

Assessment of Stability in Partitional Clustering Using Resampling Techniques

Hans-Joachim Mucha

Abstract The assessment of stability in cluster analysis is strongly related to the main difficult problem of determining the number of clusters present in the data. The latter is subject of many investigations and papers considering different resampling techniques as practical tools. In this paper, we consider non-parametric resampling from the empirical distribution of a given dataset in order to investigate the stability of results of partitional clustering. In detail, we investigate here only the very popular K -means method. The estimation of the sampling distribution of the adjusted Rand index (ARI) and the averaged Jaccard index seems to be the most general way to do this. In addition, we compare bootstrapping with different subsampling schemes (i.e., with different cardinality of the drawn samples) with respect to their performance in finding the true number of clusters for both synthetic and real data.

1 Introduction

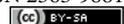
Originally, nonparametric bootstrapping is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from

Hans-Joachim Mucha
Weierstrass Institute for Applied Analysis and Stochastics (WIAS), 10117 Berlin, Mohrenstraße 39,
Germany,
✉ mucha@wias-berlin.de

ARCHIVES OF DATA SCIENCE, SERIES A
KIT SCIENTIFIC PUBLISHING
Vol. 1, No. 1, S. 21–39, 2016

DOI 10.5445/KSP/1000058747/02

ISSN 2363-9881



the original sample (Efron, 1979, 1981). Many authors such as Mammen (1992) derived asymptotic results of parametric bootstrapping. Others presented simulation results of parametric/nonparametric bootstrapping (Efron, 1981; Efron and Tibshirani, 1993).

This very simple technique allows estimation of the sampling distribution of almost any statistic. Bootstrapping falls in the broader class of resampling methods and simulation schemes. Some alternative resampling methods are subsampling (draw a subsample to a smaller size without replacement) and jittering (add noise to every single observation), and a combination of both simulation schemes.

In hierarchical cluster analysis (HCA), we found out that bootstrapping performs best for finding the number of clusters (Mucha and Bartel, 2014, 2015). In all cases (toy and real data), it outperforms subsampling. In subsampling, the choice of the parameter “resampling rate” p causes an additional problem. A subsampling rate of 90% (i.e., $p = 0.9$: this corresponds in some sense to tenfold-cross-validation) or greater performs very bad in HCA methods such as Ward and Average Linkage. The question arises: Is bootstrapping also the best choice for stability investigations of results of partitional clustering?

2 Partitional and hierarchical cluster analysis

A recent survey of partitional and hierarchical clustering algorithms is given by Reddy and Vinzamuri (2014). Here we will emphasize the differences of these two families of cluster analysis methods with respect to the results that have to be assessed by resampling methods. Hierarchical clustering looks fit and proper for resampling because of the (usual) unique and parallel clustering of the I observations into partitions of $K = 2, K = 3, \dots$ clusters. (Here, a partition $P(I, K)$ is simply the exhaustive partitioning of the set of I observations into K subsets (clusters).) In addition, pairwise distances, the usual starting point of hierarchical cluster analysis, are not affected by bootstrapping/subsampling.

The results of partitional (iterative) clustering methods are dependent on the initial partition into a fixed number of clusters K . That’s quite different from hierarchical clustering. In addition, the results of some exchange algorithms are also dependent on the sequence of the observation (Mucha, 2009). For instance, Fig. 1 shows a quite bad result of clustering of a dataset of three two-dimensional randomly generated normal subpopulations. The three Gaussian

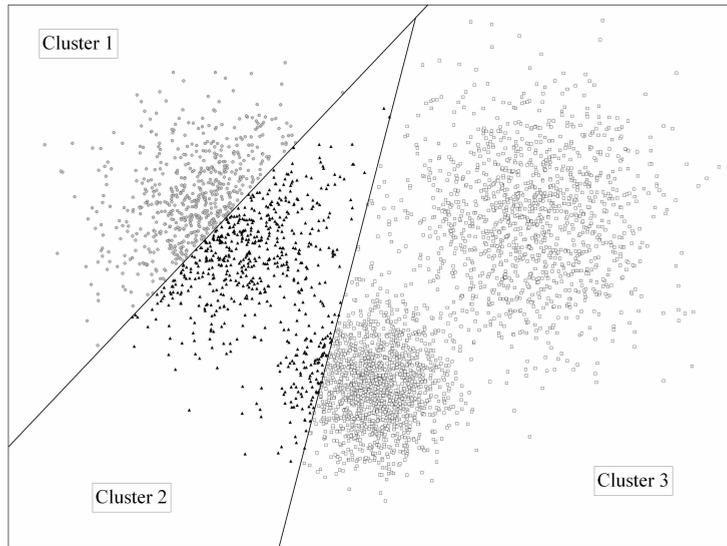


Fig. 1 Result of *Quickcluster* of SPSS applied to 4000 observations.

