



Universiteit
Leiden
The Netherlands

Improved strategies for distance based clustering of objects on subsets of attributes in high-dimensional data

Kampert, M.M.D.

Citation

Kampert, M. M. D. (2019, July 3). *Improved strategies for distance based clustering of objects on subsets of attributes in high-dimensional data*. Retrieved from <https://hdl.handle.net/1887/74690>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/74690>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/74690> holds various files of this Leiden University dissertation.

Author: Kampert, M.M.D.

Title: Improved strategies for distance based clustering of objects on subsets of attributes in high-dimensional data

Issue Date: 2019-07-03

Index

- 1SE rule, 73
- ApoE3 data, 76
- attribute
 - dispersion, 56, 100
 - distance, 10, 26
 - importance, 18, 174
 - weighting, 8, 186
 - weights, 26, 34, 100
- attribute dispersion, 56
 - mean-based, *see* non-robust
 - median-based, *see* robust
 - non-robust, 57, 100
 - robust, 57, 100
- attribute weights, 26, 100
 - for validation, 171
 - pairwise, 28, 99
 - zero-valued, 102
 - pre-specified, 101
- cluster analysis, 2
 - hierarchical, 6
 - partitional, 4
- cluster happy assumption, 28
- cluster validation, 169
- clustering
 - L -groups, 150
 - algorithm, 4
 - enhanced soft subspace , *see* ESSC
 - entropy weighted K -means, *see* EWKM
 - EWKM
 - hierarchical, 6
 - partitional, 4
 - sparse K -means, *see* SPARCL
 - sparse alternate sum, *see* SAS
 - subspace, 8, 153, 154
 - problem, 2
- computational costs
 - for COSA, 181
- convergence
 - for COSA, 182
- COSA, 25
 - λ NN algorithm, 115
 - K NN algorithm, 44
 - K NN criterion, 33, 99
 - K NN₀ algorithm, 100
 - K NN₁ algorithm, 103
 - and missing data, 183
 - attribute weights, 34
 - criterion, 29, 67
 - distance, 13, 27, 28, 103, 186
 - pseudo algorithm, 34
 - robust criterion, 69
 - within-cluster distance, 27, 99
- COSA- λ NN, 110, 115
 - algorithm, 110
 - complexity, 181
 - tuning, 160
- COSA- K NN
 - algorithm, 44
 - complexity, 181
 - tuning, 84, 160
- COSA- K NN₀, 100
 - algorithm, 100
- COSA- K NN₁, 103
 - algorithm, 103
 - tuning, 160

- PAM, 128
- partitioning, 4
 - around medoids, *see* PAM
- points of view analysis, 188
- pre-specified attribute weights, 101
- prototype model, 13, 104

- Rand Index, 142
- rate parameter, 39
- relaxation parameter, 44
- robust COSA-KNN criterion, 69

- SAS, 151
 - default grid search, 157, 159
 - golden section search, 157, 159
- scale parameter, 35, 46
- self-organizing maps, 189
- shared nearest neighbors, 137
 - similarity between two clusters, 137
 - similarity between two objects, 137
- silhouette, 130
- SMACOF, 16
- softmax function, 35
- SPARCL, 153
 - tuning, 159

- sparse K -means, *see* SPARCL
- sparse alternate sum clustering, *see* SAS

- STRAIN, 16
- STRESS, 15

- targeted distance, 187
 - dual, 188
 - single, 187
- triangular inequality, 10
 - violation, 29
- Type-I error, 87

- validation
 - attribute importance, 174
 - attribute weights, 171
 - clustering, 169

- Ward
 - linkage, 7, 134
 - minimum variance method, 134
- weak spot model, 106
- weights
 - attribute, *see* attribute weights

- zero-valued attribute weights, 102
- zigzag tendency, 69

