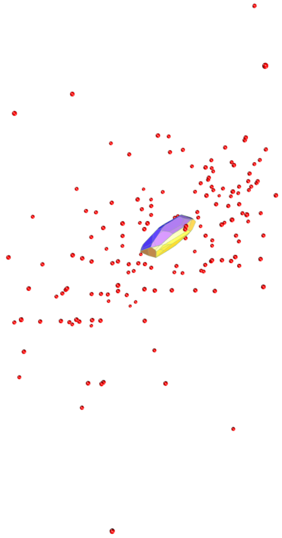
Tukey (halfspace, location) depth region: $\alpha = 41/161$



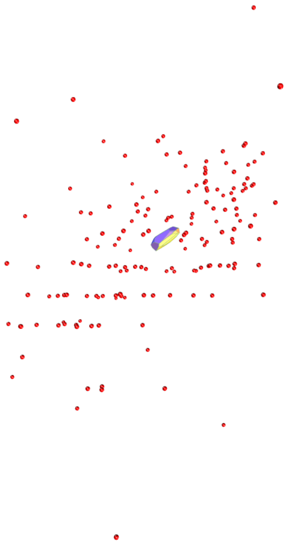
Tukey (halfspace, location) depth region: $\alpha = 49/161$



Tukey (halfspace, location) depth region: $\alpha = 57/161$



Tukey (halfspace, location) depth region: $\alpha = 65/161$



Tukey (halfspace, location) depth region: $\alpha = 68/161$

Applications

- **Multivariate data analysis** (Liu, Parelius, Singh '99);
- **Statistical quality control** (Liu, Singh '93);
- **Cluster analysis and classification** (Mosler, Hoberg '06; Li, Cuesta-Albertos, Liu '12; Mozharovskyi, Mosler, Lange '15);
- **Tests for multivariate location, scale, symmetry** (Liu '92; Dyckerhoff '02; Dyckerhoff, Ley, Paindaveine '15);
- **Outlier/anomaly detection** (Hubert, Rousseeuw, Segaert '15);
- **Multivariate risk measurement** (Cascos, Mochalov '07);
- **Robust linear programming** (Bazovkin, Mosler '15);
- **Missing data imputation** (Mozharovskyi, Josse, Husson '18);
- etc.

R-package **ddalpha** (Pokotylo, Mozharovskyi, Dyckerhoff, Nagy):
calculates a number of multivariate and functional depths;
performs depth-based classification of multivariate and functional data;
contains 50 multivariate and 5 functional data sets.

Contents

Projection property (Dyckerhoff, 2004)

- Let D be a depth. D satisfies the (**weak**) **projection property**, if for each point $\boldsymbol{x} \in \mathbb{R}^d$ and each random vector X it holds:

$$D(\boldsymbol{x}|X) = \inf_{\boldsymbol{u} \in S^{d-1}} D(\boldsymbol{u}^T \boldsymbol{x} | \boldsymbol{u}^T X) \, .$$

Projection property (Dyckerhoff, 2004)

- Let D be a depth. D satisfies the (**weak**) **projection property**, if for each point $\boldsymbol{x} \in \mathbb{R}^d$ and each random vector X it holds:

$$D(\boldsymbol{x}|X) = \inf_{\boldsymbol{u} \in S^{d-1}} D(\boldsymbol{u}^T\boldsymbol{x}|\boldsymbol{u}^T X)\,.$$

- If a depth satisfies the weak projection property, then for each direction $\boldsymbol{u} \in S^{d-1}$ we have:

$$D(\boldsymbol{x}|X) \leq D(\boldsymbol{u}^T\boldsymbol{x}|\boldsymbol{u}^T X)\,.$$

Projection property (Dyckerhoff, 2004)

- Let D be a depth. D satisfies the (**weak**) **projection property**, if for each point $\boldsymbol{x} \in \mathbb{R}^d$ and each random vector X it holds:

$$D(\boldsymbol{x}|X) = \inf_{\boldsymbol{u} \in S^{d-1}} D(\boldsymbol{u}^T\boldsymbol{x}|\boldsymbol{u}^T X)\,.$$

- If a depth satisfies the weak projection property, then for each direction $\boldsymbol{u} \in S^{d-1}$ we have:

$$D(\boldsymbol{x}|X) \leq D(\boldsymbol{u}^T\boldsymbol{x}|\boldsymbol{u}^T X)\,.$$

- If a depth D is **quasiconcave** (*i.e.* satisfies **D4con**), we say that D is a **convex depth**.

Projection property (Dyckerhoff, 2004)

- Let D be a depth. D satisfies the (**weak**) **projection property**, if for each point $\boldsymbol{x} \in \mathbb{R}^d$ and each random vector X it holds:

$$D(\boldsymbol{x}|X) = \inf_{\boldsymbol{u} \in S^{d-1}} D(\boldsymbol{u}^T \boldsymbol{x} | \boldsymbol{u}^T X) \,.$$

- If a depth satisfies the weak projection property, then for each direction $\boldsymbol{u} \in S^{d-1}$ we have:

$$D(\boldsymbol{x}|X) \leq D(\boldsymbol{u}^T \boldsymbol{x} | \boldsymbol{u}^T X) \,.$$

- If a depth D is **quasiconcave** (*i.e.* satisfies **D4con**), we say that D is a **convex depth**.

- Let D be a depth that satisfies the weak projection property. Then D is a convex depth.

Projection property (Dyckerhoff, 2004)

- Let D be a depth. D satisfies the (**weak**) **projection property**, if for each point $\boldsymbol{x} \in \mathbb{R}^d$ and each random vector X it holds:

$$D(\boldsymbol{x}|X) = \inf_{\boldsymbol{u} \in S^{d-1}} D(\boldsymbol{u}^T \boldsymbol{x}|\boldsymbol{u}^T X)\,.$$

- If a depth satisfies the weak projection property, then for each direction $\boldsymbol{u} \in S^{d-1}$ we have:

$$D(\boldsymbol{x}|X) \leq D(\boldsymbol{u}^T \boldsymbol{x}|\boldsymbol{u}^T X)\,.$$

- If a depth D is **quasiconcave** (*i.e.* satisfies **D4con**), we say that D is a **convex depth**.

- Let D be a depth that satisfies the weak projection property. Then D is a convex depth.

- **Mahalanobis**, **Tukey**, **zonoid**, **projetion** depths **satisfy** the weak projection property.

Contents

Contents

Maximum depth classifier (Ghosh and Chaudhuri, 2005)

- For X, let $X_1, ..., X_Q$ be the Q conditional random vectors of Q classes with $\boldsymbol{X}_1, ..., \boldsymbol{X}_Q$ being the corresponding (thus labeled) training samples containing $n_1, ..., n_Q$ observations, respectively, and let $\boldsymbol{n} = (n_1, ..., n_Q)$.

