Evaluation of technology clubs by clustering: A cautionary note

Andres, Antonio Rodriguez and Otero, Abraham and Amavilah, Voxi Heinrich

VSB-Technical University of Ostrava Faculty of Economics Department of Economics Sokolska trida 33, 702 00 Ostrava Czech Republic, Polytechnic School, University San Pablo CEU. Urbanización de Montepríncipe s/n 28668 Boadilla del Monte Madrid, Spain, Economics/ Division of Social and Behavioral Sciences Estrella Mountain College 3000 N. Dysart Road, Avondale, Arizona, USA

15 May 2021

Evaluation of technology clubs by clustering: A cautionary note

Antonio Rodríguez Andrés (Corresponding Author)

VSB-Technical University of Ostrava

Faculty of Economics

Department of Economics

Sokolska trida 33, 702 00 Ostrava

Czech Republic

E-mail address: antonio.rodriguez.andres@vsb.cz

Voxi Heinrich Amavilah

Economics/ Division of Social and Behavioral Sciences

Estrella Mountain College

3000 N. Dysart Road, Avondale, Arizona, USA

E-mail: vhsamavilah@gmail.com (preferred)

voxi.amavilah@estrellamountain.edu

Abraham Otero

Polytechnic School, University San Pablo CEU.

Urbanización de Montepríncipe s/n

28668 Boadilla del Monte

Madrid, Spain

E-mail: abraham.otero@gmail.com

May 2021

Abstract

Applications of machine learning techniques to economic problems are increasing. These are powerful techniques with great potential to extract insights from economic data. However, care must be taken to apply them correctly, or the wrong conclusions may be drawn. In the technology clubs literature, after applying a clustering algorithm, some authors train a supervised machine learning technique, such as a decision tree or a neural network, to predict the label of the clusters. Then, they use some performance metric (typically, accuracy) of that prediction as a measure of the quality of the clustering configuration they have found. This is an error with potential negative implications for policy, because obtaining a high accuracy in such a prediction does not mean that the clustering configuration found is correct. This paper explains in detail why this *modus operandi* is not sound from theoretical point of view and uses computer simulations to demonstrate it. We caution policy and indicate the direction for future investigations.

Keywords: Machine learning; clustering, technological change; technology clubs; knowledge economy; cross-country

JEL: *C45; C53; O38; P41; O57*

1. Introduction

Applications to nearly all fields of economics of machine learning, especially clustering, and neural networking, and deep learning, are on the rise. A simple word search in this journal (*Applied Economics*) for the phrases, or combinations thereof, machine learning, clustering techniques, neural networking, technology clubs, and technology club convergence reveals no less than 2590 counts. While we cannot be sure what the counts refer to specifically, we know for certain that with advances in big data science the increase in applications has been so much so that Athey and Imbens (2019) call attention to machine learning methods economists should know, especially those at the intersection of theoretic econometric models and applied algorithmic approaches (Athey, 2017; Varian, 2014, Bajari, Dalton, Hang, and Khwaja, 2014).[1] Specifically, Basturk, Paap, and van Dijk (2012) use a clustering technique to group countries according to their level of economic growth, and thereafter apply regression models to explain what factors might drive the composition of each cluster. Usage of clustering techniques is rising with the growth of the literature on technology clubs and technology club convergence, not surprisingly because of the rising role of technology in the competitiveness of nations (Porter, 1990; Fagerberg, Srholec, and Knell, 2007). In this literature researchers use different types of clustering techniques to classify countries according to their technological prowess and to characterize their technological progress. *How well are technology clubs classified and how correctly is the classification evaluated?* This is the research question for this paper, and one that we should care about for policy and future research reasons. Regarding the former, the effectiveness of policy depends on whether a country is classified well, and its classification correctly evaluated. This is important because the fact that technology clubs exist is a clear indication of technological heterogeneity among countries even when they are classified in one cluster (club). Indeed, one can argue that the "multiple regimes" which Durlauf and Johnson (1995) observe in the economic growth of countries are due to the underlying technological differences. Consequently, a one-size-fits-all policy is likely both inappropriate and ineffective. In other words, this study helps decision-makers (policy-makers and researchers alike) correctly identify, design, and target policy to appropriate technology clubs. The increasing popularity of these new techniques is a measure of their importance to both policy and research (Kreiner and Duca, 2020; Cerulli, 2020: Mueller, 2020; Liu, and Xie, 2019; Curie, Kleven and Zwiers, 2020; Rambachan, Kleinberg, Ludwig, and Mullainathan, 2020; Athey, Bryan, and Gans, 2020; Cowgill and Stevenson, 2020).

In the past, countries were grouped in "clubs" according to one or a few simple metrics such as the level of income (low vs. high income countries), geography (South vs. North), political ideology (East v. West), and so on. The observation that technology clubs perform differently within and across each other, gave rise to questions about technology club convergence, leading to the introduction of more quantitative approaches like clustering. Rezankova (2014), for example, evaluates different clustering algorithms, and she points out that mixture clustering

[1] FYI Professor Susan Athey once served as a consultant chief economist for the Microsoft Corporation, Hal Varian is Google Chief Economist, and Pat Bajari is a chief economist and vice-president for Amazon.com.

methods are a more promising tool for uncovering hidden patterns in our data. Wolfson, Madjd-Sadjadi, and James (2004) employ hierarchical clustering and characterize the clusters without validation. While the applications of cluster analysis and machine learning techniques have made possible quantitative classification of technology clubs using more than just a few characteristics, such improvement still leaves unanswered the question about the evaluation of technology club classification as our research question clearly shows. For example, we observed that a number of papers in the literature on technology clubs have accepted uncritically the clustering classification and evaluation of technology clubs and proceeded to analyze the performance of technology clubs and characterize their contribution to economic growth and development as such. Consequently, *the literature inadvertently formalized an error, because it has assumed that if the clusters found really reflect a true structure in the data, a classifier will be able to make good predictions (high accuracy) about the cluster's label (Castellacci, 2008; Castellacci, 2011).* The consequence of the mistaken approach is not a minor methodological issue; it is a matter of relevance to the correct use of the results obtained from clustering. Hence, the objective of this paper is to demonstrate that this assumption is incorrect, and by doing so enhance the utility of the previous literature.