subpopulations were generated with the following parameters: cardinalities 1100, 1600, and 1300, mean values $(-3, 3)$, $(0, 0)$, and $(3, 3)$, and standard deviations $(1, 1)$, $(0.7, 0.7)$, and $(1.2, 1.2)$. Here the procedure *Quickcluster* of SPSS is applied with the option *running means*: the clusters are updated after each observation is assigned to a new cluster. In this two-dimensional setting, one can check the validity of cluster analysis results visually by eye. In a high-dimensional setting, there is a need for a general validation approach that works in almost all situations (see the next subsection). In this paper, the partitional clustering methods of our software *ClusCorr98* are used (Mucha, 2009). Here, a random access to the observations is realized. This is in order to avoid such bad solutions as shown in Fig. 1. Usually, many different initial partitions, say around 50, are needed to get many different locally optimal solutions. In practice, the best solution is taken for the investigation of stability. Moreover, you have to do this for each K ($K = 2, 3, \dots$). Finally, you have to do all the things outlined above also for each bootstrap sample (or subsample). Obviously, resampling of partitional clustering looks much more costly in terms of computational complexity than hierarchical clustering. The good news is that some partitional methods such as K -means clustering can work with pairwise distances which are not affected by bootstrapping/subsampling.

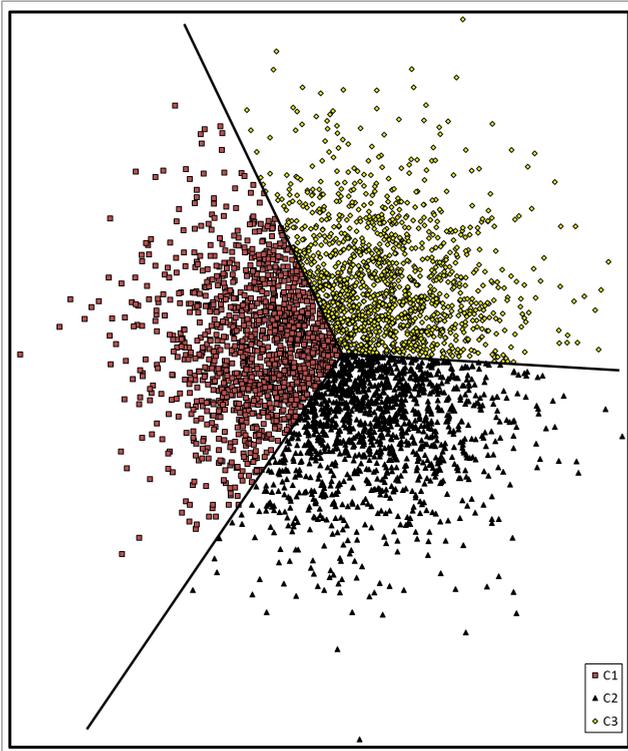


Fig. 2 K -means clustering of the two-dimensional no-structure data into three clusters (marked by black borderlines and color).

Concerning interpretation/comparison of the assessment of stability of two partitions $P(I, K)$ and $P(I, K + 1)$ of a hierarchy one has to keep in mind that exactly $K - 1$ clusters are identical, i.e., only one cluster is changed when going from $P(I, K)$ to $P(I, K + 1)$. That means, theoretically, the lower K the more the stability of the partition $P(I, K)$ depends on the stability of the partition $P(I, K + 1)$. This is different from partitional clustering where, usually, all clusters of the two partitions are different.

Even though both clustering techniques, the well-known hierarchical Ward's method and the partitional K -means method, have the same underlying statistical model (Banfield and Raftery, 1993), the results are usually different. Both methods minimize the same criterion (Eq. 1) below but they do this in another way. The K -means clustering method produces the well-known Voronoi tessellation, where the objects have minimum distance to their centroid and,

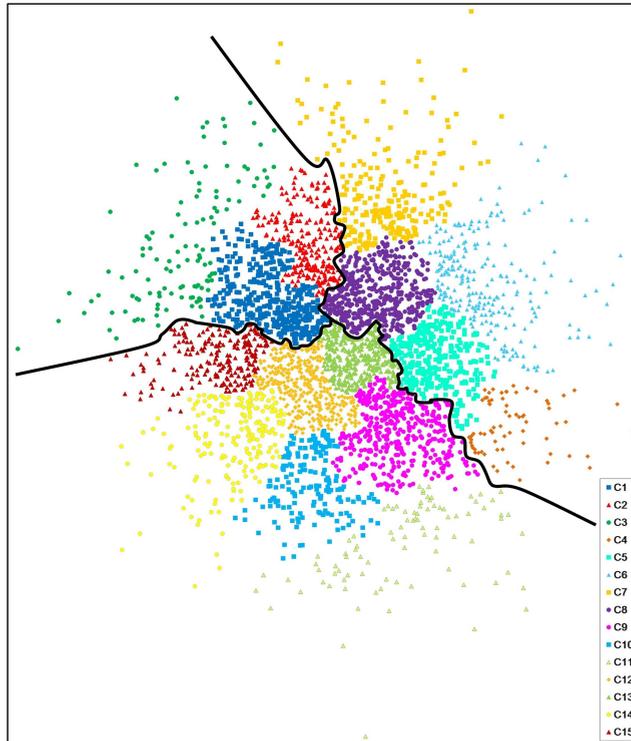


Fig. 3 Ward's clustering of the two-dimensional no-structure data into three clusters (marked by black borderlines) and into 15 clusters (marked by color), respectively.

thus, the borderlines between clusters are hyperplanes as shown in Fig. 2. There are 4000 random generated points in R^2 coming from a standard normally distributed population. In detail: a K -means clustering was done here based on pairwise proximities (squared Euclidean distances, see equations (4) and (5) below). By contrast, the Ward method does not create hyperplanes as borderlines between clusters as illustrated in Fig. 3 for the three cluster solution. Both the hierarchical Ward method and the partitional K -means method minimize the within-cluster sum of squares criterion

$$W_K(\mathbf{G}) = \sum_{k=1}^K \text{tr}(\mathbf{W}_k) \quad (1)$$

with respect to a Boolean assignment matrix \mathbf{G} for a fixed K (for details see below).

