EUSFLAT 2023

Fuzzy measures for metric learning and data-driven models

Fuzzy measures and distances

Vicenç Torra

September, 2023

Dept. CS, Umeå University, Sweden

Outline

Preliminaries

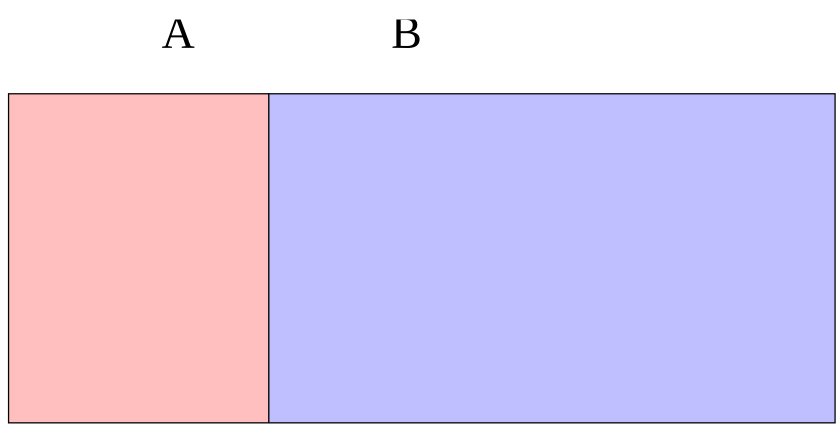
Measures

Measures

Measures:

- A measure (mathematics) as a generalization of geometric measures (e.g., area)
- Used to express size, importance, and
- probabilities

Key property: additivity:



Additive measures

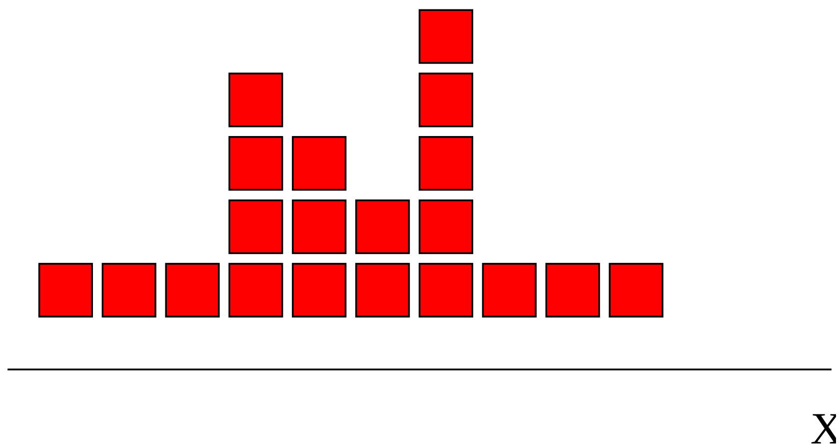
Additive measures: Formally (reference set X)

- $\mu(\emptyset) = 0$
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$ for disjoint S_1, S_2

Additive measures in statistics/probability theory

Measures: A typical example, probabilities!!
(on X and subsets of X, assume X finite)

- $\mu(\emptyset) = 0$
- $\mu(X) = 1$
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$ for disjoint S_1, S_2



Additive measures in decision making

Measures: or standard weights of sets of criteria/variables
(on X and subsets of X, assume X finite)

- $\mu(\emptyset) = 0$
- $\mu(X) = 1$
- $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$ for disjoint S_1, S_2

That is,

- the importance of the set of criteria/variables (price, comfort, size)
- equals to
- importance(price) + importance(comfort) + importance(size)

implicit assumption in problems using weighted means

Also, because of additivity for any disjoint S_1, S_2, C,

- if $\mu(S_1) < \mu(S_2)$ then also $\mu(S_1 \cup C) < \mu(S_2 \cup C)$.

Fuzzy Measures

Fuzzy measures

Non-additive measures:

- Replace the additivity condition by a monotonicity condition

$$S_1 \subseteq S_2 \text{ then } \mu(S_1) \leq \mu(S_2)$$

- This allows for interactions:
  - $\mu(S_1 \cup S_2) > \mu(S_1) + \mu(S_2)$
  - $\mu(S_1 \cup S_2) < \mu(S_1) + \mu(S_2)$

Fuzzy measures

Non-additive measures:

- Replace the additivity condition by a monotonicity condition

$$S_1 \subseteq S_2 \text{ then } \mu(S_1) \leq \mu(S_2)$$

- This allows for interactions:
  - $\mu(S_1 \cup S_2) > \mu(S_1) + \mu(S_2)$
  - $\mu(S_1 \cup S_2) < \mu(S_1) + \mu(S_2)$
- positive/negative interactions !
  - the importance of the set of criteria/variables (price, comfort, size) does not need to equal
  - importance(price) + importance(comfort) + importance(size)
- This allows for inversing inequalities for any disjoint S_1, S_2, C, it is possible
  - $\mu(S_1) < \mu(S_2)$ but also $\mu(S_1 \cup C) > \mu(S_2 \cup C)$.

Fuzzy integrals and aggregation

Aggegation functions

Aggregation

- Given variables / information sources / criteria $X = \{x_1, \ldots, x_n\}$
- and values $f : X \rightarrow [0, 1]$
- So, $f(x_i)$ value associated to x_i
- We combine them $\mathbb{C}(f(x_1), \ldots, f(x_n))$

Examples

- Arithmetic mean $\sum_{i=1}^{n}(1/n)f(x_i)$
- Weighted mean $\sum_{i=1}^{n} w_i f(x_i)$
- ... other aggregation, and also fuzzy integrals
  to combine the values (the data $f(x_i)$) w.r.t. a fuzzy measure μ.

Aggegation functions

Aggregation

- Given variables / information sources / criteria $X = \{x_1, \ldots, x_n\}$
- and values $f : X \to [0, 1]$
- So, $f(x_i)$ value associated to x_i
- We combine them $\mathbb{C}(f(x_1), \ldots, f(x_n))$

Examples

- Arithmetic mean $\sum_{i=1}^{n}(1/n)f(x_i)$
- Weighted mean $\sum_{i=1}^{n} w_i f(x_i)$
- ... other aggregation, and also fuzzy integrals
  to combine the values (the data $f(x_i)$) w.r.t. a fuzzy measure μ.
  - Choquet and Sugeno integrals
  - Generalizations and variants: Murofushi & Sugeno fuzzy t-conorm integral, Bustince & Fernandez & Mesiar etc.

