Soft Comput
DOI 10.1007/s00500-016-2478-2

FOUNDATIONS

A quantum-inspired version of the nearest mean classifier

Giuseppe Sergioli[1] · **Enrica Santucci**[1] · **Luca Didaci**[1] · **Jarosław A. Miszczak**[2] · **Roberto Giuntini**[1]

Abstract We introduce a framework suitable for describing standard classification problems using the mathematical language of quantum states. In particular, we provide a one-to-one correspondence between real objects and pure density operators. This correspondence enables us: (1) to represent the nearest mean classifier (NMC) in terms of quantum objects, (2) to introduce a quantum-inspired version of the NMC called quantum classifier (QC). By comparing the QC with the NMC on different datasets, we show how the first classifier is able to provide additional information that can be beneficial on a classical computer with respect to the second classifier.

Keywords Bloch sphere · Quantum classifier · Non-standard application of quantum formalism

Communicated by A. Di Nola.

✉ Giuseppe Sergioli
giuseppe.sergioli@gmail.com

Enrica Santucci
enrica.santucci@gmail.com

Luca Didaci
didaci@diee.unica.it

Jarosław A. Miszczak
miszczak@iitis.pl

Roberto Giuntini
giuntini@unica.it

[1] Università di Cagliari, Via Is Mirrionis 1, 09123 Cagliari, Italy

[2] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, Poland

1 Introduction

Quantum machine learning aims at merging the methods from quantum information processing and pattern recognition to provide new solutions for problems in the areas of pattern recognition and image understanding (Schuld et al. 2014a; Wittek 2014; Wiebe et al. 2015). In the first aspect, the research in this area is focused on the application of the methods of quantum information processing (Miszczak 2012) for solving problems related to classification and clustering Trugenberger (2002), Caraiman and Manta (2012). One of the possible directions in this field is to provide a representation of computational models using quantum mechanical concepts. From the other perspective, the methods for classification developed in computer engineering are used to find solutions for problems such as quantum-state discrimination (Helstrom 1976; Chefles 2000; Hayashi et al. 2005; Lu and Braunstein 2014), which are tightly connected with the recent developments in quantum cryptography.

Using quantum states for the purpose of representing patterns is naturally motivated by the possibility of exploiting quantum algorithms to boost the computational intensive parts of the classification process. In particular, it has been demonstrated that quantum algorithms can be used to improve the time complexity of the *k-nearest neighbor* (*k*NN) method. Using the algorithms presented in Wiebe et al. (2015), it is possible to obtain polynomial reductions in query complexity in comparison with the corresponding classical algorithm. Such an approach has been exploited by various authors. In Tanaka and Tsuda (2008), the authors propose an extension of Gaussian mixture models by using the statistical mechanics point of view. In their approach, the probability density functions of conventional Gaussian mixture models are expressed by density matrix representations. On the other hand, in Ostaszewski et al. (2015), the authors utilize the quantum

representation of images to construct measurements used for classification. Such approach might be particularly useful for the physical implementation of the classification procedure on quantum machines.

In the last few years, many efforts have been made to apply the quantum formalism to non-microscopic contexts (Aerts and D'Hooghe 2009; Aerts et al. 2013; Chiang et al. 2013; Eisert et al. 1999; Nagel 1963; Nagy and Nagy 2016; Ohya and Volovich 2011; Schwartz et al. 2005; Sozzo 2015; Stapp 1993) and to signal processing (Eldar and Oppenheim 2002). Moreover, some attempts to connect quantum information to pattern recognition can be found in Schuld et al. (2014a), Schuld et al. (2014b), Schuld et al. (2014c). An exhaustive survey and bibliography of the developments concerning the applications of quantum computing in computational intelligence are provided in Manju and Nigam (2014), Wittek (2014). Even if these results seem to suggest some possible computational advantages of an approach of this sort, an extensive and universally recognized treatment of the topic is still missing (Schuld et al. 2014a; Lloyd et al. 2014, 2013).

Also, this paper is motivated by the idea of using quantum formalism in a non-standard domain that consists in solving classification problems on datasets of classical objects. The main contribution of our work is the introduction of a new framework to encode the classification process by means of the mathematical language of density matrices (Beltrametti et al. 2014b, a). We show that this representation leads to two different developments: (i) It enables us to provide a representation of the nearest mean classifier (NMC) in terms of quantum objects; (ii) it can be used to introduce a quantum-inspired version of the NMC, that we call quantum classifier (QC), which can be considered (similarly as the NMC) to be a minimum distance classifier. This allows us a detailed comparison between NMC and QC. In particular, we will show that QC provides additional information about the data distribution and, in different cases, an improvement in the performance on a classical computer.

The paper is organized as follows. In Sect. 2, the basic notions of quantum information and pattern recognition are introduced. In Sect. 3, we formalize a correspondence between arbitrary two-feature patterns and pure density operators and we define the notion of *density pattern*. In Sect. 4, we provide a representation of NMC by using density patterns and by the introduction of an ad hoc definition of the distance between quantum states. Section 5 is devoted to the description of a new quantum classifier QC that does not have a classical counterpart in the standard classification process. Numerical simulations for both QC and NMC are presented, and particular benefits in favor of the first classifier are exploited. In Sect. 6, a geometrical idea to generalize the model to arbitrary n-feature patterns is proposed. Finally, Sect. 7 presents concluding remarks and suggests further developments.

2 Representing classical and quantum information quantities

In the standard quantum information theory (Bennett and Shor 1998; Shannon 1948), the states of physical systems are described by unit vectors and their evolution is expressed in terms of unitary matrices (i.e., quantum gates). However, this representation can be applied for an ideal case only, because it does not take into account some unavoidable physical phenomena, such as interactions with the environment and irreversible transformations. In the modern quantum information theory (Jaeger 2007, 2009; Wilde 2013), another approach is adopted. The states of physical systems are described by density operators—also called *mixed states* (Aharonov et al. 1998; Chiara et al. 2004; Freytes et al. 2010)—and their evolution is described by quantum operations. The space Ω_n of density operators for n-dimensional system consists of positive semidefinite matrices with unit trace.

A quantum state can be *pure* or *mixed*. We say that a state of a physical system is pure if it represents "maximal" information about the system, i.e., information that cannot be improved by further observations. A probabilistic mixture of pure states is said to be a *mixed* state. Generally, both pure and mixed states are represented by density operators that are positive and Hermitian operators (with unitary trace) living in a n-dimensional complex Hilbert space $\mathcal{H}$. Formally, a density operator ρ is pure iff $\mathrm{tr}(\rho^2) = 1$ and it is mixed iff $\mathrm{tr}(\rho^2) < 1$.

If we confine ourselves to the two-dimensional Hilbert space $\mathcal{H}$, a suitable representation of an arbitrary density operator $\rho \in \Omega_2$ is provided by

$$\begin{aligned}\rho &= \frac{1}{2}(I + r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3)\\ &= \frac{1}{2}\begin{pmatrix} 1+r_3 & r_1 - ir_2 \\ r_1 + ir_2 & 1-r_3 \end{pmatrix},\end{aligned} \tag{1}$$

where σ_i are the Pauli matrices. This expression is useful when providing a geometrical representation of ρ. Indeed, each density operator $\rho \in \Omega_2$ can be geometrically represented as a point of a radius-one sphere centered in the origin (the so-called *Bloch sphere*), whose coordinates (i.e., *Pauli components*) are r_i (with $\sum_i r_i^2 \leq 1$). By using the generalized Pauli matrices (Bertlmann and Krammer 2008; Kimura 2003), it is also possible to provide a geometrical representation for an arbitrary n-dimensional density operator, as it will be showed in Sect. 6. Again, by restricting to a two-dimensional Hilbert space, points on the surface of the Bloch sphere represent pure states, while inner points represent mixed states.

Quantum formalism turns out to be very useful not only in the microscopic scenario but also to encode classical

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

data. This has naturally suggested several attempts to represent the standard framework of machine learning through the quantum formalism (Lloyd et al. 2013; Schuld et al. 2014a). In particular, pattern recognition (Webb and Copsey 2011; Duda et al. 2000) is the scientific discipline which deals with theories and methodologies for designing algorithms and machines capable of automatically recognizing "objects" (i.e., patterns) in noisy environments. Some typical applications are multimedia document classification, remote-sensing image classification, and people identification using biometrics traits such as fingerprints.

A pattern is a representation of an object. The object can be a concrete one (i.e., an animal), and the pattern recognition task could be to identify the kind of animal or an abstract one (i.e., a facial expression), and the task could be to identify the emotion expressed by the facial expression. The pattern is characterized via a set of measurements called *features*.[1] Features can assume the forms of categories, structures, names, graphs, or, most commonly, a vector of real numbers (feature vector) $\mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$. Intuitively, a class is the set of all similar patterns. For the sake of simplicity, and without loss of generality, we assume that each object belongs to one and only one class, and we will limit our attention to two-class problems. For example, in the domain of "cats and dogs," we can consider the classes C_{cats} (the class of all cats) and C_{dogs} (the class of all dogs). The pattern at hand is either a cat or a dog, and a possible representation of the pattern could consist of the height of the pet and the length of its tail. In this way, the feature vector $\mathbf{x_1} = (x_{11}, x_{12})$ is the pattern representing a pet whose height and length of the tail are x_{11} and x_{12}, respectively.

Now, let us consider an object $\mathbf{x}_t$ whose membership class is unknown. The basic aim of the classification process is to establish which class $\mathbf{x}_t$ belongs to. To achieve this goal, standard pattern recognition designs a *classifier* that, given the feature vector $\mathbf{x}_t$, has to determine the true class of the pattern. The classifier should take into account all available information about the task at hand (i.e., information about the statistical distributions of the patterns and information obtained from a set of patterns whose true class is known). This set of patterns is called "training set," and it will be used to define the behavior of the classifier.

