Holistic AI × The Alan Turing Institute

Trade-offs of Bias with other verticals in Trustworthy AI Part II

Turing Course

Contents

- Part I – Regression and Multiclass
- **Part II – Clustering**
- Part III – Recommender Systems

Clustering algorithms – a reminder

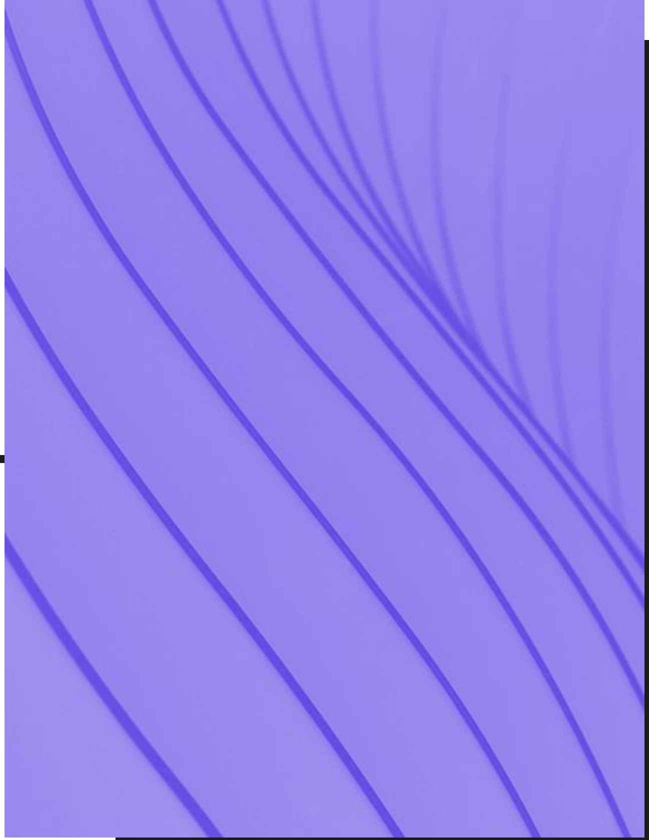
Partitional: determine all the groups at once (e.g. K-means).

Hierarchical: successively identify groups that split from or join groups that were established previously.

- agglomerative (“bottom-up”)
- divisive (“top-down”).

I – Explainability & Clustering

- Motivation
- Explainability by design: DReaM
- Post-processing: IMM & ExKMC
- Further readings



Motivation

- Ex1. Clustering candidates. How can it explain a candidate belonging to a particular cluster?

- Ex2. Clustering patients based on their scans from medical imaging techniques in order to identify cancer, tumors or other conditions.

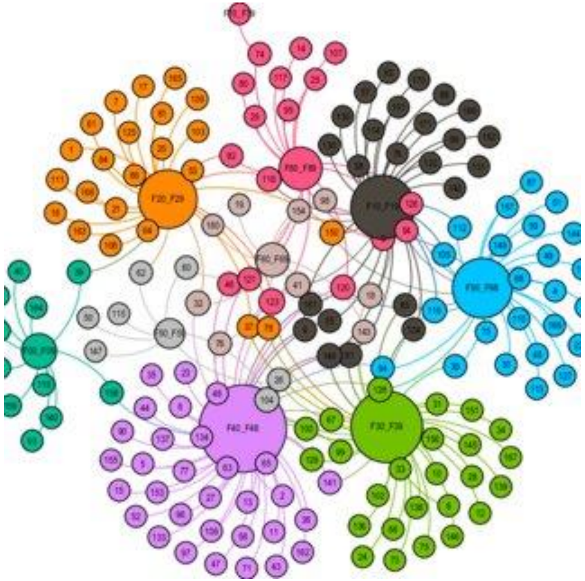
Explainability in Clustering Overview

Explainability by design.

– DReaM (Chen et al., 2016)

– CLTrees (Liu et al., 2005)

– CUBT - Clustering using unsupervised binary trees. (Fraiman et al., 2011)



Post-processing explainability.

