technische universität dortmund

Computational Intelligence

Winter Term 2025/26

Prof. Dr. Günter Rudolph

Computational Intelligence

Fakultät für Informatik

TU Dortmund

Plan for Today

Lecture 5

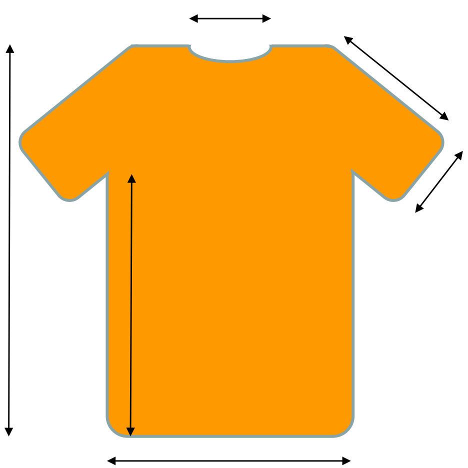
- Fuzzy Clustering

Cluster Formation and Analysis

Lecture 5

Introductory Example: Textile Industry

→ production of T-shirts (for men)



best for producer : one size

vs.

best for consumer: made-to-measure

⇒ compromize: $\underbrace{\text{S, M, L, XL, 2XL}}_{\text{5 sizes}}$

→ OK, but which lengths for which size?

Cluster Formation and Analysis

Lecture 5

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

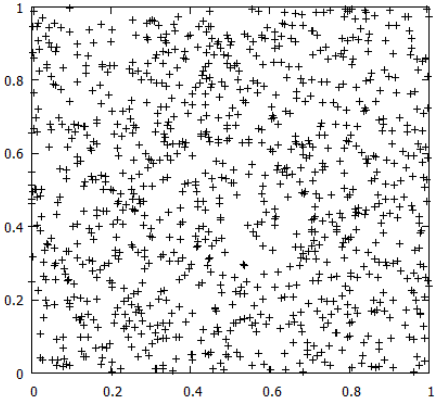
such that

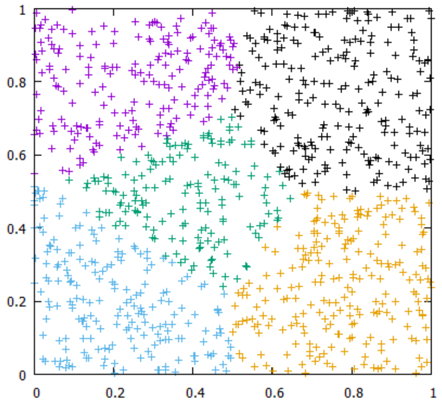
- – elements within a cluster are as homogeneous as possible
- – elements across clusters are as heterogeneous as possible

Cluster Formation and Analysis

Lecture 5

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





Hard / Crisp Clustering

Lecture 5

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

objective: group data points into cluster such that
- *points within cluster* are as *homogeneous* as possible
- *points across clusters* are as *heterogeneous* as possible

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

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

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

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

Hard / Crisp Clustering

Lecture 5

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

more precisely:
$\rightarrow$ objects are distinguishable / labeled
$\rightarrow$ clusters are nondistinguishable / unlabeled **and** nonempty

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

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

$$\begin{aligned} S(100,5) &= 6.6 \times 10^{67} \\ S(1000,5) &= 7.8 \times 10^{696} \\ S(2000,5) &= 7.3 \times 10^{1395} \end{aligned}$$

$\Rightarrow$ enumeration hopeless! $\Rightarrow$ iterative improvement procedure required!

Hard / Crisp Clustering

Lecture 5

idea: define objective function

that measures compactness of clusters *and* quality of partition

$\rightarrow$ elements in cluster C_j should be as homogeneous as possible!

$\rightarrow$ sum of squared distances to unknown center y should be as small as possible

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

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

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

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

Hard / Crisp Clustering

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

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

Definition

A partition C^* is optimal if

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

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

Theorem

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

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

Crisp K-Means Clustering

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

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

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

`repeat`

  $t = t + 1$

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

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

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

    assign x_i to C_{k^*}

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

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

From Crisp to Fuzzy Clustering

objective for **crisp** clustering:

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

→ rewrite objective:

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

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

objective for **fuzzy** clustering:

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

$u_{ij} \in [0,1] \subset \mathbb{R}, m > 1$

Fuzzy K-Means Clustering

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

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

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

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

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

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

Fuzzy K-Means Clustering

<u>two questions:</u>

(a) how to define and calculate centers $\bar{x}_j$?

