On the groups number estimation for unsupervised clustering

C. Castro, A. E. Castro, D. H. Peluffo, G. Castellanos

Abstract—Category 2. Clustering techniques usually requires manually set parameters so the classification task may be correctly carried out, one of the most common being the number of groups or clusters in which data should be separated, yet this relies in a prior knowledge of the data nature. In this work a comparison among different approaches for finding the number of groups is shown, such as singular value decomposition (SVD), analysis of the multiplicity of the greatest eigenvalues from the affinity matrix, and the percentage of the cumulative sum of the singular values of the affinity matrix. The spectral nature of the estimation process as well as the different datasets used, infers that the results rely only in the internal information of data matrixes. Results exhibits both limitations and advantages for each method, weather directly related with the nature of the data, or limited by the process structure and definition. Nonetheless these guidelines will be helpful for deciding which estimation technique best applies for clustering data regardless its origin.

Index Terms—Affinity matrix, clustering, eigenvalues, number of groups, singular values, spectral analysis.

I. INTRODUCTION

UNSUPERVISED clustering stands as one of the building blocks in data analysis, however automatic estimation of the correct or suitable number of groups remains as an open issue in most of the clustering techniques, since no universal satisfactory solution has been achieved [1]. Yet several methods for automatically finding the number of clusters, have shown good performance [2] for clustering, results are limited to the model-based nature of the process and this is not suitable for most unsupervised learning methods [3].

A most general approach arises from the spectral analysis, in dimensionality reduction processes [4], [5], analysis of the magnitude and multiplicity of the eigenvalues and singular values is used for dimensionality reduction and extraction of relevant features, moreover this analysis suggest that the eigen-analysis of the data can be also used for determine which number of clusters best fits the distribution of the data. In [6] and [7], (dis)similarity information extracted from the computation of affinity matrixes is used to reveal internal information of the data structure as well as the numbers of groups.

In this work, spectral information such as the eigen-values and singular value decomposition is studied as a method for

Cristian Castro-Hoyos: ccastroh@unal.edu.co, Master student, Universidad Nacional de Colombia, Sede Manizales.

Andrés Eduardo Castro-Ospina: aecastroo@unal.edu.co, Master student, Universidad Nacional de Colombia, Sede Manizales.

Diego Hernán Peluffo-Ordoñez: dhpeluffo@hotmail.com, Ph.D. student, Universidad Nacional de Colombia, Sede Manizales.

Germán Castellanos-Dominguez: cgcastellanosd@unal.edu.co, Ph.D., Professor Department of electrical engineering, electronic engineering and computer sciences of Universidad Nacional de Colombia, Sede Manizales.

finding the proper number of groups from a data-set in spite of its nature; this comparison attempt to illustrate independence of the estimation technique from the data structure.

The simplicity in the used algorithms, makes necessary heuristic tuning of the decision parameters; and even though non universal parameter can be found, the behavior consistency of the studied methods showed that a relative small homogeneous sample from an large data-set provides enough information for the so called manual tuning, hence reducing the computation cost of the posterior clustering technique.

II. THEORETICAL FRAMEWORK

A. Singular Value Decomposition

Let $\mathbf{X}$ be an $n \times p$ matrix, where n is the number of observations, and p the number of features to be a full rank matrix, this is

$$\mathrm{rk}\left(\mathbf{X}\right) = \min\left(n, p\right)$$

We define the normalized feature matrix $\hat{\mathbf{X}}$ as follows

$$\hat{\mathbf{X}} = \frac{\sqrt{p}\mathbf{X}}{\|\mathbf{X}\|_{\mathrm{F}}^2} \tag{1}$$

where $\|\cdot\|_{\mathrm{F}}$, denotes the frobenius norm.

The SVD of a normalized matrix $\hat{\mathbf{X}}$ is the decomposition of $\hat{\mathbf{X}}$ into the product of three matrixes as follows[8]

$$\hat{\mathbf{X}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \sum_{k=1}^{r} \boldsymbol{\sigma}_{\boldsymbol{k}}\mathbf{u}_k\mathbf{v}_k^\top \tag{2}$$

where $r = \min\left(n, p\right)$, $\mathbf{U}_{n \times n}$ and $\mathbf{V}_{p \times p}$ are the left and right singular matrixes respectively, and $\boldsymbol{\Sigma}_{n \times p}$ is a matrix with the singular values $\boldsymbol{\sigma}_{\boldsymbol{k}}$ along its diagonal and zeros everywhere else.