If no information about the statistical distributions of the patterns is available, an easy classification algorithm that could be used is the nearest mean classifier (NMC) (Manning et al. 2008; Hastie et al. 2001) or minimum distance classifier. The NMC

- computes the centroids of each class, using the patterns on the training set $\mu_i^* = \frac{1}{n_i}\sum_{\mathbf{x}\in C_i}\mathbf{x}$, where n_i is the number of patterns of the training set belonging to the class C_i;
- assigns the unknown pattern $\mathbf{x}_t$ to the class with the closest centroid.

In the next section, we provide a representation of arbitrary 2D patterns by means of density matrices, while in Sect. 4, we introduce a representation of NMC in terms of quantum objects.

3 Representation of two-dimensional patterns

Let $\mathbf{x_i} = (x_{i1}, \ldots, x_{ik})$ be a generic pattern, i.e., a point in $\mathbb{R}^k$. By means of this representation, we consider all the k features of $\mathbf{x_i}$ as perfectly known. Therefore, $\mathbf{x_i}$ represents a maximal kind of information, and its natural quantum counterpart is provided by a pure state. For the sake of simplicity, we will confine ourselves to an arbitary two-feature pattern indicated by $\mathbf{x} = (x, y)$.[2] In this section, a particular one-to-one correspondence between each pattern and its corresponding pure density operator is provided.

The pattern $\mathbf{x}$ can be represented as a point in $\mathbb{R}^2$. The stereographic projection (Coxeter 1969) allows to unequivocally map any point $r = (r_1, r_2, r_3)$ on the surface of a radius-one sphere $\mathbb{S}^2$ (except for the north pole) onto a point $\mathbf{x} = (x, y)$ of $\mathbb{R}^2$ as

$$SP : (r_1, r_2, r_3) \mapsto \left(\frac{r_1}{1-r_3}, \frac{r_2}{1-r_3}\right). \tag{2}$$

The inverse of the stereographic projection is given by

$$SP^{-1} : (x, y) \mapsto \left(\frac{2x}{x^2+y^2+1}, \frac{2y}{x^2+y^2+1}, \frac{x^2+y^2-1}{x^2+y^2+1}\right). \tag{3}$$

Therefore, by using the Bloch representation given by Eq. (1) and placing

$$r_1 = \frac{2x}{x^2+y^2+1}, \quad r_2 = \frac{2y}{x^2+y^2+1}, \quad r_3 = \frac{x^2+y^2-1}{x^2+y^2+1}, \tag{4}$$

we obtain the following definition.

[1] Hence, as a pattern is an object characterized by the knowledge of its features, analogously, in quantum mechanics a state of a physical system is represented by a density operator, characterized by the knowledge of its observables.

[2] In the standard pattern recognition theory, the symbol y is generally used to identify the label of the pattern. In this paper, for the sake of simplicity, we agree with a different notation.

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

Definition 1 (*Density pattern*) Given an arbitrary pattern $\mathbf{x} = (x, y)$, the density pattern (DP) $\rho_{\mathbf{x}}$ associated with $\mathbf{x}$ is the following pure density operator

$$\rho_{\mathbf{x}} = \frac{1}{2}\begin{pmatrix} 1+r_3 & r_1 - ir_2 \\ r_1 + ir_2 & 1-r_3 \end{pmatrix} = \frac{1}{x^2+y^2+1}\begin{pmatrix} x^2+y^2 & x-iy \\ x+iy & 1 \end{pmatrix}. \quad (5)$$

It is easy to check that $\mathrm{tr}(\rho_{\mathbf{x}}^2) = 1$. Hence, $\rho_{\mathbf{x}}$ always represents a pure state for any value of the features x and y. Following the standard definition of the Bloch sphere, it can be verified that $r_i = \mathrm{tr}(\rho_{\mathbf{x}} \cdot \sigma_i)$, with $i \in \{1, 2, 3\}$ and σ_i are Pauli matrices.

Example 1 Let us consider the pattern $\mathbf{x} = (1, 3)$. The corresponding $\rho_{\mathbf{x}}$ reads

$$\rho_{\mathbf{x}} = \frac{1}{11}\begin{pmatrix} 10 & 1-3i \\ 1+3i & 1 \end{pmatrix}.$$

One of the advantages of this encoding is based on the fact that it allows an easy visualization of an arbitrary two-feature dataset on the Bloch sphere, as it will be showed in the next section. The manner to encode a real pattern onto the space of the density operators is not unique and there is a risk of losing some information during the encoding. Taking into account the recent debates (Schuld et al. 2014a; Lloyd et al. 2013; Rebentrost et al. 2014), in order to encode a real vector to a quantum states without losing information, it is necessary to normalize the vector but maintain some information about the norm of the same vector at the same time. By following this procedure, we briefly show that it is alternatively possible to recover the stereographic encoding proposed in Eq. (5) also by simple analytical considerations.

Let $\mathbf{x} = (x, y)$ be an arbitrary real vector.

1. First, we map $\mathbf{x}$ onto a vector $\mathbf{x}'$ whose first component is given by $x + iy$ and the second component is given by the norm of $\mathbf{x}$ ($|\mathbf{x}| = \sqrt{x^2+y^2}$); i.e., $\mathbf{x} = (x, y) \mapsto \mathbf{x}' = (x + iy, \sqrt{x^2+y^2})$.
2. Now, we consider a second map: $\mathbf{x}' \mapsto \mathbf{x}'' = \left(\frac{x+iy}{\sqrt{x^2+y^2}}, \sqrt{x^2+y^2}\right)$, obtained by normalizing the first component of $\mathbf{x}'$.
3. Then, we consider the norm of $\mathbf{x}''$, i.e., $|\mathbf{x}''| = \sqrt{x^2+y^2+1}$ and we normalize the vector $\mathbf{x}''$, i.e., $\mathbf{x}'' \mapsto \mathbf{x}''' = \frac{\mathbf{x}''}{|\mathbf{x}''|} = \left(\frac{x+iy}{\sqrt{(x^2+y^2)(x^2+y^2+1)}}, \sqrt{\frac{x^2+y^2}{x^2+y^2+1}}\right)$.
4. Now, we consider the projector: $\mathbb{P} = \mathbf{x}''' \cdot (\mathbf{x}''')^{\dagger} = \frac{1}{x^2+y^2+1}\begin{pmatrix} 1 & x+iy \\ x-iy & x^2+y^2 \end{pmatrix}$.
5. Finally, we apply the operator $Not = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ to $\mathbb{P}$ and we recover: $Not(\mathbb{P}) = \frac{1}{x^2+y^2+1}\begin{pmatrix} x^2+y^2 & x-iy \\ x+iy & 1 \end{pmatrix}$, that is the same expression of the density pattern ρ_x introduced in Eq (5).

The introduction of the density pattern leads to two different developments. The first one is shown in the next section and consists of the representation of the NMC in quantum terms. Moreover, in Sect. 5, starting from the framework of density patterns, it will be possible to introduce a quantum classifier that exhibits an improvement in the performance (in terms of decreasing of the error in the classification process) with respect to the NMC.

4 Classification process for density patterns

As introduced in Sect. 2, the NMC is based on the computation of the minimum Euclidean distance between the pattern to classify and the centroids of each class. In the previous section, a quantum counterpart of an arbitrary (two feature) "classical" pattern was provided. In order to obtain a quantum counterpart of the standard classification process, we need to provide a suitable definition of distance d between DPs. In addition to satisfy the standard conditions of metric, the distance d also needs to satisfy the *preservation of the order*: Given three arbitrary patterns a, b, c such that $d_E(a, b) \leq d_E(b, c)$, if ρ_a, ρ_b, ρ_c are the DPs related to a, b, c, respectively, then $d(\rho_a, \rho_b) \leq d(\rho_b, \rho_c)$. In order to fulfill all the previous conditions, we obtain the following definition.

Definition 2 (*Normalized trace distance*) The normalized trace distance $\overline{d}_{\mathrm{tr}}$ between two arbitrary density patterns ρ_a and ρ_b is given by formula

$$\overline{d}_{\mathrm{tr}}(\rho_a, \rho_b) = K_{a,b} d_{\mathrm{tr}}(\rho_a, \rho_b), \quad (6)$$

where $d_{\mathrm{tr}}(\rho_a, \rho_b)$ is the standard trace distance, $d_{\mathrm{tr}}(\rho_a, \rho_b) = \frac{1}{2}\sum_i |\lambda_i|$, with λ_i representing the eigenvalues of $\rho_a - \rho_b$ (Barnett 2009; Nielsen and Chuang 2000), and $K_{a,b}$ is a normalization factor given by $K_{a,b} = \frac{2}{\sqrt{(1-r_{a_3})(1-r_{b_3})}}$, with r_{a_3} and r_{b_3} representing the third Pauli components of ρ_a and ρ_b, respectively.

Proposition 1 *Given two arbitrary patterns $a = (x_a, y_a)$ and $b = (x_b, y_b)$ and their respective density patterns, ρ_a and ρ_b, we have that*

$$\overline{d}_{tr}(\rho_a, \rho_b) = d_E(a, b). \quad (7)$$

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

Proof It can be verified that the eigenvalues of the matrix $\rho_a - \rho_b$ are given by

$$\mathrm{Eig}(\rho_a - \rho_b) = \pm \frac{d_E(a,b)}{\sqrt{(1+x_a^2+y_a^2)(1+x_b^2+y_b^2)}}. \tag{8}$$

Using the definition of trace distance, we have

$$\mathrm{tr}\sqrt{(\rho_a - \rho_b)^2} = \frac{d_E(a,b)}{\sqrt{(1+x_a^2+y_a^2)(1+x_b^2+y_b^2)}}. \tag{9}$$

By applying formula (4) to both r_{a_3} and r_{b_3}, we obtain that

$$K_{a,b} = \frac{2}{\sqrt{(1-r_{a_3})(1-r_{b_3})}} = \sqrt{(1+x_a^2+y_a^2)(1+x_b^2+y_b^2)}. \tag{10}$$

□

Using Proposition 1, one can see that the normalized trace distance $\overline{d}_{\mathrm{tr}}$ satisfies the standard metric conditions and the preservation of the order.

