Local Approach to Quantum-inspired Classification

Local approach to quantum-inspired classification

Enrico Blanzieri[1†], Roberto Leporini[2†] and Davide Pastorello[1*†]

[1]Department of Information Engineering and Computer Science, University of Trento, via Sommarive 9, Povo, I-38123, TN, Italy.
TIFPA-INFN, via Sommarive 14, Povo, I-38123, TN, Italy.
[2]Department of Economics, University of Bergamo, via dei Caniana 2, Bergamo, I-24127, BG, Italy.

*Corresponding author(s). E-mail(s): d.pastorello@unitn.it;
Contributing authors: e.blanzieri@unitn.it;
roberto.leporini@unibg.it;
†These authors contributed equally to this work

Abstract

In the context of quantum-inspired machine learning, remarkable mathematical tools for solving classification problems are given by some methods of quantum state discrimination. In this respect, quantum-inspired classifiers based on nearest centroid and Helstrom discrimination have been efficiently implemented on classical computers. We present a local approach combining the kNN algorithm to some quantum-inspired classifiers. We also compare the performance with respect to well-known classifiers applied to benchmark datasets.

Keywords: quantum state discrimination, k-nearest neighbors, machine learning.

Introduction

Quantum-inspired machine learning is a recent branch of machine learning based on the application of the mathematical formalism of quantum mechanics to devise novel schemes of information storing and processing for classical computers. Some quantum-inspired binary classification algorithms has been characterized from a geometric perspective and tested in a recent work [1] comparing the performances to those of well-known classical methods. This work is devoted to study some quantum-inspired classification algorithms, based on quantum state discrimination, within a local approach in the sense of a notion of locality in the feature domain w.r.t. a metric function. Thus, the approach does not rely to any notion of *quantum non-locality* or *local hidden variable theories* but refers to the locality in the feature space by selection of a neighborhood of a test point to be classified. More specifically, we reconsider the quantum-inspired algorithms based on Helstrom discrimination following an approach close to that proposed by Blanzieri and Melgani [2] where an unlabeled data instance is classified by finding its k nearest training points before running a support vector machine (SVM) over the considered k training points. The local approach turns out to be a resource in improving the accuracy in classification considering the SVM, this partially motivates the investigation of the present paper since the quantum-inspired Helstrom classifier can be interpreted as a SVM with linear kernel [3].

The considered quantum-inspired classifiers are structured on the encoding of the feature vectors into density operators and on techniques for estimating the distinguishability of quantum states like the Helstrom state discrimination and the Pretty-Good measurement (PGM). Moreover, we consider the quantum encoding in terms of Bloch vectors in order to take advantage from geometric properties of quantum states which saves space and time resources enabling the efficiency of these methods also on classical machines. Once represented data into different quantum encodings, corresponding to different feature maps, we run the classification algorithms within the local approach selecting the k nearest neighbors to execute the models. In the experimental part, we present a comparison of the performances of the local quantum-inspired classifiers against well-known classical algorithms in order to show that the local approach can be a valuable tool for increasing the performances of this kind of classifiers. A seminal research on a particular local approach based on the Voronoi tessellation of the training set to a quantum-inspired classifier defined by PGM is presented in [4].

In Section 1, we review the notion of quantum encoding of data vectors into density operators and the quantum-inspired classification based on the well-known machinery of quantum state discrimination [5–7]. Section 2 is focused on the encoding of feature vectors into Bloch vectors as a useful geometric tool to obtain a data representation that scales efficiently increasing the dimension of the feature space. In Section 3, we introduce the k-nearest neighbors algorithm (kNN) as a procedure to restrict the training set to the nearest points around the considered test points enabling the local execution of the quantum-inspired

classifiers. In section 4, there are the concluding remarks about the efficiency of local quantum-inspired classifiers and the related future developments towards innovative techniques in machine learning.

1 Quantum-inspired classification

The first building block of quantum-inspired classification is the *quantum encoding* (as well as in quantum computing), that is any procedure to encode classical information (e.g., a list of symbols) into quantum states. In this paper, we consider encodings of data vectors into density matrices on a Hilbert space H whose dimension depends on the dimension of the input space, in particular we use different quantum encodings to implement different feature maps for data representation. *Density matrices*, or density operators, are the mathematical objects used to describe the physical states of quantum systems. A density matrix on H is a positive semidefinite operator ρ such that $\mathrm{tr}\rho = 1$. *Pure states* are all the density matrices of the form $\rho = |\psi\rangle\langle\psi|$, with $\| \psi \| = 1$, which are the rank-1 projectors that can be directly identified with unit vectors up to a phase factor (according to the Dirac notation: $|\psi\rangle$ is a normalized vector, $\langle\psi|\phi\rangle$ is the inner product and $|\psi\rangle\langle\phi|$ is the outer product). Let ρ be a density operator on a d-dimensional Hilbert space (identified to $\mathbb{C}^d$), it can be written in the following form:

$$\rho = \frac{1}{d}\Big(\mathtt{I}_d + \sqrt{\frac{d(d-1)}{2}} \sum_{j=1}^{d^2-1} b_j^{(\rho)} \sigma_j\Big), \tag{1}$$

where $\{\sigma_j\}_{j=1,\dots,d^2-1}$ are the standard generators of the special unitary group $SU(d)$, also called *generalized Pauli matrices*, and $\mathtt{I}_d$ is the $d \times d$ identity matrix. The vector $\mathbf{b}^{(\rho)} = (b_1^{(\rho)}, \dots, b_{d^2-1}^{(\rho)})$, with $b_j^{(\rho)} = \sqrt{\frac{d}{2(d-1)}}\mathrm{tr}(\rho\,\sigma_j) \in \mathbb{R}$, is the *Bloch vector* associated to ρ which lies within the hypersphere of radius 1 in $\mathbb{R}^{d^2-1}$. For $d = 2$, the qubit case, the density matrices are in bijective correspondence to the points of the unit ball in $\mathbb{R}^3$ where the pure states are in one-to-one correspondence with the points of the spherical surface, the so-called *Bloch sphere*. For $d > 2$, the points contained in the unit hypersphere of $\mathbb{R}^{d^2-1}$ are not in bijective correspondence with density matrices on $\mathbb{C}^d$ such as in the case of a single qubit, so the Bloch vectors do not form a ball but a complicated convex body. However, any vector within the closed ball of radius $\frac{2}{d}$ gives rise to a density operator [8]. We apply the Bloch representation of density matrices as an efficient quantum encoding as clarify below.

Complex vectors of dimension n can be encoded into density matrices of a $(n+1)$-dimensional Hilbert space H in the following way:

$$\mathbb{C}^n \ni \mathbf{x} \mapsto |\mathbf{x}\rangle = \frac{1}{\sqrt{\| \mathbf{x} \|^2 + 1}} \left(\sum_{\alpha=0}^{n-1} x_\alpha |\alpha\rangle + |n\rangle \right) \in \mathsf{H}, \tag{2}$$

where $\{|\alpha\rangle\}_{\alpha=0,\dots,n}$ is the computational basis of H, identified as the standard basis of $\mathbb{C}^{n+1}$. The map defined in (2), called *amplitude encoding*, encodes $\mathbf{x}$ into the pure state $\rho_{\mathbf{x}} = |\mathbf{x}\rangle\langle\mathbf{x}|$ where the additional component of $|\mathbf{x}\rangle$ stores the norm of $\mathbf{x}$. Nevertheless the quantum encoding $\mathbf{x} \mapsto \rho_{\mathbf{x}}$ can be realized in terms of the Bloch vectors $\mathbf{x} \mapsto \mathbf{b}^{(\rho_{\mathbf{x}})}$ saving space resources. The improvement of memory occupation within the Bloch representation is evident when we take multiple tensor products $\rho\otimes\cdots\otimes\rho$ of a density matrix ρ constructing a feature map to enlarge the dimension of the representation space [1].

Quantum-inspired classifiers that we consider in the present work are based on quantum encoding of data vectors into density matrices, calculations of centroids, and various criteria of quantum state distinguishability such as: the Helstrom state discrimination [9, 10], the Pretty-Good measurement [6] (PGM), and the geometric construction of a minimum-error measurement [7] that we call *geometric Helstrom discrimination*. Let us briefly introduce the notion of quantum state discrimination. Given a set of arbitrary quantum states with respective a priori probabilities $R = \{(\rho_1, p_1), ..., (\rho_N, p_N)\}$, in general there is no a measurement process that discriminates the states without errors. More formally, there does not exist a POVM, i.e. a collection $E = \{E_i\}_{i=1,\dots,N}$ of positive semidefinite operators such that $\sum_{i=1}^{N} E_i = \mathtt{I}$, satisfying the following property: $\mathrm{tr}(E_i\rho_j) = 0$ when $i \neq j$ for all $i, j = 1, ..., N$. The probability of a successful state discrimination of the states in R performing the measurement E is:

