Probabilistic Noise Clustering as M-Estimators

Frank Klawonn

f.klawonn@fh-wolfenbuettel.de

University of Applied Sciences Braunschweig/Wolfenbuettel

Department of Computer Science

Data Analysis and Pattern Recognition Lab

38302 Wolfenbuettel

http://public.rz.fh-wolfenbuettel.de/∼klawonn

Overview

- Probabilistic clustering
- Noise clustering
- Robust regression and M-estimators
- Noise clustering as M-estimators
- Conclusions

Fuzzy/probabilistic clustering

- given: a data set $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^p$
- fix a number of clusters c
- Minimize the objective function

$$f = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}$$

under the constraints

$$\sum_{i=1}^{c} u_{ij} = 1 \quad \text{for all } j = 1, \ldots, n.$$

Fuzzy c-means algorithm

solution strategy: alternating optimisation

Fuzzy c-means algorithm

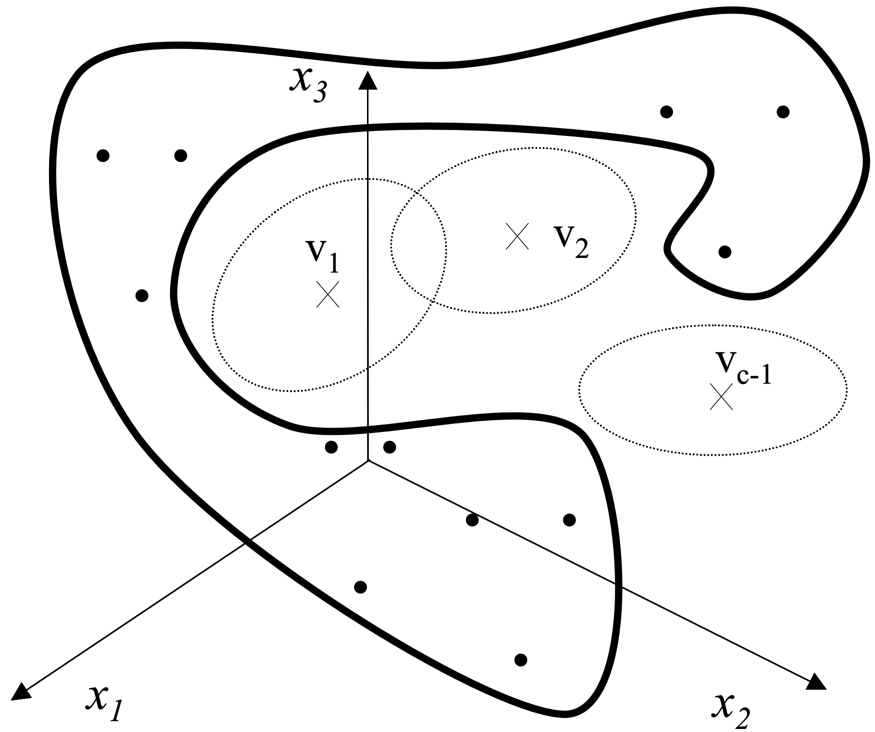
$$\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \| x_j - v_i \|^2 \quad \rightarrow \quad u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{1}{m-1}}}$$

$$\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \| x_j - v_i \|^2 \quad \rightarrow \quad v_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}}$$

Other cluster shapes

- ellipsoidal clusters (Gustafson/Kessel 1979)
- adaptable cluster volumes (Keller/Klawonn 1999)
- clusters as lines/planes/hyperplanes (Bock 1979, Bezdek 1981)
- cluster as shells of circles (Davé 1990, Krishnapuram/Nasraoui/Frigui 1992)
- clusters in the form of arbitrary quadrics (Krishnapuram/Frigui/Nasraoui 1991-1995)
- context sensitive clustering (Keller/Klawonn 1999)
- . . .

Noise clustering



Add one noise cluster with a fixed (large) distance to all data

Robust regression

linear model:

$$\begin{aligned} y_i &= \alpha + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i \\ &= \mathbf{x}_i^\top \beta + \varepsilon_i \end{aligned}$$

computed model:

$$\begin{aligned} y_i &= a + b_1 x_{i1} + \ldots + b_k x_{ik} + e_i \\ &= \mathbf{x}_i^\top \mathbf{b} + e_i \end{aligned}$$

objective function:

$$\sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i^\top \mathbf{b})$$

M-estimators

least squares method: $\rho(e) = e^2$.