The novelty of the paper is that it illustrates clearly that even when there are no clusters in the data, or when the number of clusters found by the clustering algorithm does not coincide with those really present in the data, there is a very high probability that the classifier will predict with very high accuracy the cluster's label. The *modus operandi* of this mistaken approach is based on the supposition that if the clusters found really reflect a true structure in the data, the classifier will be able to make good predictions (high accuracy) about the cluster's label. As we will show in this paper, the assumption is incorrect. From a theoretical point of view, it does not make sense to use a classifier (such a decision tree) to evaluate the quality of a clustering configuration. Consequently, the main contribution of the paper is a caution that it is an error that some papers on the literature on "technology clubs" did not employ a correct approach to validate the significance/precision of the resulting clusters (country clubs). Using an incorrect technique to assess the results of the clustering will yield an apparently good performance of the model (good metric), and yet a false impression that the "right" technology clubs have been found, leading to false and misleading policy implications from the viewpoint of decision-makers. This paper explains why it is not theoretically adequate to use a classifier to evaluate the quality of a clustering configuration. It uses computational simulations to show how this approach leads to very misleading results, and proposes the use of clustering techniques based on information theory, such as Bayesian information criterion (BIC), to find the optimum number of technology clubs in a given problem.

The rest of this paper is organized as follows. In the next section below, we briefly summarize some concepts, including the key references. Section 3 describes the operation of clustering algorithms and classifiers, from which it can be inferred why it does not make sense to use a classifier to evaluate a clustering configuration. In the Results section, we show computational simulations to demonstrate this effect. Finally, we discuss and conclude in Section 5 with some indications of directions for future investigations.

2. Some Concepts about Clustering and the Error

As used in this paper, the clustering technique is an unsupervised machine learning technique whose purpose is to find groups (clusters) of data similar to each other, and different from the data of other groups (Xu, 2015). These groups are often used as a summary of the original data. It is an exploratory data mining technique. If the algorithm has found a proper set of clusters, instead of reasoning over many, perhaps even hundreds or thousands of, individual data points, the analyst may reason over a small number of clusters. In the specific case of cross-country analysis, this means that instead of having to reason over dozens or hundreds of individual countries, we can reason over a much smaller number (typically, 2-6) of clusters of countries (clubs) that ideally are reasonably homogeneous. A technology club, for example, is a cluster of economic units (countries, industries, etc.) whose data on key variables identifies them as having the same level of technology at any given point in time. Presumably, there are many such clusters in the world (Antonelli, 1999, p. 182; Castellacci, 2008, Hypothesis 1, p. 303).

Durlauf and Johnson (1995) use regression trees and some sort of supervised machine learning technique and then use decision trees for classification (categorical variable) or regression (continuous variable), the outcome variable of which predicted "multiple" economic performance "regime". This approach is different from cluster analysis where there is no variable to be predicted and the cluster analysis determines endogenously the potential groups without making any *a priori* assumption. The problem here is that to validate the clusters found some authors train a classifier to predict the cluster label, and then they use some performance metric of the classifier, such as accuracy, to evaluate the quality of the clustering configuration. We are unable to trace the specific origins of the error and how it seeped into the literature on technology clubs. However, the error is apparent in Castellacci (2006) who shows that differences in technology are the basis for technology clubs and technology clubs have different growth dynamics; they progress (converge or diverge) towards the technology frontier at different rates. The labels for these are "advanced" technology clubs which tend to converge more rapidly than "followers" technology clubs which converge gradually but still faster than the "marginalized" technology clubs which diverge. Castellacci (2008) used a hierarchical clustering to obtain and label the three technology clusters exactly the same: "advanced", "followers", and "marginalized." The author then uses "classification and regression tree techniques (CART) in order to check the robustness of the hierarchical analysis results, and to identify the threshold values of the input variables that determine what club each country belongs to" (Castellacci, 2008, p. 304). Castellacci and Archibugi (2008) used for similar purposes an ordinal classifier (multinomial logistic regression) which for some of the growth clubs predicted correctly 100% of cases. More recently Castellacci (2011) examined the dynamics of technological innovation in a cross-country setting. Technological change is measured by the formation of growth clubs. We stress that our paper does not downplay the value of the previous literature; in fact, it improves (adds values) to that literature. For example, Castellacci's (2008) regression tree algorithm sought to identify technology-related variables that explain the nexus between technology and economic growth. There is no *a priori* reason

to believe that the country classification and dynamics of the exercise-growth relationship the paper observes are unimportant. Even so, how the clusters a validated should matter for both policy and future research, since the formation of clubs depends on technological knowledge which is a function of structural properties and initial conditions of individual countries. It is not enough for the literature to assume the validation is correct.

In a sample of 54 high income countries De la Paz Marín et al. (2015) investigated the dynamics of the knowledge economy over the period 2007-2009. They employed a hierarchical clustering to find "advanced," "medium," "initiated," and "early stage" clusters. Then, they used "three ordinal classifiers" to validate the four clusters" (De la Paz Marín et al., 2015, p.563). A similar approach is followed in De la Paz Marín et al. (2012) to predict R&D expenditures for a sample of 25 European countries. Although, the studies by Paz Marín et al. (2012, 2015), Castellaci (2006, 2011), and Arcihibugi and Castellacci (2008) are well designed and executed, the problem is the results are still questionable because they are based on an inappropriate approach. They use classifiers as a validation method of clusters algorithms. Economic policy will be ill-informed if researchers and policymakers continue to deal with clusters and validate them with classifiers. Even, Stöllinger (2013) whose study draws upon sound economic models of Nelson-Phelps (1966) on absorptive capacity and Benhabib-Spiegel (1994) on human capital in the spirit of Romer (1990), himself assume the classification by Castellaci (2008). His results show that the strongest effects of the technology gap on economic growth occurs in the intermediate group which we associate with the imitation club. Shaaba and Olalekan (2020) employ a clustering algorithm to show the patterns of ICT convergence for 205 countries. The clusters reveal equal convergence globally, but unequal regionally, except for Sub-Saharan Africa (SSA) and South Asia which diverge, with SSA having seven technology clubs, indicating that policy should consider regional idiosyncrasies. The conclusion is well-taken. However, how reliable are these results if the validation on which they are based is flawed?

3. Methods: Classifier Algorithms

The goal of cluster analysis is to find out how subjects or instances are similar on a number of variables (Xu, 2015). After a successful clustering, the analyst may study and interpret aggregate statistics describing the variables of the set of objects that have fallen in each cluster, such as the mean and standard deviation, instead of the values of the individual objects. In most clustering algorithms (especially those known as partitional clustering algorithms, which include K-means, K-medoids, or a mixture of models), the analyst needs to specify in advance how many clusters the algorithm will return (for example, via the K parameter in K-means) (that is not a trivial issue). When a given number of clusters is specified, the algorithm will always return that number of clusters, regardless of how many clusters really exist in the data (see, Figure 1). Even if the number of clusters specified by the analyst as a parameter matches the number of clusters existing in the data, most clustering algorithms do not guarantee that the algorithm will always converge to the proper solution (see Figure 1F). In the case of hierarchical clustering algorithms implemented using agglomerative algorithms, while the

analyst needs not to specify in advance the number of clusters, any number of clusters between 1 and *n*, where *n* being the number of data points, can be obtained by cutting the dendrogram at the appropriate height (see for instance Kaufman and Rousseuw, 1990).