$$\mathbb{P}_E(R) = \sum_{i=1}^{N} p_i \mathrm{tr}(E_i\rho_i). \tag{3}$$

A complete characterization, due to C.W. Helstrom [5], of the optimal measurement E_{opt} that maximizes the probability (3) for $R = \{(\rho_1, p_1), (\rho_2, p_2)\}$ can be done as follows. E_{opt} can be constructed as follows. Let $\Lambda := p_1\rho_1 - p_2\rho_2$ be the *Helstrom observable* whose positive and negative eigenvalues are, respectively, collected in the sets D_+ and D_-. Consider the two orthogonal projectors:

$$P_{\pm} := \sum_{\lambda\in D_{\pm}} P_{\lambda}, \tag{4}$$

where P_λ projects onto the eigenspace of λ. The measurement $E_{opt} := \{P_+, P_-\}$ maximizes the probability (3) that attains the *Helstrom bound*:

$$h_b(\rho_1, \rho_2) = p_1\mathrm{tr}(P_+\rho_1) + p_2\mathrm{tr}(P_-\rho_2). \tag{5}$$

Helstrom quantum state discrimination can be used to implement a quantum-inspired binary classifier with promising performances [10]. Let $\{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_M, y_M)\}$ be a training set with $\mathbf{x}_i \in \mathbb{C}^n$, $y_i \in \{1, 2\}$ $\forall i = 1, ..., M$. Assume to encode the data points into quantum states by means of $\mathbb{C}^n \ni \mathbf{x} \mapsto \rho_{\mathbf{x}} \in \mathfrak{S}(\mathsf{H})$, one can construct the quantum centroids ρ_1 and ρ_2 of

the two classes $C_{1,2} = \{\mathbf{x}_i : y_i = 1, 2\}$:

$$\rho_{1,2} = \frac{1}{|C_{1,2}|} \sum_{\mathbf{x} \in C_{1,2}} \rho_{\mathbf{x}} \tag{6}$$

Let $\{P_+, P_-\}$ be the Helstrom measurement defined by the set $R = \{(\rho_1, p_1), (\rho_2, p_2)\}$, where the probabilities attached to the centroids are $p_{1,2} = \frac{|C_{1,2}|}{|C_1|+|C_2|}$. The *Helstrom classifier* applies the optimal measurement for the discrimination of the two quantum centroids to assign the label y to a new data instance $\mathbf{x}$, encoded into the state $\rho_{\mathbf{x}}$, as follows:

$$y(\mathbf{x}) = \begin{cases} 1 \text{ if } & \mathrm{tr}(P_+\rho_{\mathbf{x}}) \geq \mathrm{tr}(P_-\rho_{\mathbf{x}}) \\ 2 & \text{otherwise} \end{cases} \tag{7}$$

A strategy to increase the accuracy in classification is given by the construction of the tensor product of q copies of the quantum centroids $\rho_{1,2}^{\otimes q}$ enlarging the Hilbert space where data are encoded. The corresponding Helstrom measurement is $\{P_+^{\otimes q}, P_-^{\otimes q}\}$, and the Helstrom bound satisfies [10]:

$$h_b(\rho_1^{\otimes q}, \rho_2^{\otimes q}) \leq h_b\left(\rho_1^{\otimes (q+1)}, \rho_2^{\otimes (q+1)}\right) \qquad \forall q \in \mathbb{N}. \tag{8}$$

Increasing the dimension of the Hilbert space of the quantum encoding, one increases the Helstrom bound obtaining a more accurate classifier. The corresponding computational cost is evident; however, in the next section, we observe that in the case of real input vectors, the space can be enlarged saving time and space by means of the encoding into Bloch vectors.

A method of quantum state discrimination for distinguishing more than two states $\{(\rho_1, p_1), ..., (\rho_N, p_N)\}$ is the square-root measurement, also known as *Pretty-Good measurement* (PGM), defined by:

$$E_i = p_i \rho^{-\frac{1}{2}} \rho_i \rho^{-\frac{1}{2}}, \tag{9}$$

where $\rho = \sum_i p_i \rho_i$, PGM is the optimal minimum-error when states satisfy certain symmetry properties [6]. Clearly to distinguish between n centroids we need a measurement with at most n outcomes. It is sometimes optimal to avoid measurement and simply guess that the state is the a priori most likely state.

