

Computational Intelligence

Winter Term 2019/20

Prof. Dr. Günter Rudolph

Lehrstuhl für Algorithm Engineering (LS 11)

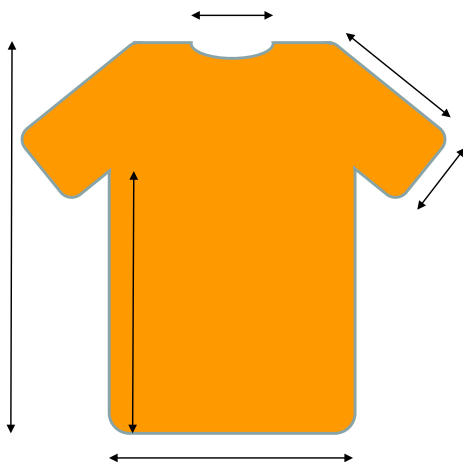
Fakultät für Informatik

TU Dortmund

- Fuzzy Clustering

Introductory Example: Textile Industry

→ production of T-shirts (for men)



best for producer : one size

vs.

best for consumer: made-to-measure

⇒ compromise: S, M, L, XL, 2XL

5 sizes

→ OK, but which lengths for which size?

idea:

- select, say, 2000 men at random and measure their “body lengths“
- arrange these 2000 men into five disjoint groups

arm’s length,
collar size,
chest girth, ...

such that

- deviations from mean of group as small as possible
- differences between group means as large as possible

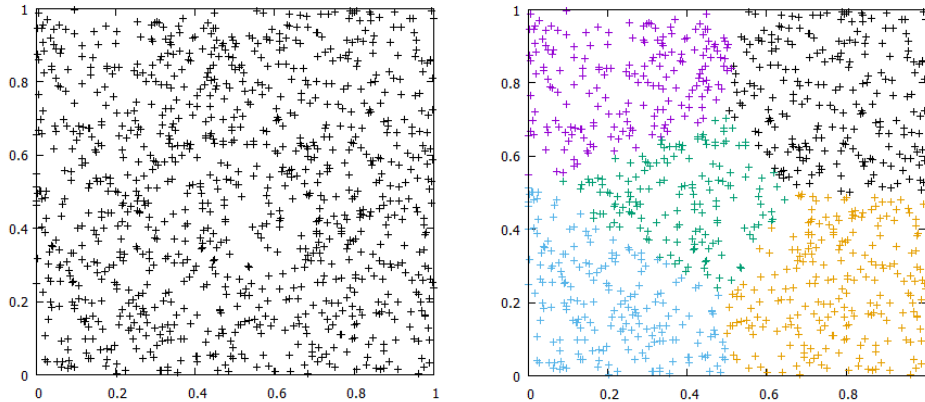
in general:

arrange objects into groups / clusters

such that

- elements within a cluster are as homogeneous as possible
- elements across clusters are as heterogeneous as possible

numerical example: 1000 points uniformly sampled in [0,1] x [0,1] → form 5 cluster



given data points $x_1, x_2, \dots, x_N \in \mathbb{R}^n$

objective: group data points into cluster such that

- points within cluster are as homogeneous as possible
- points across clusters are as heterogeneous as possible

⇒ crisp clustering is just a partitioning of data set $\{x_1, x_2, \dots, x_N\}$, i.e.,

$$\bigcup_{k=1}^K C_k = \{x_1, x_2, \dots, x_N\} \quad \text{and} \quad \forall j \neq k : C_j \cap C_k = \emptyset$$

where C_k is Cluster k and K denotes the number of clusters.

Constraint: $\forall k = 1, \dots, K : |C_k| \geq 1$ hence $1 \leq K \leq N$

Complexity: How many choices to assign N objects into K clusters?

more precisely:

→ objects are distinguishable / labeled

→ clusters are nondistinguishable / unlabeled **and** nonempty

⇒ Stirling number of 2nd kind $S(N, K) = \frac{1}{K!} \sum_{i=1}^K (-1)^{K-i} \binom{K}{i} \cdot i^N \sim \frac{K^N}{K!}$

N/K	1	2	3	4	5
10	1	511	9,330	34,105	42,525
11	1	1,023	28,501	145,750	246,730
12	1	2,047	86,526	611,501	1,379,400
13	1	4,095	261,625	2,532,530	7,508,501
14	1	8,191	788,970	10,391,745	40,075,035
15	1	16,383	2,375,101	42,355,950	210,766,920

$$S(100, 5) = 6.6 \times 10^{67}$$

$$S(1000, 5) = 7.8 \times 10^{696}$$

$$S(2000, 5) = 7.3 \times 10^{1395}$$

⇒ enumeration hopeless! ⇒ iterative improvement procedure required!

idea: define objective function that measures compactness of clusters *and* quality of partition

→ elements in cluster C_j should be as homogeneous as possible!