Maximum depth classifier (Ghosh and Chaudhuri, 2005)

- For X, let $X_1, ..., X_Q$ be the Q conditional random vectors of Q classes with $\boldsymbol{X}_1, ..., \boldsymbol{X}_Q$ being the corresponding (thus labeled) training samples containing $n_1, ..., n_Q$ observations, respectively, and let $\boldsymbol{n} = (n_1, ..., n_Q)$.
- The **maximum depth classifier** is constructed as:

$$cl_{\boldsymbol{n}}^{MD}(\boldsymbol{x}) = \underset{q \in \{1,...,Q\}}{\arg\max}\, D^{(n)}(\boldsymbol{x}|\boldsymbol{X}_q)\,.$$

Maximum depth classifier (Ghosh and Chaudhuri, 2005)

- For X, let $X_1, ..., X_Q$ be the Q conditional random vectors of Q classes with $\boldsymbol{X}_1, ..., \boldsymbol{X}_Q$ being the corresponding (thus labeled) training samples containing $n_1, ..., n_Q$ observations, respectively, and let $\boldsymbol{n} = (n_1, ..., n_Q)$.
- The **maximum depth classifier** is constructed as:

$$cl_{\boldsymbol{n}}^{MD}(\boldsymbol{x}) = \operatorname*{arg\,max}_{q \in \{1,...,Q\}} D^{(n)}(\boldsymbol{x}|\boldsymbol{X}_q)\,.$$

- π_q being priors, define its misclassification rate as

$$\Delta_{\boldsymbol{n}} = \sum_{q=1}^{Q} \pi_q P\big(cl_{\boldsymbol{n}}^{MD}(X) \neq q \,:\, X \in q\text{-th class}\big)\,.$$

Maximum depth classifier (Ghosh and Chaudhuri, 2005)

- For X, let $X_1, ..., X_Q$ be the Q conditional random vectors of Q classes with $\boldsymbol{X}_1, ..., \boldsymbol{X}_Q$ being the corresponding (thus labeled) training samples containing $n_1, ..., n_Q$ observations, respectively, and let $\boldsymbol{n} = (n_1, ..., n_Q)$.
- The **maximum depth classifier** is constructed as:

$$cl_{\boldsymbol{n}}^{MD}(\boldsymbol{x}) = \underset{q \in \{1,...,Q\}}{\arg\max}\ D^{(n)}(\boldsymbol{x}|\boldsymbol{X}_q)\,.$$

- π_q being priors, define its misclassification rate as

$$\Delta_{\boldsymbol{n}} = \sum_{q=1}^{Q} \pi_q P\big(cl_{\boldsymbol{n}}^{MD}(X) \neq q \,:\, X \in q\text{-th class}\big)\,.$$

- Suppose the densities $f_{X_1}, ..., f_{X_Q}$ are elliptically symmetric, and $f_{X_q}(\boldsymbol{x}) = g(\boldsymbol{x} - \boldsymbol{\mu}_q)$ for some location parameters $\boldsymbol{\mu}_q$ and a common density function g with $g(k\boldsymbol{x}) \leq g(\boldsymbol{x})$ for every $\boldsymbol{x}$ and $k > 1$. Then, in equal prior cases, for D^{Tuk} and D^{prj}, $\Delta_{\boldsymbol{n}}$ converges to the optimal (Bayes) risk as $\min\{n_1, ..., n_Q\} \to \infty$.

Contents

DD-plot (Li, Cuesta-Albertos, Liu, 2012)

Given: $\boldsymbol{X}_0 = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_m\}$ from P_0 and $\boldsymbol{X}_1 = \{\boldsymbol{x}_{m+1}, ..., \boldsymbol{x}_{m+n}\}$ from P_1,

DD-plot (Li, Cuesta-Albertos, Liu, 2012)

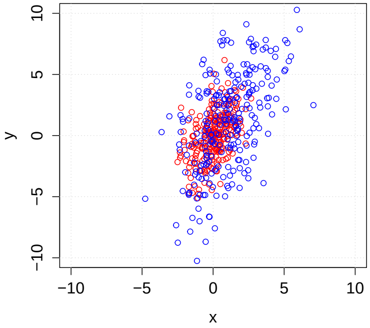
Given: $\boldsymbol{X}_0 = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_m\}$ from P_0 and $\boldsymbol{X}_1 = \{\boldsymbol{x}_{m+1}, ..., \boldsymbol{x}_{m+n}\}$ from P_1, consider the *DD*-plot (Li, Cuesta-Albertos, Liu, 2012),

$$Z = \{\boldsymbol{z}_i | \boldsymbol{z}_i = \big(D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_0), D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_1)\big), i = 1, ..., m+n\}.$$

DD-plot (Li, Cuesta-Albertos, Liu, 2012)

Given: $\boldsymbol{X}_0 = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_m\}$ from P_0 and $\boldsymbol{X}_1 = \{\boldsymbol{x}_{m+1}, ..., \boldsymbol{x}_{m+n}\}$ from P_1, consider the *DD*-plot (Li, Cuesta-Albertos, Liu, 2012),

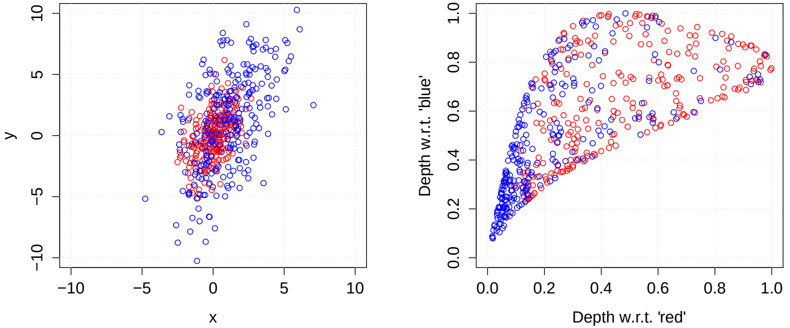
$$Z = \{\boldsymbol{z}_i | \boldsymbol{z}_i = \left(D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_0), D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_1)\right), i = 1, ..., m+n\}.$$



DD-plot (Li, Cuesta-Albertos, Liu, 2012)

Given: $\boldsymbol{X}_0 = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_m\}$ from P_0 and $\boldsymbol{X}_1 = \{\boldsymbol{x}_{m+1}, ..., \boldsymbol{x}_{m+n}\}$ from P_1, consider the *DD*-plot (Li, Cuesta-Albertos, Liu, 2012),