So the norm of the normalized feature matrix can be also expressed as

$$\left\|\hat{\mathbf{X}}\right\|_{\mathrm{F}}^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} \hat{x}_{ij}^2 = p = \sum_{l=1}^{\mathrm{rk}(\hat{\mathbf{X}})} \sigma_l^2 \tag{3}$$

and the value of the number of groups k, can be selected so that

$$k = \arg\min \left\{ \alpha \cdot p \leq \sum_{l=1}^{k} \sigma_l^2 \right\} \tag{4}$$

where α represents a free parameter to be tuned according to the complexity of the data. From (3) it is possible to infer that α must be a value between 0 and 1.

B. Analysis of the affinity matrix: Eigenvalues

Let $\mathbf{X}$ be an $n \times p$ matrix, where n is the number of observations, and p the number of features, we define the trivial affinity matrix ($\hat{\mathbf{A}}$) as,

$$\hat{\mathbf{A}} = \mathbf{X}\mathbf{X}^\top \tag{5}$$

and the exponential affinity matrix with local scaling factor as,

$$\hat{\mathbf{A}}_{ij} = \exp\left(\frac{-d^2(\mathbf{x}_i, \mathbf{x}_j)}{\sigma_i \sigma_j}\right) \tag{6}$$

where $\mathbf{x}_i$ is each data vector, $\text{dist}(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between $\mathbf{x}_i$ and $\mathbf{x}_i$. σ_i is the local scale,

$$\sigma_i = \text{dist}(\mathbf{x}_i, \mathbf{x}_N) \tag{7}$$

where $\mathbf{x}_N$ is the $N'th$ neighbor of $\mathbf{x}_i$. Then define $\mathbf{D}$ to be the diagonal matrix

$$\mathbf{D}_{ii} = \sum_{j=1}^{n} \hat{\mathbf{A}}_{ij}$$

and construct the normalized affinity matrix

$$\mathbf{L} = \mathbf{D}^{-1/2}\hat{\mathbf{A}}\mathbf{D}^{-1/2}$$

Finally the eigenvalues λ_i of this matrix are found and sorted in descending order and the multiplicity of those equal to one is chosen as the number of groups. Since the complexity of the data may cause that the most significant eigenvalues slightly differ from 1, a tolerance parameter β is introduced, thus choosing the number of groups as

$$k = \text{numel}\,(1-\beta < \lambda_i < 1+\beta) \tag{8}$$

C. A different approach: $\mathbf{A}_{Q-\alpha}$ affinity matrix

$Q - \alpha$ algorithm [9], is used to select relevant features from an affinity matrix containing information from the inner products of the observations and a weighting vector. $A_{Q-\alpha}$ is the affinity matrix, $\boldsymbol{\alpha}$ is the weighting vector and Q is a rotational orthonormal matrix.

Let $\mathbf{M}$ be a $p \times n$ matrix, defined as $\mathbf{M} = \mathbf{X}^\top = (\mathbf{m}_1, \mathbf{m}_2, ..., \mathbf{m}_p)^\top$, where $\mathbf{X}$ is the feature matrix, n the number of observations and p the number of features, with each row vector ($\mathbf{m}_i$) preprocessed so that it has zero mean, and unit norm. Affinity matrix can be obtained as:

$$\mathbf{A}_\alpha = \mathbf{M}^\top \text{diag}(\boldsymbol{\alpha})\mathbf{M} \tag{9}$$

The solution of the following optimization problem gives information about the relevant features (This method iteratively find and tune $\mathbf{Q}$ and α):

$$\max_{\alpha, Q} \text{tr}(\mathbf{Q}^\top \mathbf{A}_\alpha A_\alpha Q) = \sum_{i=1}^{p} \lambda_i^2 \tag{10}$$

$$\text{s.t.} \quad \mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_q, \quad \boldsymbol{\alpha}^\top \boldsymbol{\alpha} = 1$$