[Figure 1 near here]

Furthermore, when applied to a dataset the vast majority of clustering algorithms (including partitional and hierarchical clustering) always return either the specified number of clusters or a dendrogram that can be cut to any height, regardless of true clusters being present in the data or not (see Figure 2). The algorithms when grouping the data separate the different clusters by borders. In the case of the K-means algorithm, probably the most popular in the economic literature, these borders delimit linear regions called Voronoi regions (see, Figures 1 and 2). The algorithm will find these borders, even if the data is not distributed in groups (see Figure 2) or if the number of groups obtained does not match the true number of groups present in the data (see Figure 1). Thus, after executing a clustering algorithm, no matter how bad the cluster configuration found may be, we will always have borders that perfectly separate the data in clusters. Given this situation, there is an obvious need of having metrics to evaluate how good a set of clusters that have been found are.

[Figure 2 near here]

Classifiers are algorithms that, implicitly or explicitly, search for boundaries that separate the data in such a way that all the data that fall in the same region belong to the same class (Kuncheva, 2014). If the class is made up by the different clusters obtained by a clustering algorithm, by definition, there are borders that perfectly separate the data, irrespective of how good or bad the cluster configuration is. *If a classifier is used to find these borders (predict the labels of the clusters) what is being tested is either how powerful that given classifier is, or how much skill the data analyst has in configuring the classifier's parameters, but not if the borders found by the classifier correspond to a true underlying structure in the data (which is what we would like to evaluate in a clustering algorithm)*. Hence, from a theoretical point of view, it does not make sense to use a classifier to evaluate a clustering configuration.

4. Results

We will illustrate using computational simulations in R, version 3.6.0 (R Core Team, 2020), what happens when a classifier is used to evaluate a clustering configuration (all the R code is available in the supplementary materials). The code is made by two scripts. The first script simulates scenarios where there are no groups in the data. It generates uniformly distributed random datasets of different sizes (n = 100, 200, 500, 1000, 2000, 5000, and 10,000). For each size, it generates 100 times a random dataset with a random number of dimensions (variables), between 5 and 15. Then, K-means is applied to each dataset with a random *K* between 3 and 6. Finally, four different classifiers are trained to predict the cluster labels: a decision tree (Therneau, 2016) and a neural network (Günther, 2010) with default parameters (thus no tuning has been applied); and the same decision tree and neural network where some parameters have

been tuned (although they are the same for all the datasets). Neural networks are more powerful classifiers than decision trees; they permit finding non-linear highly complex borders, while decision trees can only find linear borders that have to be orthogonal to the axes of the feature space (see for instance, Samarasinghe, 2016).

The second script simulates scenarios in which there are clusters in the data by using a random number (between 3 and 6) of multidimensional Gaussians to generate the data. The Gaussians have a random number of dimensions between 5 and 15 and they all generate the same proportion of data. The same dataset sizes are simulated as in the previous case, and for each size 100 different simulations are performed. Then K-means is executed with a random value of *K* between 3 and 6 that must be different from the number of Gaussians that generated that dataset. Hence "the wrong number of clusters" is always obtained. Finally, the same four classifiers from the previous script are then trained to predict the cluster's labels.

Figures 3 and 4 plot the accuracy of each classifier for each size of the data set, and Tables 1 and 2 show the mean accuracy, where accuracy means the percentage of observations correctly classified divided by the total number of instances. When using a powerful classifier with proper parameters (the neural network with optimized parameters) over uniformly distributed random datasets the accuracy is > 99%, except for n =10,000, where the accuracy is 98.8%, For the Gaussian datasets, where the *K* used is always from the number of Gaussians, the accuracy is always > 99.8%. In the case of the decision tree with tuned parameters, the accuracy for both datasets is always > 90%. Note also that choosing the appropriate parameters for the classifier has more impact on its accuracy than the presence or absence of true clusters in the data.

[Figure 3 near here]

[Figure 4 near here]

[Table 1 and Table 2 near here]

5. Discussion and Conclusions

The simulation results show that it is possible to train a classifier to make a very accurate prediction of the label of a cluster (unsupervised learning algorithm), even if there are no groups in the data (Figure 3 and Table 1), or if a number of clusters different from the true number of groups present in the data has been found during the clustering (Figure 4 and Table 2). They also show that, rather than the presence or absence of clusters in the data, the tuning of the classifier has more impact on its performance.

Some classifiers, such as decision trees, have the advantage of being easy to interpret. Therefore, it may make sense to use a decision tree to predict the labels of a cluster, in order to **understand/interpret** a given clustering configuration (Fraiman, 2013). For example, the decision trees of Figure 5 permit understanding of the clusters found with K-means data on the left better than the mean vectors of the centroids of each cluster, or the mathematical

expressions of the Voronoi regions. The fact that those decision trees have a high accuracy means that the borders they found are very close to those created by the clustering algorithm; hence the trees could be an adequate tool to try to understand the clustering configuration found. However, as it can be seen in Figures 5A and 5B, this does not mean that the clustering configuration found reflects a true structure in the data. This seems to be a fact ignored by some papers in literature on technology clubs.

While this paper does not offer one specific solution, it recommends that authors and policymakers looking for a suitable way to evaluate the clustering configurations may check (Halkidi, 2002), especially when the problem in question is determining the proper number of clusters, we recommend abandoning the old and venerable K-means in favor of mixture model algorithms and using the Bayesian Information Criterion (BIC) to select the number of clusters (Ahlquist, 2012). For a recent application in the field of Economics of mixture models see, Sulkowski and White (2016), Scharfenaker and Schneider (2020), and Clement (2020).

[Figure 5 near here]

BIC provides a quantitative measure that favors cluster configurations that represent well the data (high likelihood) but at the same time it favors those clusters configurations that are as simple as possible (those that have a smaller number of clusters, and therefore of parameters). If the structure of a set contains clusters, when executing mixture model algorithm with a number of clusters lower than the true number of clusters present in the data, the optimal likelihood has not been achieved since the data would be better modeled with more components in the mixture. Hence, if we increase the number of components the likelihood, and therefore BIC, tend to increase until the number of clusters in the mixture matches the number of clusters truly present in the data (Figure 6 B). If we continue to increase the number of components in the mixture, the likelihood will not change significantly, but the model becomes more complex (more mixtures means more parameters), which will cause BIC to begin to decrease. Therefore, if the data contains true clusters, when running the mixture model with an increasing number of clusters (from 3 to 10 in Figure 6), a clear maximum should occur BIC when the number of mixes matches the number of actual clusters. On the other hand, if there are no clusters present in the data, it should not appear a clear maximum in BIC when the number of components of the mixture progressively increases (Figure 6 A). We invite economists to explore the use of sound techniques, such as BIC, to measure the quality of a cluster configuration and to determine the number of clusters present in their data, rather than the erroneous approach of using a classifier to predict the labels of clusters.