Herein

$$\mathbf{W}_k = \sum_{i=1}^I g_{ik} (\mathbf{x}_i - \bar{\mathbf{x}}_k) (\mathbf{x}_i - \bar{\mathbf{x}}_k)^T \quad (2)$$

is the sample cross-product matrix for the k th cluster \mathcal{C}_k of a given data matrix $\mathbf{X} = (x_{ij})$ consisting of I rows and J columns (variables), and

$$\bar{\mathbf{x}}_k = \frac{1}{g_{\cdot k}} \sum_{i=1}^I g_{ik} \mathbf{x}_i \quad (3)$$

is the usual maximum likelihood estimate of expected values in cluster \mathcal{C}_k . Further, $g_{\cdot k}$ is the cardinality of cluster \mathcal{C}_k , that is, $g_{\cdot k} = \sum_i g_{ik}$.

The Boolean assignment matrix \mathbf{G} formalizes the simplest (elementary) solution to the clustering problem with a fixed number of clusters K : $\mathbf{G} \in \{0, 1\}^{I \times K}$ (that is, $\mathbf{G} = (g_{ik})$) with the restriction of uniqueness and exhaustive assignment (completeness) $\sum_{k=1}^K g_{ik} = 1$ for every object i . Formally, the mapping is:

$$G: \mathcal{C} \times \{1, 2, \dots, K\} \longrightarrow \{0, 1\}$$

with

$$g_{ik} = \begin{cases} 1 & \text{if observation } i \text{ comes from the cluster (subset) } \mathcal{C}_k \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, the cluster mapping G induces a partition $P(I, K) = \{\mathcal{C}_1, \dots, \mathcal{C}_K\}$ of \mathcal{C} . Here, by definition, $\bigcup_{k=1}^K \mathcal{C}_k = \mathcal{C}$ and $\mathcal{C}_k \cap \mathcal{C}_l = \emptyset$ for every pair of clusters \mathcal{C}_k and \mathcal{C}_l , $k, l = 1, 2, \dots, K, k \neq l$. This cluster mapping yields exactly K clusters (subsets), where the numbering of the clusters is arbitrary because it usually depends on the applied clustering algorithm. Alternatively, let $\mathbf{g} = (g_1, \dots, g_I)^T$ denote the identifying labels for the clustering and thus for the cluster mapping G , where $g_i = k$ if the i th object \mathbf{x}_i comes from the k th cluster. One can understand \mathbf{g} as a categorical variable or partition variable with K different nominal states $\{1, 2, \dots, K\}$. Formally, $\mathbf{g} = \mathbf{G}\mathbf{e}$, where the vector $\mathbf{e} = (1, 2, 3, \dots, K)^T$ has K entities.

It is well known that the criterion (1) can be written in the following equivalent form without the explicit specification of cluster centers (centroids) $\bar{\mathbf{x}}_k$ (Späth, 1982):

$$W_K(\mathbf{G}) = \sum_{k=1}^K \frac{1}{2g_{\cdot k}} \sum_{i=1}^I \sum_{h=1}^I g_{ik} g_{hk} d_{ih}, \quad (4)$$

and

$$d_{ih} = d(\mathbf{x}_i, \mathbf{x}_h) = (\mathbf{x}_i - \mathbf{x}_h)^T (\mathbf{x}_i - \mathbf{x}_h) = \|\mathbf{x}_i - \mathbf{x}_h\|^2 \quad (5)$$

is the squared Euclidean distance between two observations i and h .

In practice, it is not possible to know how good our best (sub-optimum) result matches both the true (but unknown) classes and the global optimum. We start with many different initial partitions (usually 50), and we select the one that gives the best criterion value. Cluster ensemble methods are another approach in order to find a better cluster analysis result (see, for instance, Minaei-Bidgoli et al, 2014; Fischer and Buhmann, 2003).

3 Resampling techniques in cluster analysis

Nonparametric bootstrapping is resampling taken with replacement from the original data. Equivalently, bootstrapping can be formulated by choosing the following random weights of the observations:

$$m_i = \begin{cases} n & \text{if observation } i \text{ is drawn } n \text{ times} \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Here we suppose that the original weights of the observations are $m_i = 1, i = 1, 2, \dots, I$ (“unit mass”). Then, obviously, $I = \sum_i m_i$ holds in resampling with replacement. Bootstrapping generates multiple observations. When clustering “small” datasets, this can cause problems. The meaning of “small” depends on several factors of influence such as the number of dimensions (variables) and the complexity of the cluster analysis model. Small can be, for instance in the case of simple models such as K -means clustering or Ward’s method, a relation $I/K < 5$ with regard to the number of expected clusters K , or a number of observations $I < 20$. In the last situation, soft bootstrapping is recommended by Mucha and Bartel (2014). All statistical methods that make use (directly or indirectly) of weights of the observations can do bootstrapping based on (6). Concerning the K -means method based on pairwise distances, the “centers-free” criterion (4) can be generalized by introducing the weights of the observations to

$$W_K(\mathbf{G}) = \sum_{k=1}^K \frac{1}{2M_k} \sum_{i=1}^I m_i \sum_{h=1}^I g_{ik} g_{hk} m_h d_{ih}. \quad (7)$$

Obviously, it allows a computationally efficient bootstrapping because the pairwise distances (5) remain unchanged in the K -means clustering.