Other choices of ρ:
Define $\psi = \rho'$, $w(e) = \psi(e)/e$ and $w_i = w(e_i)$.

Leads to the minimisation problem

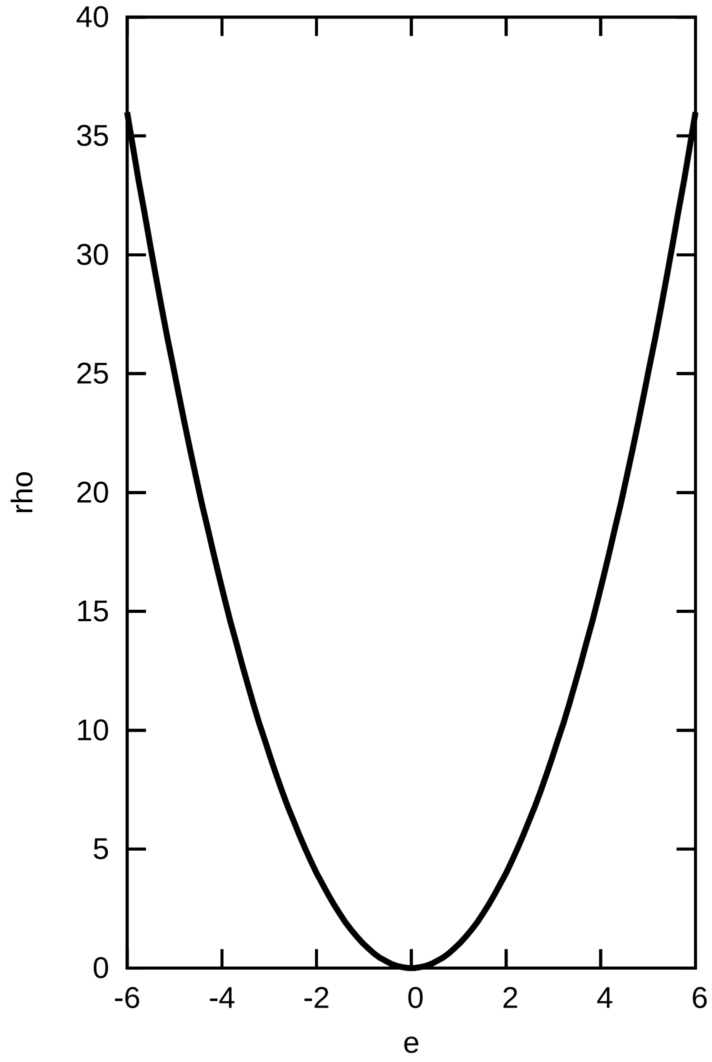
$$\sum_{i=1}^{n} w_i e_i^2 .$$

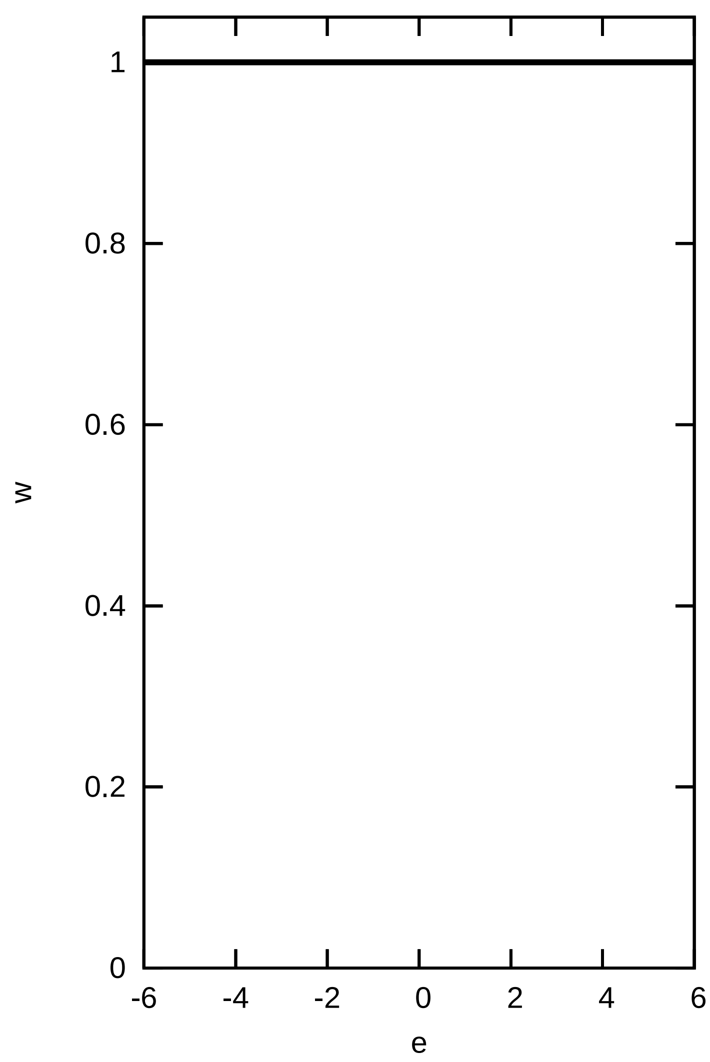
Solution strategy: Alternating optimisation.

Robust regression

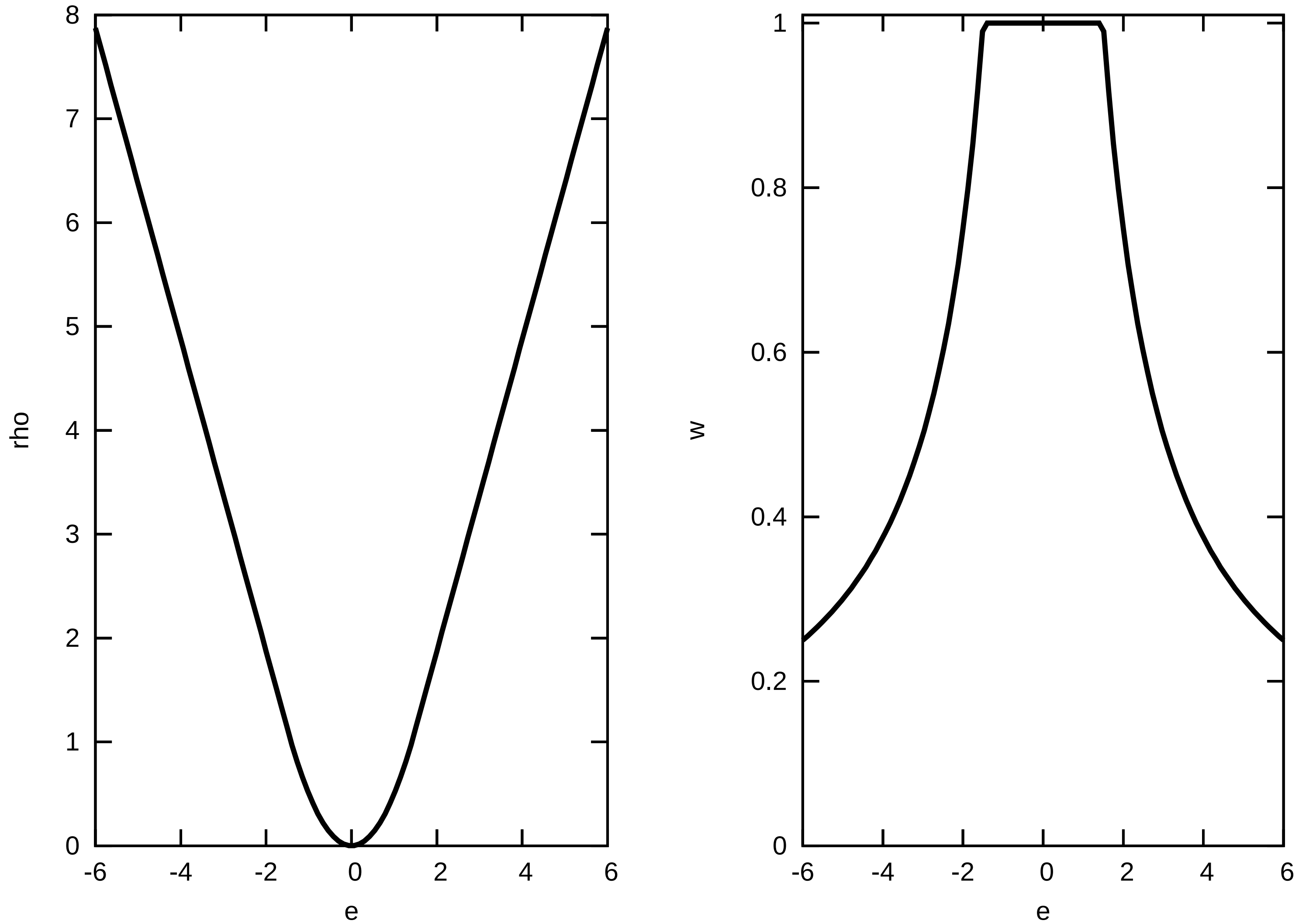
Method	$\rho(e)$
least squares	e^2
Huber	$\begin{cases} \frac{1}{2}e^2, & \text{if } \|e\| \leq k \\ k\|e\| - \frac{1}{2}k^2, & \text{if } \|e\| > k \end{cases}$
Bisquare	$\begin{cases} \frac{k^2}{6}\left(1 - \left(1 - \left(\frac{e}{k}\right)^2\right)^3\right) e^2, & \text{if } \|e\| \leq k \\ \frac{k^2}{6}, & \text{íf } \|e\| > k \end{cases}$