From equation (10) we can write the rotated affinity matrix as $\mathbf{A}_{Q-\alpha} = \mathbf{Q}^\top \mathbf{A}_\alpha \mathbf{A}_\alpha \mathbf{Q}$. As the trace of $\mathbf{A}_{Q-\alpha}$ represents the sum of the squared eigenvalues of $\mathbf{A}_\alpha$ and this matrix holds the information among observations, it can be inferred than its diagonal could give information to determine the number of groups in which data should be separated.
The value of k can be estimated from the diagonal of $\mathbf{A}_{Q-\alpha}$ and an accumulated value over the trace of this matrix, it is introduced a cumulative valor vector as:

$$z = \frac{\text{diag}(\mathbf{A}_{Q-\alpha})}{\text{tr}(\mathbf{A}_{Q-\alpha})}$$

and the value of k is chosen considering that it satisfies the following:

$$\sum_{i=1}^{k} z_i \approx \frac{\beta}{100} \tag{11}$$

where β is an accepted value of the cumulative percentage for the trace of the matrix $\mathbf{A}_{Q-\alpha}$.

III. EXPERIMENTAL SETUP

Public domain databases formatted into numeric arrays were used for the analysis and study of the estimations methods previously proposed. Table I[1] summarizes the real data used for the analysis, and exhibits the high variability between data nature.

TABLE I: Database Summary

Dataset	Source	n	p	k	Description
Auto_mpg	[10]	398	6	2	Multivariate real attributes on automobile description concerning to city-cycle fuel consumption.
Biomed	[11]	194	5	2	Multivariate real attributes on blood measurements for "normal" and "carrier" samples.
Breast	[10]	683	9	2	Multivariate integer attributes on breast cancer samples.
Diabetes	[10]	768	8	2	Multivariate integer attributes on glucose and insulin doses for diabetes patients.
Glass	[10]	214	9	4	Multivariate real attributes on 6 types of glass; defined in terms of their oxide content.
Heart	[10]	297	13	2	Multivariate real and integer attributes on heart symptoms and ECG wave variations.
Imox	[11]	192	8	4	Multivariate real attributes for digital characters I,M,O,X.
Iris	[10]	150	4	3	Multivariate real attributes from size measurements on iris plants.
Liver	[10]	345	6	2	Multivariate real and integer attributes on blood tests sensitive to liver disorders.
Mfeat-kar	[11]	2000	64	10	Karhunen-Love coefficients of handwritten numerals (0-9).
Mfeat-zer	[11]	2000	47	10	Zernike coefficients of handwritten numerals (0-9).
Satellite	[10]	6435	36	6	Multivariate integer attributes on multi-spectral values of pixels in 3x3 neighborhoods in a satellite image.
Soybean1	[10]	266	35	15	Multivariate categorical attributes on Michalski's famous soybean disease.
Texturel	[11]	81920	7	5	N/A
Wine	[10]	178	13	3	Multivariate real and integer attributes derived from the chemical analysis of wine samples.

[1]Where n, p, k refers to the number of instances, attributes and classes respectively

In addition, to avoid errors due to data redundancy, duplicated samples of each dataset where removed during each process as well as feature columns with most zero elements, in order to avoid errors in distance measures, and eigenvalue accuracy.

Taking into account that some datasets contains high number of observations, random homogeneous subsets, each with 200 instances are chosen, this skewness allows to observe the direct relationship among the spectral information given by the internal structure and the number of classes.

A. Methods and Algorithms

Algorithm 1 Singular Value Decomposition, α tuning

1) Initialization: define the $\boldsymbol{\alpha}$ vector parameter for tuning with a small enough step size, from a suggested 0.75 start point to 1-step size;
2) Normalize the data matrix with respect to its frobenius norm and rank as shown in (1).
3) Compute the singular value decomposition for the normalized matrix.
4) Calculate the sum of each singular value, until que condition from (4) is accomplished for each of the elements in the $\boldsymbol{\alpha}$ vector.
5) Establish which $\boldsymbol{\alpha}$ subset vector results in the nearest or equal value of the desired number of groups and compute its mean and standard deviation values.