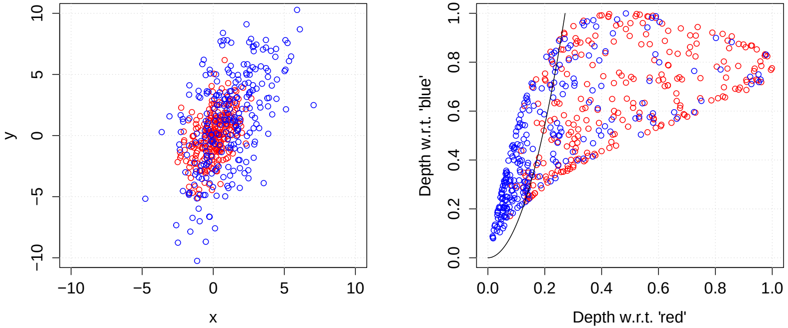
$$Z = \{\boldsymbol{z}_i | \boldsymbol{z}_i = \left(D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_0), D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_1)\right), i = 1, ..., m+n\}.$$



DD-plot (Li, Cuesta-Albertos, Liu, 2012)

Given: $\boldsymbol{X}_0 = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_m\}$ from P_0 and $\boldsymbol{X}_1 = \{\boldsymbol{x}_{m+1}, ..., \boldsymbol{x}_{m+n}\}$ from P, consider the *DD*-plot (Li, Cuesta-Albertos, Liu, 2012),

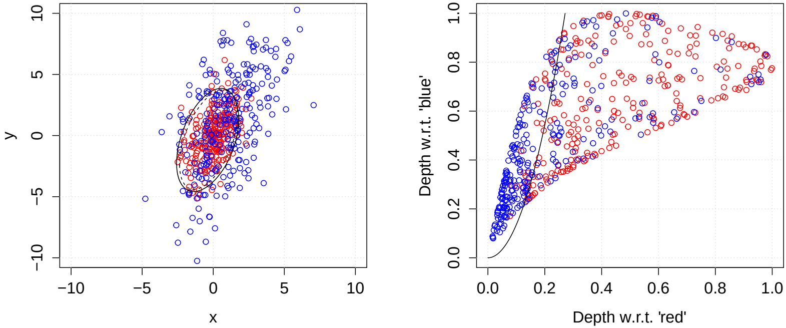
$$Z = \{\boldsymbol{z}_i | \boldsymbol{z}_i = \left(D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_0), D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_1)\right), i = 1, ..., m+n\}.$$



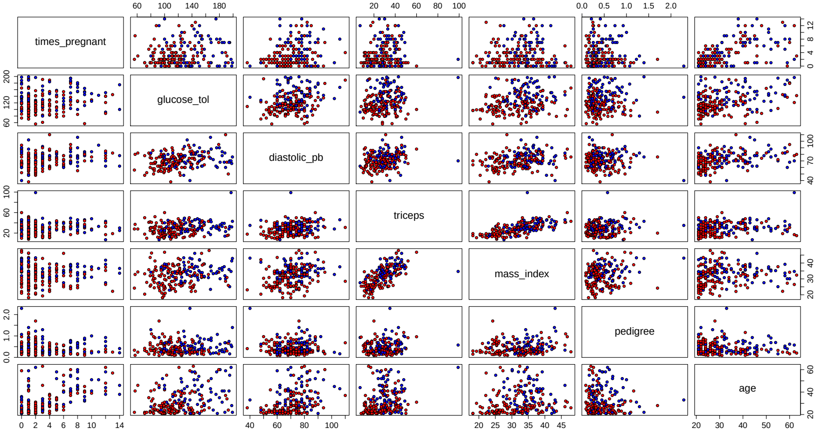
DD-plot (Li, Cuesta-Albertos, Liu, 2012)

Given: $\boldsymbol{X}_0 = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_m\}$ from P_0 and $\boldsymbol{X}_1 = \{\boldsymbol{x}_{m+1}, ..., \boldsymbol{x}_{m+n}\}$ from P, consider the *DD*-plot (Li, Cuesta-Albertos, Liu, 2012),

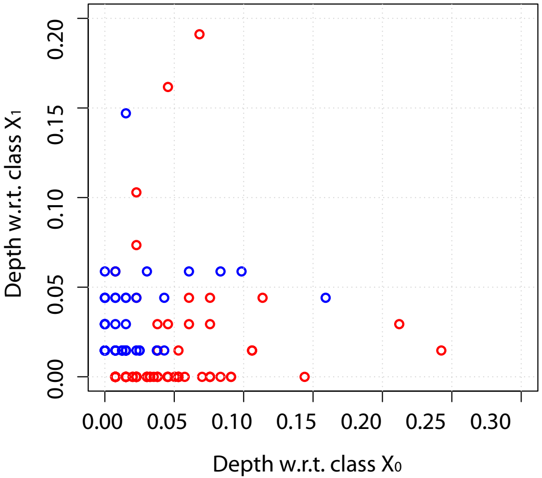
$$Z = \{\boldsymbol{z}_i | \boldsymbol{z}_i = \left(D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_0), D^{(n)}(\boldsymbol{x}_i|\boldsymbol{X}_1)\right), i = 1, ..., m + n\}.$$



Pima Indians Diabetes (Subset: $m + n = 200$, $d = 7$)



Pima Indians Diabetes: *DD*-Plot



Pima Indians Diabetes: Extended *DD*-Plot

With $q = 2$ classes (*DD*-plot) and maximum power $p = 2$ we obtain $r = 5$ features.

Pima Indians Diabetes: Extended *DD*-Plot

With $q = 2$ classes (*DD*-plot) and maximum power $p = 2$ we obtain $r = 5$ features.

In this case $Z = \{\boldsymbol{z}_i | \boldsymbol{z}_i = (D_{\boldsymbol{X}_0}(\boldsymbol{x}_i), D_{\boldsymbol{X}_1}(\boldsymbol{x}_i), D_{\boldsymbol{X}_0}(\boldsymbol{x}_i) \cdot D_{\boldsymbol{X}_1}(\boldsymbol{x}_i), D^2_{\boldsymbol{X}_0}(\boldsymbol{x}_i), D^2_{\boldsymbol{X}_1}(\boldsymbol{x}_i)); i = 1, ..., m+n\}$.

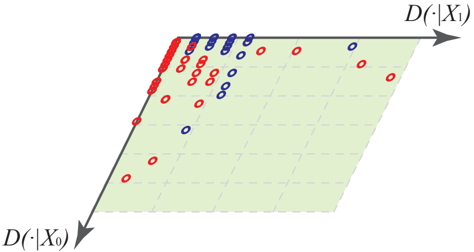
Pima Indians Diabetes: Extended *DD*-Plot

With $q = 2$ classes (*DD*-plot) and maximum power $p = 2$ we obtain $r = 5$ features.