– IMM (Dasgupta et al.,)

– ExKMC (Frost et al., 2020)

Dive-in. DReaM (Chen et al., 2016)

- Discriminative Rectangle Mixture

- Interpretable clustering that uses prior domain knowledge.
  Ex. clinical guidelines, to separate samples into groups (rules of thumb).

- Rules usually involve thresholding a set of features, which can be represented using rectangles in the feature space

- Split the feature set in:
  - ***Rule-generating features***. Used in combination with informative prior distributions (domain-knowledge) to define soft thresholds to define cluster boundaries.
  - ***Cluster-preserving features.*** clusters structure's preservation.

Dive-in. IMM & ExKMC

IMM

– Finding a clustering solution using some non-explainable clustering algorithm (like K-means)

– Labelling each example according to its cluster

– Calling a supervised algorithm that learns a decision tree

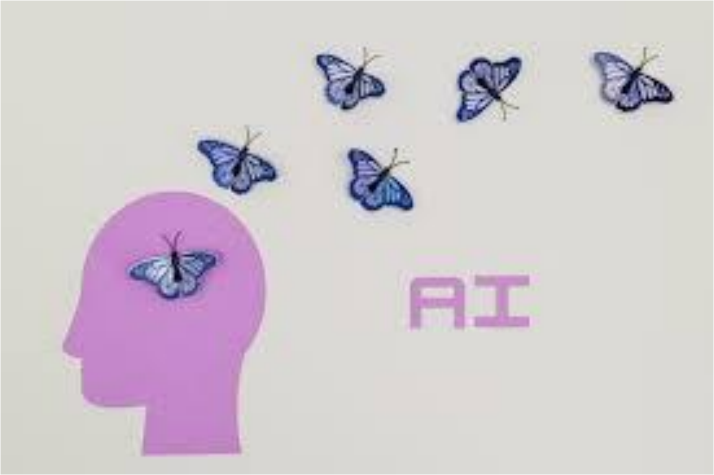
ExKMC

– Extension of IMM where the trees have more leaves than the number of clusters, to achieve better partitioning.

Interaction with Fairness

- As always -> Greater explainability go hand-in-hand with greater fairness

- Post-processing explanation (e.g. ExKMC) on a fair clustering method.