Due to the computational advantage of a quantum algorithm able to faster calculate the Euclidean distance (Wiebe et al. 2015), the equivalence between the normalized trace distance and the Euclidean distance turns out to be potentially beneficial for the classification process we are going to introduce.

Let us now consider two classes, C_A and C_B and the respective centroids[3] $a^* = (x_a, y_a)$ and $b^* = (x_b, y_b)$. The classification process based on NMC consists of finding the space regions given by the points closest to the first centroid a^* or to the second centroid b^*. The patterns belonging to the first region are assigned to the class C_A, while patterns belonging to the second region are assigned to the class C_B. The points equidistant from both the centroids represent the *discriminant function* (DF), given by

$$f_{\mathrm{DF}}(x,y) = 2(x_a - x_b)x + 2(y_a - y_b)y + (|b^*|^2 - |a^*|^2) = 0. \tag{11}$$

Thus, an arbitrary pattern $c = (x, y)$ is assigned to the class C_A (or C_B) if $f_{\mathrm{DF}}(x,y) > 0$ (or $f_{\mathrm{DF}}(x,y) < 0$).
Let us notice that the Eq. (11) is obtained by imposing the equality between the Euclidean distances $d_E(c, a^*)$ and $d_E(c, b^*)$. Similarly, we obtain the quantum counterpart of the classical discriminant function.

[3] Let us remark that, in general, a^* and b^* do not represent true centroids, but centroids estimated on the training set.

Proposition 2 *Let ρ_{a^*} and ρ_{b^*} be the DPs related to the centroids a^* and b^*, respectively. Then, the* quantum discriminant function (QDF) *is defined as*

$$f_{\mathrm{QDF}}(r_1, r_2, r_3) = \mathbf{F}(r_{a^*}, r_{b^*})^T \cdot \mathbf{r} + \tilde{K}^2 - 1 = 0 \tag{12}$$

where $\mathbf{r} = (r_1, r_2, r_3)$, $\{r_{a_i^*}\}$, $\{r_{b_i^*}\}$ *are Pauli components of ρ_{a^*} and ρ_{b^*}, respectively,* $\tilde{K} = \tilde{K}(r_{a_3^*}, r_{b_3^*}) = \frac{K_{c,a^*}}{K_{c,b^*}} = \sqrt{\frac{1-r_{a_3^*}}{1-r_{b_3^*}}}$, $\mathbf{F}(r_{a^*}, r_{b^*}) = (r_{a_1^*} - \tilde{K}^2 r_{b_1^*}, r_{a_2^*} - \tilde{K}^2 r_{b_2^*}, r_{a_3^*} - \tilde{K}^2 r_{b_3^*})$.

Proof In order to find the QDF, we use the equality between the normalized trace distances $K_{c,a^*} d_{\mathrm{tr}}(\rho_c, \rho_{a^*})$ and $K_{c,b^*} d_{\mathrm{tr}}(\rho_c, \rho_{b^*})$, where ρ_c is a generic DP with Pauli components r_1, r_2, r_3. We have

$$K_{c,a^*} d_{\mathrm{tr}}(\rho_c, \rho_{a^*}) = \sqrt{\frac{(r_1 - r_{a_1^*})^2 + (r_2 - r_{a_2^*})^2 + (r_3 - r_{a_3^*})^2}{(1 - r_{a_3^*})(1 - r_3)}},$$
$$K_{c,b^*} d_{\mathrm{tr}}(\rho_c, \rho_{b^*}) = \sqrt{\frac{(r_1 - r_{b_1^*})^2 + (r_2 - r_{b_2^*})^2 + (r_3 - r_{b_3^*})^2}{(1 - r_{b_3^*})(1 - r_3)}}. \tag{13}$$

The equality $K_{c,a^*} d_{\mathrm{tr}}(\rho_c, \rho_{a^*}) = K_{c,b^*} d_{\mathrm{tr}}(\rho_c, \rho_{b^*})$ reads

$$\sum_{i=1}^{3} r_i^2 + \sum_{i=1}^{3} r_{a_i^*}^2 - 2\sum_{i=1}^{3} r_i r_{a_i^*} = \frac{1 - r_{a_3^*}}{1 - r_{b_3^*}} \left(\sum_{i=1}^{3} r_i^2 + \sum_{i=1}^{3} r_{b_i^*}^2 - 2\sum_{i=1}^{3} r_i r_{b_i^*} \right). \tag{14}$$

In view of the fact that ρ_{a^*}, ρ_{b^*} and ρ_c are pure states, we use the conditions $\sum_{i=1}^{3} r_{a_i^*}^2 = \sum_{i=1}^{3} r_{b_i^*}^2 = \sum_{i=1}^{3} r_i^2 = 1$ and we get

$$\sum_{i=1}^{3} \left(r_{a_i^*} - \frac{1 - r_{a_3^*}}{1 - r_{b_3^*}} r_{b_i^*} \right) r_i + \frac{1 - r_{a_3^*}}{1 - r_{b_3^*}} - 1 = 0. \tag{15}$$

□

This completes the proof.

Similarly to the classical case, we assign the DP ρ_c to the class C_A (or C_B) if $f_{\mathrm{QDF}}(r_1, r_2, r_3) > 0$ (or $f_{\mathrm{QDF}}(r_1, r_2, r_3) < 0$). Geometrically, Eq. (12) represents the surface equidistant from the DPs ρ_{a^*} and ρ_{b^*}.

Let us remark that, if we express the Pauli components $\{r_{a_i^*}\}$, $\{r_{b_i^*}\}$ and $\{r_i\}$ in terms of classical features by Eq. (4), then Eq. (12) exactly corresponds to Eq. (11). As a consequence, given an arbitrary pattern $c = (x, y)$, if $f_{\mathrm{DF}}(c) > 0$ (or $f_{\mathrm{DF}}(c) < 0$), then its relative DP ρ_c will satisfy $f_{\mathrm{QDF}}(\rho_c) > 0$ (or $f_{\mathrm{QDF}}(\rho_c) < 0$, respectively).

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

As an example, the comparison between the classical and quantum discrimination functions for the *Moon* dataset (composed of 200 patterns equally allocated in two classes) is made in Fig. 1. Plots in Fig. 1a, b present the classical and quantum discrimination, respectively.

It is worth noting that the correspondence between the pattern expressed as a feature vector (according to the standard pattern recognition approach) and the pattern expressed as a density operator is quite general. Indeed, it is not related to a particular classification algorithm (NMC, in the previous case) nor to the specific metric at hand (the Euclidean one). Therefore, it is possible to develop a similar correspondence by using other kinds of metrics and/or classification algorithms, different from NMC, adopting exactly the same approach.

This result suggests potential developments which consist of finding a quantum algorithm able to implement the normalized trace distance between density patterns. So, it would correspond to implement the NMC on a quantum computer with the consequent well-known advantages (Wiebe et al. 2015). The next section is devoted to the exploration of another development that consists of using the framework of density patterns in order to introduce a "purely" quantum classification process (having no direct classical correspondence) called QC. It can be considered as a quantum-inspired version of the classical NMC because it substantially is a minimum distance classifier among quantum objects. The main difference between them, as we will show by numerical simulations, is that the NMC is a linear classifier which does not take into account the data dispersion, while the QC is not linear, and conversely, it seems sensitive to the data dispersion. As it will be showed in the next section by involving some datasets, this fact seems to be particularly beneficial (with respect to the NMC) mostly in the cases where the classes are quite mixed, and hence, the NMC generates a considerable error.

5 Quantum classification procedure

In Sect. 4, we have shown that the NMC can be expressed by means of quantum formalism, where each pattern is replaced by a corresponding density pattern and the Euclidean distance is replaced by the normalized trace distance. Representing classical data in terms of quantum objects seems to be particularly promising in quantum machine learning. Quoting Lloyd et al. (2013) "Estimating distances between vectors in N-dimensional vector spaces takes time $O(logN)$ on a quantum computer. Sampling and estimating distances between vectors on a classical computer is apparently exponentially hard". This convenience has already been exploited in machine learning for similar tasks (Wiebe et al. 2015; Giovannetti et al. 2008). Hence, finding a quantum algorithm for pattern classification using the proposed encoding could be particularly beneficial to speed up the classification process and it can suggest interesting developments. However, they are beyond the scope of this paper.

What we propose in this section is to exhibit some explicative examples to show how, on a classical computer, our classification procedure, based on the minimum distance, gives additional information with respect to the standard NMC.

5.1 Description of the quantum classifier (QC)

In order to get a real advantage in the classification process, we need to be not confined in a pure representation of the classical procedure in quantum terms. For this reason, we introduce a purely quantum representation where we consider a new definition of centroid. The basic idea is to define a *quantum centroid* not as the stereographic projection of the classical centroid, but as a convex combination of density patterns.

Trivially, given two real points x and y, the point $z = \frac{1}{2}(x+y)$ has the property to minimize the quantity $d_E(x,z)+d_E(z,y)$. In this case, $d_E(x,z) = d_E(z,y) = \frac{1}{2}d_E(x,y)$. Similarly, let us consider two density operators ρ and σ and let $\tau = \frac{1}{2}(\rho+\sigma)$. It is straightforward to show that $d_{\mathrm{tr}}(\rho,\tau) = d_{\mathrm{tr}}(\tau,\sigma) = \frac{1}{2}d_{\mathrm{tr}}(\rho,\sigma)$. In fact, $d_{\mathrm{tr}}(\rho,\tau) = d_{\mathrm{tr}}(\rho,\frac{1}{2}(\rho+\sigma)) = \frac{1}{2}\sum|\mathrm{Eigenvalues}(\rho-\frac{1}{2}\rho-\frac{1}{2}\sigma)| = \frac{1}{2}\sum|\mathrm{Eigenvalues}(\frac{1}{2}(\rho-\sigma))| = \frac{1}{2}\cdot\frac{1}{2}\sum|\mathrm{Eigenvalues}(\rho-\sigma)| = \frac{1}{2}d_{\mathrm{tr}}(\rho,\sigma)$. Analogously, we prove that $d_{\mathrm{tr}}(\sigma,\tau) = \frac{1}{2}d_{\mathrm{tr}}(\rho,\sigma)$.