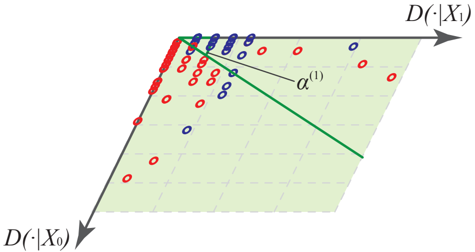
In this case $Z = \{\boldsymbol{z}_i | \boldsymbol{z}_i = (D_{\boldsymbol{X}_0}(\boldsymbol{x}_i), D_{\boldsymbol{X}_1}(\boldsymbol{x}_i), D_{\boldsymbol{X}_0}(\boldsymbol{x}_i) \cdot D_{\boldsymbol{X}_1}(\boldsymbol{x}_i), D^2_{\boldsymbol{X}_0}(\boldsymbol{x}_i), D^2_{\boldsymbol{X}_1}(\boldsymbol{x}_i)); i = 1, ..., m+n\}$.

Object number	Extended properties				
	p_1 $D_{\boldsymbol{X}_0}(x_i)$	p_2 $D_{\boldsymbol{X}_1}(x_i)$	p_3 $D_{\boldsymbol{X}_0}(x_i) \cdot D_{\boldsymbol{X}_1}(x_i)$	p_4 $D^2_{\boldsymbol{X}_0}(x_i)$	p_5 $D^2_{\boldsymbol{X}_1}(x_i)$
1	$D_{\boldsymbol{X}_0}(x_1)$	$D_{\boldsymbol{X}_1}(x_1)$	$D_{\boldsymbol{X}_0}(x_1) \cdot D_{\boldsymbol{X}_1}(x_1)$	$D^2_{\boldsymbol{X}_0}(x_1)$	$D^2_{\boldsymbol{X}_1}(x_1)$
2	$D_{\boldsymbol{X}_0}(x_2)$	$D_{\boldsymbol{X}_1}(x_2)$	$D_{\boldsymbol{X}_0}(x_2) \cdot D_{\boldsymbol{X}_1}(x_2)$	$D^2_{\boldsymbol{X}_0}(x_2)$	$D^2_{\boldsymbol{X}_1}(x_2)$
...					
i	$D_{\boldsymbol{X}_0}(x_i)$	$D_{\boldsymbol{X}_1}(x_i)$	$D_{\boldsymbol{X}_0}(x_i) \cdot D_{\boldsymbol{X}_1}(x_i)$	$D^2_{\boldsymbol{X}_0}(x_i)$	$D^2_{\boldsymbol{X}_1}(x_i)$
...					
$m+n$	$D_{\boldsymbol{X}_0}(x_{m+n})$	$D_{\boldsymbol{X}_1}(x_{m+n})$	$D_{\boldsymbol{X}_0}(x_{m+n}) \cdot D_{\boldsymbol{X}_1}(x_{m+n})$	$D^2_{\boldsymbol{X}_0}(x_{m+n})$	$D^2_{\boldsymbol{X}_1}(x_{m+n})$

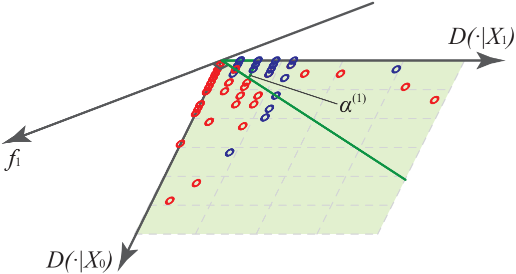
Pima Indians Diabetes: $DD\alpha$-classifier



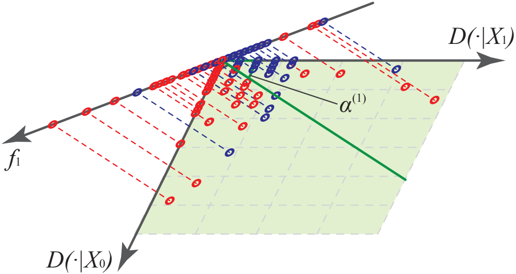
Pima Indians Diabetes: $DD\alpha$-classifier



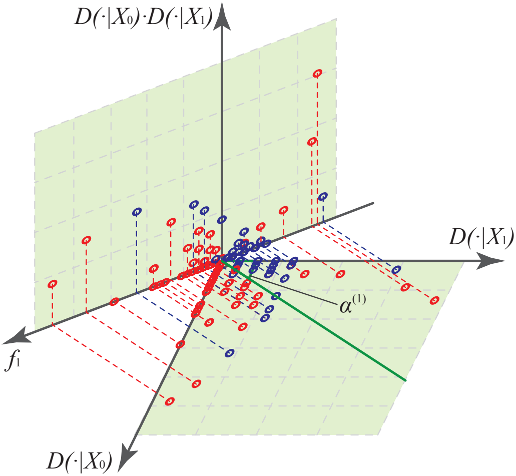
Pima Indians Diabetes: $DD\alpha$-classifier



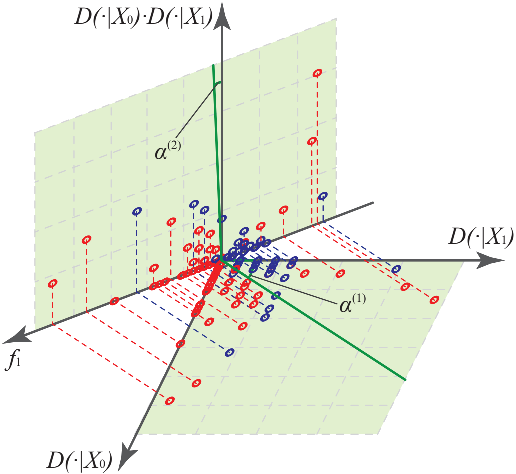
Pima Indians Diabetes: $DD\alpha$-classifier



Pima Indians Diabetes: $DD\alpha$-classifier



Pima Indians Diabetes: $DD\alpha$-classifier



$DD\alpha$-classifier: Remarks

- The **low-dimensional depth space** (even with, say polynomial, extension) is **dense**, so that the empirical-risk-minimizing hyperplane can be well approximated;

$DD\alpha$-classifier: Remarks

- The **low-dimensional depth space** (even with, say polynomial, extension) is **dense**, so that the empirical-risk-minimizing hyperplane can be well approximated;
- It is robust:
  - **in sense of points:** by its empirical-risk-minimizing nature with **zero-or-one penalty**;

$DD\alpha$-classifier: Remarks

- The **low-dimensional depth space** (even with, say polynomial, extension) is **dense**, so that the empirical-risk-minimizing hyperplane can be well approximated;

- It is robust:
  - **in sense of points:** by its empirical-risk-minimizing nature with **zero-or-one penalty**;
  - **in sense of features:** its model complexity is kept low, as only (a few) most important features are selected;

$DD\alpha$-classifier: Remarks

- The **low-dimensional depth space** (even with, say polynomial, extension) is **dense**, so that the empirical-risk-minimizing hyperplane can be well approximated;
- It is robust:
  - **in sense of points:** by its empirical-risk-minimizing nature with **zero-or-one penalty**;
  - **in sense of features:** its model complexity is kept low, as only (a few) most important features are selected;
  - operating on a **compact space**: "not too much freedom to mess up".