[Figure 6 near here]

Cluster validation and labeling should proceed cautiously because they require careful considerations. Besides the use of BIC, other machine learning techniques that would be benefit policy and future research on this topic are emerging and economists should take note of them. For example, Onan (2019) shows that a classification may be correct in terms of the number of clusters, but there may be a “minority class” phenomenon by which one cluster has very small instances relative to others because of an imbalance in the dataset, a probable situation when

dealing with technology clubs in developing countries where data is normally both inadequate and imprecise. In such cases other machine learning techniques must be brought to bear. Ensemble approaches enable (text) mining for specific information within the clubs themselves before clustering (Onan, 2016). These methods have proved useful in characterizing "sentiment" in the data, leading to efficient classification and accurate prediction of phenomena (Onan, 2020; Onan and Tocoglu, 2020). Hybrid ensemble permits supervised and unsupervised analysis of structured and unstructured data by combining search algorithms and k-means into a "classifier ensemble" with higher analytical and predictive power than conventional classifier algorithms that are not diverse (Onan, 2017). The objective is not just classification, but to utilize data mining for strategic planning (Onan, Bul, and Yaner, 2016), identify the subjective content of the data (Onan, Korukoglu, and Bulut, 2016), and or promote multiple objectives that are implicit in individual technology clusters as (a) Onan and Tocoglu (2021) illustrate with the example of the use of neural language network and deep neural networks to uncover deep meaning such as as sarcasm, (b) Onan, Korukoglu, and Bulut, 2016) show with respect to "multiobjective weighted rating ensemble classifier," and (c) Onan and Korukoglu, (2015) advocate in selecting the appropriate model based on a specific feature and rank aggregation. All this literature supports the need to use machine learning techniques carefully. The cautionary note which this paper presents is a good example of the issues that can arise when techniques are not used properly.

Disclosure statement

No potential conflict of interest was reported by the authors

Funding

Financial support from the Czech Science Foundation (No. GA19-25280S) and the Ministry of Science, Innovation and Universities of Spain (No. RTI2018-095324-B-I00) is gratefully acknowledged.

Acknowledgements

This paper is dedicated to the memory of both my parents for always loving and supporting me during my academic career: Francisco Rodríguez Gude and Esperanza Andrés González who died from COVID-19 in March 2020 in Madrid. The usual caveat applies.

References

Ahlquist, J.S., and Breunig, C. (2012). Model-based clustering and typologies in the social sciences. *Political Analysis*, 20(1): 92—112.

Antonelli, C. (1999). *Evolution of Technological Cooperation in the Microdynamics of Technological Change*. London and New York: Routledge Frontiers of Political Economy. Chapter 9.

Athey, Susan C., Kevin A. Bryan, and Gans, J.S. (2020). The allocation of decision authority to human and artificial intelligence. *AEA Papers and Proceedings,* 110: 80—84. *DOI: 10.1257/pandp.20201034*

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science* 355(6324): 483—485.

Athey, S., Imbens, G. (2019). Machine learning methods economists should know about. Available at https://arxiv.org/abs/1903.10075

Bajari, P., Dalton, C., Hong, H., and Khwaja, A. (2014). Moral hazard, adverse selection and health expenditures: A Semiparametric Analysis. *RAND Journal of Economics* 45: 747-763.

Basturk, N., Paap, R., and van Dijk, D. (2012) Structural differences in economic growth: an endogenous clustering approach. *Applied Economics*, 44(1): 119-134.

Castellacci, F., Archibugi, D. (2008). The technology clubs: The distribution of knowledge across nations. *Research Policy*, 37: 1659—1673

Castellacci, F. (2008). Technology clubs, technology gaps, and growth trajectories. *Structural Change and Economics Dynamics*, 19: 301—314

Castellacci, F. (2011). Closing the technology gap? *Review of Development Economics*, 15(1): 180—197.

Cerulli, G. (2020). Improving econometric prediction by machine learning. *Applied Economics Letters*. Forthcoming.

Clement, J. (2020). Social protection clusters in sub-Saharan Africa. *International Journal of Social Welfare*, 29: 20—28.

Cowgill, Bo., Stevenson, M.T. (2020). Algorithmic social engineering. *AEA Papers and Proceedings,* 110: 96—100. *DOI: 10.1257/pandp.20201037*

Currie, J., Kleven, H., and Zwiers, E. (2020). Technology and big data are changing economics: Mining text to track methods. *AEA Papers and Proceedings,* 110: 42—48. *DOI: 10.1257/pandp.20201058*

De la Paz-Marín, M., Campoy-Muñoz, P., and Hervás-Martínez, C. (2012). Non-linear multi-classifier model based on artificial intelligence to predict research and development performance in European countries. *Technological Forecasting and Social Change*, 79(9): 1731—1745.

De la Paz-Marín, M., Gutiérrez, P.A., and Martínez, C.H. (2015). Classification of countries' progress toward a knowledge economy based on machine learning classification techniques. *Expert Systems with Applications*, 42: 562—572.

Durlauf, S. N., Johnson, P. A. (1995). Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics*, 10: 365—384.

Fagerberg, J., Srholec, M., and Knell, M/ (2007). The competitiveness of nations: Why some countries prosper while others fall behind. *World Development*, 35 (10): 1595-1620.

Fraiman, R., Ghattas, B., and Svarc, M. (2013). Interpretable clustering using unsupervised binary trees. *Advances in Data Analysis and Classification*, 7(2): 125—145.

Günther, F., and Fritsch, S. (2010). neuralnet: Training of neural networks. *The R Journal*, 2(1): 30—38.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002). Cluster validity methods: part I. *ACM Sigmod Record*, 31(2): 40—45.

Kaufman, L., and Rousseeuw. P. (1990*). Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey, NJ: Wiley.

Kreiner, A., and Duca, J.V. (2020). Can machine learning on economic data better forecast the unemployment rate? *Applied Economics Letters*, 27 (17): 1434—1437

Kuncheva, L.I. (2014). *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons.

Liu, Y., and Xie, T. (2019). Machine learning versus econometrics: Prediction of box office. *Applied Economics Letters*, 26(2): 124—130.

Mueller, S.Q. (2020). Pre-and within season attendance forecasting in Major League Basketball: A random forest approach. *Applied Economics*, 52(41): 4512—4528

Nelson, R., and Phelps, E. (1966). Investment in humans, technological diffusion, and economic growth. *American Economic Review*, 56(2): 67—75.

Onan, A. (2019). Consensus Clustering-Based Undersampling Approach to Imbalanced Learning. Scientific Programming PB - Hindawi.