Following this argument, we reasonably introduce the following definition.

Definition 3 (*Quantum centroid*) Given a dataset $\{P_1,\ldots,P_n\}$ with $P_i = (x_i, y_i)$, let us consider the respective set of density patterns $\{\rho_1,\ldots,\rho_n\}$. The Quantum centroid is defined as:

$$\rho_{QC} = \frac{1}{n}\sum_{i=1}^{n}\rho_i.$$

Obviously, the reasonable ways to define a quantum version of the classical centroid are not unique. We accord with this definition because, as we show in the rest of the section, it turns out to be beneficial in the reduction of the error during some typical classification process. The reasons of this convenience are intuitively contained in the following argument. Let us notice that ρ_{QC} is a mixed state that does not have any counterpart in the standard pattern recognition. Indeed, the quantum centroid may include some further information that the classical centroid generally discards. In fact, the classical centroid does not involve all the information about the dispersion of a given dataset, i.e., the classical centroid is invariant

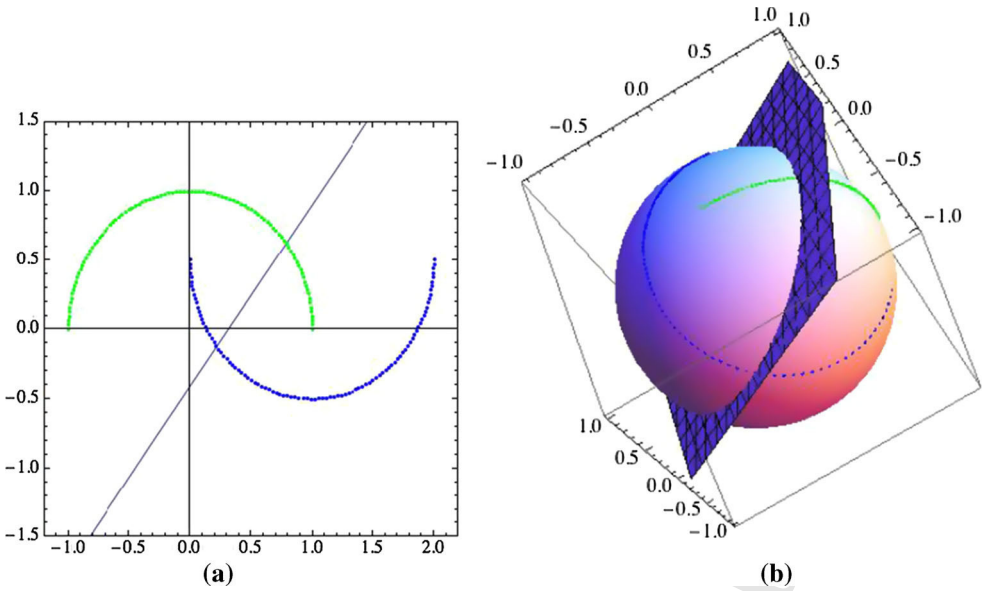

Fig. 1 Comparison between the discrimination procedures for the *Moon* dataset in $\mathbb{R}^2$ (**a**) and in the Bloch sphere $\mathbb{S}^2$ (Banana). *Green* and *blue points* represent the two classes in the real space and in the Bloch sphere, respectively. The *straight line* in **b** represents the classical discriminant function given by (11); on the other hand, the plane that intersects the Bloch sphere in **b** represents the quantum discriminant function given by (12)

under uniform scaling transformations of the data. Consequently, the classical centroid does not take into account any dispersion phenomena. Standard pattern recognition compensates for this lack by involving the covariance matrix (Duda et al. 2000).

On the other hand, the quantum centroid is not invariant under uniform scaling. In what follows, we show how the general expression of the quantum centroid is dependent on an arbitrary rescaling of a given dataset.

Let us consider the set of n points $\{P_1, \ldots, P_n\}$, where $P_i = (x_i, y_i)$ and let $C = (c_x, c_y) = (\frac{1}{n}\sum_{j=1}^{n} x_j, \frac{1}{n}\sum_{j=1}^{n} y_j)$ be the respective classical centroid. A uniform rescaling of n points of the dataset corresponds to move each point P_i along the line joining itself with C, whose generic expression is given by: $y_{x_i} = \frac{x-c_x}{x_i-c_x}(y_i - c_y) + c_y$. Let $\tilde{P}_i = (\tilde{x}_i, y_{\tilde{x}_i})$ be a generic point on this line. Obviously, a uniform rescaling of P_i by a real factor α is represented by the map: $\tilde{P}_i = (\tilde{x}_i, y_{\tilde{x}_i}) \mapsto \alpha\tilde{P}_i = (\alpha\tilde{x}_i, y_{\alpha\tilde{x}_i})$. Even if the classical centroid is not dependent on the rescaling factor α, on the other hand the expression of the quantum centroid is:

$$\rho_{QC} = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^{n} \frac{(\alpha\tilde{x}_i)^2+(y_{\alpha\tilde{x}_i})^2}{(\alpha\tilde{x}_i)^2+(y_{\alpha\tilde{x}_i})^2+1} & \sum_{i=1}^{n} \frac{\alpha\tilde{x}_i - i y_{\alpha\tilde{x}_i}}{(\alpha\tilde{x}_i)^2+(y_{\alpha\tilde{x}_i})^2+1} \\ \sum_{i=1}^{n} \frac{\alpha\tilde{x}_i + i y_{\alpha\tilde{x}_i}}{(\alpha\tilde{x}_i)^2+(y_{\alpha\tilde{x}_i})^2+1} & \sum_{i=1}^{n} \frac{1}{(\alpha\tilde{x}_i)^2+(y_{\alpha\tilde{x}_i})^2+1} \end{pmatrix}$$

that, clearly, is dependent on α.

According to the same framework used in Sect. 4, given two classes C_A and C_B of real data, let ρ_{QCa} and ρ_{QCb} be the respective quantum centroids. Given a pattern P and its respective density pattern, ρ_P, P is assigned to the class C_A (or C_B) if $d_{tr}(\rho_P, \rho_{QCa}) < d_{tr}(\rho_P, \rho_{QCb})$ (or $d_{tr}(\rho_P, \rho_{QCa}) > d_{tr}(\rho_P, \rho_{QCb})$, respectively). Let us remark that we no longer need any normalization parameter to be added to the trace distance d_{tr}, because the exact correspondence with the Euclidean distance is no more a necessary requirement in this framework. From now on, we refer to the classification process based on density patterns, quantum centroids, and trace distances as the *Quantum Classifier* (QC).

We have shown that the quantum centroid is not independent from the dispersion of the patterns and it could contain some additional information with respect to the classical centroid. Consequently, it is reasonable to expect that in different cases QC could provide some better performance than the NMC. The next subsection will be devoted to the exploitation of the difference between the classification procedures by means of numerical simulations on different datasets.
Before presenting the experimental results, let us briefly introduce the main statistical indices widely used to evaluate the performance of a supervised learning algorithm.

In particular, for each class, it is typical to refer to the respective confusion matrix (Fawcet 2006). It is based on four possible kinds of outcome after the classification of a certain pattern:

- True positive (TP): pattern correctly assigned to its class;
- True negative (TN): pattern correctly assigned to another class;

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

– False positive (FP): pattern uncorrectly assigned to its class;
– False negative (FN): pattern uncorrectly assigned to another class.

According to above, it is possible to recall the following definitions use to evaluate the performance of an algorithm.[4]

True positive rate (TPR), or sensitivity or recall: $\text{TPR} = \frac{\text{TP}}{\text{TP+FN}}$; false positive rate (FPR): $\text{FPR} = \frac{\text{FP}}{\text{FP+TN}}$; true negative rate (TNR): $\text{TNR} = \frac{\text{TN}}{\text{TN+FP}}$; false negative rate (FNR): $\text{FNR} = \frac{\text{FN}}{\text{FN+TP}}$.

Let us consider a dataset of C elements allocated in m different classes. We also recall the following basic statistical notions:

– Error: $E = 1 - \frac{\text{TP+TN}}{C}$;
– Precision: $\text{Pr} = \frac{\text{TP}}{\text{TP+FP}}$.

Moreover, another statistical index that is very useful to indicate the reliability of a classification process is given by the Cohen's k, that is $k = \frac{\text{Pr}(a)-\text{Pr}(e)}{1-\text{Pr}(e)}$, where $\text{Pr}(a) = \frac{\text{TP+TN}}{C}$ and $\text{Pr}(e) = \frac{(\text{TP+FP})(\text{TP+FN})+(\text{FP+TN})(\text{TN+FN})}{C^2}$. The value of k is such that $-1 \leq k \leq 1$, where the case $k = 1$ corresponds to a perfect classification procedure and the case $k = -1$ corresponds to the worst classification procedure.

5.2 Implementing the quantum classifier

In this subsection, we implement the QC on different datasets and show the difference between QC and NMC in terms of the values of error, accuracy, precision, and other probabilistic indices summarized above.

We will show how our quantum classification procedure exhibits partial or significant convenience with respect to the NMC on a classical computer.

In particular, we consider four datasets: Two of them (following Gaussian distributions) called *Gaussian* and *3Class-Gaussian*, where the first one is composed of 200 patterns allocated in two classes and the second one is composed of 450 patterns allocated in three classes, and the other two, called *Moon* and *Banana*, composed of 200 and 5300 patterns (respectively) allocated in two different classes.

The experiments have been conducted by randomly subdividing each dataset into a training set made up of 80% of instances and a test set containing the remaining instances. The results are reported in terms of averages of the computed statistical indices over 100 runs of the experiments.

