A quantum inspired methodology for enhancement of data discrimination power.

Rosilda B. Souza, Emeson J. S. Pereira and Tiago A. E. Ferreira
Department of Statistics and Informatics
Federal Rural University of Pernambuco, Recife, Brazil,
Email: rosildabenicio@yahoo.com.br, emesonsantana@gmail.com, tiago@deinfo.ufrpe.br

***Abstract*—The proposed methodology is a process for enhancement the data power discrimination based on the Cover theorem with a quantum inspiration. Given a set of data with several class, the proposed process consists in increasing its dimension in order to try become the non linearly separable problem into a linearly separable problem. Also is supposed that the data are observables in the quantum world, *i.e.*, the data (real number) are expected value measurements of the given transformation with respect to a quantum state (complex numbers). Therefore, the methodology applies a Genetic Algorithm for search the inverse mapping of the expected value measurements, transforming the real number into complex number, subject to the constraint of magnitude conservation. The traditional methods of classification like K-means, KNN and LDA were applied to benchmark classification problems in two conditions: raw data set and transformed data set with the proposed methodology. The comparison of the classification results are presented, indicating a enhancement in the data power discrimination when the proposed pre-processing is applied.**

I. Introduction

The clustering methodologies are a very important and usual branch in several areas like data mining, statistics, engineering, computer science and all science that works with data analysis [1]. In general, the clustering techniques have not a tutor or teacher, characterizing an unsupervised learning process. In these clustering systems, the data are grouped in clusters according with some similarity measure.

In this sense, a clustering system can defines classes (or clusters) for a data set using a feature vector provided by the feature extractor to assign an object of the data set to one of these classes. It is the task of a classifier.

The degree of difficulty to classify a data into a several numbers of classes depends on the variability in the features values for data objects in the same category relative to the difference between the feature values for data objects in different categories. Basically, two factors can have an affect on the variability of feature values for a data object: the data complexity and the noise [1]. If the data set presents a low noise level and complexity with linearly independent relations for the features, then the classification task is a linearly separable problem and there is a optimal hyperplane that produce a perfect classification [2].

On the other hand, if the classification problem is a non linearly separable problem, there is not a optimal hyperplane that conduce to a perfect classification. In this case, because the perfect classification is often impossible, the general task is to determine the probability of one data object to belong for each of the possible classes. Therefore, the linearly separable problem is a task relatively more simple to solve than a non linearly separable problem.

However, if the data set is non linearly separable and it is a d-dimensional set (each data object is a d-dimensional vector of patterns), the Cover Theorem [3], [4] postulates that expanding the dimensionality of the representation space (space where the data set is represented) the asymptotic probability of ambiguous classification will decrease. Thus, if this dimensionality increasing is large enough, the non linearly separable problem, with high probability, may becomes a linearly separable problem.

In a quantum system [5], a state vector is a complex vector, commonly described by $|\Psi\rangle$. A feature of the quantum system can be measured applying a specific operator to the state vector. If the value of feature is a real number, then this feature is called of *observable* and this feature can be observed in the classic world (real world).

In this work is proposed a new data pre-processing based on the Cover Theorem and inspired on a quantum system. Given a data set, it is supposed that the points of the data set are measures of some feature of a quantum system. Thus, a Genetic Algorithm is used to search a transformation in the data set, where a quantum system with a state vector subject to the constraints of minimum variability of each new class in the quantum world and of the maximum distance between the centroids of each class is built. With this transformation, the representation space is dimensionally increased providing better conditions for the classification task.

Classical algorithms of classification as K-means [6], K-Nearest Neighbor (KNN) [4] and Linear Discriminant Analysis (LDA) [7] were used in the classification task. Several experiments were performed with a benchmark data set extracted of the UCI Machine Repository [8] and an artificial data set, where the classifiers were applied to the data with and without the proposed transformation. The experimental results show a excellent classification performance when the proposed transformation is applied to the data set.

This paper is organized as follows. In Section II is described the methodology proposed here to enhancement the data discrimination power in the classification task. The description of the data sets utilized in the experimental process

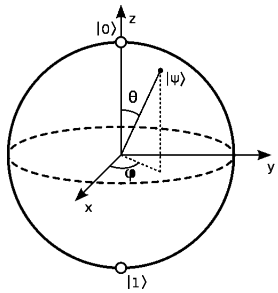


Fig. 1. The Bloch Sphere. $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

is presented in the Section II-C. The experimental set up, results, and conclusions are shown in the Section III and IV, respectively.