→ sum of squared distances to unknown center y should be as small as possible

→ find y with $\sum_{i \in C_j} d(x_i, y)^2 \rightarrow \min!$

typically, $d(x_i, y) = \|x_i - y\| = \sqrt{(x_i - y)'(x_i - y)}$ (Euclidean norm)

$$\frac{d}{dy} \sum_{i \in C_j} (x_i - y)'(x_i - y) = -2 \sum_{i \in C_j} (x_i - y) \stackrel{!}{=} 0$$

$$\Rightarrow \sum_{i \in C_j} x_i \stackrel{!}{=} \sum_{i \in C_j} y = |C_j| \cdot y \quad \Rightarrow y = \frac{1}{|C_j|} \sum_{i \in C_j} x_i =: \bar{x}_j$$

→ elements in each cluster C_j should be as homogeneous as possible!

→ find partition $C = (C_1, \dots, C_K)$ with $D(C) = \sum_{j=1}^K \sum_{i \in C_j} d(x_i, \bar{x}_j)^2 \rightarrow \min!$

Definition

A partition C^* is optimal if

$$D(C^*) = \min\{ D(C) : C \in P(N, K) \}$$

where $P(N, K)$ denotes all partitions of N elements in K clusters.

Theorem

$$\min_{C \in P(N, K)} D(C) = \max_{C \in P(N, K)} \sum_{j=1}^K |C_j| \cdot \|\bar{x}_j - \bar{x}\|^2$$

where \bar{x} is the mean of all x .

$\forall k = 1, \dots, K$: set $C_k = \emptyset$

$\forall x \in \{x_1, \dots, x_N\}$: assign x to some cluster C_k

set $t = 0$ and $D^{(t)} = \infty$

repeat

$t = t + 1$

$$\forall k = 1, \dots, K: \bar{x}_k = \frac{1}{|C_k|} \sum_{x \in C_k} x$$

$\forall i = 1, \dots, N$: $d_{ik} = d(x_i, \bar{x}_k)$ distance to center of cluster k

let k^* be such that $d_{ik^*} = \min\{d_{ik} : k = 1, \dots, K\}$

assign x_i to C_{k^*}

$$D^{(t)} = \sum_{k=1}^K \sum_{x \in C_k} d(x, \bar{x}_k)$$

until $D^{(t-1)} - D^{(t)} < \varepsilon$

objective for **crisp** clustering:

find partition $C = (C_1, \dots, C_K)$ with $D(C) = \sum_{j=1}^K \sum_{i \in C_j} d(x_i, \bar{x}_j)^2 \rightarrow \min!$

→ rewrite objective:

find partition $C = (C_1, \dots, C_K)$ with $D(C) = \sum_{j=1}^K \sum_{i=1}^N u_{ij} \cdot d(x_i, \bar{x}_j)^2 \rightarrow \min!$

expresses membership $\longrightarrow u_{ij} = \begin{cases} 1 & \text{if } x_i \in C_j \\ 0 & \text{otherwise} \end{cases}$

objective for **fuzzy** clustering:

find partition $C = (C_1, \dots, C_K)$ with $D(C) = \sum_{j=1}^K \sum_{i=1}^N u_{ij}^m \cdot d(x_i, \bar{x}_j)^2 \rightarrow \min!$
 $u_{ij} \in [0, 1] \subset \mathbb{R}, m > 1$

find partition $C = (C_1, \dots, C_K)$ with $D(C) = \sum_{j=1}^K \sum_{i=1}^N u_{ij}^m \cdot d(x_i, \bar{x}_j)^2 \rightarrow \min!$

where

$u_{ij} \in [0, 1] \subset \mathbb{R}$ denotes membership of x_i to cluster C_j

$m > 1$ denotes a fixed *fuzzifier* (controls / affects membership function)

subject to

$$\sum_{j=1}^K u_{ij} = 1 \quad \forall i = 1, \dots, N$$

each x_i distributes membership completely over clusters C_1, \dots, C_K
 → normalization

$$0 < \sum_{i=1}^N u_{ij} < N \quad \forall j = 1, \dots, K$$

at least one element belongs to some extent to a certain cluster, but not all elements to a single cluster

two questions:

- (a) how to define and calculate centers \bar{x}_j ?
- (b) how to obtain optimal memberships u_{ij} ?