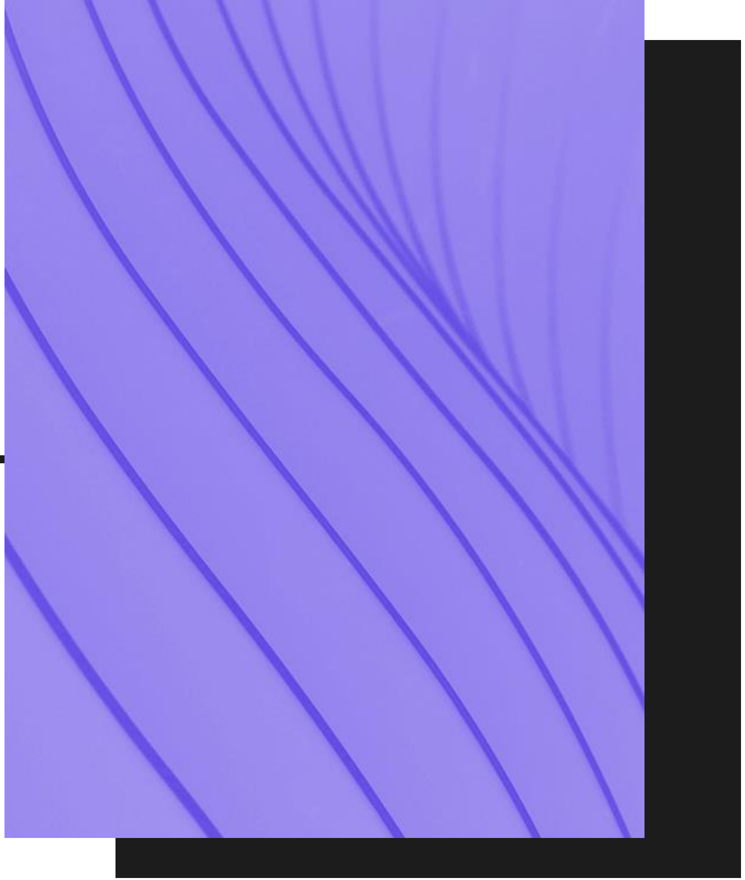
Further Readings

- Fitting a classifier to resulting cluster and using classifier's method (here SHAP): Shuyang Xiang's blog post

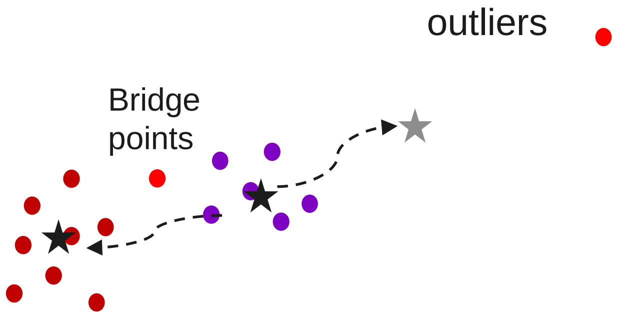
- IMM & ExKMC in practice: Mimi Dutta's article

II – Robustness & Clustering

- Motivation
- Example: Trimmed K-mean
- Robustness of Fairness
- Further readings



Motivation



k-means (k=2)

- **Breakdown point**: minimum fraction of data points which if significantly modified will also significantly modify the output of the estimator. Large breakdown point means better robustness.

- **Example**. In a class, who should repeat the year by constituting 2 clusters. One of the student has the maximum grade whilst all others are more towards the average grade. This will skew the clusters so that more students will be sent to repeat their year than if that outstanding student was not here.

Robustifying k-means (1/2)

– K-means -> least squares criterion, from which it inherits the lack of robustness

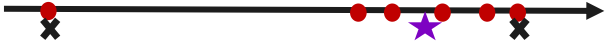
$$\sum_{i=1}^{n} (x_i - m)^2$$

$$\sum_{i=1}^{n} |x_i - m|$$

– Partitioning Around Medoids (PAM). L2 criterion replaced by L1 -> breakdown point is still equal to 0

Robustifying k-means (2/2)

Trimming. Ex. Trimmed mean



Which points to trim in a cluster setting? Trimmed k-means (García-Escudero et al, 2010). search of k centers $\{m_1, \dots, m_k\} \subset \mathbb{R}^p$ solving the double minimization problem:

$$\arg\min_{\mathbf{Y}} \min_{m_1,\dots,m_k} \sum_{x_i \in \mathbf{Y}} \min_{j=1,\dots,k} \left\| x_i - m_j \right\|^2$$

Where Y ranges on the class of subsets of size $[n(1 - \alpha)]$ within the sample $\{x_1, \dots, x_n\}$

$\alpha = 0$ --> normal k-means

Robustness of Fairness

Robust Fair Clustering (Chhabra et al, 2023)

Proposed attack. Changing a small portion of individuals’ protected group memberships => affects fairness significantly on existing fair clustering algorithms.

Proposed solution. Consensus Fair Clustering (CFC). Utilizes consensus clustering along with fairness constraints to output robust and fair clusters.

1. sample a subset of training data and run cluster analysis r times. Since attacked samples are a tiny portion of the whole training data, the probability of these being selected into the subset is also small, which decreases their negative impact. Creates r basic partitions
2. CFC fuses the basic partitions with a fairness constraint.

Further Readings

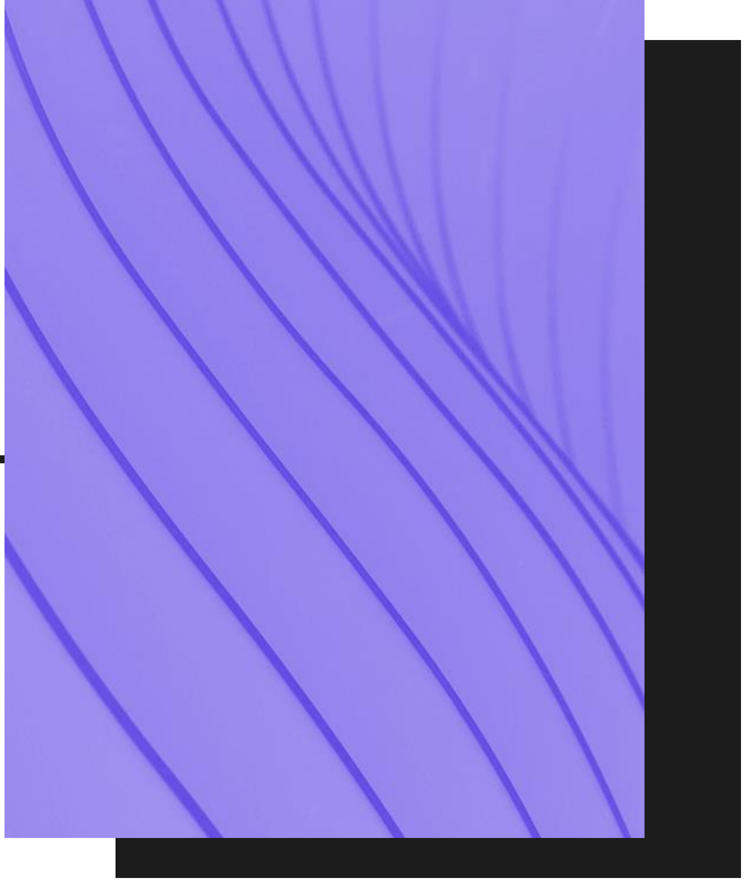
- **Review**: García-Escudero, L. A., Gordaliza, A., Matrán, C., & Mayo-Iscar, A. (2010). A review of robust clustering methods. Advances in Data Analysis and Classification, 4(2), 89-109.

- **For Hierarchical Clustering**: Robust Hierarchical Clustering by Balcan etal. 2014

- I**nteraction Robustness/Explainability** (Saisubramanian et al., 2020)

II – Privacy & Clustering

- Motivation
- Privacy preserving k-means
- Interactions with fairness
- Further readings



Motivation

- Clustering on union of dataset A and B, without revealing individual datasets.

- Ex. Bioinformatics where the data sets are owned by separate organizations, who do not want to reveal their individual data sets.

Privacy preserving k-means

Paper by Jha et al., 2005

Distributed k-means:

- With Trusted Third Party (TTP).
- Given mean μ_i for clusters, parties A & B assign data points to cluster, then send to TTP the pair:

$$\left(\boldsymbol{s}_i^A, \boldsymbol{n}_i^A\right)$$

Sum of samples in cluster i for A

Number of samples in cluster i for A

- Then TTP sends new mean to A & B: $\boldsymbol{\mu}_i = \frac{\boldsymbol{s}_i^A + \boldsymbol{s}_i^B}{\boldsymbol{n}_i^A + \boldsymbol{n}_i^B}$

Privacy preserving k-means

Privacy-preserving k-means:

- Third-party replaced by a privacy-preserving protocol
- Two parties need to jointly pick a common random vector for initial clusters means.

Interaction with Fairness

- In the k-means privacy protocol case, it is impossible to have access to the protected attributes in the combined cluster creation

- Only disaggregated fairness metrics can be evaluated. Less statistical power.

Further Readings

- **Differentially Private Clustering**: Google Research article. https://arxiv.org/pdf/1912.07820.pdf

- **Hierarchical privacy-preserving**: De, I., & Tripathy, A. (2014). A secure two party hierarchical clustering approach for vertically partitioned data set with accuracy measure. In Recent Advances in Intelligent Informatics (pp. 153-162). Springer, Cham.

Thank you!