II. Proposed Methodology

A. A Quantum System

To understand the quantum inspiration used in this work some principles about a quantum system will be introduced.

A quantum system is governed by the quantum mechanics [5]. The quantum mechanics is a branch of physics which deals with phenomena in the microscopic scale, where the variables magnitude are on the order of the Planck Constant ($h = 6.62606910^{-34} J \cdot s$). In this sense, the quantum mechanics is more fundamental theory than a classical physics, because a macroscopic system (classical system) can be understood as the composition of microscopic systems.

Mathematically, the possible states of a quantum mechanical system can be represented by unit vectors, called "state vectors" and represented in the Dirac notation by $|\Psi\rangle$. Formally, these state vectors reside in a complex Hilbert space [9], commonly called of "state space" or the "associated Hilbert space" of the quantum system. This state space is well defined up to a complex number of norm 1. A state vector can be graphically represented by the Bloch sphere [5], [10], as shown in the Figure 1. For the quantum computer [10] the canonical space base is $|0\rangle$ and $|1\rangle$, where

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1}$$

with,

$$|\alpha|^2 + |\beta|^2 = 1 \tag{2}$$

Following the Dirac notation [5], it is possible to define an operator A where the equation is valid,

$$A|\Psi\rangle = \lambda|\Psi\rangle. \tag{3}$$

The $|\Psi\rangle$ is a eigenvector and λ is a eigenvalue. In general, $\lambda \in \mathbb{C}$, but if $\lambda \in \mathbb{R}$ implies that the operator A is Hermitian and the value of λ can be measured in the real world. In the last case, the operator A is called *observable* and the its value can be obtained by a measure of the expected value given by,

$$\langle\Psi|A|\Psi\rangle = \langle\Psi|\lambda|\Psi\rangle = \lambda\langle\Psi|\Psi\rangle = \lambda$$

where all information about the phases (Figure 1: θ and φ) is lost.

Therefore, the heuristic proposed here assumes that the data set is composed by values of an observable in the quantum world, off by a multiplicative constant. Thus, the procedure will search by a state vector $|\Psi\rangle$ where an observable projective operator applied to this state vector generates the data analyzed.

B. Methodology

Let $X = \{x_1^{(d)}, x_2^{(d)}, \ldots, x_N^{(d)}\}$ a dataset to be classified into k classes, in which each $x_j^{(d)} \in \mathbb{R}^d$.

According the theorem of Cover [3], given a training data set non linearly separable, it can be, with high probability, transformed into a training data set linearly separable by projecting it into a higher dimension space.

The purpose of this process is increase the dimensionality of a data set, for this the mapping $x_j^{(d)} \to z_j^{(d)}$ is proposed, in which $z_j^{(d)} \in \mathbb{C}^d$. Thereby, to each $x_j^{(d)}$ there is an ordinate pair $(a_j^{(d)}, b_j^{(d)})$, satisfying the constraint

$$(x_j^{(d)})^2 = |z_j^{(d)}|^2 = (a_j^{(d)})^2 + (b_j^{(d)})^2. \tag{4}$$

Therefore, z_j^d is associated to the quantum state $|\Psi\rangle$ (off a multiplicative constant), and there is an observable operation T when applied to the quantum state obtains the eigenvalue $|z_j^{(d)}|^2$,

$$T|\Psi\rangle = |z_j^{(d)}|^2|\Psi\rangle$$

To obtain the pair $(a_j^{(d)}, b_j^{(d)})$, it was used a modified genetic algorithm proposed in [11] and described in the next subsection. With the application of the proposed heuristic, one pattern $x_j^{(d)}$ d-dimensional in the real space becomes $2d$-dimensional in the complex space.

As shown in the Section II-A, after measure the value of the observable, all information about the phase of the quantum state is lost. Thus, the genetic algorithm will search by the phase values that maximize the data power discrimination. This situation is reached if,

- The variance of each class is minimized;
- The distance between the classes' centroids are mutually maximized.