Onan, A., Korukoğlu, S., Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications*, 57: 232—247.

Onan, A. (2020). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Computer Applications in Engineering Education:* 1— 18. https://doi.org/10.1002/cae.22253.

Onan, A., Toçoğlu, M.A. (2020). Weighted word embeddings and clustering-based identification of question topics in MOOC discussion forum posts. *Computer Applications in Engineering Education*; 1— 15. https://doi.org/10.1002/cae.22252.

Onan, A. (2018). An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science*, 44(1): 28—47. https://doi.org/10.1177/0165551516677911.

Onan, A. (2018). Sentiment analysis on Twitter based on ensemble of psychological and linguistic feature sets. *Balkan Journal of Electrical and Computer Engineering*, 6(2): 69—77.

Onan, A. (2016). The use of data mining for strategic management: a case study on mining association rules in student information system. *Croatian Journal of Education*: *Hrvatski časopis za odgoj i obrazovanje*, 18(1): 41—70.

Onan, A. (2017). Hybrid supervised clustering based ensemble scheme for text classification. *Kybernetes*, Vol. 46(2): 330—348.

Onan, A., Korukoglu, S., and Bulut, H. (2016). LDA based topic modelling in text sentiment classification: An empirical analysis. *International Journal of Computational Linguistics and Applications*, 7(1): 101—119.

Onan, A., and Toçoğlu, M.A. (2021). A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *IEEE Access*, 9: 7701—7722.

Onan, A., Korukoğlu, S., and Bulut, H. (2016). A multi-objective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Systems with Applications*, 62: 1—16.

Onan, A., and Korukoğlu, S. (2017). A feature selection model based on genetic rank aggregation for text sentiment classification. *Journal of Information Science*, 43(1): 25—38.

Porter, M.E. (1990). *The Competitive Advantage of Nations*. New York: Free Press.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambachan, A., Kleinberg, J., Ludwig, J., and Mullainathan, S. (2020). An economic perspective on algorithmic fairness. *AEA Papers and Proceedings,* 110: 91—95. *DOI: 10.1257/pandp.20201036*

Rezankova, H. (2014). Cluster analysis of economic data. *Statistika,* 94(1): 73—85.

Samarasinghe, S. (2016). *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*. CRC Press.

Shaaba Saba, C., Oladipo Olalekan, D. (2020). Convergence patterns in global ICT: Fresh insights from a club clustering algorithm. *Telecommunications Policy*, 44(10), 102010.

.

Scharfenaker, E., and Schneider, M.P.A. (2020). Labor market segmentation and the distribution of income: New evidence from Internal Census Bureau Data. In book: *Great Polarization: Economics, Institutions and Policies in the Age of Inequality*. Cambridge.

Stöllinger, R. (2013). International spillovers in a world of technology clubs. *Structural Change and Economic Dynamics*, 27: 19—35.

Sulkowski, A. White, D.S. (2016). A happiness Kuznets curve? Using model-based cluster analysis to group countries based on happiness, development, income, and carbon emissions. *Environment, Development and Sustainability*, 18(4): 1095—1111.

Therneau, T., Atkinson, B., Ripley, B., and Ripley, M.B. (2015). Package 'rpart'. Available online at: cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf (accessed April 2020).

Wolfson, M., Madjd-Sadjadi, Z., and James, P. (2004). Identifying national types: A cluster analysis of politics, economics, and conflict. *Journal of Peace Research*, 41(5): 607—623.

Xu, D., and Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2): 165—193.

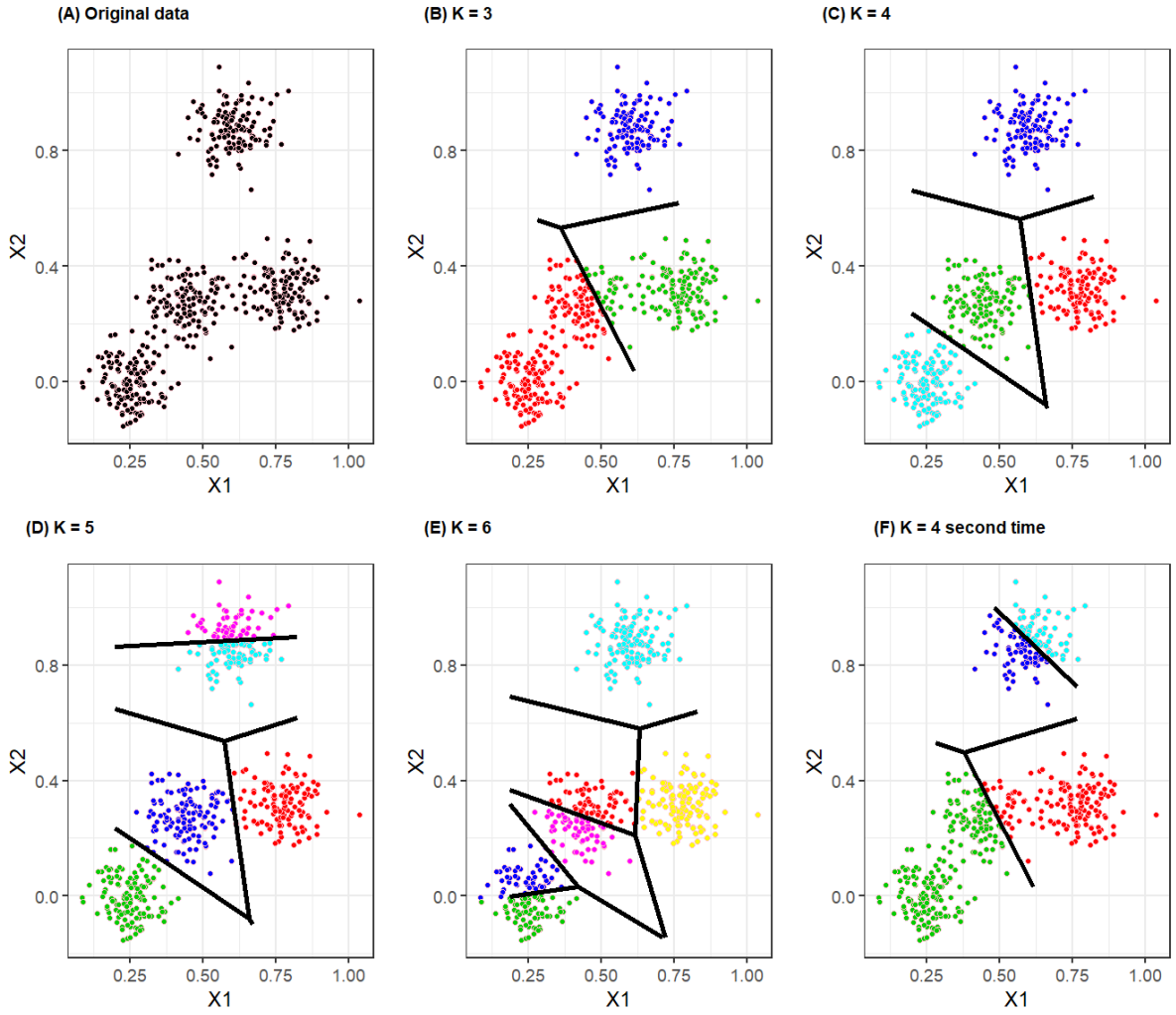

Figure 1. *Dataset made up by 500 random data points generated by four two-dimensional Gaussians. Cluster assignment after executing K-means is shown in color, and Voronoi regions are drawn in black. Source: Own elaboration*