$DD\alpha$-classifier: Remarks

- The **low-dimensional depth space** (even with, say polynomial, extension) is **dense**, so that the empirical-risk-minimizing hyperplane can be well approximated;
- It is robust:
  - **in sense of points:** by its empirical-risk-minimizing nature with **zero-or-one penalty**;
  - **in sense of features:** its model complexity is kept low, as only (a few) most important features are selected;
  - operating on a **compact space**: "not too much freedom to mess up".
- Separation **through zero**:
  - reasonable, as a point with both depths equal to zero – an **outsider** – cannot be readily classified;

$DD\alpha$-classifier: Remarks

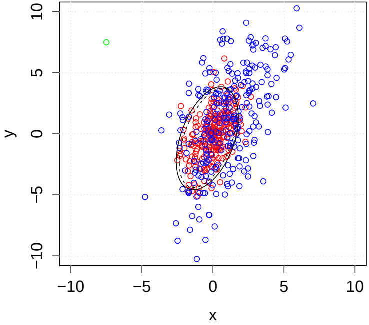
- The **low-dimensional depth space** (even with, say polynomial, extension) is **dense**, so that the empirical-risk-minimizing hyperplane can be well approximated;
- It is robust:
  - **in sense of points:** by its empirical-risk-minimizing nature with **zero-or-one penalty**;
  - **in sense of features:** its model complexity is kept low, as only (a few) most important features are selected;
  - operating on a **compact space**: "not too much freedom to mess up".
- Separation **through zero**:
  - reasonable, as a point with both depths equal to zero – an **outsider** – cannot be readily classified;
  - **fast**, as its complexity in each plane is of quick-sorting: $O\big((m+n)\log(m+n)\big)$.

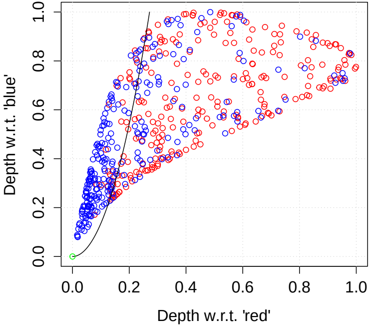
Contents

Outsiders

In the classification phase, an **outsider** is a point that have zero depth in all classes, and thus lies in the origin of the *DD*-plot.

These points need an additional **outsider treatment**.





Outsider treatments

- Avoid the problem:
  - **Ignore**/identify.

Outsider treatments

- Avoid the problem:
  - **Ignore**/identify.
  - Assign **randomly** (accounting for prior probabilities).

Outsider treatments

- Avoid the problem:
  - **Ignore**/identify.
  - Assign **randomly** (accounting for prior probabilities).

- Use an additional (simple) classifier:
  - The traditional **Linear Discriminant Analysis**.

Outsider treatments

- Avoid the problem:
  - **Ignore**/identify.
  - Assign **randomly** (accounting for prior probabilities).

- Use an additional (simple) classifier:
  - The traditional **Linear Discriminant Analysis**.
  - **Mahalanobis depth** discrimination with **cardinalities** for priors

    $$\text{class}(\boldsymbol{x}) = \arg\max_i \pi_i D_{Mah}(\boldsymbol{x}|X_i),$$

    with π_i-s being class portions. Estimates to be applied:
    - **moment** estimates or
    - **robust** estimates: (MVE, MCD).

Outsider treatments

- Avoid the problem:
  - **Ignore**/identify.
  - Assign **randomly** (accounting for prior probabilities).

- Use an additional (simple) classifier:
  - The traditional **Linear Discriminant Analysis**.
  - **Mahalanobis depth** discrimination with **cardinalities** for priors
    $$\text{class}(\boldsymbol{x}) = \arg\max_i \pi_i D_{Mah}(\boldsymbol{x}|X_i),$$
    with π_i-s being class portions. Estimates to be applied:
    - **moment** estimates or
    - **robust** estimates: (MVE, MCD).
  - **k-nearest-neighbors** discrimination using:
    - Euclidean distance normalized by Σ_X (affine invariant).

Outsider treatments

- Avoid the problem:
  - **Ignore**/identify.
  - Assign **randomly** (accounting for prior probabilities).

- Use an additional (simple) classifier:
  - The traditional **Linear Discriminant Analysis**.
  - **Mahalanobis depth** discrimination with **cardinalities** for priors
    $$\text{class}(\boldsymbol{x}) = \arg\max_i \pi_i D_{Mah}(\boldsymbol{x}|X_i),$$
    with π_i-s being class portions. Estimates to be applied:
    - **moment** estimates or
    - **robust** estimates: (MVE, MCD).
  - **k-nearest-neighbors** discrimination using:
    - Euclidean distance normalized by Σ_X (affine invariant).

- Approximate decision boundary:
  - **SVM-simplified**.

Outsider treatments: SVM-simplified. Two SVMs

Optimal hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2$ s.t.

$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1,$
$i = 1, ..., m + n$

Outsider treatments: SVM-simplified. Two SVMs

Optimal hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2$ s.t.

$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1,$
$i = 1, ..., m + n$

Maximize:

$W(\Lambda) = \Lambda'\mathbf{1} - \frac{1}{2}\Lambda'\mathbf{L}\Lambda$ w.r.t.

$\Lambda' = (\lambda_1, ..., \lambda_{m+n})$ s.t.

$\Lambda \geq \mathbf{0},$
$\Lambda' Y = 0.$

Outsider treatments: SVM-simplified. Two SVMs

Optimal hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2$ s.t.

$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1$,
$i = 1, ..., m+n$

Soft margin hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot F(\sum_{i=1}^{m+n} \xi_i^{\sigma})$ s.t.

$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1 - \xi_i, \ \xi_i \geq 0$,
$i = 1, ..., m+n$

Maximize:

$W(\Lambda) = \Lambda'\mathbf{1} - \frac{1}{2}\Lambda'\mathbf{L}\Lambda$ w.r.t.

$\Lambda' = (\lambda_1, ..., \lambda_{m+n})$ s.t.

$\Lambda \geq \mathbf{0}$,
$\Lambda' Y = 0$.

Outsider treatments: SVM-simplified. Two SVMs

Optimal hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2$ s.t.

$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1,$
$i = 1, ..., m + n$

Maximize:

$W(\Lambda) = \Lambda'\mathbf{1} - \frac{1}{2}\Lambda'\mathbf{L}\Lambda$ w.r.t.

$\Lambda' = (\lambda_1, ..., \lambda_{m+n})$ s.t.