The optimal POVM $\{E_i\}_i$ for minimum-error state discrimination over $R = \{(\rho_1, p_1), ..., (\rho_N, p_N)\}$ satisfies the following necessary and sufficient Helstrom conditions [7]:

$$\Gamma - p_i \rho_i \geq 0 \qquad \forall i = 1, \ldots, N, \tag{10}$$

where the Hermitian operator, also known as *Lagrange operator*, is defined by $\Gamma := \sum_i p_i \rho_i E_i$. It is also useful to consider the following properties which can be obtained from the above conditions:

$$E_j(p_j \rho_j - p_i \rho_i) E_i = 0 \quad \forall i, j. \tag{11}$$

For each i the operator $\Gamma - p_i\rho_i$ can have two, one, or no zero eigenvalues, corresponding to the zero operator, a rank-one operator, and a positive-definite operator, respectively. In the first case, we use the measurement $\{E_i = \mathtt{I}, E_{i\neq j} = 0\}$ for some i where $p_i \geq p_j\ \forall j$, i.e. the state belongs to the a priori most likely class. In the second case, if $E_i \neq 0$, it is a weighted projector onto the corresponding eigenstate. In the latter case, it follows that $E_i = 0$ for every optimal measurement.

Given the following Bloch representations:

$$\Gamma = \frac{1}{d}\Big(a\,\mathtt{I}_d + \sqrt{\frac{d(d-1)}{2}}\sum_{j=1}^{d^2-1} b_j\sigma_j\Big)\ , \qquad \rho_i = \frac{1}{d}\Big(\mathtt{I}_d + \sqrt{\frac{d(d-1)}{2}}\sum_{j=1}^{d^2-1} b_j^{(i)}\sigma_j\Big)\ , \tag{12}$$

in order to determine the Lagrange operator in $\mathbb{C}^d$ we need d^2 independent linear constraints:

$$2p_i\Big(a - \widehat{\mathbf{b}}^{(i)}\cdot\mathbf{b} - \frac{p_i}{2}(1 - |\widehat{\mathbf{b}}^{(i)}|^2)\Big) = a^2 - |\mathbf{b}|^2. \tag{13}$$

A measurement with more than d^2 outcomes can always be decomposed as a probabilistic mixture of measurements with at most d^2 outcomes. Therefore, if the number of classes is greater than or equal to d^2 and we get d^2 linearly independent equations, we construct the Lagrange operator and derive the optimal measurements. From the geometric point of view, we obtain the unit vectors corresponding to the rank-1 projectors $E_i = \frac{1}{d}\Big(\mathtt{I}_d + \sqrt{\frac{d(d-1)}{2}}\sum_{j=1}^{d^2-1} n_j^{(i)}\sigma_j\Big)$ where $\mathbf{n}^{(i)} = \frac{\widehat{\mathbf{b}}^{(i)} - a\mathbf{b}}{|\widehat{\mathbf{b}}^{(i)} - a\mathbf{b}|} \in \mathbb{R}^{d^2-1}$ giving the POVM of the measurement. It is also possible to further partition the classes in order to increase the number of centroids and of the corresponding equations. The classification is carried out in this way: An unlabelled point $\widehat{\mathbf{x}}$ is associated with the first label y such that $\mathbf{b}^{(\hat{\mathbf{x}})}\cdot\mathbf{n}^{(y)} = \max_i \mathbf{b}^{(\hat{\mathbf{x}})}\cdot\mathbf{n}^{(i)}$, where $d = \lceil\sqrt{length(\mathbf{x}) + 2}\rceil$. Such a geometric construction of the minimum-error state discrimination will be considered for the local approach to quantum-inspired classification tested below.

2 Bloch representation and centroid calculation

In quantum-inspired machine learning, the encoding of data instances into Bloch vectors of density operators turns out to be a useful geometric tool to reduce memory consumption in defining feature maps into higher dimensional spaces [4]. Let us consider the simplest case of encoding data vectors in $\mathbb{R}^2$ to density operators on $\mathbb{C}^2$ by means of the Bloch representation of qubit states:

$$\mathbb{R}^2 \ni (x_1, x_2) \mapsto \eta_{(x_1,x_2)} = \frac{1}{2}\Big(\mathtt{I}_2 + \sum_{j=1}^{3} b_j\sigma_j\Big), \tag{14}$$

where the Bloch vector of the density operator $\eta_{(x_1,x_2)}$ is given by $\mathbf{b} = \frac{1}{\sqrt{x_1^2+x_2^2+1}}(x_1, x_2, 1) \in \mathbb{R}^3$. Assuming data are divided into classes, the centroids of the classes can be calculated as the means of the Bloch vectors encoding data vectors. Thus, the centroids are points inside the Bloch sphere of a qubit then correspond to density operators.