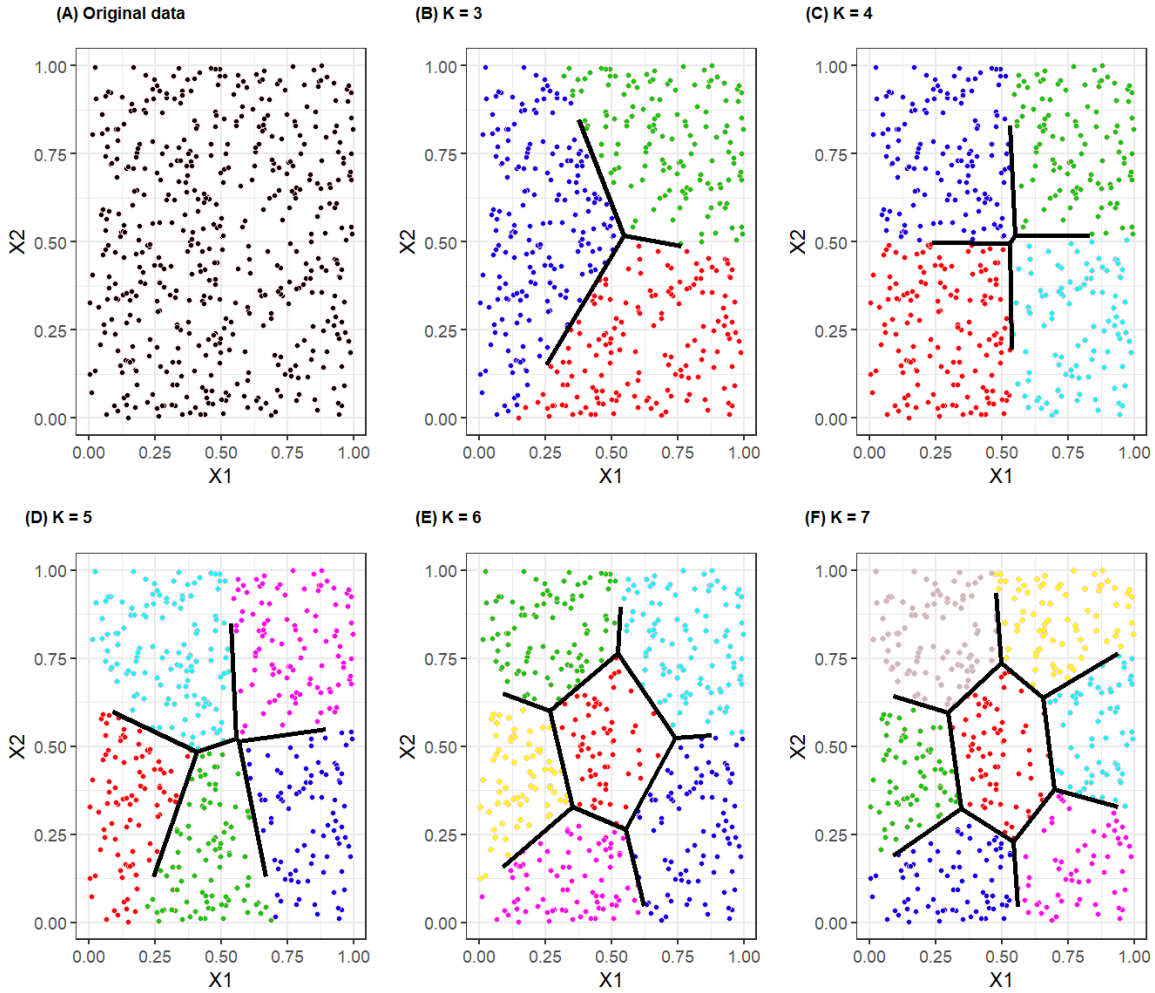

Figure 2. *Two-dimensional dataset made up by 500 data generated by random uniform distribution (hence there are no clusters). Cluster assignment after executing K-means is shown in color, and Voronoi regions are drawn in black. Source: Own elaboration.*

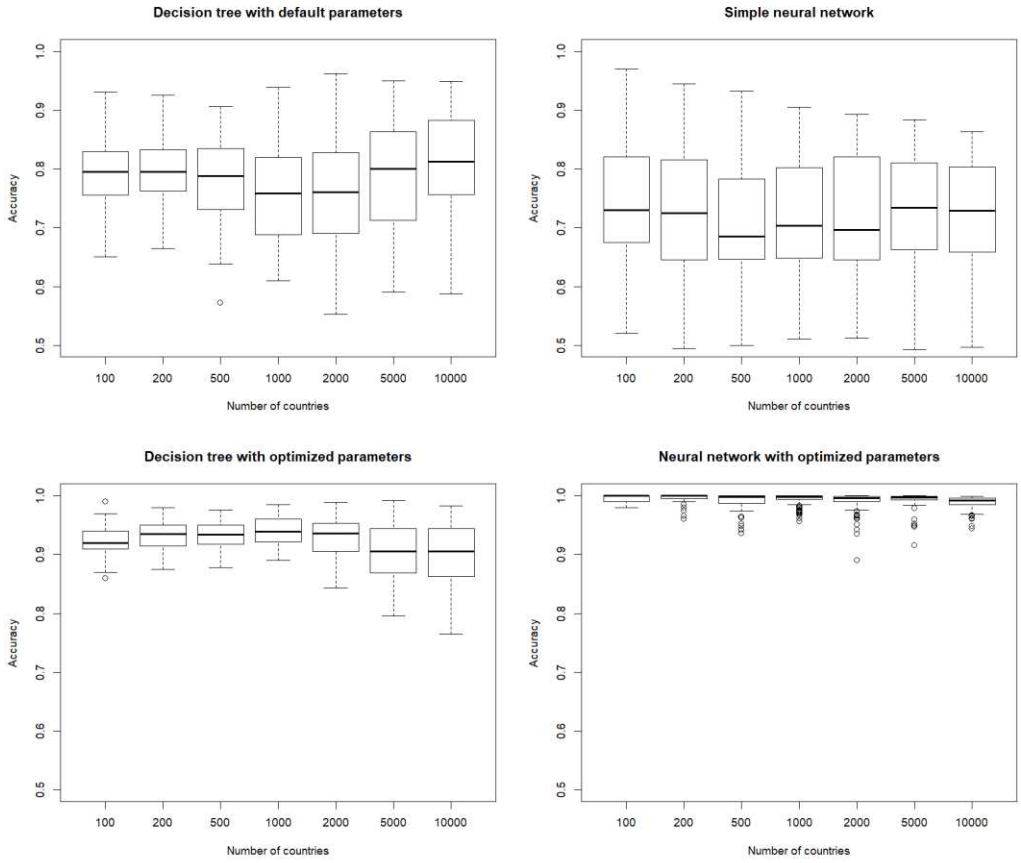

Figure 3. *Boxplots of the accuracy of the uniformly distributed random datasets. Source: Own elaboration*