$\Lambda \geq \mathbf{0},$
$\Lambda' Y = 0.$

Soft margin hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot F(\sum_{i=1}^{m+n} \xi_i^\sigma)$ s.t.

$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0,$
$i = 1, ..., m + n$

Maximize:

$W(\Lambda) = \Lambda'\mathbf{1} - \frac{1}{2}\Lambda'\mathbf{L}\Lambda$ w.r.t.

$\Lambda' = (\lambda_1, ..., \lambda_{m+n})$ s.t.

$\mathbf{0} \leq \Lambda \leq C\mathbf{1},$
$\Lambda' Y = 0.$

Outsider treatments: SVM-simplified. Two SVMs

Optimal hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2$ s.t.

$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \geq 1,$
$i = 1, ..., m+n$

Maximize:

$W(\Lambda) = \Lambda'\mathbf{1} - \frac{1}{2}\Lambda'\mathbf{L}\Lambda$ w.r.t.

$\Lambda' = (\lambda_1, ..., \lambda_{m+n})$ s.t.

$\Lambda \geq \mathbf{0},$
$\Lambda' Y = 0.$

Soft margin hyperplane

$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot F(\sum_{i=1}^{m+n} \xi_i^\sigma)$ s.t.

$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0,$
$i = 1, ..., m+n$

Maximize:

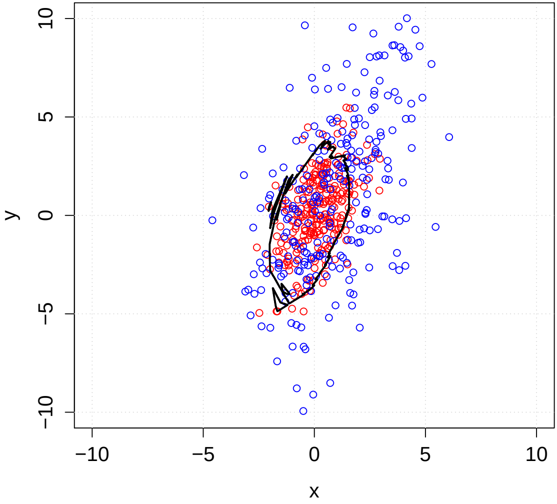
$W(\Lambda) = \Lambda'\mathbf{1} - \frac{1}{2}\Lambda'\mathbf{L}\Lambda$ w.r.t.

$\Lambda' = (\lambda_1, ..., \lambda_{m+n})$ s.t.

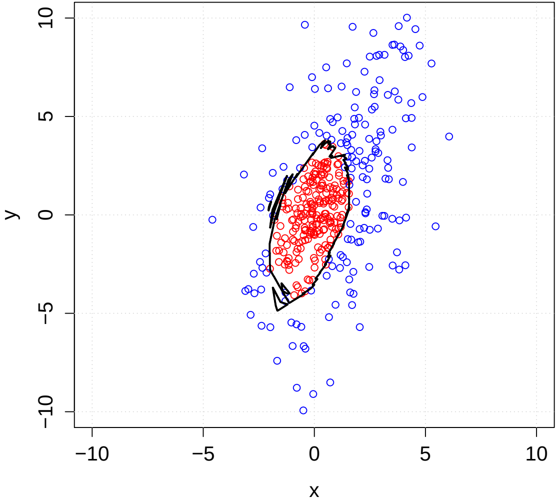
$\mathbf{0} \leq \Lambda \leq C\mathbf{1},$
$\Lambda' Y = 0.$

$$L_{ij} = y_i y_j K_\gamma(\boldsymbol{x}_i, \boldsymbol{x}_j),\ \ i, j = 1, ..., m+n;$$
$$K_\gamma(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2).$$