Distances

From the Euclidean distance to CI-based distances

Distances: $d : [0,1]^{|X|} \times [0,1]^{|X|} \to \mathbb{R}^+$.

$$d(A = (a_1, \ldots, a_n), B = (b_1, \ldots, b_n))$$

- Euclidean distance (squared)

$$d(A = (a_1, \ldots, a_n), B = (b_1, \ldots, b_n)) = \sum (a_i - b_i)^2$$

- Weighted Euclidean (with weights w)

From the Euclidean distance to CI-based distances

Distances: $d : [0,1]^{|X|} \times [0,1]^{|X|} \to \mathbb{R}^+$.

$$d(A = (a_1, \ldots, a_n), B = (b_1, \ldots, b_n))$$

- Euclidean distance (squared)

$$d(A = (a_1, \ldots, a_n), B = (b_1, \ldots, b_n)) = \sum (a_i - b_i)^2$$

- Weighted Euclidean (with weights w)

$$\begin{aligned} d_w(A, B) &= \sum w_i (a_i - b_i)^2 \\ &= WM(d(V_1(A), V_1(B)), \ldots, d(V_n(A), V_n(B))) \end{aligned}$$

where $d(V_i(A), V_i(B)) = (a_i - b_i)^2$.

From the Euclidean distance to CI-based distances

Distances: $d : [0,1]^{|X|} \times [0,1]^{|X|} \to \mathbb{R}^+$.

- Choquet integral-based (with measure μ)

From the Euclidean distance to CI-based distances

Distances: $d : [0,1]^{|X|} \times [0,1]^{|X|} \to \mathbb{R}^+$.

- Choquet integral-based (with measure μ)

$$d_\mu(A,B) = CI_\mu(d(V_1(A),V_1(B)),\ldots,d(V_n(A),V_n(B)))$$

where $d(V_i(A),V_i(B)) = (a_i - b_i)^2$.

From the Euclidean distance to CI-based distances

Distances: $d : [0,1]^{|X|} \times [0,1]^{|X|} \to \mathbb{R}^+$.

- Choquet integral-based (with measure μ)

$$d_\mu(A,B) = CI_\mu(d(V_1(A),V_1(B)),\ldots,d(V_n(A),V_n(B)))$$

where $d(V_i(A),V_i(B)) = (a_i - b_i)^2$.

CI generalizes WM, and WM generalizes Euclidean distance,
So, appropriate μ and w make d_w and d_μ the Euclidean distance

When μ submodular, d_μ a metric (triangle inequality)

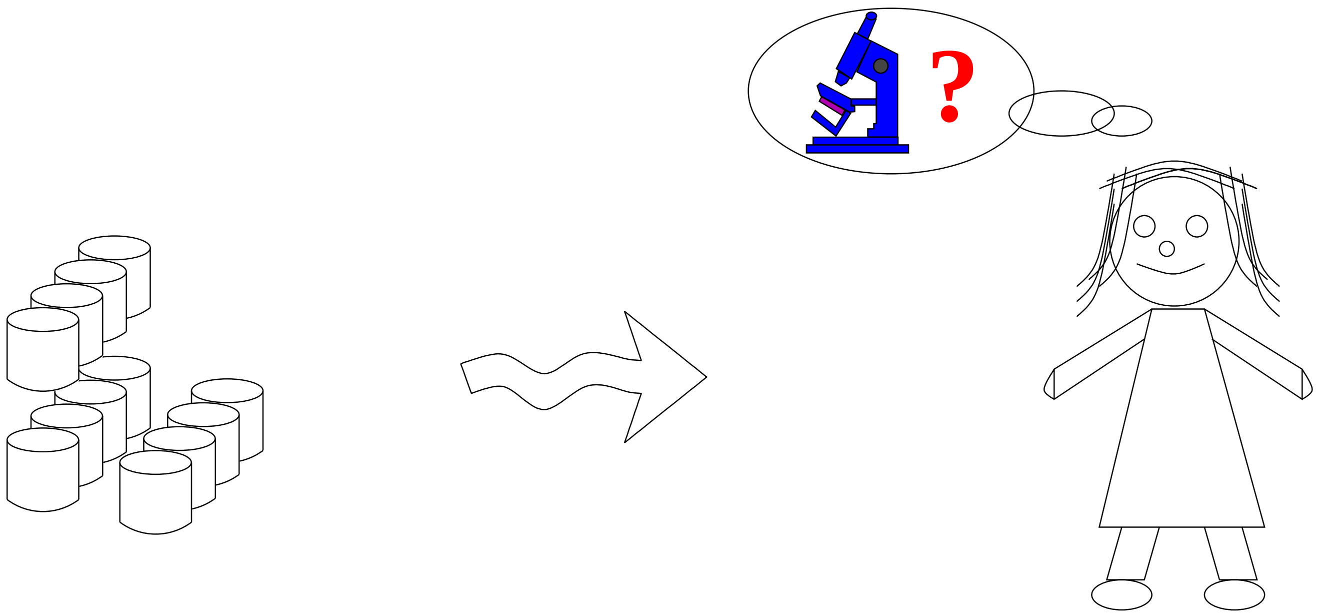
An application:

data sharing and data privacy

metric learning for d_μ

Context: Data privacy

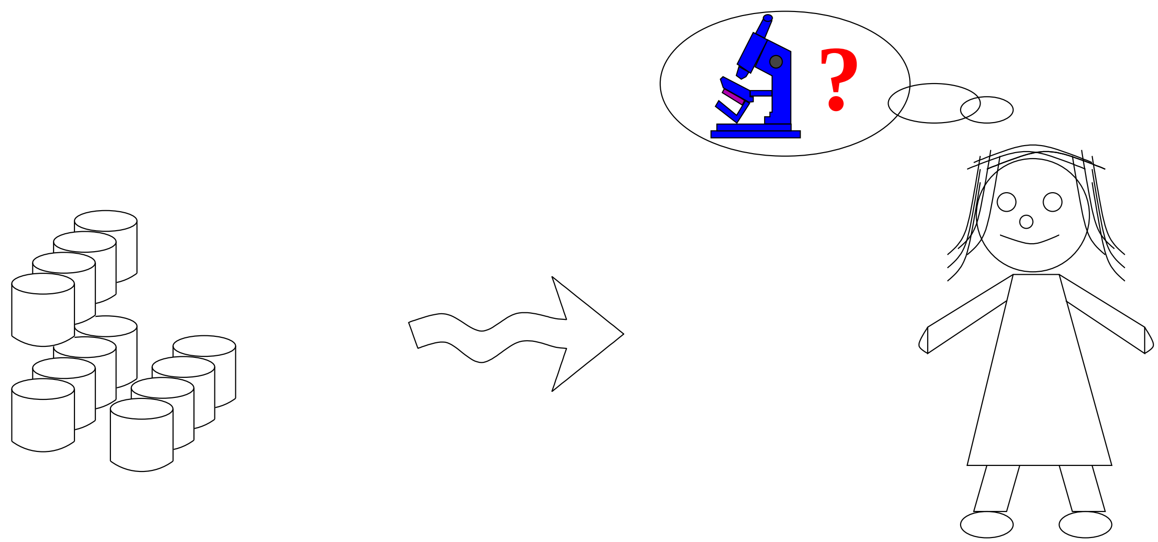
Data privacy in context. A researcher wants to analyze data



DB = {(*Aina*, *Age* = *40*, *Street*=*Llucmajor*, *salary*=*1800 EUR*), ...}

Context: Data privacy

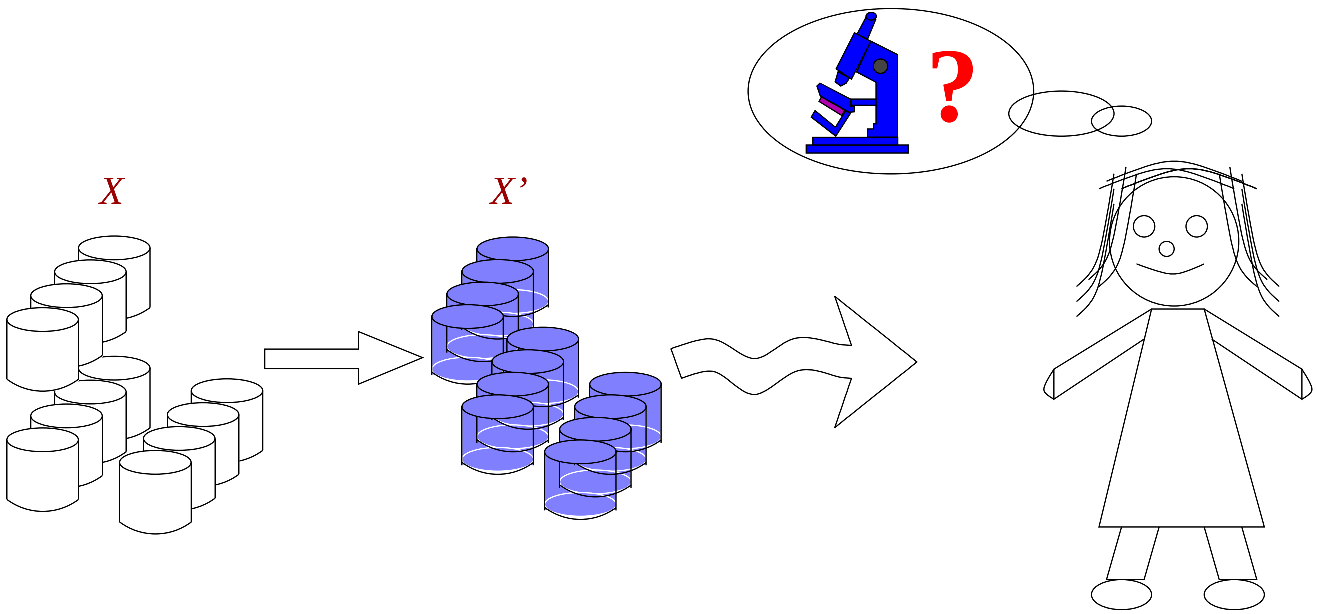
- Disclosure from the data themselves
  - Identity disclosure: find Aina in the database
  - Attribute disclosure: learn Aina's salary
- Usual: identity disclosure leads to attribute disclosure



DB = {(~~*Aina*~~, *Age* = *40*, *Street*=*Llucmajor*, *salary*=*1800 EUR*), ...}

Context: Data privacy

- To avoid disclosure, remove identifiers, anonymize records / modify records



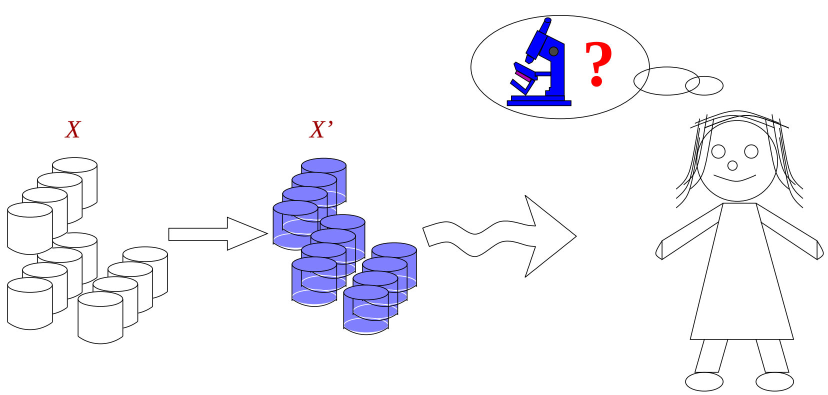
DB = {(~~Aina~~, Age = 41, Street/Neigh.=El Molinar,
salary=1800 EUR), ...}