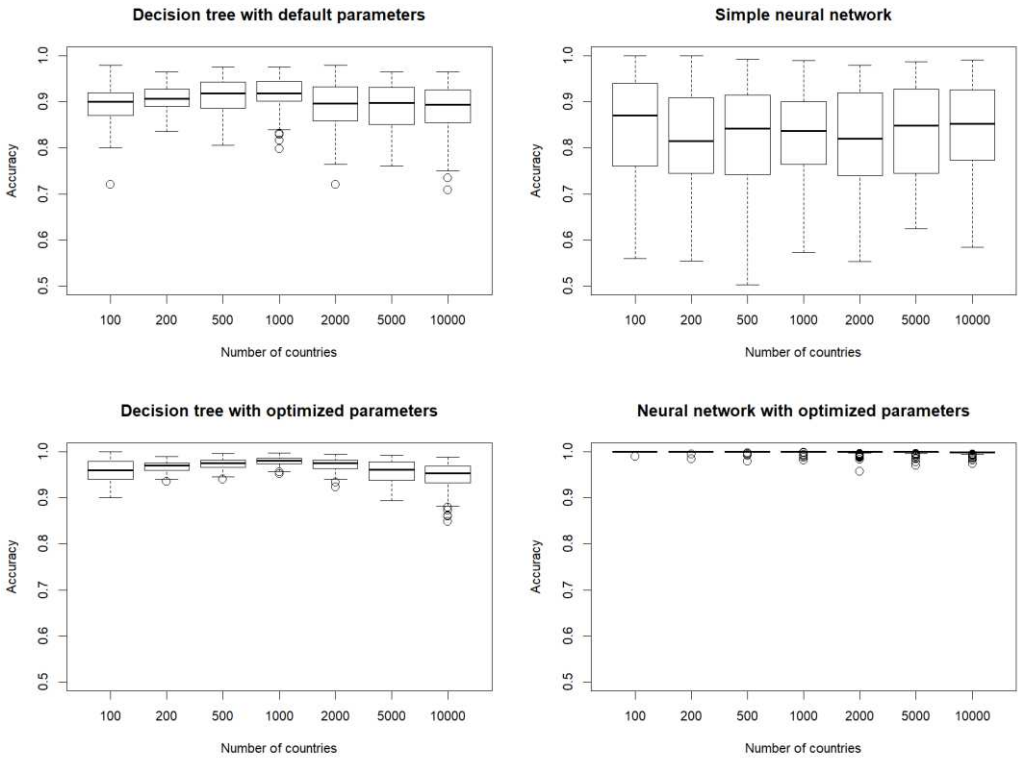

Figure 4. *Boxplots of the accuracy of the Gaussian datasets where K is always different from the number of Gaussians. Source: Own elaboration*.

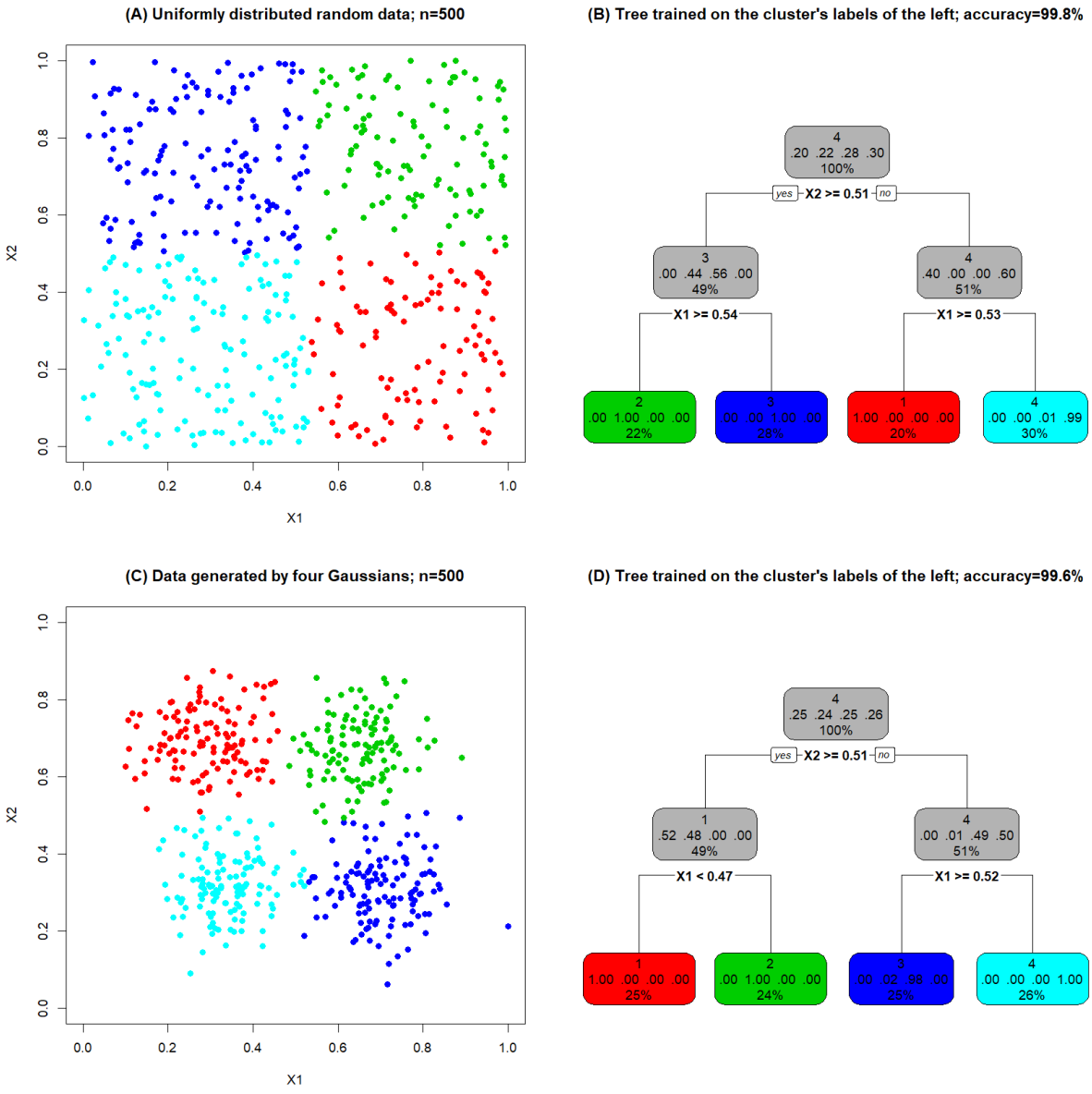

Figure 5. *Decision trees may help in explaining a clustering configuration, but not in assessing if it reflects the true structure in the data. Source: Own elaboration.*