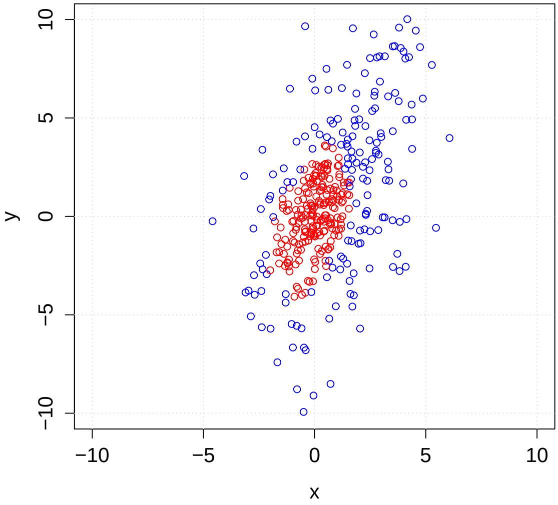
Outsider treatments: SVM-simplified



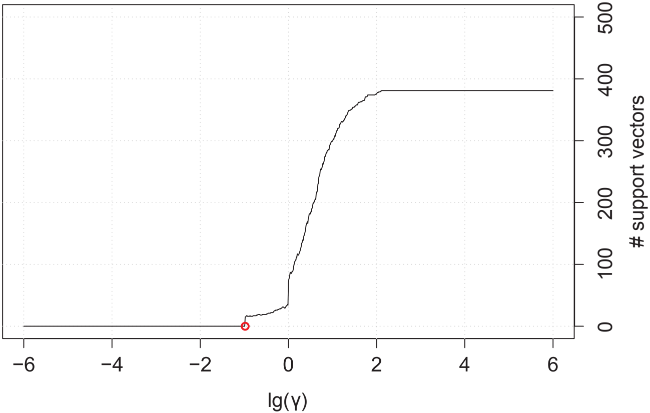
Outsider treatments: SVM-simplified



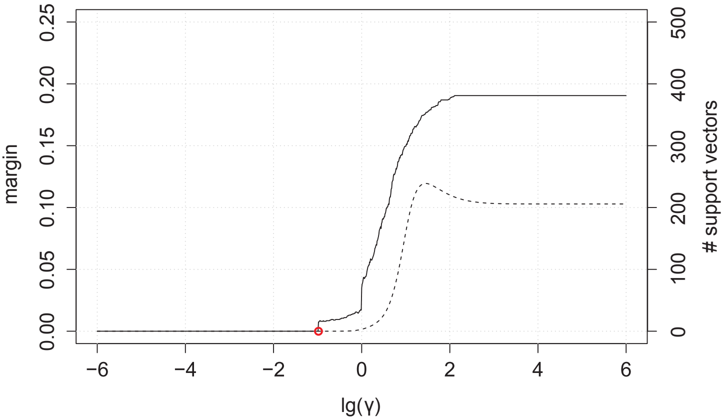
Outsider treatments: SVM-simplified



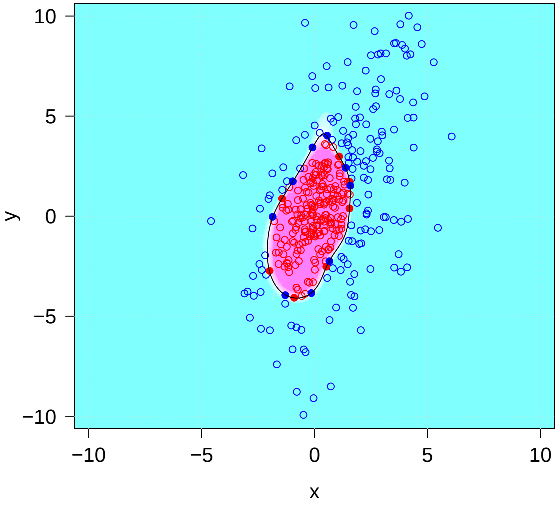
Outsider treatments: SVM-simplified



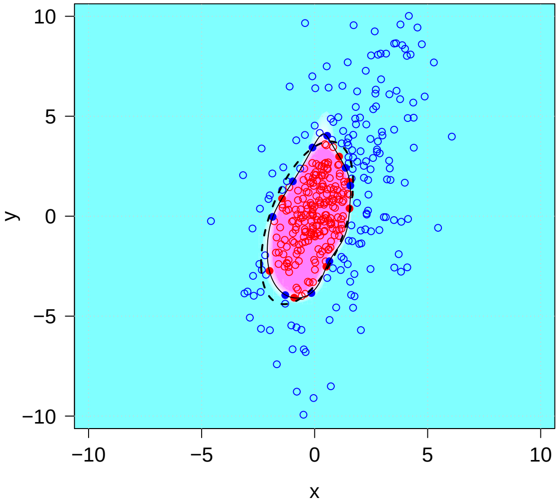
Outsider treatments: SVM-simplified



Outsider treatments: SVM-simplified



Outsider treatments: SVM-simplified



Contents

The depth kNN (Paindaveine & Van Bever, 2015)

- Let $Q = 2$.

The depth kNN (Paindaveine & Van Bever, 2015)

- Let $Q = 2$.
- For an observation $\boldsymbol{x} \in \mathbb{R}^d$, consider a new data set

$$\boldsymbol{Y} = \boldsymbol{X}_0 \cup \boldsymbol{X}_1 \cup \{2\boldsymbol{x} - \boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_0 \cup \boldsymbol{X}_1\}\,.$$

The depth kNN (Paindaveine & Van Bever, 2015)

- Let $Q = 2$.
- For an observation $\boldsymbol{x} \in \mathbb{R}^d$, consider a new data set

$$\boldsymbol{Y} = \boldsymbol{X}_0 \cup \boldsymbol{X}_1 \cup \{2\boldsymbol{x} - \boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_0 \cup \boldsymbol{X}_1\}.$$

- For $\beta = \frac{k}{n_1+n_2}$, define the **depth-based neighborhood** $R_{\boldsymbol{x}}^{\beta(\boldsymbol{n})}$ as the smallest depth region $D_\alpha^{(n)}(\boldsymbol{Y})$ (*i.e.* the one with the highest α) containing k observations from $\boldsymbol{X}_0 \cup \boldsymbol{X}_1$.

The depth kNN (Paindaveine & Van Bever, 2015)

- Let $Q = 2$.
- For an observation $\boldsymbol{x} \in \mathbb{R}^d$, consider a new data set

$$\boldsymbol{Y} = \boldsymbol{X}_0 \cup \boldsymbol{X}_1 \cup \{2\boldsymbol{x} - \boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_0 \cup \boldsymbol{X}_1\}\,.$$

- For $\beta = \frac{k}{n_1+n_2}$, define the **depth-based neighborhood** $R_{\boldsymbol{x}}^{\beta(\boldsymbol{n})}$ as the smallest depth region $D_\alpha^{(n)}(\boldsymbol{Y})$ (*i.e.* the one with the highest α) containing k observations from $\boldsymbol{X}_0 \cup \boldsymbol{X}_1$.
- Let $k_{\boldsymbol{X}_i} = \#\{\boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_i \cap R_{\boldsymbol{x}}^{\beta(\boldsymbol{n})}\}$.

The depth kNN (Paindaveine & Van Bever, 2015)

- Let $Q = 2$.
- For an observation $\boldsymbol{x} \in \mathbb{R}^d$, consider a new data set

$$\boldsymbol{Y} = \boldsymbol{X}_0 \cup \boldsymbol{X}_1 \cup \{2\boldsymbol{x} - \boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_0 \cup \boldsymbol{X}_1\}\,.$$

- For $\beta = \frac{k}{n_1+n_2}$, define the **depth-based neighborhood** $R_{\boldsymbol{x}}^{\beta(\boldsymbol{n})}$ as the smallest depth region $D_\alpha^{(n)}(\boldsymbol{Y})$ (*i.e.* the one with the highest α) containing k observations from $\boldsymbol{X}_0 \cup \boldsymbol{X}_1$.
- Let $k_{\boldsymbol{X}_i} = \#\{\boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_i \cap R_{\boldsymbol{x}}^{\beta(\boldsymbol{n})}\}$.
- The **depth-based** k**-nearest-neighbors** classifier can then be defined as:

$$cl_{\boldsymbol{n}}^{DkNN}(\boldsymbol{x}) = \begin{cases} 0 & \text{if } k_{\boldsymbol{X}_0} \geq k_{\boldsymbol{X}_1}\,, \\ 1 & \text{if } k_{\boldsymbol{X}_0} < k_{\boldsymbol{X}_1}\,. \end{cases}$$

The depth kNN (Paindaveine & Van Bever, 2015)

- Let $Q = 2$.
- For an observation $\boldsymbol{x} \in \mathbb{R}^d$, consider a new data set

$$\boldsymbol{Y} = \boldsymbol{X}_0 \cup \boldsymbol{X}_1 \cup \{2\boldsymbol{x} - \boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_0 \cup \boldsymbol{X}_1\}\,.$$

- For $\beta = \frac{k}{n_1+n_2}$, define the **depth-based neighborhood** $R_{\boldsymbol{x}}^{\beta(\boldsymbol{n})}$ as the smallest depth region $D_\alpha^{(n)}(\boldsymbol{Y})$ (*i.e.* the one with the highest α) containing k observations from $\boldsymbol{X}_0 \cup \boldsymbol{X}_1$.
- Let $k_{\boldsymbol{X}_i} = \#\{\boldsymbol{z} \,:\, \boldsymbol{z} \in \boldsymbol{X}_i \cap R_{\boldsymbol{x}}^{\beta(\boldsymbol{n})}\}$.
- The **depth-based** k**-nearest-neighbors** classifier can then be defined as:

$$cl_{\boldsymbol{n}}^{DkNN}(\boldsymbol{x}) = \begin{cases} 0 & \text{if } k_{\boldsymbol{X}_0} \geq k_{\boldsymbol{X}_1}\,, \\ 1 & \text{if } k_{\boldsymbol{X}_0} < k_{\boldsymbol{X}_1}\,. \end{cases}$$

- Parameter k should be tuned,
  e.g. chosen through cross-validation.