Therefore, a description of the proposed methodology can be given by,

1) For each pattern $x_j^{(d)}$ of the training data set, the genetic algorithm finds an ordinate pair $(a_j^{(d)}, b_j^{(d)})$ which satisfies the constraint given by the Equation 4;
2) After the genetic algorithm adjust all pairs $(a_j^{(d)}, b_j^{(d)})$, a new data set is created with the double dimensionality of the original data set;
3) Some Classification/Clustering Algorithm is applied to the new data set in the complex space.

In this way, it is possible to generate a complex data set representation where the classification task has a lower complexity compared to the same classification task in real space. The experimental results (Section III) will demonstrate that the proposed mapping $x_j^{(d)} \to z_j^{(d)}$ is possible and really provides an enhancement of data power discrimination.

1) The Genetic Algorithm Used: The modified genetic algorithm (GA) used here was utilized by Ferreira [11].

The population is composted of individuals, or chromosomes, each one representing a possible set of pairs $(a_j^{(d)}, b_j^{(d)})$ corresponding to the patterns $x_j^{(d)}$ of the original data set. In this way, a individual is a set of N 3-tuple like $(a_j^{(d)}, b_j^{(d)}, k_j)$, where k_j is a label indicating the class of the pattern $x_j^{(d)}$ and N is the number of the patterns in the original data set.

In each GA generation, two chromosomes in the population are selected to undergo genetics operations. The selection process is done by the method of spinning roulette wheel [11], [12].

The genetic operators include crossover and mutation operators. The crossover operator exchange information from two parents ($\mathbf{p}_1$ and $\mathbf{p}_2$). Here, four crossover operators were used,

$$C_1 = \frac{\mathbf{p}_1 + \mathbf{p}_2}{2} \tag{5}$$

$$C_2 = \mathbf{P}_{max}(1-w) + \max(\mathbf{p}_1, \mathbf{p}_2)w \tag{6}$$

$$C_3 = \mathbf{P}_{min}(1-w) + \min(\mathbf{p}_1, \mathbf{p}_2)w \tag{7}$$

$$C_4 = \frac{(\mathbf{P}_{max} + \mathbf{P}_{Min})(1-w) + (\mathbf{p}_1 + \mathbf{p}_2)w}{2} \tag{8}$$

where $\mathbf{P}_{max}$ and $\mathbf{P}_{Min}$ are the vectors with the maximum and minimum possible values for each gen of a chromosome, respectively. The function $\max$ returns the maximum values for each gen of $\mathbf{p}_1$ and $\mathbf{p}_2$, and the function $\min$ the minimum values. $w \in \mathbb{R}$ is a weight, where in the experiments was used the value of $w = 0.9$.

After the potential offspring is selected by the crossover operator, the best offspring is chosen. If this best offspring is better than the worst chromosome from the old population, then this offspring replaces the worst chromosome.

The four new chromosomes generated by the crossover process (C_1, C_2, C_3, C_4) are cloned and its clones undergo the mutation operation, where the features inherited from their parents can be changed. For each offspring cloned, three new chromosomes are generated by the mutation operation,

$$\begin{aligned} MC_{l,\alpha} =& [c_l^1 \; c_l^2 \; \cdots \; c_l^{no\_Vars}] + \\ & [\delta_1 mc_l^1 \; \delta_2 mc_l^2 \; \cdots \delta_{no\_Vars} mc_l^{no\_Vars}] \end{aligned} \tag{9}$$

where $\alpha = 1, 2, 3$ is the mutation index, $l = 1, 2, 3, 4$ is the offspring index, δ_u ($u = 1, 2, ..., no\_Vars$ and no_Vars is the number of genes in a chromosome) can only take values 0 or 1, and mc_l^u ($u = 1, 2, ..., no\_Vars$) are randomly generated numbers that satisfy the constraint $parameter_{min}^u \leq c_l^u + mc_l^u \leq parameter_{max}^u$.

The first mutation operation ($\alpha = 1$) is such that only one δ_u is one (u being generated randomly within the valid range) and all others are zero in Equation 9. The second mutation operation ($\alpha = 2$) is obtained by Equation 9, where some δ_u (randomly chosen) are set to one and others are set to zero. The third mutation operation ($\alpha = 3$) is obtained with all δ_u equal to one in Equation 9.