Within the quantum encoding (2), a real vector $\mathbf{x} \in \mathbb{R}^{d-1}$ is encoded into a projection operator $\rho_{\mathbf{x}} = |\mathbf{x}\rangle \langle \mathbf{x}|$, on a d-dimensional Hilbert space where $d \geq 2$. For simplicity, we consider again an input vector $(x_1, x_2) \in \mathbb{R}^2$ and the corresponding projection operator $\rho_{(x_1,x_2)}$ on $\mathbb{C}^3$ given applying the encoding (2):

$$\rho_{(x_1,x_2)} = \frac{1}{x_1^2 + x_2^2 + 1} \begin{pmatrix} x_1^2 & x_1x_2 & x_1 \\ x_1x_2 & x_2^2 & x_2 \\ x_1 & x_2 & 1 \end{pmatrix}. \tag{15}$$

The explicit calculation shows that the Bloch vector of $\rho_{(x_1,x_2)}$ has null components:

$$\mathbf{b}^{(x_1,x_2)} = \frac{1}{1 + x_1^2 + x_2^2} \Big(2x_1x_2, 2x_1, 2x_2, 0, 0, 0, x_1^2 - x_2^2, \frac{x_1^2 + x_2^2 - 2}{\sqrt{3}} \Big). \tag{16}$$

Instead of using the matrix (15) with nine real elements to represent $\rho_{(x_1,x_2)}$, memory occupation can be improved by storing $\rho_{(x_1,x_2)}$ only in terms of the non-zero components of the Bloch vector. In general, the technique of removing the components that are zero or are repeated several times (whose positions in the Bloch vector are known a priori) allows reducing the space and the calculation time considering only the significant values that allow to carry out the classification.

Generally speaking, defining a quantum encoding is equivalent to select a feature map to represent feature vectors into a space of higher dimension. In this sense, data representation into quantum states can be considered a way to perform kernel tricks. In the case of the considered quantum encoding $\mathbb{R}^2 \ni (x_1, x_2) \mapsto \rho_{(x_1,x_2)} \in \mathfrak{S}(\mathbb{C}^3)$, in view of (16) the nonlinear explicit injective function $\varphi : \mathbb{R}^2 \to \mathbb{R}^5$ to encode data into Bloch vectors can be defined as follows:

$$\varphi(x_1, x_2) := \frac{1}{x_1^2 + x_2^2 + 1} \Big(2x_1x_2, 2x_1, 2x_2, x_1^2 - x_2^2, \frac{x_1^2 + x_2^2 - 2}{\sqrt{3}} \Big). \tag{17}$$

From a geometric point of view, the mapped feature vectors are points on the surface of a hyper-hemisphere. The centroids of the classes, calculated as the means of these feature vectors, are points inside the hypersphere that, in general, do not have an inverse image in terms of a density operators, however they can be rescaled to a Bloch vector as discussed below.

In order to improve the accuracy of the classification, one can increase the dimension of the representation space providing k copies of the quantum states, in terms of a tensor product, encoding data instances and centroids

into "redundant" density matrices $\rho^{\otimes q}$. According to the quantum formalism, multiple copies of the states are described in a tensor product Hilbert space with a strong impact in terms of computational resources because of the exponential growth of the dimension. However the Bloch encoding allows an efficient implementation of feature maps: consider two copies of the density operator $\rho_{(x_1,x_2)}$ defined in (15) and calculate the Bloch vector of $\rho_{(x_1,x_2)} \otimes \rho_{(x_1,x_2)}$, by removing null and repeated entries from the Bloch vector we obtain the following injective function for data encoding:

$$\varphi(x_1,x_2) := \frac{1}{(x_1^2+x_2^2+1)^2}\Big(2x_1^3x_2, 2x_1^3, 2x_1^2x_2^2, 2x_1^2x_2, 2x_1^2, 2x_1x_2^3, 2x_1x_2^2, 2x_1x_2, 2x_1, 2x_2^3,$$
$$2x_2^2, 2x_2, x_1^2(x_1-x_2)(x_1+x_2), \frac{x_1^2(x_1^2+x_2^2-2)}{\sqrt{3}}, \frac{x_1^2(x_1^2-2x_2^2+1)}{\sqrt{6}},$$
$$\frac{x_1^4-4x_2^4+x_1^2(2x_2^2+1)}{\sqrt{10}}, \frac{x_1^2+x_1^4-5x_2^2+2x_1^2x_2^2+x_2^4}{\sqrt{15}}, \frac{x_1^4+x_2^2+x_2^4+x_1^2(2x_2^2-5)}{\sqrt{21}},$$
$$\frac{x_1^4-6x_2^2+x_2^4+2x_1^2(x_2^2+1)}{2\sqrt{7}}, \frac{1}{6}(x_1^2+x_2^2-2)(x_1^2+x_2^2+4)\Big). \tag{18}$$