Context: Data privacy

Privacy models. A computational definition for privacy. Publish a DB

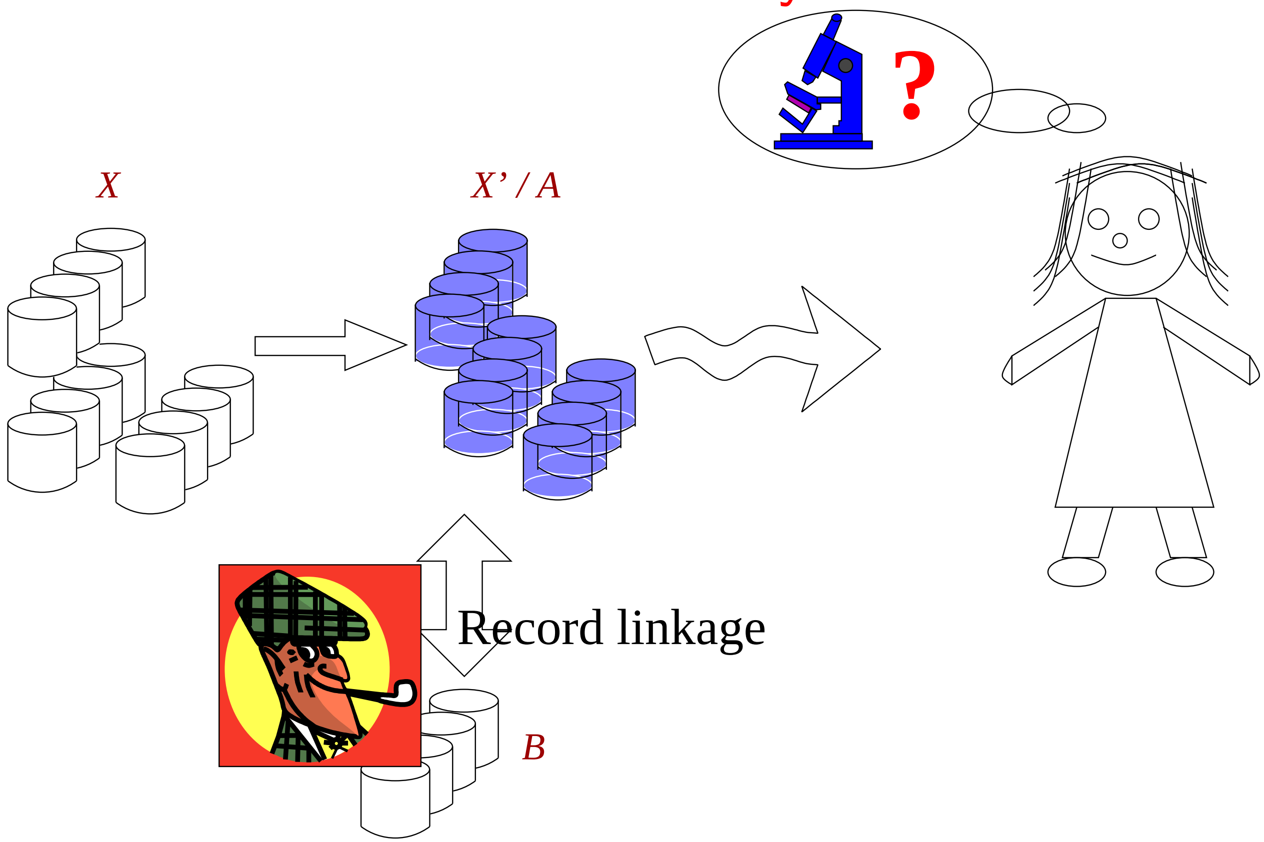
- **Reidentification privacy.** Avoid finding a record in a database.
- **k-Anonymity.** A record indistinguishable with $k-1$ other records.
- **Interval disclosure.** The value for an attribute is outside an interval computed from the protected value: values different enough.
- **Result privacy.** We want to avoid some results when an algorithm is applied to a database.

Privacy measures. Measures to assess the privacy level of e.g. protected database.



Context: Identity disclosure risk in data privacy

- Identity disclosure risk by modeling an intruder attack
  - How many records in B can be correctly linked to X'



Context: Identity disclosure risk in data privacy

- Identity disclosure risk measure
  - Worst case scenario = the most conservative estimation of risk
  - Worst case scenario / maximum knowledge:
    - Best information $B = X$
    - Best knowledge on the protection process: transparency attacks
    - Best record linkage algorithm:
      - Best record linkage algorithm: distance-based record linkage
      - Best parameters: distance
  - Best means: the most possible number of reidentifications
    The more the better (for an intruder)