A real number is randomly generated and compared to a user defined number $p_{Mut} \in [01]$ (accepted mutational probability, here $p_{Mut} = 0.1$). If the real number is smaller than p_{Mut} then the mutated chromosome replaces the chromosome with the smallest fitness in the population. However, if the real number is larger than p_{Mut}, then the mutated chromosome replaces the chromosome with the smallest fitness of the population if and only if its fitness is greater than the fitness of the worst chromosome in the population.

All individuals (ind) are evaluated according to the fitness function given by,

$$f(ind) = \frac{\sum_{t=1}^{k} \sum_{j>t}^{k} \left(Z_{\mathbf{C}}^{(t)} - Z_{\mathbf{C}}^{(j)} \right)^2}{1 + \sum_{g=1}^{k} \sum_{v>g}^{k} \sqrt{\frac{1}{n_v - 1} \sum_{w=1}^{n_v - 1} \left(Z_{\mathbf{C}}^{(t)} - z_w^{(v)} \right)^2}} \tag{10}$$

in which $Z_{\mathbf{C}}^{(l)} = \frac{1}{n_l} \sum_{j=1}^{n_l} a_j^{(l)} - \imath b_j^{(l)}$ is the centroid of the classe l, n_l is the elements number of the class l, $\imath = \sqrt{-1}$ is the complex constant and $z_w^{(l)} = a_w^{(l)} + \imath b_w^{(l)}$ is a point in the complex space corresponding to the w-th real point of the class l in the original data set.

The objective of the fitness function defined in the Equation 10 is maximize the distance between clusters (or classes) and minimize the variance of each cluster in the complex space.

C. Data

The description of the data sets used to test the proposed methodology are related in the Tabel I with the domain (or nature of the data set), the size, the dimensionality d, and the number of classes.

TABLE I. DESCRIPTION SYNTHETIC OF THE DATA SETS USED.

Data	Domain	Size	Dimensionality	Classes
Concentric Circles	Artificial Data	600	2	2
Concentric Spirals	Artificial Data	600	2	2
Iris	Botanic	150	4	3

For the concentric circles were generated artificially 600 two dimensional points, divided into two classes: in the first class were generated randomly 300 points in a radius $0 \leq r \leq 0.5$, the inner circle. The second class, the rest of the points were generated randomly in a radius $1.0 \leq r \leq 1.5$, the outer circle. The two classes can be visualized in the Figure 2a. In particular for classification procedures based on calculate of distance between a point and a class centroid, this data set will generate an ambiguous classification.

For the concentric spirals also were generated artificially 600 two dimensional points, divided into two classes. We adopted the Archimedean spiral wherein the polar coordinates (r, θ) can be described by the equation $r = a + b\theta$, with $a, b \in \mathbb{R}$. Changing the parameter a will turn the spiral, while b controls the distance between successive turnings. We take $b = 0.5$ for the inner spiral and $b = -0.5$ for outer spiral and $a = 0$ for both cases. In the Figure 3a the two classes can be visualized. For classification procedures based on calculate of distance between a point and a class centroid, for data set concentric spirals also will generate an ambiguous classification.

The data set about the Iris flowers is a classification benchmark data set and it was obtained of UCI Machine Learning Repository [8], widely used to test new proposed methods in problems of classification. The data contains information about three species of Iris flower, namely Iris Setosa, Iris Versicolor and Iris Virginica. This data set consists of 150 examples with four attributes by species, lenght and widht of petals and sepal. One class is well separable of the others, while the others two are overlapping, generating ambiguous classification.

For the three data set used here, the data were normalized in the interval [0,1] by following equation

$$x_n^{(d)} = \frac{x_j^{(d)} - x_{min}^{(d)}}{x_{max}^{(d)} - x_{min}^{(d)}} \tag{11}$$

in which $x_j^{(d)}, x_{min}^{(d)}, x_{max}^{(d)}$ are respectively, the original value of the pattern of the data, minimum and maximum value of the pattern $x_j^{(d)}$.

III. Experimental Results

For each one of the three data set, Concentric circles, Concentric spirals and Iris, was applied the proposed methodology with the genetic algorithm to build the complex representation of the respective data set. After that, three different classification algorithm were employed to clustering the data in the real (raw data) and transformed data (complex space). The classification algorithm chosen were K-means [6], K-Nearest Neighbor (KNN) [4] and Linear Discriminant Analysis (LDA), or Fisher's linear discriminant [7], that are quite used to classification and clustering of a dataset.