In (18), we consider only 20 values instead of 81 matrix elements of $\rho_{(x_1,x_2)} \otimes \rho_{(x_1,x_2)}$, 51 values instead of 729 for $\rho_{(x_1,x_2)} \otimes \rho_{(x_1,x_2)} \otimes \rho_{(x_1,x_2)}$ and so on. Therefore, the Bloch representation allows an efficient storing of redundant density matrices $\rho^{\otimes q}$.

Let us consider a training set divided into the classes $C_1, ..., C_M$, assume we have any training point $\mathbf{x}$ encoded into the Bloch vector $\mathbf{b}^{(\mathbf{x})}$ of a pure state on $\mathbb{C}^d$. The calculation of the centroid of the class C_i, within this quantum encoding, must take into account that the mean of the Bloch vectors $\mathbf{b}^{(i)} := \frac{1}{|C_i|}\sum_{\mathbf{x}\in C_i}\mathbf{b}^{(\mathbf{x})}$ does not represent a density operator in general. In fact, for $d > 2$ the points contained in the unit hypersphere of $\mathbb{R}^{d^2-1}$ are not in bijective correspondence with density matrices on $\mathbb{C}^d$. However, since any vector within the closed ball of radius $\frac{2}{d}$ gives rise to a density operator, a centroid can be defined in terms of a meaningful Bloch vector by a rescaling:

$$\widehat{\mathbf{b}}^{(i)} := \frac{2}{d|C_i|}\sum_{\mathbf{x}\in C_i}\mathbf{b}^{(\mathbf{x})}. \tag{19}$$

In the following, we choose $\widehat{\mathbf{b}}^{(i)}$ as the definition of centroid in order to represent the centroids as density matrices in order to perform a meaningful quantum state discrimination.

3 Local quantum-inspired classifiers

In this section, we introduce the local approach to quantum-inspired classification. More precisely, we consider the execution of the classifiers based on quantum state discrimination described in section 1 after a selection of the k training points that are closest to a considered unclassified instance.

The k-nearest neighbors algorithm [11] (kNN) is a really simple classification algorithm, and consists of the following steps:

1. the computation of the chosen distance metric between the test element and all training data points;
2. the extraction of the k elements closest to the test instance;
3. the assignment of the class label through a majority voting based on the labels of the k nearest neighbors.

In the following, we apply the kNN for the extraction of the closets elements to the test point then the classification is performed by a quantum-inspired algorithm instead of majority voting. On the one hand, given a test point, the kNN can be executed over the data vectors in the input space, e.g. considering the Euclidean distance, then the k neighbors can be encoded into density matrices and used for a quantum-inspired classification. On the other hand, the entire dataset can be encoded into density matrices and the kNN selects the k neighbors evaluating an operator distance among quantum states. In the latter case, we consider the *Bures distance* that is a quantum generalization of the Fisher information, it is defined by:

$$d_B(\rho_1, \rho_2) = \sqrt{2\left(1 - \sqrt{\mathcal{F}(\rho_1, \rho_2)}\right)}, \tag{20}$$

where the fidelity between density operators is given by $\mathcal{F}(\rho_1, \rho_2) = \left(\mathrm{tr}\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\right)^2$. Let us note that the fidelity reduces to $\mathcal{F}(\rho_1, \rho_2) = \langle\psi_1|\rho_2|\psi_1\rangle$ when $\rho_1 = |\psi_1\rangle\langle\psi_1|$. Therefore the Bures distance between the pure state ρ_1 and the arbitrary state ρ_2 can be expressed in term of the Bloch representation as follows:

$$d_B(\rho_1, \rho_2) = \sqrt{2\left(1 - \sqrt{\frac{1}{d}\left(1 + (d-1)\mathbf{b}^{(1)} \cdot \mathbf{b}^{(2)}\right)}\right)} \equiv D_B\left(\mathbf{b}^{(1)}, \mathbf{b}^{(2)}\right), \tag{21}$$

where $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$ are the Bloch vectors of ρ_1 and ρ_2 respectively and d is the dimension of the Hilbert space of the quantum encoding. The special form of the Bures distance, expressed in terms of Bloch vectors as in (21), is relevant for our purpose because data vectors are encoded into pure states and the quantum centroids are calculated as Bloch vectors of mixed states in general.