We denote the variables listed in the tables as follows: E = Error; E_i = Error on the class i; Pr = Precision; k = Cohen's k; TPR = True positive rate; FPR = False positive rate; TNR = True negative rate; FNR = False negative rate. Let us remark that i) the values listed in the table are referred to the mean values over the classes; ii) in the case where the number of the classes is equal to two, a pattern that is correctly classified as belonging to a class corresponds to a pattern that is correctly classified as not belonging to the other class; on this basis, the mean value of TPR is equal to the mean value of TNR, and similarly, the mean value of FPR is equal to the mean value of FNR.

In order to provide a complete visualization of the difference between the two classification procedures, we also represent in the figures below the results of both classifications by considering the whole dataset for both training and test dataset.[5]

We stress that for some of the following datasets, the NMC is clearly not the optimal classifier and there exist classifiers that overcome it in terms of accuracy and performance. However, our aim in this context is confined the comparison of two minimum distance classifiers (classical and quantum-inspired version) trying to capture the main differences.

5.2.1 Gaussian dataset

This dataset consists of 200 patterns allocated in two classes (of equal size), following Gaussian distributions whose means are $\mu_1 = (1, 1)$, $\mu_2 = (2, 2)$ and covariance matrices are $\Sigma_1 = \text{diag}(20, 50)$, $\Sigma_2 = \text{diag}(5, 5)$, respectively.

As depicted in Fig. 2, the classes appear particularly overlapped and the QC is able to classify a number of true positive patterns that is significantly larger than the NMC. Hence, see Table 1, the error of the QC is (about 20%) smaller than the error of the NMC. In particular, the QC turns out to be strongly beneficial in the classification of the patterns of the second class. Moreover, the values related to accuracy, precision, and the other statistical indices also exhibit relevant improvements with respect to the NMC.[6]

5.2.2 The 3ClassGaussian dataset

In this example, we consider an equally distributed three-class dataset, consisting of 450 patterns. The classes are distributed as Gaussian random variables whose means are $\mu_1 = (-3, -3)$, $\mu_2 = (5, 5)$, $\mu_3 = (7, 7)$ and covariance matrices are $\Sigma_1 = \text{diag}(50, 100)$, $\Sigma_2 = \text{diag}(10, 5)$, and $\Sigma_3 = \text{diag}(30, 70)$, respectively. 3

[4] For the sake of the simplicity, from now on, we indicate $\sum_{j=1}^{C} \text{TP}_j$ with TP. Similarly for TN, FP, and FN.

[5] This make sense because it can be seen that for all the datasets we deal with, the classification error is very similar both with and without splitting training and test sets.

[6] Let us remark that there are some patterns correctly classified by the NMC which are neglected by the QC. On this basis, exploiting their complementarity, in principle it also makes sense to consider a combination of both classifiers.

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

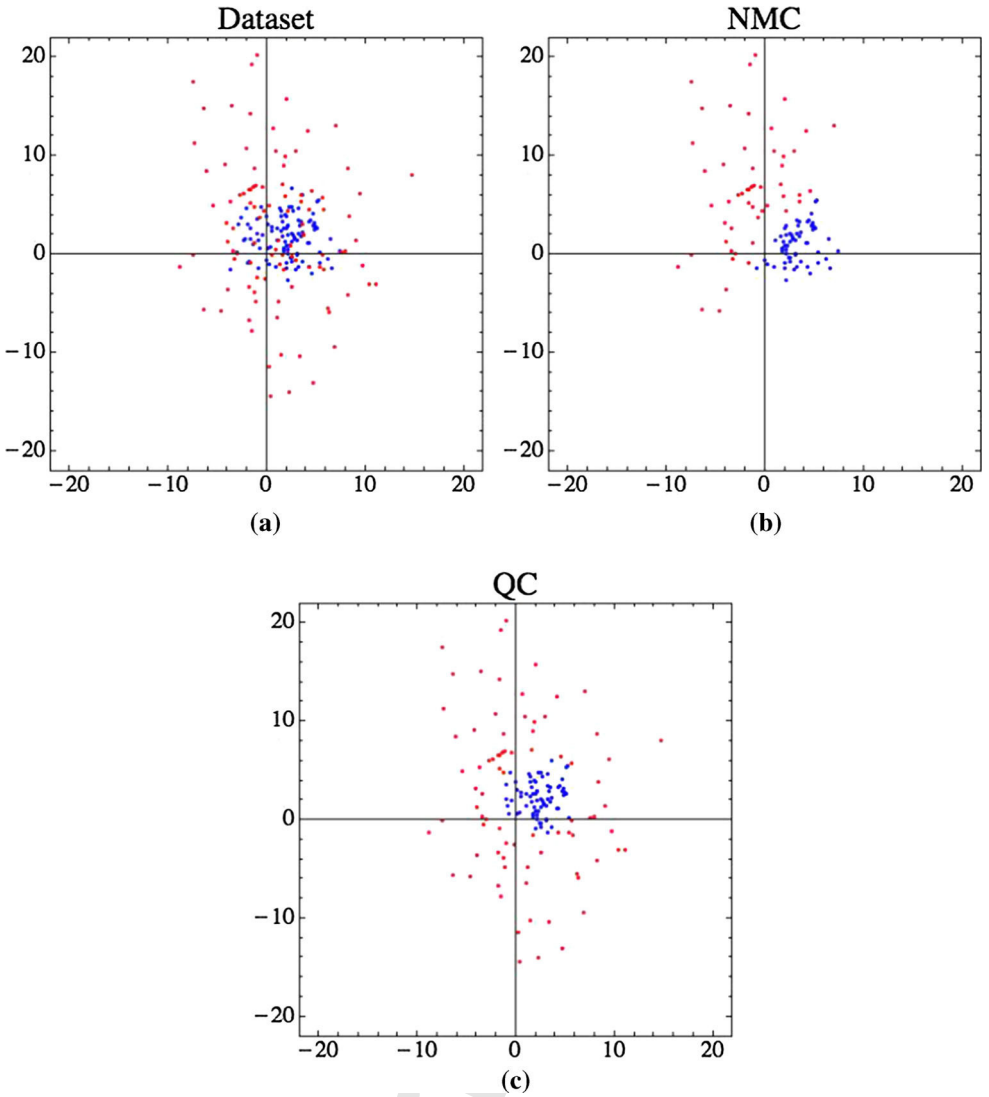

Fig. 2 Experimental results obtained for the Gaussian dataset: **a** dataset used in the experiments, **b** classification obtained using NMC, **c** classification obtained using QC

Table 1 Gaussian dataset

	E	$E1$	$E2$	Pr	k	TPR	FPR
NMC	0.457 ± 0.065	0.409 ± 0.108	0.503 ± 0.096	0.544	0.085	0.544	0.456
QC	0.246 ± 0.064	0.291 ± 0.096	0.203 ± 0.088	0.756	0.502	0.753	0.247

As for the two-class Gaussian dataset, the three classes appear quite overlapped, and once again, the computation of the error and the other statistical indices evaluated for both QC and NMC shows that the first one is more convenient especially for the classification of the first- and second-class patterns (Table 2).[7]

5.2.3 The Moon dataset

This dataset consists of 200 patterns equally distributed in two classes. In this case, we observe a mean error reduction of about 4%. In particular, the classification error related to the second class is very similar for both NMC and QC, while we can note that the QC turns out to be particularly beneficial in the classification of the first-class patterns (for which the error decreases by about 10%).

[7] In this case, by combining QC and NMC together, the mean error decreases up to about 0.247 (±4.280).

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

Table 2 3Gaussian dataset

	E	$E1$	$E2$	$E3$	Pr	k	TPR	FPR	TNR	FNR
NMC	0.358 ± 0.046	0.385 ± 0.085	0.422 ± 0.094	0.272 ± 0.077	0.651	0.462	0.640	0.179	0.821	0.360
QC	0.288 ± 0.045	0.296 ± 0.097	0.310 ± 0.084	0.259 ± 0.083	0.723	0.567	0.712	0.144	0.856	0.288

5.2.4 The Banana dataset

The Banana dataset presents a particularly complex distribution that is very hard to deal with the NMC. Indeed, the classification error we get by using the NMC is high and we would not use it in practice. Anyway, as we have already explained, we consider this particular dataset in order to show the substantial difference between two approaches (i.e., minimum distance classifiers) having the same nature (Fig. 3; Table 3).

This dataset consists of 5300 patterns unequally distributed between the two classes (2376 patterns belonging to the first class and 2924 belonging to the second one). In this case, the QC turns out to be beneficial in terms of all statistical indices and for both classes[8] it exhibits a mean error reduction of about 3% (Figs. 4, 5; Table 4).

Let us notice that, in accordance with the well known *No Free Lunch Theorem* (Duda et al. 2000), even if the previous examples exhibit a (particular or partial) benefit of the QC with respect to the NMC, in general there is no classifier whose performance is better than all the others for any dataset. This paper is focused on the comparison between the NMC and the QC because these methods are exclusively based on the pattern-centroid distance. Anyway, a wide comparison among the QC and other commonly used classifiers (such as the LDA—Linear Discriminant Analysis—and the QDA—Quadratic Discriminant Analysis) will be proposed for future works, where also other quantum metrics (such as the Fidelity, the Bures distance etc) instead of the trace distance and alternative definitions of quantum centroids will be considered to provide an adaptive version of the quantum classifier.

6 Geometrical generalization of the model

In Sect. 3, we provided a representation of an arbitrary two-feature pattern $\mathbf{x}$ in terms of a point on the surface of the Bloch sphere $\mathbb{S}^2$, i.e., a density operator $\rho_{\mathbf{x}}$. A geometrical extension of this model to the case of n-feature patterns inspired by quantum framework is possible.

[8] Similarly to the Gaussian case, also for the Banana dataset, the NMC is able to correctly classify some points unclassified by the QC. Indeed, by considering the combination of both classifiers, the mean error can decrease up to 10%.