Algorithm 2 Higher eigenvalues of the affinity matrix

1) Compute each σ_i for each $\mathbf{x}_i \in \mathbf{X}$ using the equation 7.
2) Compute the affinity matrix
$\hat{A}_{ij} = \exp\left(\frac{-d^2(\mathbf{x}_i,\mathbf{x}_j)}{\sigma_i \sigma_j}\right)$
with $\hat{A}_{ii} = 0$.
3) Define $\mathbf{D}$ to be a diagonal matrix with $D_{ii} = \sum\limits_{j=1}^{n} \hat{A}_{ij}$
4) Construct the normalized affinity matrix $\mathbf{L} = \mathbf{D}^{-1/2}\hat{A}\mathbf{D}^{-1/2}$
5) Find the eigenvalues of $\mathbf{L}$ and sort them in descending order.
6) Count the number of eigenvalues in the range $[1-\beta, 1+\beta]$, where β is a tolerance parameter close to 0.

$$\hat{A}_{ij} = \exp\left(\frac{-d^2(\mathbf{x}_i,\mathbf{x}_j)}{\sigma_i \sigma_j}\right)$$

Algorithm 3 $Q - \alpha$

1) Initialization $\mathbf{M} = \mathbf{X}^\top$, random $\mathbf{Q}^{(0)}$, $\mathbf{m}_i \leftarrow (\mathbf{m}_i - \mu(\mathbf{m}_i))/\|\mathbf{m}_i\|$
2) Make $\mathbf{G}$:
$$\mathbf{G} = \left(\mathbf{M}\mathbf{M}^\top\right)\mathbf{M}\mathbf{Q}\mathbf{Q}^\top\mathbf{M}^\top$$
3) Compute $\boldsymbol{\alpha}$ as the eigenvector associated with the greatest eigenvalue of $\mathbf{G}$.
4) Compute $\mathbf{A}_\alpha = \mathbf{M}^\top \text{diag}(\boldsymbol{\alpha})\mathbf{M}$
5) Compute the orthonormal transformation:
$$\mathbf{Z}^{(r)} = \mathbf{A}_\alpha^{(r)}\mathbf{Q}^{(r-1)}$$
6) Compute QR decomposition: $\left[Q^{(r)}, R\right] = \text{qr}\left(Z^{(r)}\right)$
7) Increase r: $r \leftarrow r + 1$ and return to step 2.
8) Compute $\mathbf{A}_{Q-\alpha}$: $\mathbf{A}_{Q-\alpha} = \mathbf{Q}^\top \mathbf{A}_\alpha \mathbf{A}_\alpha \mathbf{Q}$
9) Compute z: $z = \frac{\text{diag}(A_{Q-\alpha})}{\text{tr}(A_{Q-\alpha})}$
10) Find the value of k that satisfies: $\sum\limits_{i=1}^{k} z_i \approx \frac{\beta}{100}$, were β is the accepted percentage manually tuned.

IV. RESULTS AND DISCUSSION

Parameter tuning was performed for each method and datasets from section III. The resulting values were averaged when necessary, and summarized in Table II, exhibiting the mean value and the respective standard deviation for each found parameter, for studying the stability of the methods.

TABLE II: Results Summary

Database	Method	k	Parameter		
			Name	Mean	STD
Auto_mpg	Eigenvalues	3	N	35.5	3.02765
	SVD	2	α	0.999804	$1.00995E{-}4$
	$Q-\alpha$	2	β	0.995	N/A
Biomed	Eigenvalues	2	N	30.5	5.91607
	SVD	2	α	0.910268	$4.6677E{-}2$
	$Q-\alpha$	2	β	0.98	N/A
Breast	Eigenvalues	2	N	21.5	11.11305
	SVD	2	α	0.876986	$9.060357E{-}3$
	$Q-\alpha$	2	β	0.97	N/A
Diabetes	Eigenvalues	2	N	30	6.20483
	SVD	2	α	0.873526	$4.4149E{-}2$
	$Q-\alpha$	2	β	0.85	N/A
Glass	Eigenvalues	4	N	7.5	0.70710
	SVD	4	α	0.999848	$3.86E{-}05$
	$Q-\alpha$	4	β	0.99	N/A
Heart	Eigenvalues	2	N	16	5.62731
	SVD	2	α	0.989342	$3.365373E{-}3$
	$Q-\alpha$	2	β	0.85	N/A
Imox	Eigenvalues	4	N	3	0
	SVD	4	α	0.964052	$3.93695E{-}3$
	$Q-\alpha$	4	β	0.98	N/A
Iris	Eigenvalues	3	N	15	3.89444
	SVD	3	α	0.999	$3.66606E{-}4$
	$Q-\alpha$	3	β	0.99996	N/A
Liver	Eigenvalues	2	N	15	5.04975
	SVD	2	α	0.935752	$1.999652E{-}2$
	$Q-\alpha$	2	β	0.97	N/A
Mfeat-kar	Eigenvalues	1	N	21	11.40175
	SVD	10	α	0.731372	$7.978403E{-}3$
	$Q-\alpha$	10	β	0.89	N/A
Mfeat-zer	Eigenvalues	2	N	3.5	1.29099
	SVD	10	α	0.988146	$8.41194E{-}4$
	$Q-\alpha$	10	β	0.9968	N/A
Satellite	Eigenvalues	6	N	2	0
	SVD	6	α	0.997882	$2.06094E{-}4$
	$Q-\alpha$	6	β	0.9997	N/A
Soybean1	Eigenvalues	4	N	2	0
	SVD	15	α	0.979428	$9.62439E{-}4$
	$Q-\alpha$	15	β	0.9988	N/A
Texturel	Eigenvalues	4	N	4.5	0.70710
	SVD	5	α	0.989742	$2.0395425E{-}3$
	$Q-\alpha$	5	β	0.99998	N/A
Wine	Eigenvalues	3	N	33	4.47213
	SVD	3	α	0.99997	$1.41E{-}05$
	$Q-\alpha$	3	β	0.97	N/A