Subsampling is resampling taken without replacement from the original data. It can also be formulated by choosing the following random weights of the observations:

$$m_i^* = \begin{cases} 1 & \text{if observation } i \text{ is drawn randomly} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Here $I > L = \sum_i m_i^*$ holds in resampling without replacement. The parameter $p = L/I$ is needed which causes an additional problem, i.e., setting the cardinality L of the drawn sample. This is different from bootstrapping where no parameter is needed because here the cardinality of the drawn sample always equals I . Below we will investigate subsampling with different p values, say $p = 0.6$ (“Sub60%”), $p = 0.75$ (“Sub75%”), and $p = 0.9$ (“Sub90%“). A practical way out from choosing the parameter p would be to discard multiple points in a bootstrap scheme (named “Boot2Sub“ in the investigations below). Concretely, the random bootstrap-weights m_i in (6) have to be modified simply to

$$m_i^* = \begin{cases} 1 & \text{if observation } i \text{ is drawn } n \text{ times} \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

As a consequence of subsampling via (9), the cardinality of such a subsample “Boot2Sub“ is around 63.2% of the I observations (see Efron and Tibshirani, 1997). Clearly, “Boot2Sub“ (based on (9)) and bootstrapping (based on (6)) lead to identical results for all the cluster analysis methods that make no use (directly or indirectly) of the weights of the observations m_i such as the hierarchical Single Linkage or Complete Linkage method.

For instance, the resampling method can be used to investigate the variations of the centroids of the clusters, see Mucha and Bartel (2014). As an application, Fig. 5 shows the estimates of the location parameters that are the result of hierarchical Ward’s clustering of 250 non-parametric subsamples of the toy dataset presented in Fig. 4. Here three clusters were investigated (for details see Mucha and Bartel, 2014). But, in clustering, the estimation of parameters such as the expected values is not the main task. However, in the case of quantitative data, an estimation of the confidence regions around the cluster centroids can be of interest. The final aim of clustering is the formation of groups either as a partition or a hierarchy of a given set of observations. Therefore, here the focus is on a general investigation of the stability based on partitions. This covers also hierarchies because they can be considered as a set of partitions (Mucha,

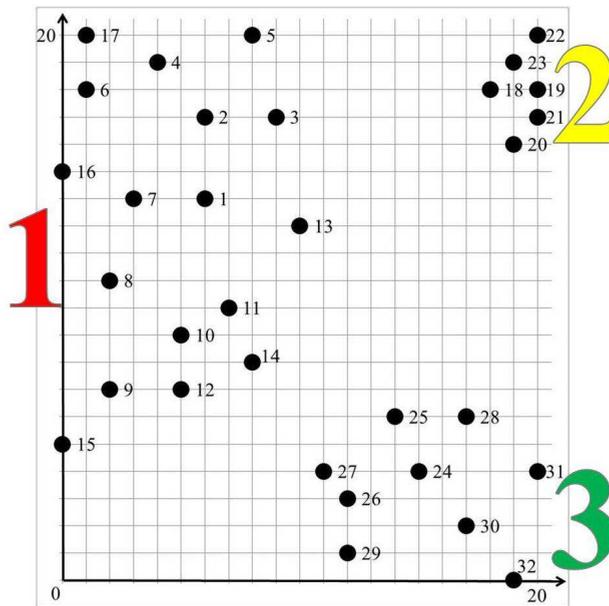


Fig. 4 Plot of the two-dimensional toy dataset divided into three classes by eye. The latter can be found exactly by the partitional K -means clustering. The data values are integers. They can be taken directly from the plot. The observations are numbered.

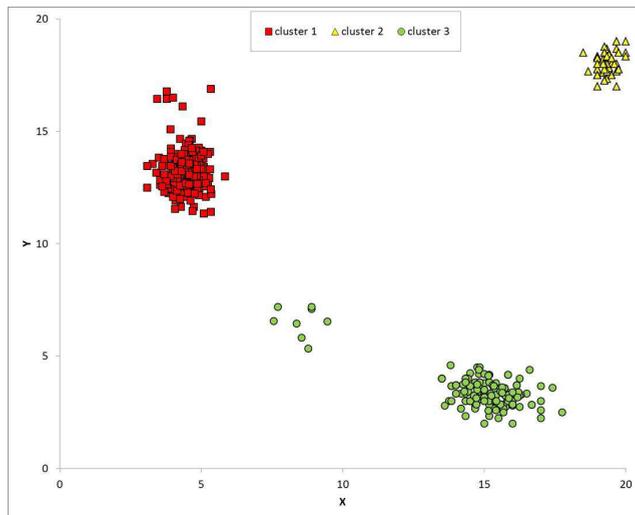


Fig. 5 Plot of the estimates of the location parameter of clusters. They are the result of Ward's HCA of 250 subsamples (75% resampling rate) into three clusters.