ad a) let $d(x_i, \bar{x}_j) = \|x_i - \bar{x}_j\|_2$

$$\frac{d}{d\bar{x}_j} \sum_{i=1}^N u_{ij}^m \cdot (x_i - \bar{x}_j)'(x_i - \bar{x}_j) = -2 \sum_{i=1}^N u_{ij}^m \cdot (x_i - \bar{x}_j) \stackrel{!}{=} 0$$

$$\Leftrightarrow \sum_{i=1}^N u_{ij}^m x_i \stackrel{!}{=} \sum_{i=1}^N u_{ij}^m \bar{x}_j \Leftrightarrow \bar{x}_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \rightarrow \text{weighted mean!}$$

ad b) let $d_{ij} := d(x_i, \bar{x}_j) = \|x_i - \bar{x}_j\|_2$

apply **Lagrange multiplier** method:

$$\frac{\partial}{\partial u_{ij}} \sum_{j=1}^K \sum_{i=1}^N u_{ij}^m \cdot d_{ij}^2 - \sum_{i=1}^N \lambda_i \left(\sum_{j=1}^K u_{ij} - 1 \right) = m u_{ij}^{m-1} \cdot d_{ij}^2 - \lambda_i \stackrel{!}{=} 0$$

without constraints $\rightarrow u_{ij}^* = 0$

$$u_{ij}^* = \left(\frac{\lambda_i}{m \cdot d_{ij}^2} \right)^{\frac{1}{m-1}}$$

$$\sum_{j=1}^K u_{ij} = \sum_{j=1}^K \left(\frac{\lambda_i}{m \cdot d_{ij}^2} \right)^{\frac{1}{m-1}} = \sum_{j=1}^K \frac{\lambda_i^{\frac{1}{q}}}{(m \cdot d_{ij}^2)^{\frac{1}{q}}} = \lambda_i^{\frac{1}{q}} \sum_{j=1}^K \frac{1}{(m \cdot d_{ij}^2)^{\frac{1}{q}}} \stackrel{!}{=} 1$$

set $q = m - 1$

$$\Rightarrow \lambda_i^* = \left[\sum_{k=1}^K \frac{1}{(m \cdot d_{ik}^2)^{\frac{1}{q}}} \right]^{-q}$$

after insertion:

$$u_{ij}^* = \left(\frac{1}{m \cdot d_{ij}^2} \left[\frac{1}{\sum_{k=1}^K \left(\frac{1}{m \cdot d_{ik}^2} \right)^{\frac{1}{m-1}}} \right]^{m-1} \right)^{\frac{1}{m-1}} = \left[\sum_{k=1}^K \left(\frac{d_{ik}}{d_{ij}} \right)^{\frac{2}{m-1}} \right]^{-1}$$

choose $K \in \mathbb{N}$ and $m > 1$
choose u_{ij} at random (obeying constraints)
repeat

$\forall j = 1, \dots, K$: calculate centers \bar{x}_j

$\forall i = 1, \dots, N$:

let $J_i = \{j : x_i = \bar{x}_j\}$

if $J_i = \emptyset$ determine memberships u_{ij}

else

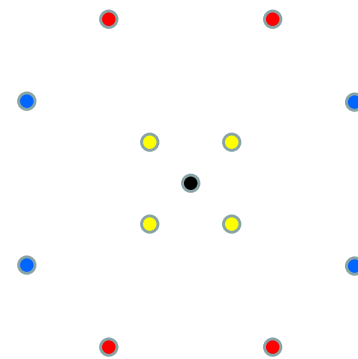
choose u_{ij} such that $\sum_{j \in J_i} u_{ij} = 1$

and $u_{ij} = 0$ for $j \notin J_i$

until $D(C^{(t)}) - D(C^{(t+1)}) < \varepsilon$ or $t = t_{max}$

problems:

- choice of K
calculate quality measure for each #cluster; then choose best
- choice of m
try some values; typical: $m=2$; use interval \rightarrow fuzzy type-2



black dot is center of
- red cluster
- blue cluster
- yellow cluster
in case of equal weights

$u_{ij} = 1 / |J_i|$ for $j \in J_i$ appears plausible

but: different values algorithmically better

\rightarrow cluster centers more likely to separate again (\rightarrow tiny randomization?)

- **Partition Coefficient**

$$PC(C_1, \dots, C_K) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K u_{ij}^2 \quad (\text{"larger is better"})$$

$$\left. \begin{array}{l} \text{maximum if } u_{ij} \in \{0, 1\} \rightarrow \text{crisp partition} \\ \text{minimum if } u_{ij} = \frac{1}{K} \rightarrow \text{entirely fuzzy} \end{array} \right\} \frac{1}{K} \leq PC(C_1, \dots, C_K) \leq 1$$

- **Partition Entropy**

$$PE(C_1, \dots, C_K) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K u_{ij} \cdot \log_2(u_{ij}) \quad (\text{"smaller is better"})$$

$$\left. \begin{array}{l} \text{maximum if } u_{ij} = \frac{1}{K} \rightarrow \text{entirely fuzzy} \\ \text{minimum if } u_{ij} \in \{0, 1\} \rightarrow \text{crisp partition} \end{array} \right\} 0 \leq PE(C_1, \dots, C_K) \leq \log_2(K)$$

- **Silhouette Values (crisp version)**

$$PC(C_1, \dots, C_K) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K u_{ij}^2 \quad (\text{"larger is better"})$$