In this section, by generalizing the encoding proposed in Sect. 2, we introduce a method for representing an arbitrary n-dimensional real pattern as a point in the radius-one hypersphere $\mathbb{S}^n$, centered in the origin.

A quantum system described by a density operator ρ in an n-dimensional Hilbert space $\mathcal{H}$ can be represented by a linear combination of the n-dimensional identity I and n^2-1 $(n \times n)$-square matrices $\{\sigma_i\}$ [i.e., *generalized Pauli matrices* (Bertlmann and Krammer 2008; Kimura 2003)]:

$$\rho = \frac{1}{n} I + \frac{1}{2} \sum_{i=1}^{n^2-1} r_i \sigma_i, \tag{16}$$

where the real numbers $\{r_i\}$ are the Pauli components of ρ. Hence, by Eq. (16), a density operator ρ acting on an n-dimensional Hilbert space can be geometrically represented as a (n^2-1)-dimensional point $P = (r_1, r_2, \ldots, r_{\tilde{n}})$ in the Bloch hypersphere $\mathbb{S}^{\tilde{n}-1}$, with $\tilde{n} = n^2 - 1$. Therefore, by using the generalization of the stereographic projection (Karlığa 1996), we obtain the vector $\mathbf{x} = (x_1, x_2, \ldots, x_{\tilde{n}-1})$ that is the correspondent of P in $\mathbb{R}^{n^2-2}$. In fact, the generalization of Eqs. (2–3) is given by

$$SP_{(\tilde{n})} : (r_1, r_2, \ldots, r_{\tilde{n}}) \mapsto \left(\frac{r_1}{1-r_{\tilde{n}}}, \frac{r_2}{1-r_{\tilde{n}}}, \ldots, \frac{r_{\tilde{n}-1}}{1-r_{\tilde{n}}} \right)$$
$$= (x_1, x_2, \ldots, x_{\tilde{n}-1}) \tag{17}$$

$$SP^{-1}_{(\tilde{n})} : (x_1, x_2, \ldots, x_{\tilde{n}-1})$$
$$\mapsto \left(\frac{2x_1}{\sum_{i=1}^{\tilde{n}} x_i^2 + 1}, \ldots, \frac{2x_{\tilde{n}-1}}{\sum_{i=1}^{\tilde{n}} x_i^2 + 1}, \frac{\sum_{i=1}^{\tilde{n}} x_i^2 - 1}{\sum_{i=1}^{\tilde{n}} x_i^2 + 1} \right)$$
$$= (r_1, r_2, \ldots, r_{\tilde{n}}). \tag{18}$$

Hence, by Eq. (17), a two-dimensional density matrix is determined by three Pauli components and it can be mapped onto a two-dimensional real vector. Analogously, a three-dimensional density matrix is determined by eight Pauli components and it can be mapped onto a seven-dimensional real vector. Generally, an n-dimensional density matrix is determined by n^2-1 Pauli components and it can be mapped onto an n^2-2-dimensional real vector.

Now, let consider an arbitrary vector $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ with $(n-1)^2 - 1 < m < n^2 - 2$. In this case, Eq. (18) cannot be applied because $m \neq n^2 - 2$. In order to represent a in an n-dimensional Hilbert space, it is sufficient to involve only $m+1$ Pauli components (instead of all the n^2-1 Pauli components of the n-dimensional space). Hence, we need to

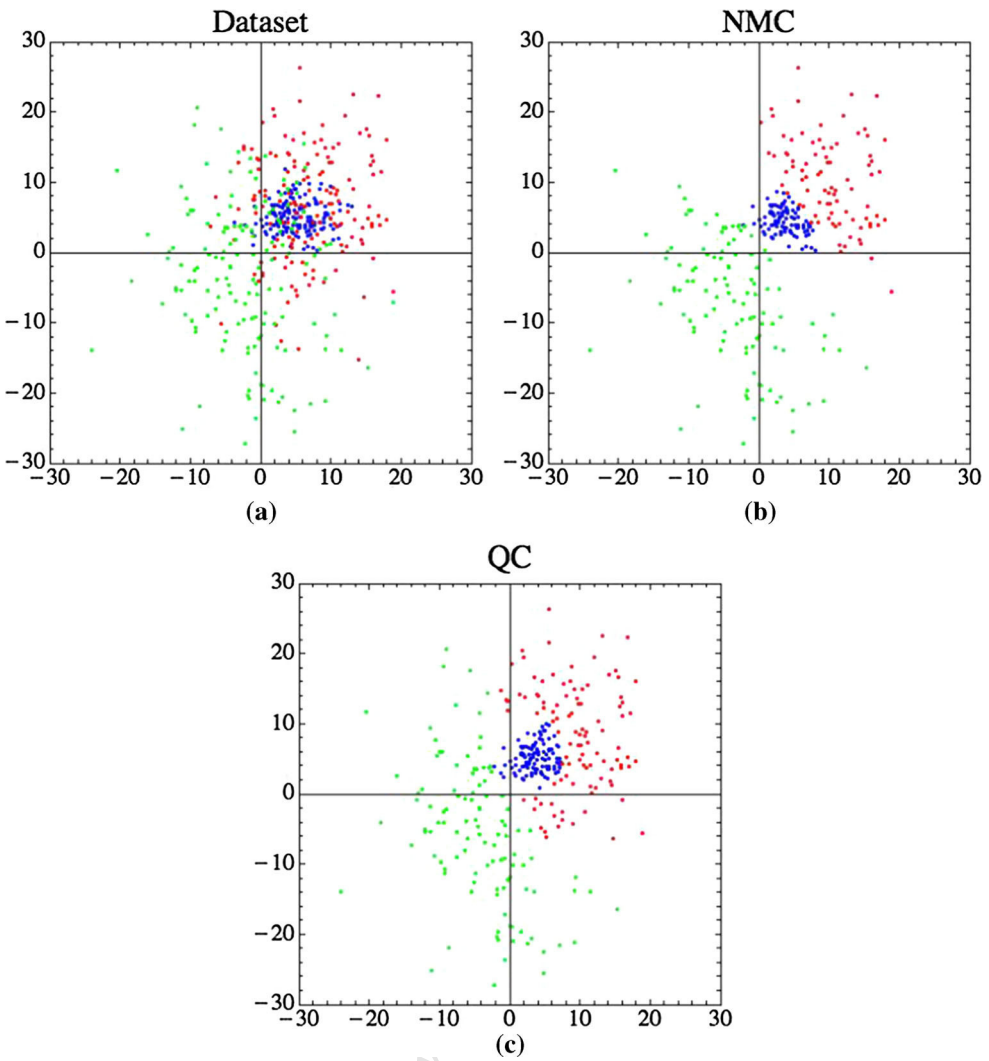

Fig. 3 Experimental results obtained for the 3ClassGaussian dataset: **a** dataset used in the experiments, **b** classification obtained using NMC, **c** classification obtained using QC

Table 3 Moon dataset

	E	$E1$	$E2$	Pr	k	TPR	FPR
NMC	0.211 ± 0.057	0.212 ± 0.094	0.210 ± 0.076	0.789	0.572	0.789	0.211
QC	0.174 ± 0.047	0.118 ± 0.075	0.226 ± 0.079	0.831	0.649	0.828	0.172

project the Bloch hypersphere $\mathbb{S}^{n^2-2}$ onto the hypersphere $\mathbb{S}^m$. We perform this projection by using Eq. (18) and by assigning some fixed values to a number of Pauli components equal to $n^2 - m - 2$. In this way, we obtain a representation in $\mathbb{S}^m$ that involves $m + 1$ Pauli components and it finally allows the representation of an m-dimensional real vector.

Example 2 Let us consider a vector $\mathbf{x} = (x_1, x_2, x_3)$. By Eq. (18), we can map $\mathbf{x}$ onto a vector $r_{\mathbf{x}} = (r_1, r_2, r_3, r_4) \in \mathbb{S}^3$. Hence, we need to consider a three-dimensional Hilbert space $\mathcal{H}$. Then, an arbitrary density operator $\rho \in \Omega_3$ can be written as

$$\rho = \frac{1}{3}\left(I + \sqrt{3}\sum_{i=1}^{8} r_i \sigma_i\right) \tag{19}$$

with $\{r_i\}$ Pauli components such that $\sum_{i=1}^{8} r_i^2 \leq 1$ and $\{\sigma_i\}$ generalized Pauli matrices. In this case, $\{\sigma_i\}$ is the set of eight 3×3 matrices also known as *Gell-Mann matrices*, namely