2007). To assess the stability of a cluster in the most general way, resampling techniques can be used.

Xiong and Li (2013) investigated many measures of stability with reference to cluster analysis. Here our focus is on two measures, namely the adjusted Rand index (ARI) R and the Jaccard index γ . Why is validation of clustering so important? That is because cluster analysis presents clusters in almost any case. Real clusters should be stable, i.e., they should be confirmed and reproduced to a high degree if the dataset is changed in a non-essential way (Hennig, 2007). Thus, clustering of a randomly drawn sample of a dataset consisting of really well-separated clusters should lead to similar results.

In clustering, usually nothing is known about the true class structure, especially about the number of clusters K . Therefore, the performance or the stability of clustering can not be assessed by counting the rate of misclassifications based on a confusion matrix. However, with the help of non-parametric bootstrapping we are able to operate also on a confusion matrix. It comes from crossing two partitions: the original one and one coming from clustering a “bootstrap” sample. Then the adjusted Rand index or other measures of stability can operate on such an “artificial” confusion matrix. Usually, hundreds of bootstrap samples are needed, see for details (Mucha and Bartel, 2015). Here we work with $B = 250$ bootstrap samples and we take the average (or median) of the B ARI values to come to a final $R_K, K = 2, 3, \dots$. The maximum R_K gives us an idea about the number of clusters K we are looking for.

In addition to ARI, bootstrapping the Jaccard coefficient can be recommended. The latter assesses the similarity between sets (clusters), for details, see Hennig (2007). It can be used to measure the stability of each individual cluster k by the corresponding Jaccard values γ_k^b with regard to the bootstrap sample $b, b = 1, 2, \dots, B$. Then we take the average (or median) of the B Jaccard values to come to γ_k that assesses the stability of an individual cluster k . Both the ARI and the averaged Jaccard measure γ_K are recommended for an investigation of the stability of a partition into K clusters. Here, the latter is the average of all Jaccard values γ_k of the K individual clusters of a partition into K clusters. An alternative proposal can be, for instance, a weighted average of all Jaccard values γ_k .

To summarize, bootstrapping of a stability measure is based on an original clustering that is compared many times to corresponding clustering results coming from a bootstrap sample. Concerning more details about bootstrapping a stability index see Mucha (2007); Hennig (2007); Mucha and Bartel (2015).

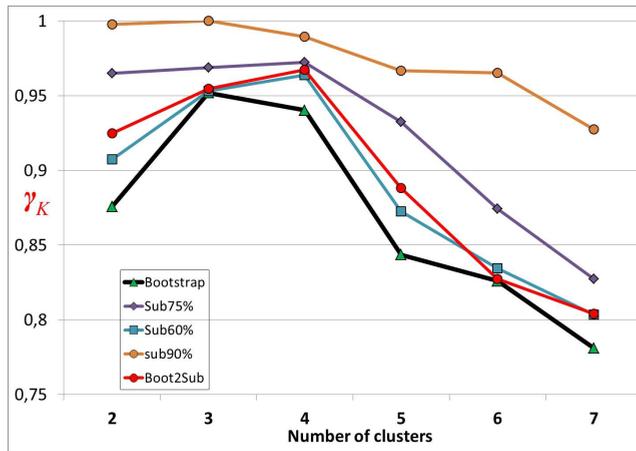


Fig. 6 Jaccard's measures of partitional K -means clustering (shown in Fig. 4) of the toy data.

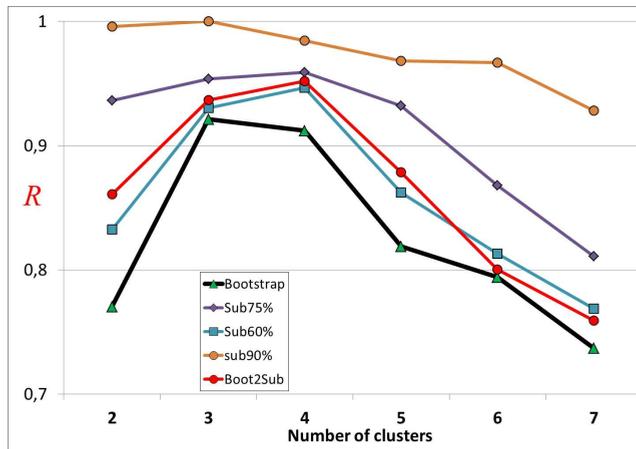


Fig. 7 ARI measures of partitional K -means clustering (shown in Fig. 4) of the toy data.

Other ways of the evaluation of cluster solutions via the bootstrap can be found, for instance, in Fang and Wang (2012), and Dolnicar and Leisch (2010).