Context: Identity disclosure risk in data privacy

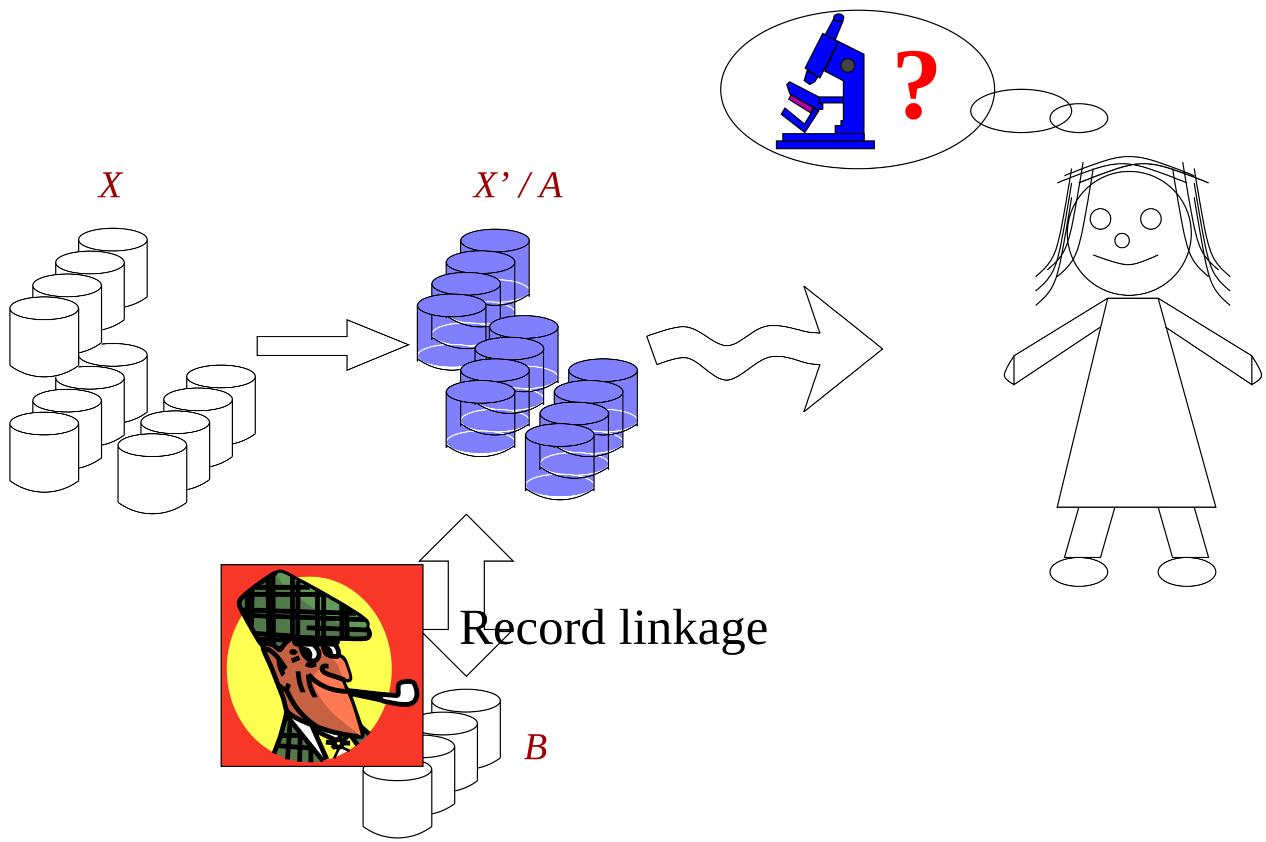
- Can we do better than with the Euclidean distance?
- Other options:
  - Weighted Euclidean distance (weights w) d_w
  - Mahalanobis distance (using covariance matrix Q)
- But also
  - Choquet integral (measure μ) d_μ
  - Bilinear forms (using positive definite matrix Q) d_Q

Context: Identity disclosure risk in data privacy

- Can we do better than with the Euclidean distance?
- Other options:
  - Weighted Euclidean distance (weights w) d_w
  - Mahalanobis distance (using covariance matrix Q)
- But also
  - Choquet integral (measure μ) d_μ
  - Bilinear forms (using positive definite matrix Q) d_Q
- Num. Reidentifications $d_\mu \geq$ Num. Reid. $d_w \geq d$

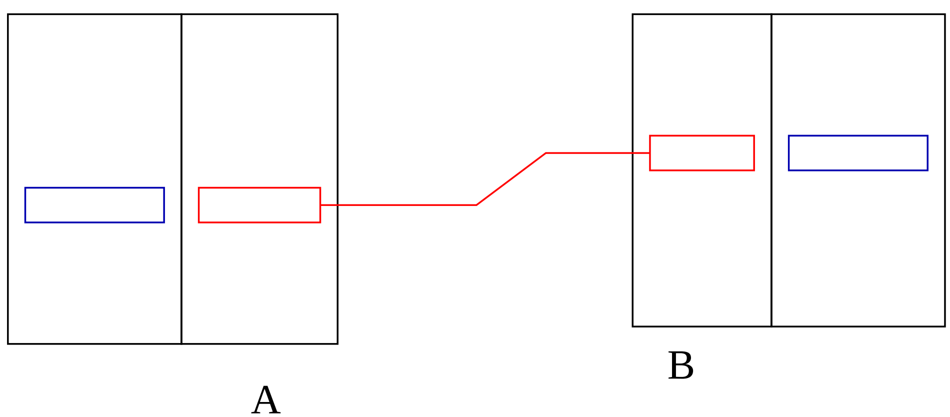
Context: Identity disclosure risk in data privacy

- How to find these parameters (μ and Q)?
- For risk analysis of a protected file X', we know both X and $A = X'$
- So, find best parameters using optimization (and $B = X$)



Context: Identity disclosure risk in data privacy

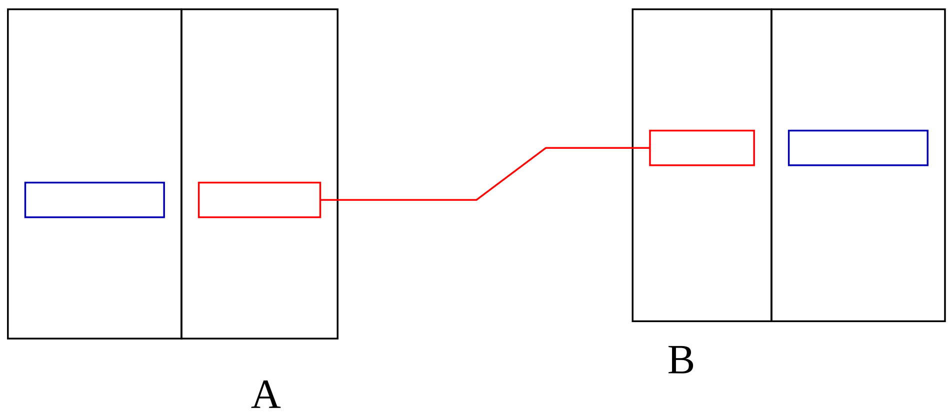
- Distance based record linkage: $d(A_i, B_i)$



- Find the *nearest* record (*nearest* in terms of a distance)
- Formally, 2 sets of vectors $A_i = (a_1, \ldots, a_N)$, (a_i protected version of b_i) $B_i = (b_1, \ldots, b_N)$
- $V_k(a_i)$: kth variable, ith record
- Distance $d(V_k(a_i), V_k(b_j))$ for all pairs (a_i, b_j).