(b) how to obtain optimal memberships u_{ij}?

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

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

$$\Leftrightarrow \sum_{i=1}^{N} u_{ij}^m \, x_i \stackrel{!}{=} \sum_{i=1}^{N} u_{ij}^m \, \bar{x}_j \quad \Leftrightarrow \quad \boxed{\bar{x}_j = \frac{\sum\limits_{i=1}^{N} u_{ij}^m \, x_i}{\sum\limits_{i=1}^{N} u_{ij}^m}}$$

→ weighted mean!

Fuzzy K-Means Clustering

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

*apply **Lagrange multiplier** method:*

$$\frac{\partial}{\partial u_{ij}} \sum_{j=1}^{K} \sum_{i=1}^{N} u_{ij}^m \cdot d_{ij}^2 - \underbrace{\sum_{i=1}^{N} \lambda_i \left(\sum_{j=1}^{K} u_{ij} - 1 \right)}_{\text{without constraints } \to \, u_{ij}^* = 0} = m \, u_{ij}^{m-1} \cdot d_{ij}^2 - \lambda_i \stackrel{!}{=} 0$$

$$\Updownarrow$$

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

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

set $q = m - 1$

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

Fuzzy K-Means Clustering

after insertion:

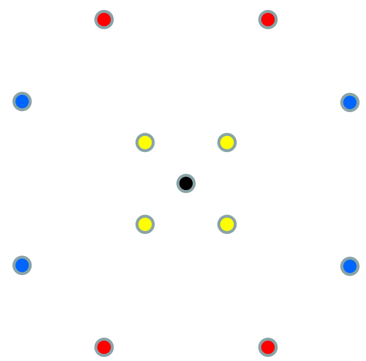
$$u_{ij}^* = \left(\frac{1}{m \cdot d_{ij}^2} \left[\frac{1}{\sum\limits_{k=1}^{K} \left(\frac{1}{m \cdot d_{ik}^2} \right)^{\frac{1}{m-1}}} \right]^{m-1} \right)^{\frac{1}{m-1}} = \boxed{\left[\sum_{k=1}^{K} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}} \right]^{-1}}$$

choose $K \in \mathbb{N}$ and $m > 1$
choose u_{ij} at random (obeying constraints)
repeat
  $\forall j = 1, \ldots, K$: calculate centers $\bar{x}_j$
  $\forall i = 1, \ldots, N$:
    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:

- <u>choice of K</u>
  calculate quality measure for each #cluster;
  then choose best
- <u>choice of m</u>
  try some values;
  typical: *m*=2;
  use interval → fuzzy type-2

Example: Special Case $|J_i| > 1$



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

Graphical Representation of Fuzzy Cluster

n = 2

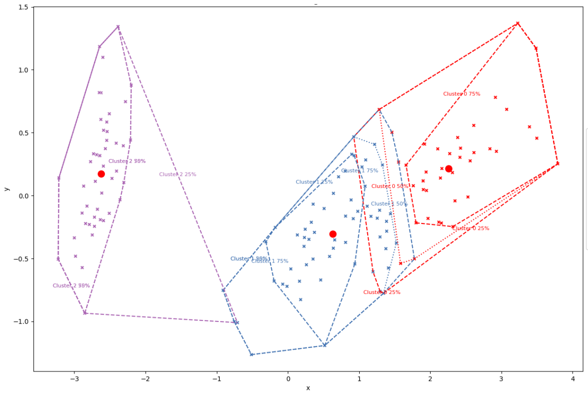
idea: plot data points in 2-D, draw isolines with degree of membership

example:

requested: 3 cluster

red bullet = cluster center
dashed lines = isolines

observations:
- red and blue cluster overlap
- purple cluster almost isolated



Graphical Representation of Fuzzy Cluster

n > 2

idea: reduction of data dimension from n to 2

→ several methods available:
principal component analysis (PCA), multi-dimensional scaling (MDS),
singular value decomposition (SVD), spectral embedding, ...