4 Bootstrapping versus subsampling in partitional cluster analysis

Fig. 4 introduces a toy dataset consisting of three classes $\mathcal{C}_1 = \{1, 2, \dots, 17\}$, $\mathcal{C}_2 = \{18, 19, \dots, 23\}$, and $\mathcal{C}_3 = \{24, 25, \dots, 32\}$, i.e., it seems to be plausible that there are three classes when looking at the scatterplot. In Fig. 6, different resampling techniques are compared based on the averaged Jaccard measure γ_K for the validation of results of the toy data shown in Fig. 4. The three cluster solution of K -means clustering matches exactly the three classes shown in Fig. 4. Fig. 7 shows similar results as Fig. 6 but based on the ARI R_K .

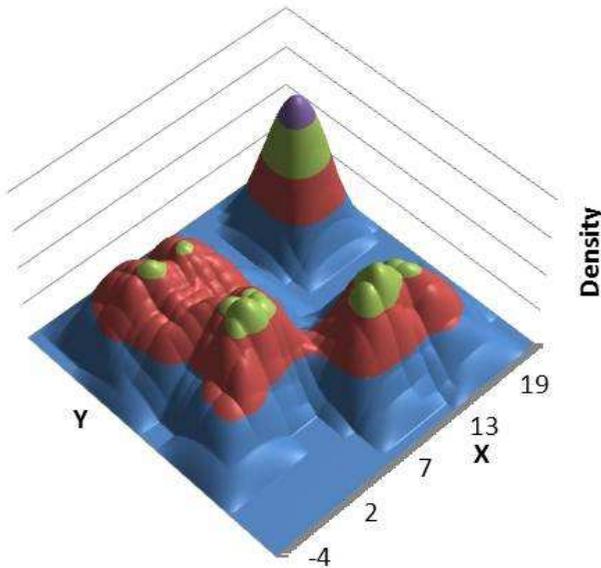


Fig. 8 Plot of the bivariate density estimate of the toy data.

Without much doubt, in this experiment only bootstrapping finds out that there are three clusters. In addition, the ARI “outperforms“ Jaccard with respect to the steepest rise when going from $K = 2$ to $K = 3$ clusters. But both present similar results and especially both vote clearly for three clusters and for at most four clusters. The latter because of the steep decrease when going further on to five clusters. Almost all subsampling versions fail in finding the three clusters. In addition, “Sub90%“ doesn’t indicate any partition clearly. Fig. 8 shows a continuous representation of the toy data. Only class 2 looks homogeneous and well separated (see also Fig. 4), and, maybe, there are more than three peaks.

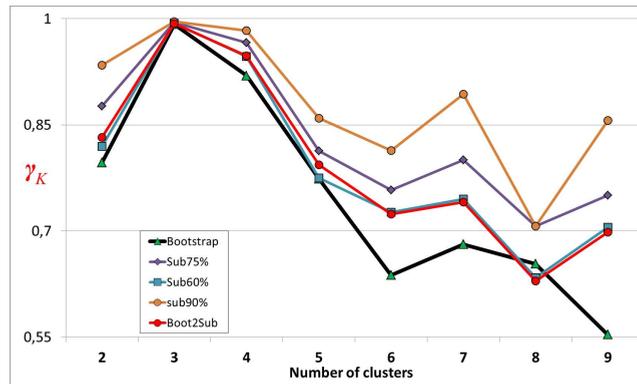


Fig. 9 Jaccard's measures of K -means clustering of the Gaussian data.

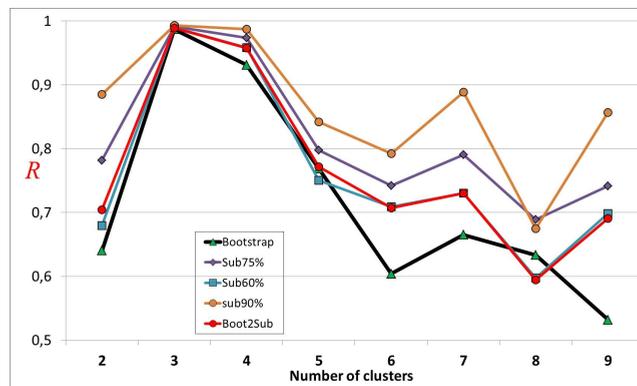


Fig. 10 ARI measures of partitional clustering of the Gaussian data.

Similar to Figs. 6 and 7, Figs. 9 and 10 show the validation results of the partitional K -means clustering of the randomly generated two-dimensional three class data based on the averaged Jaccard measure γ_K and the ARI R_K , respectively. The three Gaussian sub-populations were generated with the following parameters: cardinalities 80, 130, and 90, mean values $(-3, 3)$, $(0, 0)$, and $(3, 3)$, and standard deviations $(1, 1)$, $(0.7, 0.7)$, and $(1.2, 1.2)$. K -means clustering is successful in dividing (decomposing) the data into three subsets: only five errors are counted.



Fig. 11 Variance of ARI measures of partitional clustering of the Gaussian data.

Here bootstrapping performs also best in finding the three classes because it has

1. the maximum value at $K = 3$,
2. the most steeply rising when coming from $K = 2$ and going to $K = 3$, and
3. the most steeply sloping when going further on to $K = 4$.

As before, “Sub90%“ performs worst. Subsampling “Boot2Sub“ looks most similar to bootstrapping. However, looking at the variances of ARI, say for $K = 2$, bootstrapping has nearly three times more variance (Fig. 11). Bootstrapping looks very instable for the $K = 2$ solution in contrast to “Boot2Sub“, and, thus, bootstrapping excludes the wrong solution much clearer.