The cited classification methods were applied in the raw data set (d-dimensional real space) as is widely viewed in the literature, and as much as equally applied to transformed data set ($2d$-dimensional complex space) obtained from the methodology presented here.

The results of both conditions are compared for assure the robustness of the process to become the data separable with the increasing of the dimensionality of the representation space, creating the ideal conditions for an enhancement the data discriminant power.

The first algorithm used to classify the data sets is the K-means algorithm [6], [1]. For the concentric circle, the K-means was applied to the raw data set and reached 82% of correct classification for the class 1 (inner circle) and 46% of correct classification for the class 2 (outer circle). The cause for this ambiguous classification is because the concentric circle raw data set has two class with centroids localized in the same place in the real representation space.

TABLE II. Experimental results of the K-means method for the real data and transformed data of Concentric circles, Concentric spirals and of Iris data set.

Dataset	Classe	K-MEANS Raw	K-MEANS Transf
Concentric circles	Classe 1	82%	100%
	Classe 2	46%	100%
Concentric spirals	Classe 1	43%	100%
	Classe 2	60%	100%
Iris	Classe 1	100%	100%
	Classe 2	94%	100%
	Classe 3	72%	100%

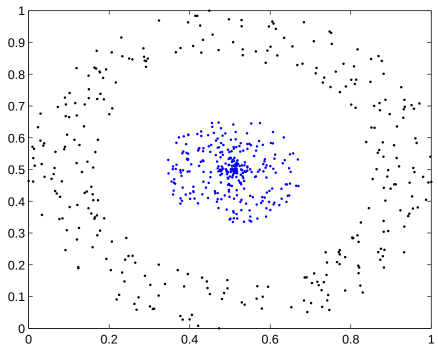
For the Concentric spiral dataset, the Kmeans was applied to the raw data set and for the class 1 (inner spiral) reached 43% of correct classification and for the class 2 (outer spiral) 60% of correct classification. The ambiguous classification is because the concentric spiral raw data set has two class with centroids localized in the same place in the real representation space, like happen with the concentric circles.

And for the Iris raw data set, the K-means algorithm reached a perfect classification for the first class (Iris Setosa) since this class is totally separated of two other class (Iris Versicolor and Iris Virginica). But, for these two other classes, Iris Versicolor (class 2) and Iris Virginica (class 3), which are overlapped, the K-means reached 94% of correct classification for the class 2 and 72% of correct classification for class 3. Again, the non separability of the data classes degrades the classification performance.

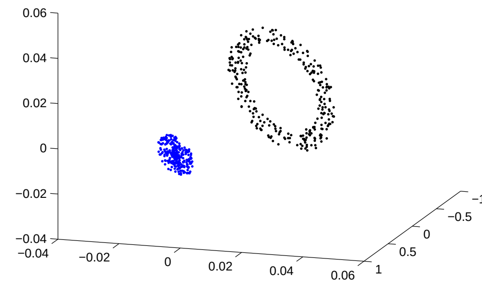
When is applied the proposed methodology to increasing the dimensionality of the representation space, passing of a d-dimensional real representation space for a $2d$-dimensional complex representation space, the classes will become separable with high probability, as postulated by Cover [3]. In this way, the K-means algorithm was applied to the transformed concentric circles, concentric spirals and Iris data set, reaching a perfect classification, demonstrating a gain in the ability of data discrimination and a possible separation of the classes of data sets. The Table II summarizes the experimental results reached by the K-means algorithms for the three data sets for raw and transformed data conditions.

The second algorithm employed for the classification task was the KNN algorithm [4], [1]. For the three data sets, concentric circles, concentric spirals and Iris, the data were partitioned in two subsets: a training subset with 80% of the data and a test subset with 20% of the data. For the concentric circles raw data set, the KNN reached a perfect classification with the parameter $k \leq 111$, where k is the number of neighbours used in the KNN algorithm to realize the classification task. For the concentric circles transformed data set, the KNN also reached a perfect classification, but in this case with $k \geqslant 1$. Thus, although there was no difference in the classification performance, there was a considerable reduction in the complexity of the problem, where the original problem is non linearly separable and after the proposed transformation the problem becomes linearly separable, as shown by the Figure 2.