Context: Identity disclosure risk in data privacy

- Distance based record linkage: $d(A_i, B_i)$



  - Find the *nearest* record (*nearest* in terms of a distance)
  - Formally, 2 sets of vectors $A_i = (a_1, \dots, a_N)$, (a_i protected version of b_i) $B_i = (b_1, \dots, b_N)$
  - $V_k(a_i)$: kth variable, ith record
  - Distance $d(V_k(a_i), V_k(b_j))$ for all pairs (a_i, b_j).

- Distance based on aggregation functions $\mathbb{C}$
  E.g., $\mathbb{C} = CI$ (Choquet integral)

- Worst-case scenario: learn weights/fuzzy measure
  $\rightarrow$ Optimization problem

Context: Identity disclosure risk in data privacy

- Distance based record linkage: $d(A_i, B_i)$
  - Main constraint: for a given i, for all j

$$\sum_{k=1}^{N} p_i d(V_k(A_i), V_k(B_j)) > \sum_{k=1}^{N} p_i d(V_k(A_i), V_k(B_i))$$

  For aligned files A and B (i.e., A_i corresponds to B_i)

- As this is sometimes impossible to satisfy for all i, introduce K_i which means $K_i = 1$ incorrect linkage, and then

$$\sum_{k=1}^{N} p_i (d(V_k(A_i), V_k(B_j)) - d(V_k(A_i), V_k(B_i))) + CK_i > 0$$

Context: Identity disclosure risk in data privacy

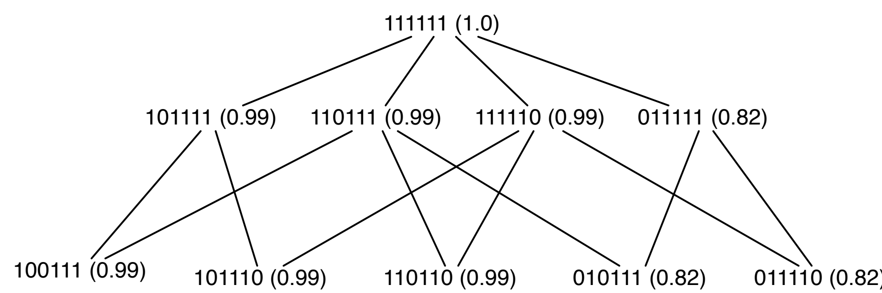
- Case $\mathbb{C} = WM$:

$$
\begin{aligned}
&Minimise \quad && \sum_{i=1}^{N} K_i \\
&Subject\ to: && \\
& && \sum_{k=1}^{N} p_i (d(V_k(a_i), V_k(b_j)) - d(V_k(a_i), V_k(b_i))) + CK_i > 0 \\
& && K_i \in \{0, 1\} \\
& && \sum_{i=1}^{N} p_i = 1 \\
& && p_i \geq 0
\end{aligned}
$$

- Similar with $\mathbb{C} = CI$ (Choquet integral) and μ
- Extensive work comparing different scenarios and $\mathbb{C}$.

Context: Identity disclosure risk in data privacy

- Results give:
  - number reidentifications in the worst-case scenario
  - Importance of weights (or sets of weights in fuzzy measures)
- Examples:
  - Choquet integral



  - Weighted Mean (WM):
    - ▷ V_1 0.016809573957189, V_2 0.00198841786482128, V_3 0.00452923777074791
    - ▷ V_4 0.138812880222131, V_5 0.835523953314578, V_6 0.00233593687053289

Distances on fuzzy measures

Classical case

How to compare fuzzy measures: using probability ones as inspiration

- f-divergence, KL-divergence, etc.
  based on Radon-Nikodym-like derivatives[1]
- Wasserstein distance/earth mover's distance
  based on optimal transport problem.

[1] Work based on Sugeno's work on Choquet calculus and in collaboration with Sugeno and Narukawa: INS 2020, FSS 2016, EUSFLAT 2013

Optimal transport for probabilities

Classical case

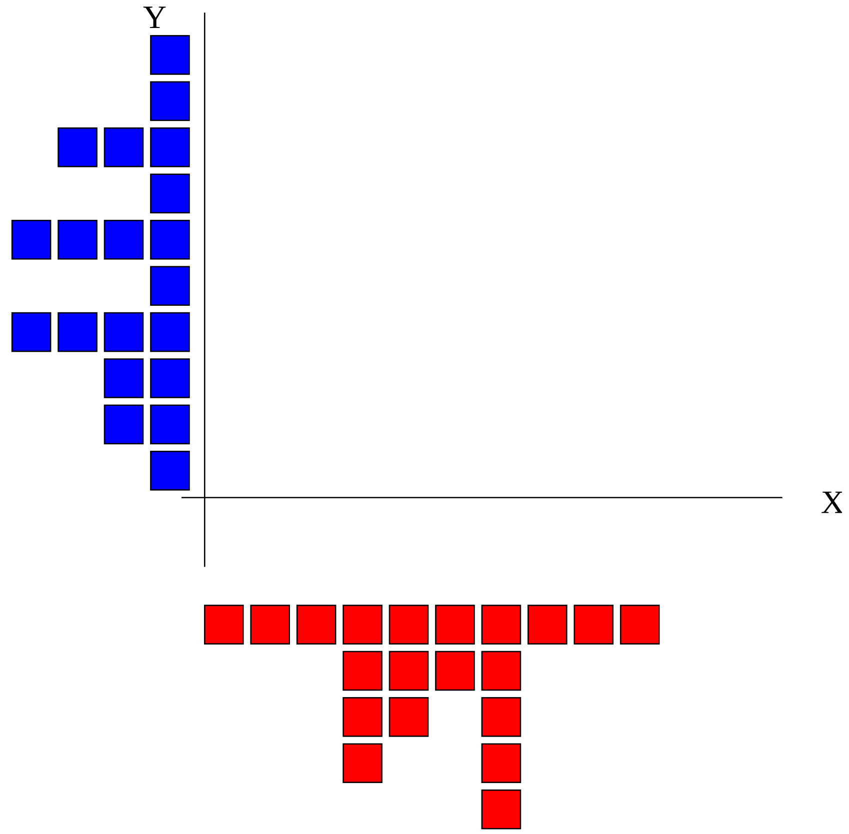
Optimal transport problem: The case of probabilities