$$\sigma_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

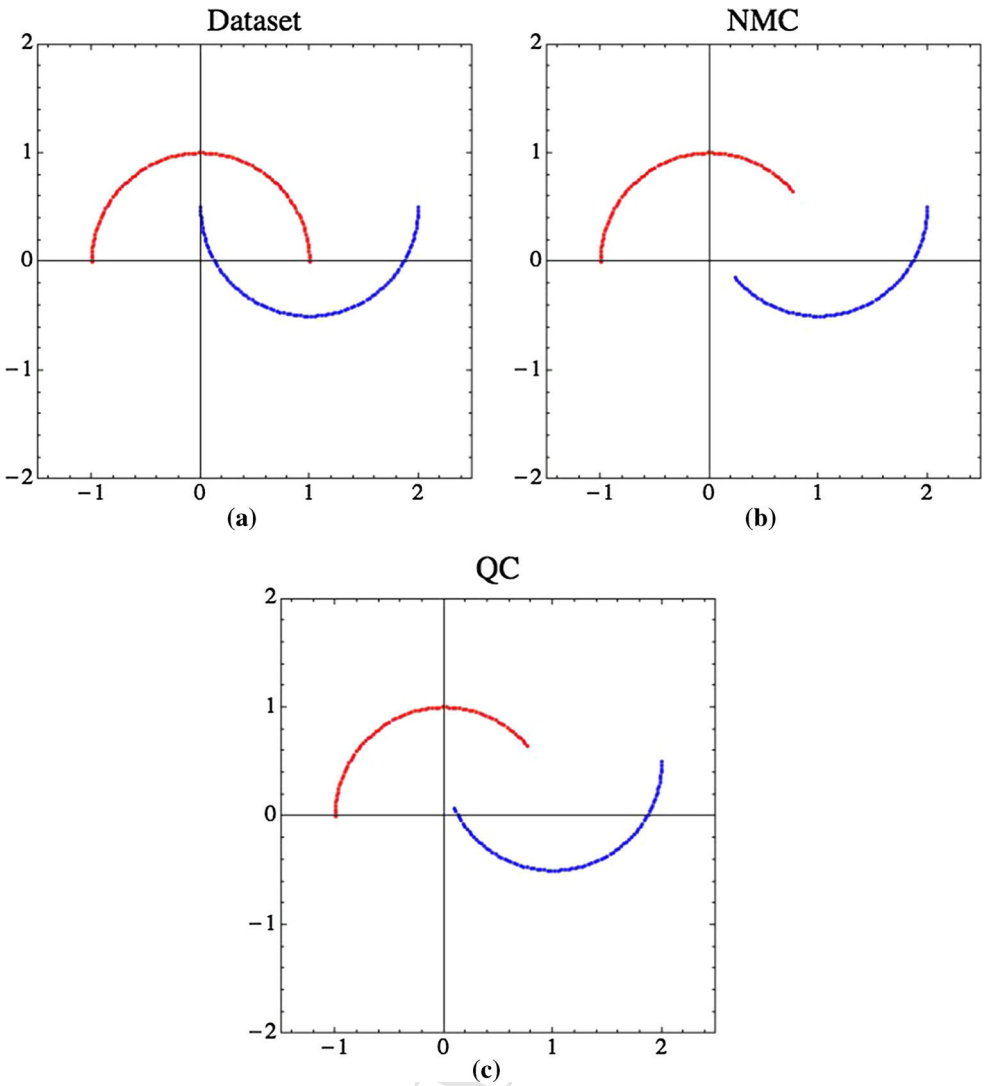

Fig. 4 Experimental results obtained for the Moon dataset: **a** dataset used in the experiments, **b** classification obtained using NMC, **c** classification obtained using QC

Table 4 Banana dataset

	E	$E1$	$E2$	Pr	k	TPR	FPR
NMC	0.448 ± 0.017	0.425 ± 0.028	0.466 ± 0.021	0.554	0.107	0.554	0.446
QC	0.419 ± 0.015	0.387 ± 0.022	0.446 ± 0.023	0.583	0.165	0.584	0.416

$$\sigma_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \sigma_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \sigma_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$
$$\sigma_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \sigma_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{20}$$

Consequently, the generic form of a density operator ρ in the three-dimensional Hilbert space is given by

$$\rho = \frac{1}{3} \begin{pmatrix} \sqrt{3}r_3 + r_8 + 1 & \sqrt{3}(r_1 - ir_2) & \sqrt{3}(r_4 - ir_5) \\ \sqrt{3}(r_1 + ir_2) & -\sqrt{3}r_3 + r_8 + 1 & \sqrt{3}(r_6 - ir_7) \\ \sqrt{3}(r_4 + ir_5) & \sqrt{3}(r_6 + ir_7) & 1 - 2r_8 \end{pmatrix}. \tag{21}$$

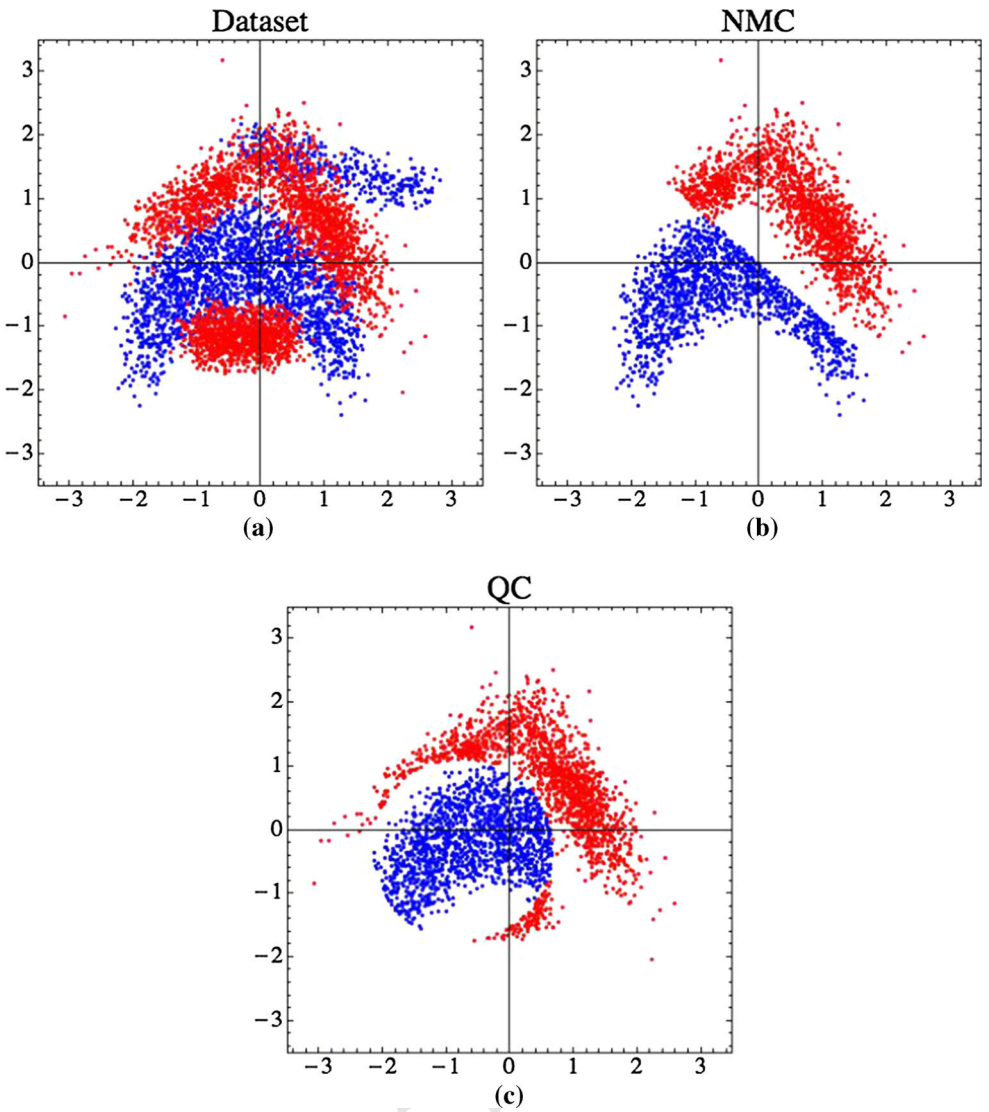

Fig. 5 Experimental results obtained for the Banana dataset: **a** dataset used in the experiments, **b** classification obtained using NMC, **c** classification obtained using QC

Then, for any ρ, it is possible to associate an eight-dimensional Bloch vector $r = (r_1, \ldots, r_8) \in \mathbb{S}^7$. However, by taking $r_j = 0$ for $j = 5, \ldots, 8$, we obtain

$$\rho_{\mathbf{x}} = \frac{1}{3}\begin{pmatrix} \sqrt{3}r_3 + 1 & \sqrt{3}(r_1 - ir_2) & \sqrt{3}r_4 \\ \sqrt{3}(r_1 + ir_2) & -\sqrt{3}r_3 + 1 & 0 \\ \sqrt{3}r_4 & 0 & 1 \end{pmatrix} \tag{22}$$

that, by Eq. (18), can be seen as a point projected in $\mathbb{S}^3$, where

$$SP_{(4)}^{-1}(\mathbf{x}) = r_{\mathbf{x}} = \left(\frac{2x_1}{\sum_{i=1}^3 x_i^2 + 1}, \frac{2x_2}{\sum_{i=1}^3 x_i^2 + 1}, \frac{2x_3}{\sum_{i=1}^3 x_i^2 + 1}, \frac{\sum_{i=1}^3 x_i^2 - 1}{\sum_{i=1}^3 x_i^2 + 1}\right). \tag{23}$$

The generalization introduced above allows the representation of arbitrary patterns $\mathbf{x} \in \mathbb{R}^n$ as points $\rho_{\mathbf{x}} \in \mathbb{S}^n$ as a natural extension of the encoding proposed in Sect. 2. Also, the classification procedure introduced in Sect. 4 can be naturally extended for an arbitrary n-feature pattern where the normalized trace distance between two DPs ρ_a and ρ_b can be expressed using Eq. (17) in terms of the respective Pauli components as

$$\overline{d}_{\mathrm{tr}}(\rho_a, \rho_b) = \frac{\sqrt{\sum_{i=1}^n [(r_{a_i} - r_{b_i}) - (r_{a_i} r_{a_{n+1}} - r_{b_i} r_{a_{n+1}})]^2}}{(1 - r_{a_{n+1}})(1 - r_{b_{n+1}})}. \tag{24}$$

Analogously, also the QC could be naturally extended to a n-dimensional problem (without loss of generality) by introducing a n-dimensional quantum centroid.

Journal: 500 MS: 2478 TYPESET DISK LE CP Disp.:2016/12/29 Pages: 15 Layout: Large

7 Conclusions and further developments

In this work, we have proposed a quantum-inspired version (QC) of the nearest mean classifier (NMC) and we have shown some convenience of the QC by comparing them on different datasets. Firstly, we have introduced a one-to-one correspondence between two-feature patterns and pure density operators by using the concept of *density patterns*. Starting from this representation, firstly we have described the NMC in terms of quantum objects by introducing an ad hoc definition of *normalized trace distance*. We have found a quantum version of the discrimination function by means of Pauli components. The equation of this surface was obtained by using the normalized trace distance between density patterns. Geometrically, it corresponds to a surface that intersects the Bloch sphere. This result could be potentially useful because it suggests to find an appropriate quantum algorithm able to provide a quantum version of the NMC in a quantum computer, with a consequent significative reduction in the computational complexity of the process (Lloyd et al. 2013; Wiebe et al. 2015).