The KNN algorithm reached a different classification performance for different neighborhood values k for the raw concentric spiral dataset. For $1 \leqslant k \leqslant 5$ we have 100% of



(a) Real Space



(b) Projection on $\mathbb{R}^3$ of the complex space

Fig. 2. The Concentric Circles, (a) 600 points distributed into two concentric circles in the original real space and (b) a projection of the complex space in the $\mathbb{R}^3$ space obtained with the application of the methodology.

TABLE III. Experimental results of the KNN method for the real data and transformed data of Concentric circles, Concentric spirals and of Iris data set (Test Set).

Dataset	Classe	KNN	
		Raw	Transf
Concentric circles	Classe 1	100%	100%
	Classe 2	100%	100%
Concentric spirals	Classe 1	74%	100%
	Classe 2	68%	100%
Iris	Classe 1	100%	100%
	Classe 2	90%	100%
	Classe 3	90%	100%

TABLE IV. Experimental results of the LDA method for the real data and transformed data of Concentric circles, Concentric spirals and of Iris data set.

Dataset	Classe	LDA	
		Raw	Transf
Concentric circles	Classe 1	46%	100%
	Classe 2	51%	100%
Concentric spirals	Classe 1	55%	100%
	Classe 2	48%	100%
Iris	Classe 1	100%	100%
	Classe 2	96%	100%
	Classe 3	98%	100%

correct classification for both classes. For $7 \leqslant k \leqslant 21$ we have in average 100% of correct classification for the class 1 (inner spiral) and 93% for class 2. These correct classification values decrease for some k values in both classes and after the class 1 return increasing while the class 2 continues decreasing. As follows, for $23 \leqslant k \leqslant 35$ in average 92% and 76% for class 1 and class 2, in that order, for $37 \leqslant k \leqslant 107$ in average 56% and 51%, for $109 \leqslant k \leqslant 121$ in average 75% and 36%, and finally for $123 \leqslant k \leqslant 151$ in average 84% and 19%. While for the transformed data the KNN algorithm also is able to generate a perfect classification, with 100% of hits for both classes and all k values tested. In the Table III is presented the average of hits for $1 \leqslant k \leqslant 75$.

For the Iris data set the KNN algorithm reached the same classification performance for $1 \leqslant k \leqslant 23$ for both situations of raw and transformed data. For raw Iris data, the KNN algorithm also is not able to generate a perfect classification, since the classes of the Iris data are overlapped. But for the transformed Iris data set the KNN algorithm reached a perfect classification, as shown in the Table III for the test set.

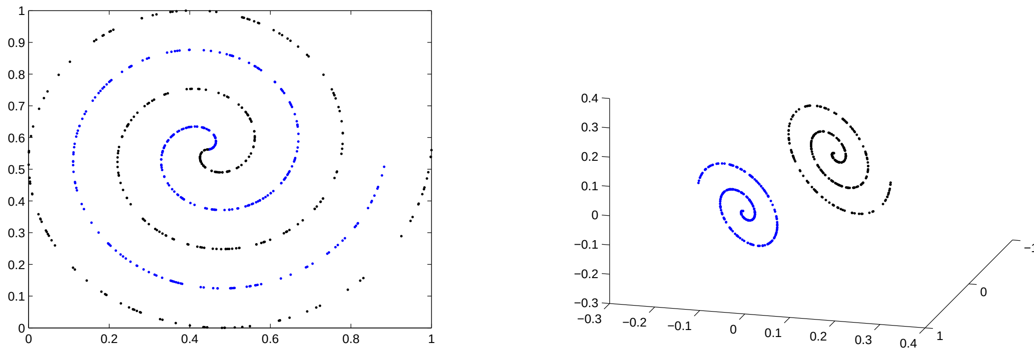
The last algorithm employed to classify the data set was the Linear Discriminant Analysis (LDA) [7], [1]. As the three raw data sets are non linearly separable, the LDA algorithm is not able to reach a perfect classification. However, for the transformed data sets, the LDA algorithm reached a perfect classification performance for all data sets, concentric circles, concentric spirals and Iris. One more time, these experimental results implies that with a high probability the proposed transformation is able to decrease the complexity of the data set and enhancement the data discrimination power, where a no linearly separable problem becomes in a linearly separable problem. The Table IV summarizes the results of the LDA algorithm for the raw and transformed data sets.