M-estimators: Least squares

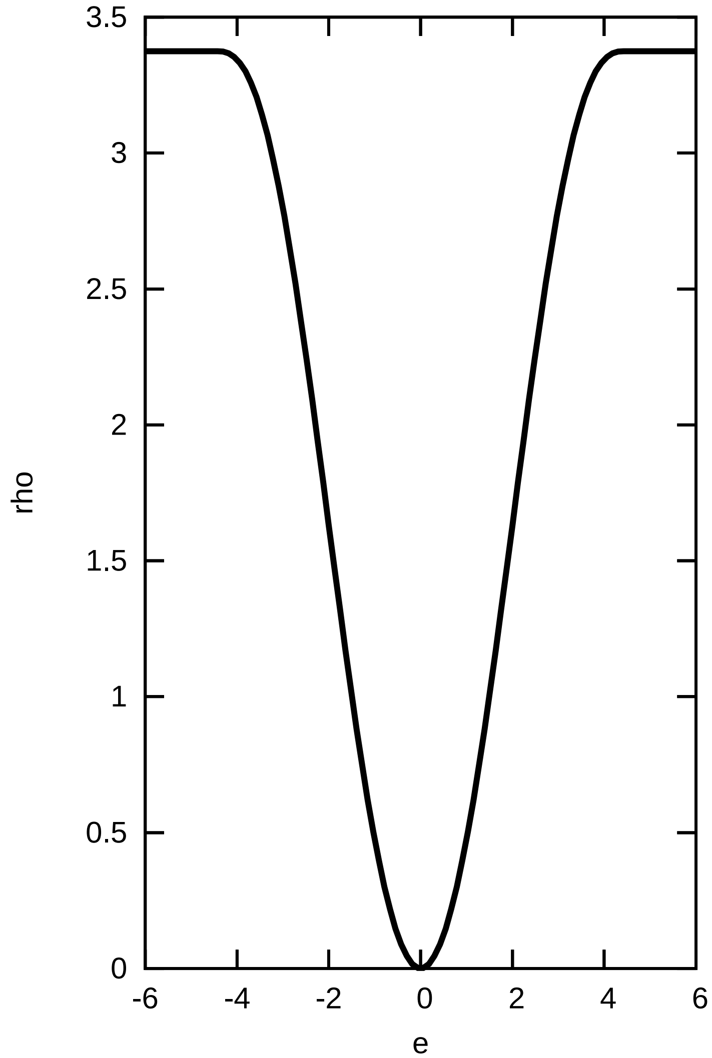


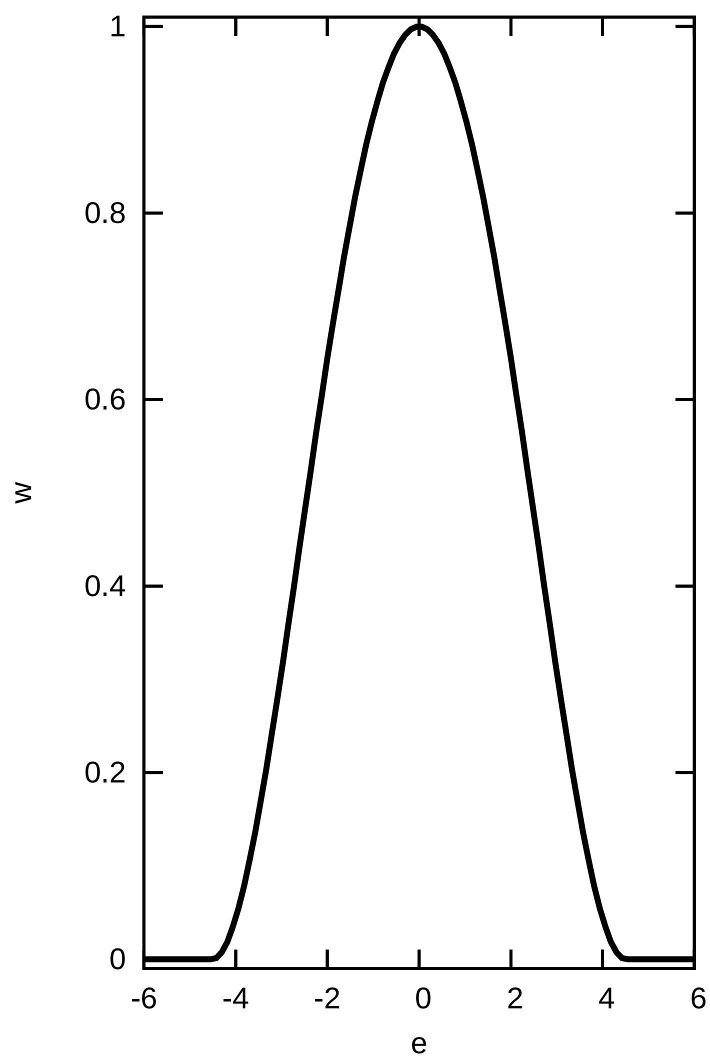


M-estimators: Huber

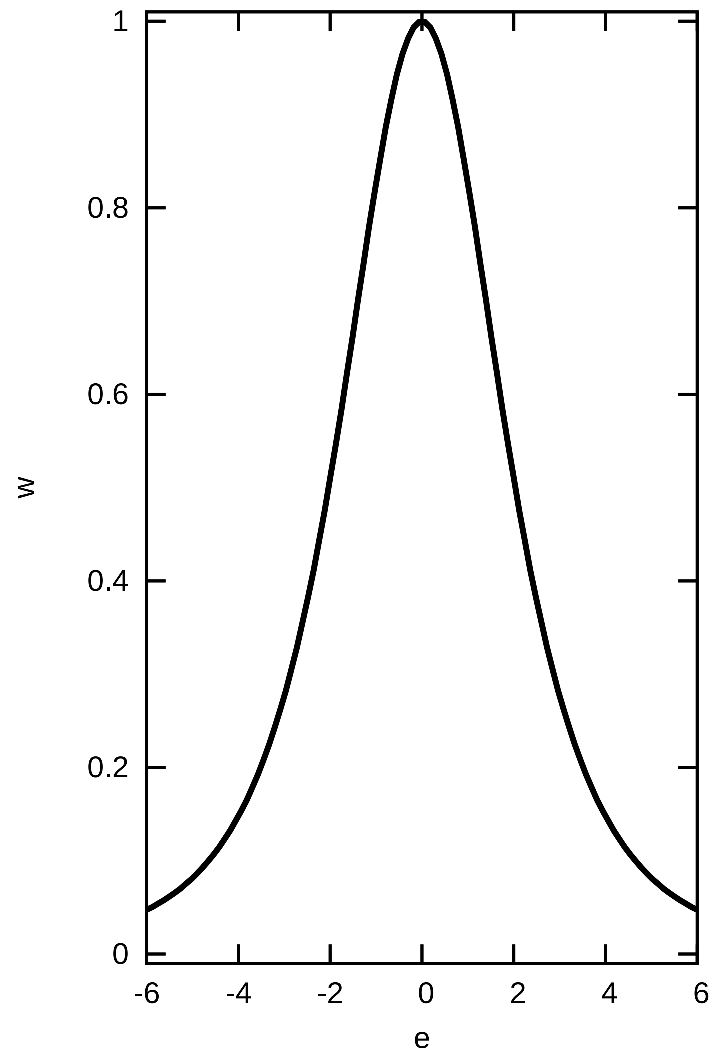


M-estimators: Bisquare





M-estimators: Fuzzy clustering



M-estimators: Fuzzy clustering

breakdown point: 0