Secondly, a suitable definition of a *quantum centroid* that does not have a classical direct correspondence permits to introduce a quantum classifier, which can be considered as a quantum-inspired version of the NMC, i.e., a minimum distance classifier.

The main implementative result of the paper consists of comparing the performance of NMC and QC on datasets with different properties. In particular, we found out that the QC may exhibit some better performance sensitive to the data dispersion. Then, the QC seems to be promising for classifying datasets whose classes have mixed distributions more difficult to treat by using the NMC. This also suggests to compare the QC with other standard classifiers as a further development. Further developments will be devoted to compare the QC with other commonly used classical classifiers.

Finally, we have presented a generalization of our model that allows to express arbitrary n-feature patterns as points on the hypersphere S^n, obtained by using the generalized stereographic projection. However, even if it is possible to associate the points of a n-hypersphere to n-feature patterns, these points do not generally represent density operators. In Kimura (2003), Jakóbczyk and Siennicki (2001), Kimura and Kossakowski (2005), the authors found some conditions that guarantee the one-to-one correspondence between the points on particular regions of the hypersphere and density matrices. A full development of our work is therefore intimately connected to the study of the geometrical properties of the generalized Bloch sphere.

Acknowledgements This work has been partly supported by the project "Computational quantum structures at the service of pattern recognition: modeling uncertainty" (CRP-59872) funded by Regione Autonoma della Sardegna, L.R. 7/2007 (2012) and the FIRB project "Structures and Dynamics of Knowledge and Cognition" (F21J120001 40001).

Funding This study was funded by Regione Autonoma della Sardegna, L.R. 7/2007, CRP-59872 (2012).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aerts D, D'Hooghe B (2009) Classical logical versus quantum conceptual thought: examples in economics, decision theory and concept theory. Quantum interaction (Lecture Notes in Computer Science), vol 5494. Springer, Berlin, pp 128–142

Aerts D, Gabora L, Sozzo S (2013) Concepts and their dynamics: a quantum-theoretic modeling of human thought. Top Cogn Sci 5(4):737–772

Aharonov D, Kitaev A, Nisan N (1998) Quantum circuits with mixed states. In: Proceedings of the 30th annual ACM symposium on theory of computing, pp 20–30. ACM

Barnett SM (2009) Quantum information, 16, Oxford Master Series in Physics. Oxford University Press, Oxford. Oxford Master Series in Atomic, Optical, and Laser Physics

Beltrametti E, Chiara ML Dalla, Giuntini R, Leporini R, Sergioli G (2014a) A quantum computational semantics for epistemic logical operators. Part II: semantics. Int J Theor Phys 53(10):3293–3307

Beltrametti E, Chiara ML Dalla, Giuntini R, Leporini R, Sergioli G (2014b) A quantum computational semantics for epistemic logical operators. Part I: epistemic structures. Int J Theor Phys 53(10):3279–3292

Bennett CH, Shor PW (1998) Quantum information theory. IEEE Trans Inf Theory

Bertlmann RA, Krammer P (2008) Bloch vectors for qudits. J Phys A 41(23):235303

Caraiman S, Manta V (2012) Image processing using quantum computing. In: IEEE 16th international conference on system theory, control and computing (ICSTCC), 2012, pp 1–6

Chefles A (2000) Quantum state discrimination. Contemp Phys 41(6):401–424. arXiv:quant-ph/0010114

Chiang H-P, Chou Y-H, Chiu C-H, Kuo S-Y, Huang Y-M (2013) A quantum-inspired Tabu search algorithm for solving combinatorial optimization problems. Soft Comput 18(9):1771–1781

Coxeter HSM (1969) Introduction to geometry, 2nd edn. Wiley, New York

Chiara ML Dalla, Giuntini R, Greechie R (2004) Reasoning in quantum theory: sharp and unsharp quantum logics, vol 22. Springer Science and Business Media, Berlin

Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley Interscience, Hoboken

Eisert J, Wilkens M, Lewenstein M (1999) Quantum games and quantum strategies. Phys Rev Lett 83(15):3077

Eldar YC, Oppenheim AV (2002) Quantum signal processing. IEEE Sig Process Mag 19(6):12–32

Fawcet T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27(8):861–874

Freytes H, Sergioli G, Aricò A (2010) Representing continuous t-norms in quantum computation with mixed states. J Phys A 43(46):465306 12

Giovannetti V, Lloyd S, Maccone L (2008) Quantum random access memory. Phys Rev Lett 100(16):160501

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Springer, Berlin

Hayashi A, Horibe M, Hashimoto T (2005) Quantum pure-state identification. Phys Rev A 72(5):052306

Helstrom CW (1976) Quantum detection and estimation theory. Academic Press, Cambridge

Jaeger G (2007) Quantum information. Springer, New York (An overview, With a foreword by Tommaso Toffoli)

Jaeger G (2009) Entanglement, information, and the interpretation of quantum mechanics. Frontiers Collection. Springer, Berlin

Jakóbczyk L, Siennicki M (2001) Geometry of Bloch vectors in two-qubit system. Phys Lett A 286(6):383–390

Karlığa B (1996) On the generalized stereographic projection. Beiträge Algebra Geom 37(2):329–336

Kimura G (2003) The Bloch vector for N-level systems. Phys Lett A 314(56):339–349

Kimura G, Kossakowski A (2005) The Bloch-vector space for N-level systems: the spherical-coordinate point of view. Open Syst Inf Dyn 12(03):207–229. arXiv:quant-ph/0408014

Kolossa D, Haeb-Umbach R (2011) Robust speech recognition of uncertain or missing data: theory and applications. Springer Science and Business Media, Berlin

Lloyd S, Mohseni M, Rebentrost P (2013) Quantum algorithms for supervised and unsupervised machine learning. arXiv:1307.0411

Lloyd S, Mohseni M, Rebentrost P (2014) Quantum principal component analysis. Nat Phys 10(9):631–633

Lu S, Braunstein SL (2014) Quantum decision tree classifier. Quantum Inf Process 13(3):757–770

Manju A, Nigam MJ (2014) Applications of quantum inspired computational intelligence: a survey. Artif Intell Rev 42(1):79–156

Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval, vol 1. Cambridge university press, Cambridge

Miszczak JA (2012) High-level structures for quantum computing, 6 synthesis (Lectures on Quantum Computing). Morgan and Claypool Publishers, San Rafael

Nagel E (1963) Assumptions in economic theory. Am Econ Rev 211–219

Nagy M, Nagy N (2016) Quantum-based secure communications with no prior key distribution. Soft Comput 20(1):87–101

Nielsen MA, Chuang IL (2000) Quantum computation and quantum information. Cambridge University Press, Cambridge

Ohya M, Volovich I (2011) Mathematical foundations of quantum information and computation and its applications to nano- and bio-systems. Theoretical and Mathematical Physics. Springer, Dordrecht

Ostaszewski M, Sadowski P, Gawron P (2015) Quantum image classification using principal component analysis. Theor Appl Inf 27:3. arXiv:1504.00580

Rebentrost P, Mohseni M, Lloyd S (2014) Quantum support vector machine for big feature and big data classification. Phys Rev Lett 113:130503

Sause MGR, Horn S (2013) Quantification of the uncertainty of pattern recognition approaches applied to acoustic emission signals. J Nondestruct Eval 32(3):242–255

Schuld M, Sinayskiy I, Petruccione F (2014a) An introduction to quantum machine learning. Contemp Phys 56(2). arXiv:1409.3097

Schuld M, Sinayskiy I, Petruccione F(2014b) Quantum computing for pattern classification. In: PRICAI 2014: trends in artificial intelligence, pp 208–220. Springer

Schuld M, Sinayskiy I, Petruccione F (2014c) The quest for a quantum neural network. Quantum Inf Process 13(11):2567–2586

Schwartz JM, Stapp HP, Beauregard M (2005) Quantum physics in neuroscience and psychology: a neurophysical model of mind-brain interaction. Philos Trans R Soc B Biol Sci 360(1458):1309–1327

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(379–423):623–656

Sozzo S (2015) Effectiveness of the quantum-mechanical formalism in cognitive modeling. Soft Comput. doi:10.1007/s00500-015-1834-y

Stapp HP (1993) Mind, matter, and quantum mechanics. Springer, Berlin

Tanaka K, Tsuda K (2008) A quantum-statistical-mechanical extension of gaussian mixture model. J Phys Conf Ser 95(1):012023

Trugenberger CA (2002) Quantum pattern recognition. Quantum Inf Process 1(6):471–493

Webb AR, Copsey KD (2011) Statistical pattern recognition, 3rd edn. Wiley, New York

Wiebe N, Kapoor A, Svore KM (2015) Quantum nearest-neighbor algorithms for machine learning. Quantum Inf Comput 15(34):0318–0358

Wilde MM (2013) Quantum Inf Theory. Cambridge University Press, Cambridge

Wittek P (2014) Quantum machine learning: what quantum computing means to data mining. Academic Press, Cambridge

Journal: 500
Article: 2478

Springer
the language of science

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details required	Author's response
1.	Please check if the corresponding author is correctly identified.	
2.	As per the information provided by the publisher, Fig. 1 will be black and white in print; hence, please confirm whether we can add "colour figure online" to the caption.	
3.	Please suggest whether the caption of Table 2 can be changed as '3ClassGaussian Dataset'.	
4.	Please check and confirm the inserted citation of Figures 3, 4, 5 and Tables 2, 3, 4 are correct. If not, please suggest an alternative citation. Please note that figures and tables should be cited in sequential order in the text.	
5.	Please provide volume and page number for reference Bennett and Shor (1998).	
6.	References Kolossa and Haeb-Umbach (2011), Sause and Horn (2013) are given in list but not cited in text. Please cite in text or delete from list.	
7.	Please provide volume number for reference Nagel (1963).	