IV. Conclusion

In this work was presented a method quantum inspired, developed for increase the dimensionality of data sets, becoming non linearly separable classes in linearly separable classes, decreasing the complexity of the data and enhancement the data discrimination power, implying in a better classification task performance.

Here, it is assumed there is a quantum observable operator that generate the observed values of the data set. Therefore, it is searched the inverse transform, that takes real data into complex data, increasing the dimensionality of the originals dataset. A genetic algorithms was used for search this inverse transformation, acting like an inverse operator. The search is the task of defines a set of ordinate numerics pairs which mapping a real element into an complex element respecting the constraint defined by the Equation 4.

The concentric circles, concentric spirals and Iris data set were used to test the propose methodology. The three data sets are non linearly separable problem. After the employ of the proposed methodology, three different classical classification algorithms, K-means, KNN and LDA, are applied to the raw and transformed data, demonstrating the viability of this methodology to enhancement the data discrimination power. The experimental results shown the robustness of the



(a) Real Space

(b) Projection on $\mathbb{R}^3$ of the complex space

Fig. 3. The Concentric Spirals, (a) 600 points distributed into two concentric spirals in the original real space and (b) a projection of the complex space in the $\mathbb{R}^3$ space obtained with the application of the methodology.

methodology for decreasing of the complexity of the data, facilitating success of the classification task.

The Figure 2 presents the raw concentric circles data set (Figure 2a) and the a projection in the $\mathbb{R}^3$ of the transformed concentric circles data set (Figura 2b). And the Figure 3 shows the same behavior for the concentric spirals data set. Clearly, it is possible notice that the non linearly separable problem becomes with the proposed methodology in a linearly separable problem, enhancement the data discrimination power.

Therefore, the proposed methodology presented here was able to become the three non linearly separable problems, Concentric circles, Concentric spirals and Iris, in linearly separable problems. Furthermore, base on the Cover theorem [3], the proposed methodology, with high probability, will becomes same problem non linearly separable in a linearly separable problem.

Thus, this work demonstrate that is possible take a non linearly separable problem in a linearly separable problem. The next step is to develop a method to classify the new point not presented in the learning process performed by genetic algorithm. This new classification method is in implementation process and has an initial set of promising results, but still need to be consolidated.

Acknowledgment

To the Foundation of Support to Science and Technology of the State of Pernambuco/FACEPE, by financial support for the development of this research project.

References

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, 2000.

[2] V. N. Vapnik, *Statistical Learnig Theory*. John Wiley & Sons, 1998.

[3] T. M. Cover, “Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition.” *IEEE Transactions on Eletronic Computers*, 1965.

[4] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.

[5] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics*. John Wiley & Sons, 1977.

[6] J. MacQueen, “Some methods for classication and analysis of multivariate observations.” in *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. Berkeley: University of California Press, 1967, pp. 281–297.

[7] R. A. Fisher, “The use of multiple measurements in taxonomic problems.” *Ann. Eugen.*, vol. 7, p. 179188, 1936.

[8] A. Frank and A. Asuncion, “Uci machine learning repository,” april 2013, irvine, CA: University of California, School of Information and Computer Science. [Online]. Available: http://archive.ics.uci.edu/ml

[9] L. Denath and P. Mikusiński, *Hilbert Spaces with Applications*, 3rd ed. Elsevier Academic Press, 2005.

[10] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 1st ed. Cambridge University Press, 2000.

[11] T. A. E. Ferreira, G. C. Vasconcelos, and P. J. L. Adeodato, “A new intelligent system methodology for time series forecasting with artificial neural networks,” *Neural Processing Letters*, vol. 28, no. 2, pp. 113–129, 2008. [Online]. Available: http://dx.doi.org/10.1007/s11063-008-9085-x

[12] F. H. F. Leung, H. K. Lam, S. H. Ling, and P. K.-S. Tam, “Tuning of the structure and parameters of a neural network using an improved genetic algorithm,” *Neural Networks, IEEE Transactions on*, vol. 14, no. 1, pp. 79–88, 2003.