In Algorihtm 1, the locality is imposed by running the kNN on the input space finding the training vectors that are closest to the test point, then there is the quantum encoding into pure states and a quantum-inspired classifier (Helstrom, PGM, geometric Helstrom) is locally executed over the restricted training set. In Algorithm 2, the test point and all the training points are encoded into Bloch vectors of pure states then a kNN is run w.r.t. the Bures

distance to find the nearest neighbors in the space of the quantum representation, then a quantum-inspired classifier is executed with the training instances corresponding to the closest quantum states.

Algorithm 1 *Local quantum-inspired classification based on kNN in the input space before the quantum encoding. The classifier can be:* Helstrom, PGM, geometric Helstrom.

Require: Dataset X of labelled instances, unlabelled point $\hat{\mathbf{x}}$
Ensure: Label of $\hat{\mathbf{x}}$
find the k nearest neighbors $\mathbf{x}_1, ..., \mathbf{x}_k$ to $\hat{\mathbf{x}}$ in X w.r.t. the Euclidean distance
encode $\hat{\mathbf{x}}$ into a pure state $\rho_{\hat{\mathbf{x}}}$
for $j = 1, ..., k$ **do**
encode $\mathbf{x}_j$ into a pure state $\rho_{\mathbf{x}_j}$
end for
run the quantum-inspired classifier with training points encoded into $\{\rho_{\mathbf{x}_j}\}_{j=1,...,k}$.

Algorithm 2 *Local quantum-inspired classification based on kNN in the Bloch representation after the quantum encoding. The classifier can be:* Helstrom, PGM, geometric Helstrom.

Require: Dataset X of labelled instances, unlabelled point $\hat{\mathbf{x}}$
Ensure: Label of $\hat{\mathbf{x}}$
encode $\hat{\mathbf{x}}$ into a Bloch vector $\mathbf{b}^{(\hat{\mathbf{x}})}$ of a pure state
for $\mathbf{x} \in X$ **do**
encode $\mathbf{x}$ into a Bloch vector $\mathbf{b}^{(\mathbf{x})}$ of a pure state
end for
find the k nearest neighbors to $\mathbf{b}^{(\hat{\mathbf{x}})}$ in $\{\mathbf{b}^{(\mathbf{x})}\}_{\mathbf{x}\in X}$ w.r.t. the distance D_B
run the quantum-inspired classifier over the k nearest neighbors.

A local quantum-inspired classifier can be defined without quantum state discrimination but considering a *nearest mean classification* like the following: after the quantum encoding we perform a kNN selection and calculate the centroid of each class considering only the nearest neighbors to the test point, finally we assign the label according to the nearest centroid as schematized in Algorithm 3.

4 Conclusions

The present paper is focused on some methods of quantum-inspired machine learning, in particular classification algorithms based on quantum state discrimination. We adopted a geometric formulation in defining quantum encodings of classical data in terms of Bloch vectors of density operators as in the previous work [1]. The novel contribution of the present paper is the local

Algorithm 3 *Local quantum-inspired nearest mean classifier.*

Require: Training set X divided into n classes C_i, unlabelled point $\hat{\mathbf{x}}$
Ensure: Label of $\hat{\mathbf{x}}$
encode $\hat{\mathbf{x}}$ into a Bloch vector $\mathbf{b}^{(\hat{\mathbf{x}})}$ of a pure state
for $\mathbf{x} \in X$ **do**
encode $\mathbf{x}$ into a Bloch vector $\mathbf{b}^{(\mathbf{x})}$ of a pure state
end for
find the neighborhood $K = \{\mathbf{b}^{(\mathbf{x}_1)}, ..., \mathbf{b}^{(\mathbf{x}_k)}\}$ of $\mathbf{b}^{(\hat{\mathbf{x}})}$ w.r.t. the distance D_B
for $i = 1, ..., n$ **do**
construct the centroid $\widehat{\mathbf{b}}^{(i)} = \frac{2}{d|C_i^k|} \sum_{\mathbf{x} \in C_i^k} \mathbf{b}^{(\mathbf{x})}$ where $C_i^k := \{\mathbf{x} \in C_i : \mathbf{b}^{(\mathbf{x})} \in K\}$
end for
find the closest centroid $\widehat{\mathbf{b}}^{(l)}$ to $\mathbf{b}^{(\hat{\mathbf{x}})}$ w.r.t. the distance D_B
return label of the class C_l