Next, a real dataset is investigated: the well-known Swiss banknotes data (Flury and Riedwyl, 1988). The data consists of 200 Swiss bank notes based on 6 measurements. There are 100 genuine bank notes and 100 forged ones. Figures 12 and 13 show the validation results of K -means clustering that finds the two classes almost perfectly except for one misclassified observation only. The two true classes are confirmed by both the averaged Jaccard index γ_K and the ARI R_K . The steepest decrease when coming from $K = 3$ and going to $K = 4$ indicates that at most three clusters have a high stability. The latter comes

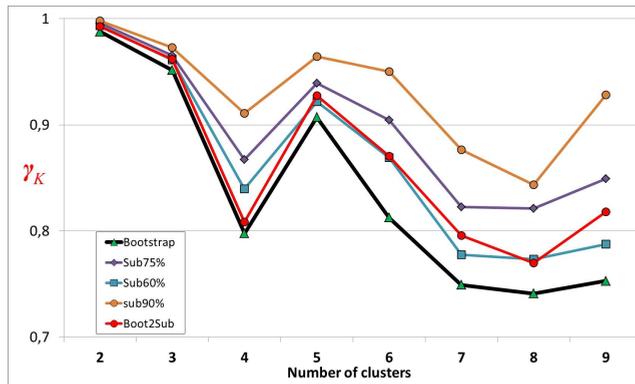


Fig. 12 Averaged Jaccard of the partitional K -means clustering of the Swiss bank notes data.

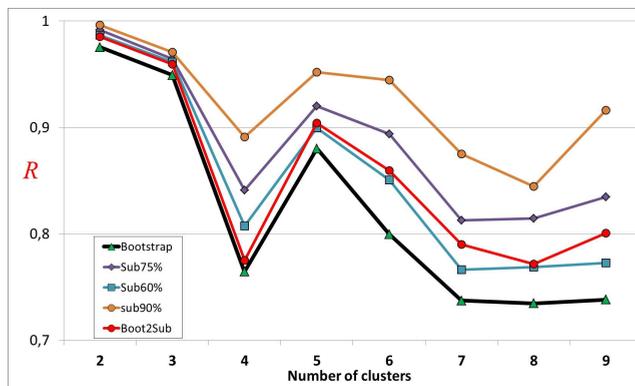


Fig. 13 ARI of the partitional K -means clustering of the Swiss bank notes data.

from the fact that the class of forged bank notes is much more heterogeneous than the class of genuine bank notes (Mucha, 1996). Maybe, the reason for this is that the forged banknotes stem from several different workshops.

The Iris flower data is another well-known real dataset (Fisher, 1936). There are 150 observations that come from three species (classes). One class (species) is easy to find because it looks well separated from the other two in a principle component analysis plot (Mucha, 1992). The other two species are not well separated of each other. 16 errors are counted when using the K -means method with $K = 3$. Fig. 14 and Fig. 15 show the validation results of K -means clustering. The true three classes partition cannot be confirmed by both the averaged Jaccard index and the ARI. The main reason for this failure may be, among

others, that K -means clustering is not the appropriate method with an error rate of more than 10%.

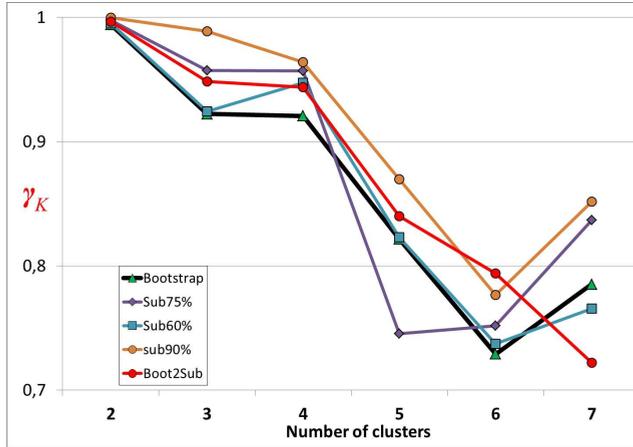


Fig. 14 Averaged Jaccard of the partitional K -means clustering of the Iris data.

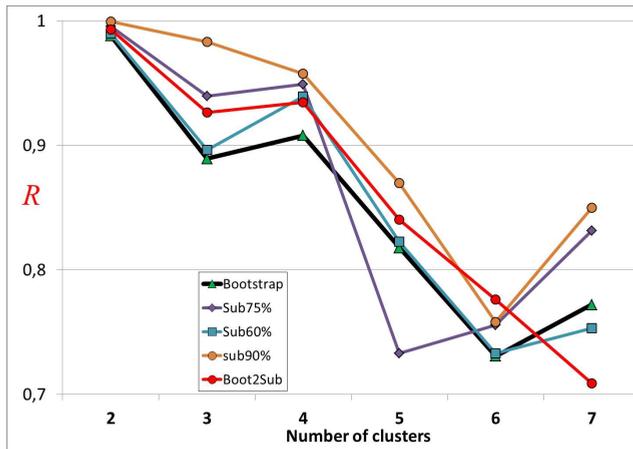


Fig. 15 ARI of the partitional K -means clustering of the Iris data..