Since Algorithms (1) and (2), go through an iterative process to find the best suitable parameter, standard deviations are computed, however even when the Algorithm (3), can be computed several times for finding the most general parameter, the random nature of the orthonormal matrix $\mathbf{Q}$ makes this process unreliable, so standard deviation its not accepted for this method.

Fig. 1 and Fig. 2, visually exhibits the behavior of the mentioned parameters α and N, for some of the data-sets, both cases show better estimation at low number of classes, and decreases performance when the number of groups is large.

A. SVD

Heuristic tuning of the α parameter, was made starting from 0.75, with a step size $2E{-}5$ to 0.99998. As expected, the majority of the spectral information is held in the greatest

singular values, of the orthogonal decomposition of a feature matrix; most of the databases allows to tune the decision parameter close to the unity.

However this method exhibits a very important limitation, since the normalization its made respect to the rank of the matrix, the maximum number of groups that it can estimate would be equal to this value.

(a)

(b)
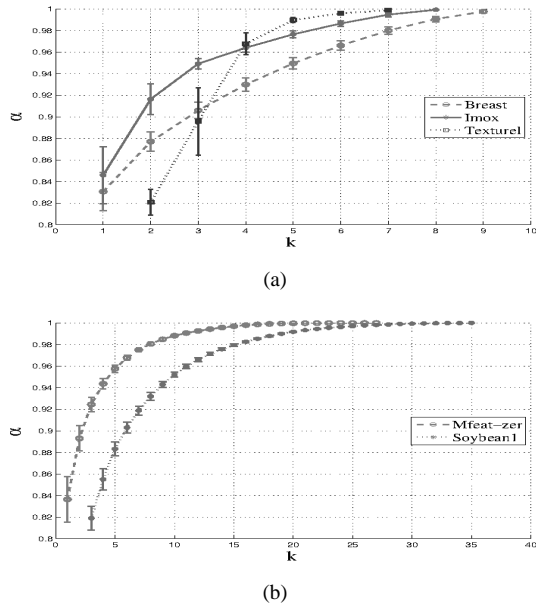
Fig. 1: Estimation via SVD: 1(a) Breast data-set with 2 classes, Imox data-set with 4 classes and Texturel data-set with 5 classes. 1(b) Mfeat-zer data-set with 10 classes and Soybean1 data-set with 15 classes.

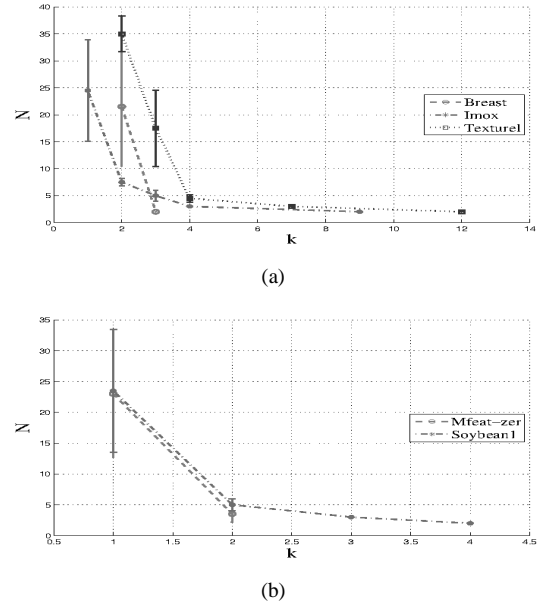### B. Eigen-Values from Affinity Matrixes