- Inputs:
  - X, and probability measure P on X (with prob. dist. p)
  - Y, and probability measure Q on Y (with prob. dist. q) (on X and subsets of X, assume X finite)
- Output:
  - Assignment from P to Q
  - A cost of the assignment: optimal

Classical case

Optimal transport problem: The case of probabilities

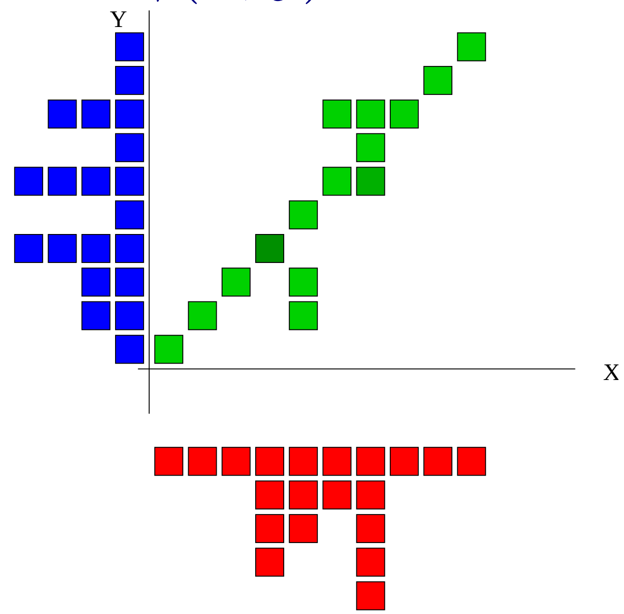
- Probability distributions on X and Y



Classical case

Optimal transport problem: The case of probabilities

- Assignment of probabilities $\gamma(x, y)$



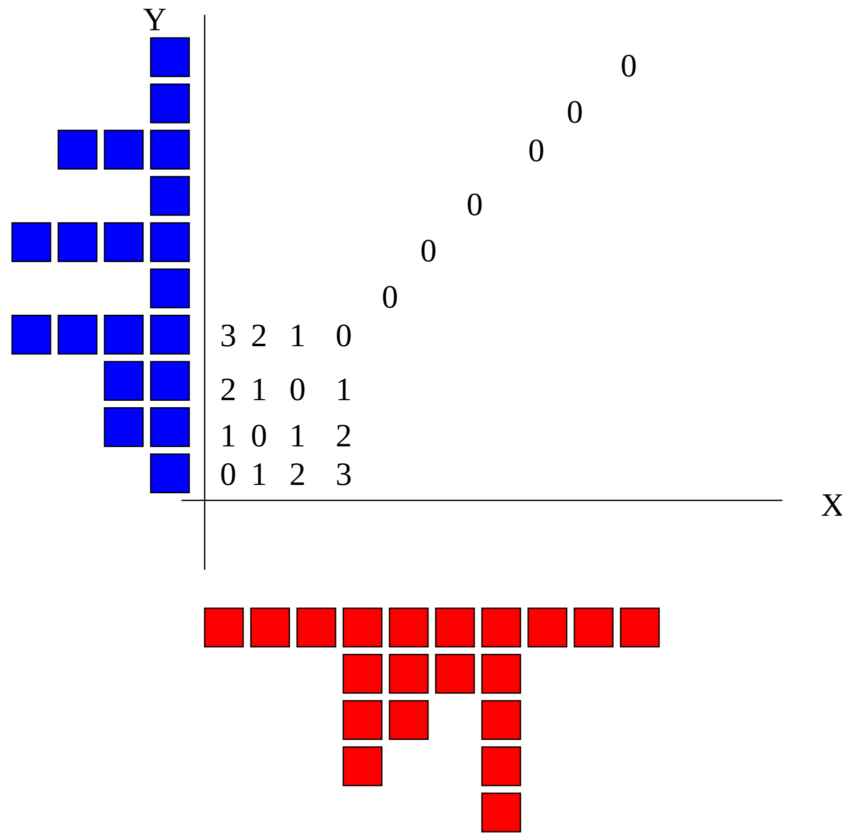
- γ positive, and marginals should be p and q

$$p(x) = \sum_{y \in Y} \gamma(x, y)$$
$$q(y) = \sum_{x \in X} \gamma(x, y)$$

Classical case

Optimal transport problem: a cost function

- $c : X \times Y \to \mathbb{R}^+$



- Cost: $\sum_{x \in X} \sum_{y \in Y} c(x,y)\gamma(x,y)$
- Distance: from the assignment with minimum cost.

Optimal transport and fuzzy measures

Optimal transport

Optimal transport problem: The case of non-additive measures

- input:
  - X, and fuzzy measure μ on X
  - Y, and fuzzy measure ν on Y
- Output:
  - Assignment from μ to ν
  - A cost of the assignment: optimal

Optimal transport

How to proceed?

- Option 0. We consider a cost function on $X \times Y$ and a Choquet integral of measures on $X \times Y$ with marginals μ and ν.
  - For all fuzzy measures in $X \times Y$, minimum CI

Optimal transport

How to proceed?

- Option 0. We consider a cost function on $X \times Y$ and a Choquet integral of measures on $X \times Y$ with marginals μ and ν.
  - For all fuzzy measures in $X \times Y$, minimum CI
- The problem seems difficult in practice
  - The Fubini theorem does not apply in general for Choquet integral
  - Margins, also Choquet integrals (?)

Measures, transforms, and optimal transport

Option 1 and 2

Transforms

Measures and transforms: Equivalent representation of a measure.

$$\mu \leftrightarrow \tau_\mu$$

They are set functions (same as μ):

$$\tau_\mu : 2^X \to \mathbb{R}$$

There are different transforms with different properties.