5 Summary

In partitional cluster analysis, bootstrapping seems to be also the first choice for both the decision about the number of clusters and the general investigation/assessment of stability. In all cases investigated so far (toy and real data), it outperforms subsampling. It seems to me that multiple observations in bootstrap samples (i.e., observations with mass $m_i > 1$ in (6)) have a great influence for finding the (true) number of clusters. Why? This question has to be answered in the future. The experience of bootstrapping as the winner is similar to the results of hierarchical cluster analysis presented in (Mucha and Bartel, 2014, 2015). But, we investigated here only the very popular K -means method. In subsampling, the choice of the parameter “resampling rate” p causes an additional problem. The simulation results based on a low subsampling rate such as 60% looks similar to bootstrapping. If it necessarily should be subsampling then the recommendation is to take the usual bootstrap scheme but discard multiple observations, i.e., “Boot2Sub“. As a consequence, approximately 63.2% of the observations will be presented in such a subsample. The advantage is that no parameter for setting the sample size is necessary.

References

- Banfield JD, Raftery AE (1993) Model-based gaussian and non-gaussian clustering. *Biometrics* 49:803–821
- Dolnicar S, Leisch F (2010) Evaluation of structure and reproducibility of cluster solutions using the bootstrap. *Marketing Letters* 21(1):83–101, DOI 10.1007/s11002-009-9083-4
- Efron B (1979) Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7(1):1–26, DOI 10.1214/aos/1176344552
- Efron B (1981) Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* 68(3):589–599, URL <http://www.jstor.org/stable/2335441>
- Efron B, Tibshirani R (1993) *An Introduction to the Bootstrap*. Chapman and Hall, Boca Raton, USA
- Efron B, Tibshirani R (1997) Improvements on cross-validation: The 632+ bootstrap method. *Journal of the American Statistical Association* 92(438):548–560, DOI 10.1080/01621459.1997.10474007

- Fang Y, Wang J (2012) Selection of the number of clusters via the bootstrap method. *Computational Statistics and Data Analysis* 56(3):468–477, DOI 10.1016/j.csda.2011.09.003
- Fischer B, Buhmann JM (2003) Bagging for path-based clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(11):1411–1415, DOI 10.1109/TPAMI.2003.1240115
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7(2):179–188, URL <http://hdl.handle.net/2440/15227>
- Flury B, Riedwyl H (1988) *Multivariate Statistics: A Practical Approach*. Chapman and Hall, London
- Hennig C (2007) Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis* 52(1):258–271, DOI 10.1016/j.csda.2006.11.025
- Mammen E (1992) *When Does Bootstrap Work? Asymptotic Results and Simulations*. Springer, New York, DOI 10.1007/978-1-4612-2950-6
- Minaei-Bidgoli B, Parvin H, Alinejad-Rokny H, Alizadeh H, Punch W (2014) Effects of resampling method and adaptation on clustering ensemble efficacy. *Artificial Intelligence Review* 41(1):27–48, DOI 10.1007/s10462-011-9295-x
- Mucha HJ (1992) *Clusteranalyse mit Mikrocomputern*. Akademie Verlag, Berlin
- Mucha HJ (1996) Cluscorr: Cluster analysis and multivariate graphics under MS EXCEL. In: Mucha HJ, Bock HH (eds) *Classification and Clustering: Models, Software and Applications*, 10, WIAS, Berlin, pp 97–106
- Mucha HJ (2007) On validation of hierarchical clustering. In: Decker R, Lenz HJ (eds) *Advances in Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, Berlin, pp 115–122, DOI 10.1007/978-3-540-70981-7_14
- Mucha HJ (2009) Cluscorr98 for Excel 2007: Clustering, multivariate visualization, and validation. In: Mucha HJ, Ritter G (eds) *Classification and Clustering: Models, Software and Applications*, 26, WIAS, Berlin, pp 14–41
- Mucha HJ, Bartel HG (2014) Soft bootstrapping in cluster analysis and its comparison with other resampling methods. In: Spiliopoulou M, Schmidt-Thieme L, Janning R (eds) *Data Analysis, Machine Learning and Knowledge Discovery*, Springer, Berlin, pp 97–104, DOI 10.1007/978-3-319-01595-8_11
- Mucha HJ, Bartel HG (2015) Resampling techniques in cluster analysis: Is subsampling better than bootstrapping? In: S Krolak-Schwerdt BL, Böhmer M (eds) *European Conference on Data Analysis*, Springer, Berlin

- Reddy CK, Vinzamuri B (2014) A survey of partitional and hierarchical clustering algorithms. In: Aggarwal CC, Reddy CK (eds) *Data Clustering. Algorithms and Applications*, Chapman and Hall/CRC, Boca Raton, USA, pp 87–126
- Späth H (1982) *Cluster analysis algorithms for data reduction and classification of objects. Computers and their applications*, Ellis Horwood, Chichester
- Xiong H, Li Z (2013) Clustering validation measures. In: Aggarwal CC, Reddy CK (eds) *Data Clustering. Algorithms and Applications*, Chapman and Hall/CRC, Boca Raton, USA, pp 571–605