Internal data structure is expected to rely in the eigenvalues and eigenvectors from feature matrixes; using the affinity matrix permits to expose (dis)similarity characteristics and extract the spectral information required for the analysis. Trivial matrix from equation (5) does not exposed real internal information, but allows to analyze data from a symmetric squared matrix, however since no define internal structure is achieved, the analysis of the eigenvalues does not provide any accuracy in the estimation of the number of groups.

Introducing local scaling into the exponential affinity matrix, has probe to be a more accurate distribution of the internal structure[7]. Thus the analysis of this structure easily can be used to estimate the number of clusters, however its necessary to find the proper number of neighbors used for scaled the elements of the affinity matrix.

Being then an iterative time consuming process, it is limited to the number of samples taken from the original data for tuning, i.e. when the number of clusters is high, greater than 10, is necessary find the best number of neighbors from a large data sample since the complete structure can not be find from a small one, thus the method becomes inefficient.

Due to the multivariate and multicomponent nature of the entire database (all data-sets used) the β from Eq. (8) was set

for providing a wide range, thus $\beta = 1.5E^-2$.

(a)

(b)

Fig. 2: Estimation via eigenvalues: 2(a) Breast data-set with 2 classes, Imox data-set with 4 classes and Texturel data-set with 5 classes. 2(b) Mfeat-zer data-set with 10 classes and Soybean1 data-set with 15 classes.

C. $\mathbf{A}_{Q-\alpha}$ affinity matrix

(a)

(b)

(c)
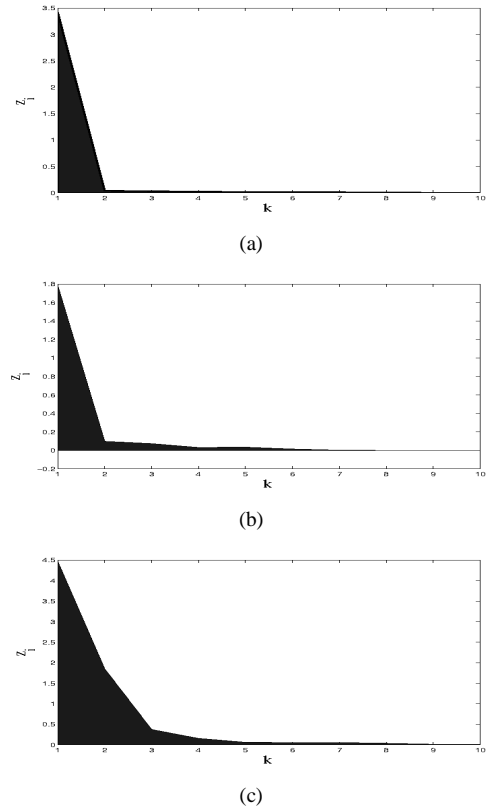
Fig. 3: Estimation via $Q-\alpha$: 3(a) Breast data-set with 2 classes. 3(b) Imox data-set with 4 classes. 3(c) Texturel data-set with 5 classes.

Analysis probed that the behavior of the proposed parameter was similar to that proposed for the SVD, however even though the behavior is similar, the orthonormal projection where it comes from, does not have the same form as the singular values, and provides a more narrow range for the parameter, yet a more much small size-step is necessary for tuning. From Table II it is seen that the smallest β its close to 0.85 while for SVD its corresponding α is 0.73. However visual analysis of this method can only be appreciated from the area of the diagonal vector used in (11). This is shown in Fig. 3, where cutting the horizontal axis at the point on which the percentage of the cumulated area is equal to $100 \cdot \beta$ must match the expected number of groups.