Transforms

Measures and transforms: $\mu \leftrightarrow \tau_\mu$

- Möbius transform

$$\tau_\mu(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \mu(B).$$

- If μ additive (probability)
  - $\tau_\mu(B) = p(x_i)$, if $B = \{x_i\}$ (singletons)
  - $\tau_\mu(B) = 0$, if $|B| > 1$ (non-singletons)
- If μ a belief function
  - $\tau_\mu(B) \in [0, 1]$

Optimal transport

- Option 1. If the measure is a belief function,
  Möbius transform is always positive
  - Probability on sets, define OT on Möbius transform
  - Marginals on the Möbius transform (addition of Möbius)
  - Cost functions on $2^X \times 2^X$
- Same problem but larger space, easy definition

Optimal transport

- Option 2. If the measure is not a belief, Möbius can be positive and negative
  - Use absolute value of the assignment

$$OF = \sum_{\emptyset \subset A \subseteq X} \sum_{\emptyset \subset B \subseteq X} c_M(A,B) |assg(A,B)|$$

- Different problem, doable: linear problem, linear constraints

Optimal transport

- Option 2. Problems:
  - Not only negative, but arbitrarily large (or small – negative).
  - For X with cardinality at least n, we can define a measure μ with
    - ▷ $\tau_\mu(A) = -n$ for sets of cardinality $n+1$, and
    - ▷ $\tau_\mu(A) = (n^2+n)/2$ for sets of cardinality $n+2$
  - In a way, we are counting the same *measure* multiple times

Measures, transforms, and optimal transport

Option 3

Transforms

Measures and transforms: $\mu \leftrightarrow \tau_\mu$

- $(\max, +)$-transform[2]

$$\tau_\mu(B) = \mu(B) - \max_{A \subset B} \mu(A)$$

  - The $(\max, +)$-transform is always positive in [0,1]
  - If μ additive $\tau_\mu(B) = \min_{x_i \in B} \mu(\{x_i\})$.

[2] V. Torra, $(Max, \oplus)$-transforms and genetic algorithms for fuzzy measure identification, Fuzzy Sets and Systems 28 (2022) 253-265

Optimal transport

- Option 3. Definition through the $(max, +)$-transform, cost $c_a : 2^X \times 2^X \to [0, 1]$
  - Find $assg : 2^X \times 2^X \to [0, 1]$ such that
    - ▷ $assg(\emptyset, \emptyset) = 0$
    - ▷ $\tau_\mu(A) = \sum_{B' \subseteq X} assg(A, B')$ for all $A \neq \emptyset$
    - ▷ $\tau_\nu(B) = \sum_{A' \subseteq X} assg(A', B)$ for all $B \neq \emptyset$
  - Cost of the assignment:
    - ▷ $cost(c_a, assg) = \sum_{A \subseteq X} \sum_{B \subseteq X} c_a(A, B) assg(A, B)$.

Optimal transport

- Option 3. Then, we can define:
  - Optimal transport: Assignment with minimal cost
  - Wasserstein-like discrepancy:

$$d_{c_a}(\mu, \nu) = \inf_{assg \in \Pi(\tau_\mu, \tau_\nu)} \sum_{A \subseteq X} \sum_{B \subseteq X} c_a(A, B) assg(A, B)$$

Optimal transport

- Option 3. Example, μ is additive, ν is not, $(\max, +)$-transforms τ_μ and τ_ν, feasible assignment:

$\nu(B)$	τ_ν	set	$lack_\nu$							
1	0.8	X	0	0	0.3	0.5	0	0	0	0
0	0	$\{x_2, x_3\}$	0	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_3\}$	0	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_2\}$	0	0	0	0	0	0	0	0
0	0	$\{x_3\}$	0	0	0	0	0	0	0	0
0	0	$\{x_2\}$	0	0	0	0	0	0	0	0
0.2	0.2	$\{x_1\}$	0	0.2	0	0	0	0	0	0
$lack_\mu$		$\emptyset$	$--$	0	0	0	0.1	0.1	0.4	0.1
set			$\emptyset$	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	X
τ_μ	$--$		$--$	0.2	0.3	0.5	0.1	0.1	0.4	0.1
$\mu(A)$			$--$	0.2	0.3	0.5	0.4	0.6	0.9	1.0

Optimal transport

- Properties[3]:
  - This is a proper generalization
  - When our FM solution is a probability solution?
    - ▷ assignment on X, Y vs. assignment on 2^X, 2^Y
    - ▷ cost on X, Y vs. cost on 2^X, 2^Y
  - Results on
    - ▷ A cost function that is independent on the measures
    - ▷ A cost function that depends on the measures

[3] V. Torra (2023) The transport problem for non-additive measures, Euro. J Oper. Res. 311 679-689

Optimal transport

- Implementation: Linear problem with linear constraints
  - 1. Linear problem with linear constraints (case belief functions)
    OT with 2^X variables
  - 2. Linear problem with linear constraints (case Möbius transform)
    Same but: transformation of $|t|$ into two additional constraints $+t \leq t'$, $-t \leq t'$. So, $2 \cdot 2^{2|X|}$ additional constraints
  - 3. Linear problem with linear constraints (case $(\max, +)$-transform)
    OT with 2^X variables.
    Software: http://www.mdai.cat/code/
    - ▷ What is an appropriate cost function?

Summary

Summary

- Results presented
  - Fuzzy measures for metric learning / distances
  - Distance for fuzzy measures
- Future directions
  - Distances for fuzzy measures
  - Foundations of optimal transport, Wasserstein distance and related topics for fuzzy measures

References

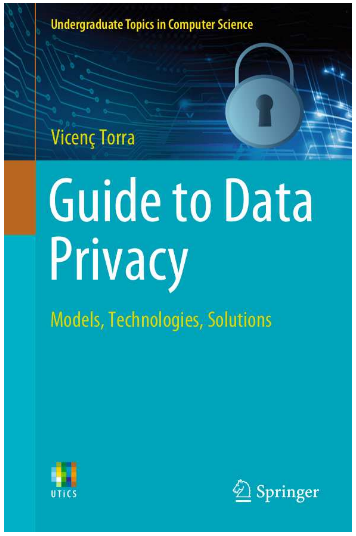
References

- D. Abril, V. Torra, G. Navarro-Arribas (2015) Supervised learning using a symmetric bilinear form for record linkage. Inf. Fusion 26: 144-153.

- V. Torra (2023) The transport problem for non-additive measures, Euro. J Oper. Res. 311 679-689

- V. Torra, G. Navarro-Arribas (2020) Fuzzy meets privacy: a short overview, Proc. INFUS 2020.

- V. Torra (2022) Guide to Data Privacy, Springer.



Thank you