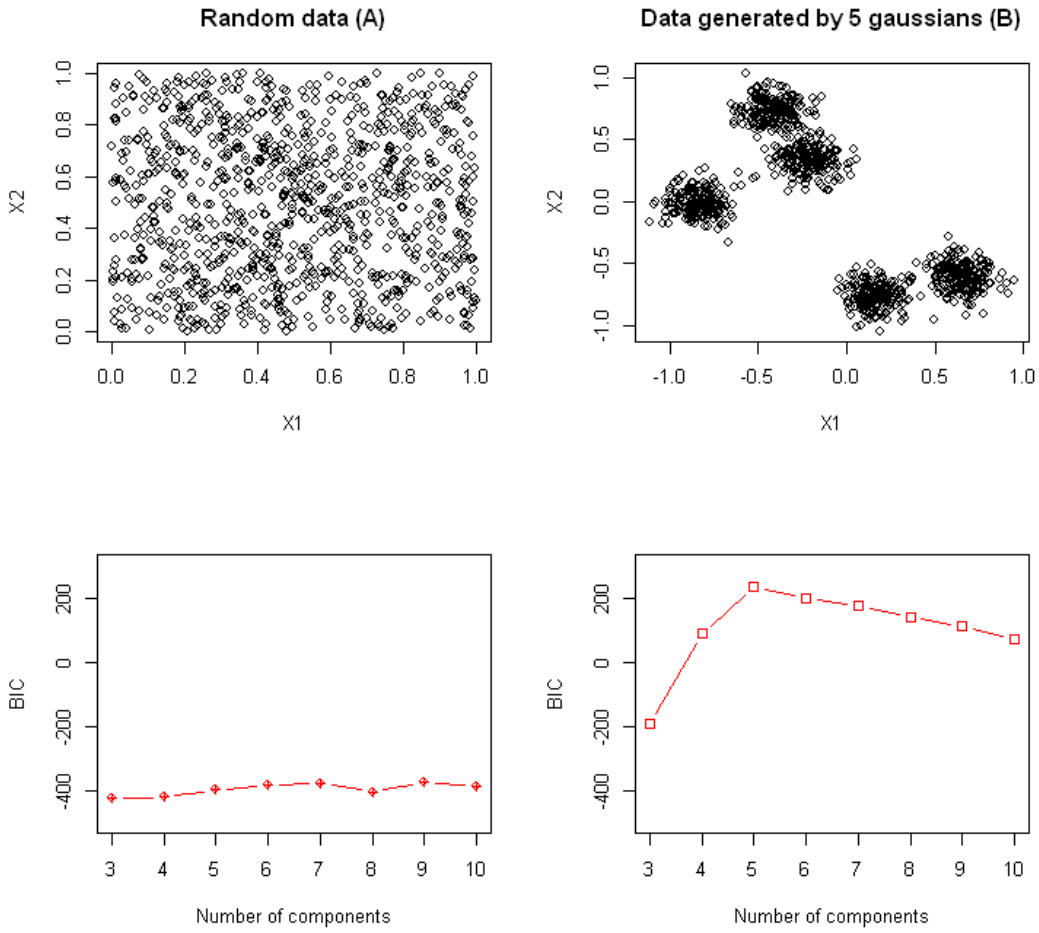

Figure 6. *On the left, a data set made up of 1000 uniform randomly distributed points, and the BIC resulting from executing a Gaussian model mixture algorithm with a number of Gaussians (components) between 3 and 10. On the right, 1000 points generated by 5 Gaussians, and the corresponding BIC as using between 3 and 10 components. Note how when clusters are present in the data, a clear maximum appears in the BIC plot. Source: Own elaboration*

Table 1. *Accuracy (mean ± standard deviation) of each classifier for the uniformly distributed random datasets (DT: Decision Tree, NN: Neural Network). Source: Own computation.*

Dataset size	100	200	500	1000	2000	5000	10000
DT default parameters	79.1 ± 5.6	79.5 ± 5.6	78.0 ± 6.8	75.8 ± 8.4	75.5 ± 9.4	79.2 ± 8.9	81.4 ± 8.4
NN default parameters	74.4 ± 10.6	73.1 ± 11.1	71.3 ± 9.9	72.2 ± 9.3	72.1 ± 9.9	72.6 ± 8.8	72.6 ± 8.0
DT optimized parameters	92.1 ± 2.5	93.2 ± 2.4	93.4 ± 2.2	94.1 ± 2.3	92.8 ± 3.2	90.5 ± 4.6	90.1 ± 5.1
NN optimized parameters	99.7 ± 0.5	99.6 ± 0.8	99.1 ± 1.4	99.4 ± 0.9	99.1 ± 1.6	99.3 ± 1.3	98.8 ± 1.1

Table 2. *Accuracy (mean ± standard deviation) of each classifier for the uniformly distributed random datasets (DT: Decision Tree, NN: Neural Network). Source: Own computation*

Dataset size	100	200	500	1000	2000	5000	10000
DT default parameters	89.6 ± 4.2	90.8 ± 2.8	91.3 ± 3.6	91.5 ± 3.8	89.0 ± 5.2	88.6 ± 5.4	88.6 ± 5.7
NN default parameters	84.8 ± 10.5	82.0 ± 10.1	82.2 ± 11.2	82.6 ± 9.9	82.1 ± 10.5	83.4 ± 9.6	84.0 ± 9.9
DT optimized parameters	95.8 ± 2.2	96.8 ± 1.3	97.4 ± 1.1	98.0 ± 0.9	97.3 ± 1.4	95.6 ± 2.6	94.6 ± 3.1
NN optimized parameters	99.9 ± 0.1	99.9 ± 0.2	99.9 ± 0.2	99.9 ± 0.2	99.8 ± 0.7	99.8 ± 0.5	99.8 ± 0.4