V. Conclusions and Future Work

Spectral information has been shown useful for finding the proper number of groups from small homogeneous subsets of data. Hence reducing computation time for iterative clustering processes, however limitations exist on each method, rank dependency and high number of classes stands as some of the major problems. Even if the methods are not capable of determine the exact number of groups, they are useful for giving a general idea of the complexity of the dataset, and moreover on the relevance of the feature vectors, since they are based on dimensionality reduction algorithms.

Standard deviation (STD) analysis of each data-set allows to determine when a method is more appropriate, when STD is to small(below 2 for the eigenvalues approach and $5E^-4$ in the SVD method), heuristic tuning and stabilization of the method becomes more difficult, and a correct number of groups may not be found.

Some of the proposed parameters seem to have constant behavior, then its possible to infer that data itself could self-tune the value of the parameter from features such as Euclidean distances, statistics measures as correlation and standard deviation, among others. As future work then is propose to find this self-tune characteristics and methods from internal spectral information, to finally automatize the estimation.

References

[1] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a dataset via the gap statistic," *The Royal Statistic Society*, vol. 63, pp. 411–423, 2000.

[2] L. Wang, C. Leckie, K. Ramamohanarao, and J. Bezdek, "Automatically determining the number of clusters in unlabeled data sets," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 3, pp. 335 –350, March 2009.

[3] N. X. Vinh and J. Epps, "A novel approach for automatic number of clusters detection in microarray data based on consensus clustering," in *Bioinformatics and BioEngineering, 2009. BIBE '09. Ninth IEEE International Conference on*, June 2009, pp. 84 –91.

[4] M. Polito and P. Perona, "Grouping and dimensionality reduction by locally linear embedding," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 1255–1262.

[5] P. Liu, S.-D. Tan, B. McGaughy, L. Wu, and L. He, "Termmerg: An efficient terminal-reduction method for interconnect circuits," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 26, no. 8, pp. 1382 –1392, aug. 2007.

[6] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 849–856.

[7] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," *Advances in Neural Information Processing Systems 17*, vol. 2, no. 1601-1608, pp. 1601–1608, 2004.

[8] S. Lee and M. Hayes, "Properties of the singular value decomposition for efficient data clustering," *Signal Processing Letters, IEEE*, vol. 11, no. 11, pp. 862 – 866, Nov 2004.

[9] L. Wolf and A. Shashua, "Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weighted-based approach," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, Oct 2003, pp. 378 –384 vol.1.

[10] A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: http://archive.ics.uci.edu/ml

[11] R. Duin, P. Juszczak, P. Paclík, E. Pekalska, D. de Ridder, D. Tax, and S. Verzakov, "PR-Tools4.1, a matlab toolbox for pattern recognition," 2007. [Online]. Available: http://prtools.org
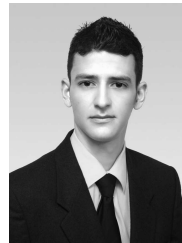
Cristian Castro-Hoyos was born in Bogota D.C., Cundinamarca, in 1989. He received his degree in electronic engineering from the Universidad Nacional de Colombia, Manizales, Colombia, in 2011. Currently, he is M.Eng. student in the same university. His main research interests are computer applied maths, pattern recognition and signal processing and their applications in biosignal analysis and telecommunication systems.

Andrés Eduardo Castro-Ospina was born in Manizales, Caldas, in 1989. He finished his bachelor degree in Electronic Engineering, from Universidad Nacional de Colombia Sede Manizales, in 2011. Currently, he is an active student of M. Eng. - Industrial automation at Universidad Nacional de Colombia Sede Manizales.

Diego Hernán Peluffo-Ordoñez was born in San Juan de Pasto, Nariño, in 1986. He received his degree in electronic engineering and the M.Eng. degree in industrial automation from the Universidad Nacional de Colombia, Manizales, Colombia, in 2008 and 2010, respectively. Currently, he is PhD student in the same university. His main research interests are applied maths and unsupervised learning and their applications in biosignals analysis.

César Germán Castellanos-Domínguez received his undergraduate degree in radiotechnical systems and his Ph.D. in processing devices and systems from the Moscow Technical University of Communications and Informatics, in 1985 and 1990 respectively. Currently, he is a Full Professor in the Department of Electrical, Electronic and Computer Engineering at Universidad Nacional de Colombia Sede Manizales. He is also the Leader of the Signal Processing and Recognition Group at the same university. His teaching and research interests include information and signal theory, digital signal processing and bioengineering.