2 options:
→ first reduce dimension of data, then clustering in 2-D
→ first clustering in n-D, then reduce dimension of cluster data

finally, apply method developed for n = 2

Measures for Cluster Quality

- **Partition Coefficient**

$$\mathrm{PC}(C_1, \ldots, C_K) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{K} u_{ij}^2$$

("larger is better")

maximum if $u_{ij} \in \{0, 1\}$ → crisp partition
minimum if $u_{ij} = \frac{1}{K}$ → entirely fuzzy

$$\frac{1}{K} \le \mathrm{PC}(C_1, \ldots, C_K) \le 1$$

- **Partition Entropy**

$$\mathrm{PE}(C_1, \ldots, C_K) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{K} u_{ij} \cdot \log_2(u_{ij})$$

("smaller is better")

maximum if $u_{ij} = \frac{1}{K}$ → entirely fuzzy
minimum if $u_{ij} \in \{0, 1\}$ → crisp partition

$$0 \le \mathrm{PE}(C_1, \ldots, C_K) \le \log_2(K)$$

Measures for Cluster Quality

Finding an appropriate number of clusters

Possible approach: (`mc` = max. number of clusters)

```
set score s* to worst possible value
for c = 2 to mc
   apply FCM with c clusters (yields membership matrix U)
   determine quality of clustering from U (yields score s)
   if s better than s* then
      s* = s; c* = c
   endif
endfor
output c*
```

Many more methods for assessing quality of clustering available:

Hong-Yu Wang, Jie-Sheng Wang, Guan Wang: A survey of fuzzy clustering validity evaluation methods, *Information Sciences 618:270-297*, 2022.
https://doi.org/10.1016/j.ins.2022.11.010

So far, the distance measure was defined via the p-norm with p = 2 (Euclidean norm)

$$d(x,y) = \|x-y\|_p = \left(\sum_{i=1}^{n}(x_i-y_i)^p\right)^{\frac{1}{p}}$$

In principle, every value of p with 1 < p < ∞ can be used with this approach.

⇒ for *given centers* $\bar{x}_j$ the expression for the memberships is unchanged

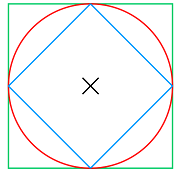
Let $d_{ij} = \|x_i - \bar{x}_j\|_p$ then $u_{ij} = \boxed{\left[\sum_{k=1}^{K}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}\right]^{-1}}$

But even for *given memberships* there is no explicit expression for the centers.

⇒ must solve K minimization problems with n variables numerically

There exist algorithms for p = 1 and p = ∞
with *explicit* expressions for centers and memberships.

L. Bobrowski and J. C. Bezdek:
c-Means clustering with the L_1 and L_∞ norms,
IEEE Transactions on Systems, Man, and Cybernetics 21(3):545-554, 1991.
DOI: 10.1109/21.97475



same distance for p = 1, 2, ∞

Mahalanobis distance

$d_{ij} = \|x_i - \bar{x}_j\|_M = \sqrt{(x_i - \bar{x}_j)^\top M^{-1}(x_i - \bar{x}_j)}$ with positive definite matrix M

⇒ leads to elliptical shape of cluster

matrix M not fixed but adapted from given data → each cluster has own matrix

⇒ *Gustafson-Kessel* fuzzy cluster algorithm (1979)

$$\min_{U,\bar{X},M} \sum_{j=1}^{K}\sum_{i=1}^{N} u_{ij}^m (x_i - \bar{x}_j)^\top M_j^{-1}(x_i - \bar{x}_j) \quad \text{s.t. } |M_j| = 1, \ \sum_{j=1}^{K} u_{ij} = 1, \ 0 \leqslant u_{ij} \leqslant 1$$

Fuzzy c-Shells Algorithm

- every cluster k = (v, r) is characterized by center v of shell and radius r
- must define an appropriate distance function

$d^2(x,(v,r)) = (\|x-v\|_2 - r)^2$ [R.N. Davé 1990]

problem: shell center not given as explicit expression;
must solve a set of nonlinear equations

$d^2(x,(v,r)) = (\|x-v\|_2^2 - r^2)^2$ [Krishnapuram et al. 1991]

shell center and memberships determined by explicit expressions;
but points outside the shells punished harder → results different