One extreme outlier will shift at least one cluster centre extremely, even in the case of noise clustering, since

$$u_{ij}^m(x_j)x_j \to \infty$$

for an extreme outlier x_j with $\| x \| \to \infty$.

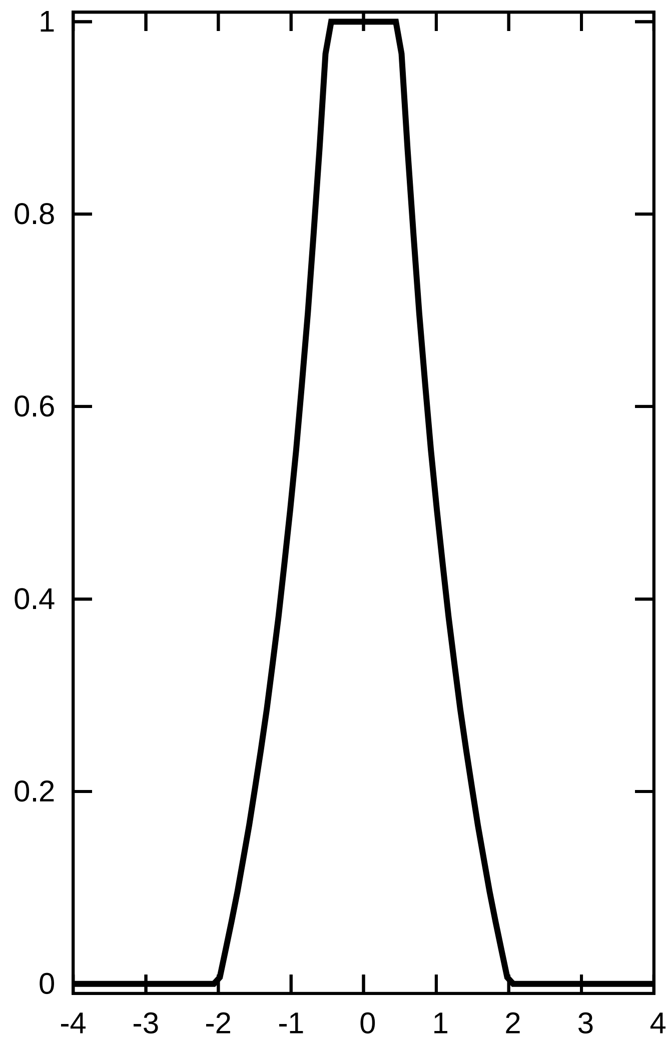
The fuzzifier transformation

$$g : [0,1] \to [0,1], \quad u \;\mapsto\; u^m$$

possible alternative

$$g(u) = \alpha u^2 + (1-\alpha)u \quad \text{with } 0 \le \alpha \le 1$$

M-estimators: Fuzzy clustering



Conclusions

- Modified fuzzy clustering offers new M-estimators also for robust regression

- Fuzzy clustering can be made robust using suitable alternatives to the fuzzifier