Contents

Thank you for your attention! Questions?

- Brunel VE (2017). Concentration of the Empirical Level Sets of Tukey's Halfspace Depth. *arXiv:1605.09456*.
- Burr MA, Fabrizio RJ (2017). Uniform Convergence Rates for Halfspace Depth. *Statistics and Probabilty Letters*, **124**, 33–40.
- Chaudhuri P (1996). On a Geometric Notion of Quantiles for Multivariate Data. *Journal of the American Statistical Association*, **91**, 862–872.
- Cui X, Lin L, Yang G (2008). An Extended Projection Data Depth and Its Applications to Discrimination. *Communications in Statistics – Theory and Methods*, **37**, 2276–2290.
- Donoho D (1982). *Breakdown Properties of Multivariate Location Estimators.* Ph.D. thesis, Harvard University.
- Donoho DL, Gasko M (1992). Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *The Annals of Statistics*, **20**, 1803–1827.
- Dutta S, Ghosh AK (2012). On Robust Classification Using Projection Depth. *Annals of the Institute of Statistical Mathematics*, **64**, 657–676.
- Dutta S, Ghosh AK (2016). Multi-Scale Classification Using Localized Spatial Depth. *Journal of Machine Learning Research*, **17**, 1–30.

Thank you for your attention! Questions?

- Dyckerhoff R (2004). Data Depths Satisfying the Projection Property. *AStA – Advances in Statistical Analysis*, **88**, 163–190.
- Dyckerhoff R (2017). Convergence of Depths and Depth-Trimmed Regions. *arXiv:1611.08721*.
- Dyckerhoff R, Mosler K (2011). Weighted-Mean Trimming of Multivariate Data. *Journal of Multivariate Analysis*, **102**, 405–421.
- Dyckerhoff R, Mosler K, Koshevoy G (1996). Zonoid Data Depth: Theory and Computation. In A Prat (ed.), *COMPSTAT '96 – Proceedings in Computational Statistics*, pp. 235–240. Springer.
- Dyckerhoff R, Mozharovskyi P (2016). Exact Computation of the Halfspace Depth. *Computational Statistics and Data Analysis*, **98**, 19–30.
- Ghosh AK, Chaudhuri P (2005). On Data Depth and Distribution-Free Discriminant Analysis Using Separating Surfaces. *Bernoulli*, **11**, 1–27.
- Ghosh AK, Chaudhuri P (2005). On Maximum Depth and Related Classifiers. *Scandinavian Journal of Statistics*, **32**, 327–350.
- Hallin M, Paindaveine D, Siman M (2010). Multivariate Quantiles and Multiple-Output Regression Quantiles. From L1 Optimization to Halfspace Depth. *The Annals of Statistics*, **38**, 635–669.

Thank you for your attention! Questions?

- Koltchinskii V (1997). M-Estimation, Convexity and Quantiles. *The Annals of Statistics*, **25**, 435–477.
- Kong L, Mizera I (2012). Quantile Tomography: Using Quantiles with Multivariate Data. *Statistica Sinica*, **22**, 1589–1610.
- Koshevoy G (2002). The Tukey Depth Characterizes the Atomic Measure. *Journal of Multivariate Analysis*, **83**, 360–364.
- Koshevoy G (2003). Lift-Zonoid and Multivariate Depths. In R Dutter, P Filzmoser, U Gather, P Rousseeuw (eds.), *Developments in Robust Statistics*, pp. 194–202. Physica-Verlag, Heidelberg.
- Koshevoy G, Mosler K (1997). Zonoid Trimming for Multivariate Distributions. *The Annals of Statistics*, **25**, 1998–2017.
- Li J, Cuesta-Albertos JA, Liu RY (2012). *DD*-Classifier: Nonparametric Classification Procedure Based on *DD*-Plot. *Journal of the American Statistical Association*, **107**, 737–753.
- Liu RY (1990). On a Notion of Data Depth Based on Random Simplices. *The Annals of Statistics*, **18**, 405–414.
- Mahalanobis PC (1936). On the Generalized Distance in Statistics. *Proceedings of the National Institute of Sciences of India*, **12**, 49–55.

Thank you for your attention! Questions?

- Masse JC (2002). Asymptotics for the Tukey Median. *Journal of Multivariate Analysis*, **81**, 286–300.
- Masse JC (2004). Asymptotics for the Tukey Depth Process, with an Application to a Multivariate Trimmed Mean. *Bernoulli*, **10**, 397–419.
- Mizera I (2002). On Depth and Deep Points: A Calculus. *The Annals of Statistics*, **30**, 1681–1736.
- Mosler K (2002). *Multivariate Dispersion, Central Regions, and Depth: The Lift Zonoid Approach.* Springer-Verlag, New York.
- Mosler K (2013). Depth Statistics. In C Becker, R Fried, S Kuhnt (eds.), *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*, pp. 17–34. Springer. Berlin.
- Nieto-Reyes A, Battey H (2016). A Topologically Valid Definition of Depth for Functional Data. *Statistical Science*, **31**, 61–79.
- Paindaveine D, Van Bever G (2015). Nonparametrically Consistent Depth-Based Classifiers. *Bernoulli*, **21**, 62–82.
- Paidaveine D, Siman M (2012). Computing Multiple-Output Regression Quantile Regions. *Computational Statistics and Data Analysis*, **56**, 840–853.

Thank you for your attention! Questions?

- Paidaveine D, Siman M (2012). Computing Multiple-Output Regression Quantile Regions from Projection Quantiles. *Computational Statistics*, **27**, 29–49.
- Rousseeuw PJ, Ruts I (1999). The Depth Function of a Population Distribution. *Metrika*, **49**, 213–244.
- Stahel W (1981). *Robust Estimation: Infinitesimal Optimality and Covariance Matrix Estimators (In German).* Ph.D. thesis, Swiss Federal Institute of Technology in Zurich.
- Struyf AJ, Rousseeuw PJ (1999). Halfspace Depth and Regression Depth Characterize the Empirical Distribution. *Journal of Multivariate Analysis*, **69**, 135–153.
- Tukey JW (1975). Mathematics and the Picturing of Data. In R James (ed.), *Proceedings of the International Congress of Mathematicians*, volume 2, pp. 523–531. Canadian Mathematical Congress.
- Vardi Y, Zhang C (2000). The Multivariate L_1-Median and Associated Data Depth. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1423–1426.
- Zuo Y, Serfling R (2000). General Notions of Statistical Depth Function. *The Annals of Statistics*, **28**, 461–482.