Algorithm 4 *Quantum-inspired nearest mean local classifier.*

Require: n classes C_i of training points, unlabelled point $\hat{\mathbf{x}}$
Ensure: Label y of $\hat{\mathbf{x}}$
encode $\hat{\mathbf{x}}$ into a Bloch vector $\mathbf{b}^{(\hat{\mathbf{x}})}$ of a pure state
for $\mathbf{x} \in C_i$ **do**
encode $\mathbf{x}$ into a Bloch vector $\mathbf{b}^{(\mathbf{x})}$ of a pure state
end for
construct the centroids $\widehat{\mathbf{b}}^{(i)} = \frac{2}{d|C_i|} \sum_{\mathbf{x} \in C_i} \mathbf{b}^{(\mathbf{x})}$ for $i = 1, \ldots, n$
construct means of (all or $h = 2$?) $\binom{n}{k}$ for $k = 1, \ldots, h$) combinations $\widehat{\mathbf{b}}^{(1)}, \ldots, \widehat{\mathbf{b}}^{(n)}, \frac{1}{2}(\widehat{\mathbf{b}}^{(1)} + \widehat{\mathbf{b}}^{(2)}), \ldots, \frac{1}{h}(\widehat{\mathbf{b}}^{(1)} + \ldots + \widehat{\mathbf{b}}^{(h)})$
for combination **do**
find the k-nearest neighbors $\mathbf{b}^{(\mathbf{x})}$ and construct the local centroids $\widehat{\mathbf{b}}_l^{(j)}$;
end for
return label of the nearest local centroid (or after the pretty good measurement)

approach adopted to execute the classifier not over the entire training set but in a neighborhood of the test point selected by a kNN. In a future paper, we will present and discuss some empirical results for evaluating the impact of locality in quantum-inspired classification comparing the performances of the proposed algorithms to classical methods over benchmarks datasets.

Acknowledgments. This work was supported by Q@TN, the joint lab between University of Trento, FBK- Fondazione Bruno Kessler, INFN-National Institute for Nuclear Physics and CNR- National Research Council.

Declarations. Conceptualization, E.B, R.L. and D.P.; software, R.L.; validation, E.B., D.P.; formal analysis, D.P.; writing—original draft preparation, R.L. and D.P.; writing—review and editing, E.B, D.P. All authors have read and agreed to this version of the manuscript. The code is also at the following repository: **github.com/leporini/classification** (accessed on: August

30, 2022).
The authors declare no conflict of interest.

References

[1] Leporini, R., Pastorello, D.: An efficient geometric approach to quantum-inspired classifications. Scientific Reports vol 12, 1 (2022) https://doi.org/10.1038/s41598-022-12392-1

[2] Blanzieri, E., Melgani, F.: An adaptive svm nearest neighbor classifier for remotely sensed imagery. In: 2006 IEEE International Symposium on Geoscience and Remote Sensing, pp. 3931–3934 (2006). https://doi.org/10.1109/IGARSS.2006.1008

[3] Leporini, R., Pastorello, D.: Support vector machines with quantum state discrimination. Quantum Reports **3(3)** (2021). https://doi.org/10.3390/quantum3030032

[4] Leporini, R., Pastorello, D.: Quantum-Inspired Classification Based on Voronoi Tessellation and Pretty-Good Measurements **4(4)** (2022). https://doi.org/10.3390/quantum4040031

[5] Helstrom, C.W.: Quantum detection and estimation theory. J. Stat. Phys. **1**, 231–252 (1969). https://doi.org/10.1007/BF01007479

[6] Mochon, C.: Family of generalized pretty good measurements and the minimal-error pure-state discrimination problems for which they are optimal. Physical Review A **93(3)** (2006). https://doi.org/10.1103/PhysRevA.73.032328

[7] Bae, J.: Structure of minimum-error quantum state discrimination. New Journal of Physics **96(7)** (2013). https://doi.org/10.1088/1367-2630/15/7/073037

[8] Kimura, G., Kossakowski, A.: The bloch-vector space for n-level systems: the spherical-coordinate point of view. Open Systems & Information Dynamics **12(3)** (2005). https://doi.org/10.1007/s11080-005-0919-y

[9] Sergioli, G., Giuntini, R., Freytes, H.: A new quantum approach to binary classification. PLoS ONE **14(5)**, 0216224 (2019). https://doi.org/10.1371/journal.pone.0216224

[10] Giuntini, R., Freytes, H., Park, D.K., Blank, C., Holik, F., Chow, K.L., Sergioli, G.: Quantum state discrimination for supervised classification. arXiv:2104.00971v1 (2021)

[11] Fix, E., Hodges, J.L.: Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties