Clustering Selection Management System Report

mk3.csv – unsupervised

Tuesday 17th November, 2020 – 18:21

Preprocessing

Table 1: Specific Metrics for all Preprocessing steps

Metric	# Rows w/ missings removed	# Converted columns to OHE	# Quantiles for Quant. Scal.	# Non Distinct rows removed
Value	0	0	60	0

Setup

Hardware

Table 2: Hardware Statistics of the underlying Hardware Setup

Statistic	*Amount of main memory*	*# CPU-Threads*	# CPU-Cores
Value	8.59	4	2

Input Parameters

Table 3: Given *General* Input-Parameter Values

Parameter	Accuracy Efficiency Preference	Prefer Finding arbitrary Cluster Shapes?	Avoid High Effort of (Hyper-) Parameter Tuning?
Value	efficiency	True	False

Table 4: Given *Distance-Metric-based* Input-Parameter Values

Parameter	Find Compact or Isolated Clusters?	Ignore Magnitude and Rotation?	Measure Distribution Differences?	Grid-based Distance?
Value	True	False	False	False

Metadata

Table 5: *General* Profiled Metadata Results regarding the Dataset

Statistic	*#Rows*	*#Columns*	#Classes	# Missing Values
Value	600	3	0	0

Table 6: *Further* Profiled Metadata Results regarding the Dataset

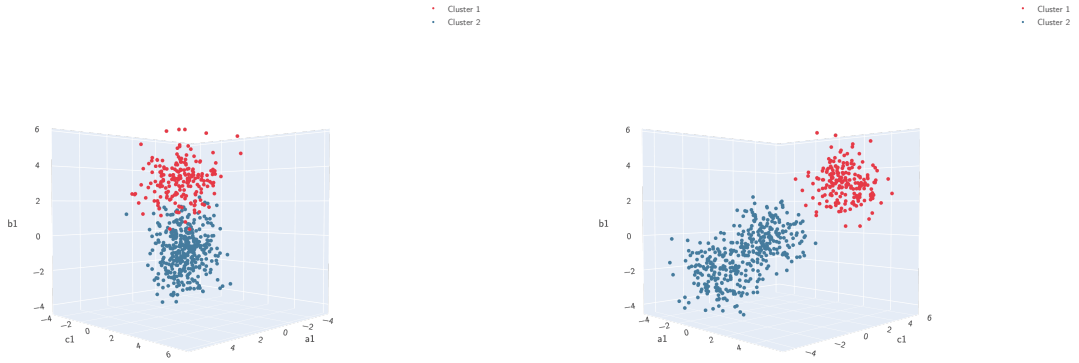
Statistic	*Outlier %*	*High Correlation %*	Class Std. Deviation
Value	0.025	1.0	-

Selection Steps

Table 7: Listing of all CSMS Iterations

Iteration	Selected Algorithm	Selection-Score	Tuned (Hyper-) Parameters	Silh. Score of Sampling
Iteration 1	vbgmm	7.71	max_n_components = 15	0.63
Iteration 2	optics	6.80	distance = mahalanobis, min_samples = 2	0.00
Iteration 3	em	6.67	n_clusters = 2	0.80
Iteration 4	kmeans	3.87	n_clusters = 2	0.80
Iteration 5	affinity	3.22		0.61
Iteration 6	spectral	3.13	n_clusters = 2	0.80
Iteration 7	agglomerative	2.27	distance = euclidean, n_clusters = 2	0.80
Iteration 8	meanshift	2.24		0.80

Results



(a) Clustering View 1

(b) Clustering View 2

Figure 1: Final Clustering result, represented in two different views of the same plot

Table 8: Final Clustering Result

Algorithm	Tuned (Hyper-) Parameters	Reached Silh. Score	Total CPU-Runtime of the CSMS
spectral	n_clusters = 2	0.80	14.95s