RESEARCH ARTICLE

Classification of data with a qudit, a geometric approach

Aikaterini Mandilara[1,2] · Babette Dellen[3] · Uwe Jaekel[3] · Themistoklis Valtinos[1] · Dimitris Syvridis[1,2]

Abstract

We propose a model for data classification using isolated quantum d-level systems or else qudits. The procedure consists of an encoding phase where classical data are mapped on the surface of the qudit's Bloch hyper-sphere via rotation encoding, followed by a rotation of the sphere and a projective measurement. The rotation is adjustable in order to control the operator to be measured, while additional weights are introduced in the encoding phase adjusting the mapping on the Bloch's hyper-surface. During the training phase, a cost function based on the average expectation value of the observable is minimized using gradient descent thereby adjusting the weights. Using examples and performing a numerical estimation of lossless memory dimension, we demonstrate that this geometrically inspired qudit model for classification is able to solve nonlinear classification problems using a small number of parameters only and without requiring entangling operations.

1 Introduction

With low-depth quantum circuits coming to pass, the interest for devising applications for these physical units has much increased. One fast-developing direction that already forms a sub-discipline of Quantum Machine Learning (Dunjko and Briegel 2018; Biamonte et al. 2017) is devising methods for

✉ Aikaterini Mandilara
mandkat@di.uoa.gr

Babette Dellen
dellen@hs-koblenz.de

Uwe Jaekel
jaekel@hs-koblenz.de

Themistoklis Valtinos
sdi0600015@di.uoa.gr

Dimitris Syvridis
dsyvridi@di.uoa.gr

[1] Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis, Ilisia 15784, Greece

[2] Eulambia Advanced Technologies Ltd, Agiou Ioannou 24, Ag. Paraskevi 15342, Greece

[3] Department of Mathematics and Technology - RheinAhrCampus Remagen, University of Applied Sciences Koblenz, Joseph-Rovan-Allee 2, Remagen 53424, Germany

addressing problems of classical machine learning (ML) with variational quantum circuits (VQCs) (Cerezo et al. 2021). These types of quantum circuits have adjustable angles in gates which can be trained in a fashion analogous to neural networks (Schuld et al. 2020; Farhi and Neven 2018; Pérez-Salinas et al. 2020; Havlíček et al. 2019). On formal level though the mathematical analogy of VQCs with neural networks is far from straightforward, mainly due to the reversibility of VQCs, and the problem of quantum neuron is usually approached with more intrigued models (Schuld et al. 2014; Tacchino et al. 2019; Wan et al. 2017; Cao et al. 2017; Verdon et al. 2017; Torrontegui and García-Ripoll 2019). In addition to neural networks, VQCs show similarities with classical kernel machines (Schuld and Killoran 2019; Schuld 2021; Havlíček et al. 2019) by generating a feature map of classical data in the Hilbert space. In general, the interpretation and most profitable use of VQCs in ML tasks remains an open topic of discussion, including the accurate evaluation of their capacity and their potential or advantages compared to classical models.

This work aims to contribute to the question whether quantum circuits are suitable for solving ML tasks and how increasing the dimension of the Hilbert space can be exploited for this purpose. There are two paths to follow: One is to employ n entangled qubits achieving an exponential

increase of space, the other, less investigated path, is to employ qudits. For a single qudit, the dimension of the Hilbert space is increasing linearly with d and without requiring entangling operations which remain demanding on practical level. Our quantum toy model consists of a single qudit operated by a low-depth quantum circuit (which we call single layer). With these limited resources, we are able to show that with a proper encoding and adjustment of d with respect to the dimension of the input one may achieve double lossless memory (LM) dimension (Friedland and Krell 2018) as compared to habitual single-layer neural networks (NN) possessing the same number of trainable parameters. This effect cannot be achieved in the absence of parameters in the encoding phase controlling the feature map on the Bloch hyper-sphere.

Going one step further, while keeping the input dimension fixed, the capacity of the quantum system (in the sense of LM dimension) can be further increased by either re-uploading the data (Pérez-Salinas et al. 2020; Wach et al. 2023), this way introducing more depth into the quantum circuit, or, alternatively, as we propose in this work, to use higher dimensional quantum systems by increasing d. We get preliminary evidence that the two methods give comparable results and therefore the selection should be done in dependence of the available resources.

The structure of the manuscript is as follows. We start by introducing qudits first as physical entities in the lab and then as mathematical objects, i.e., as vectors on the Bloch hypersphere. Based on this representation, we develop a general scheme for mapping the data on its surface and rotating them. We then evaluate different encoding-rotation models according to LM dimension and we draw conclusions on optimal methodology. We illustrate the efficiency of qubit and qutrit models by applying them to standard classification problems including both synthetic and real-world data.

2 Qudits

A qudit stands for the state of a d-level quantum system just as a qubit describes a quantum 2-level system. The quantum degrees of freedom and systems habitually used as qubits, with appropriate modifications, are suitable to accommodate qudits in most of the cases. The theoretical interest in qudits (Wang et al. 2020) has driven a considerable number of experimental results on generation and manipulation of single and entangled qudits. While neutral multilevel Rydberg atoms (Deller et al. 2023; Weggemans et al. 2022) or molecular magnets seem the most natural candidates for qudit's realization, the majority of qudit's demonstrations in the context

of quantum information have been achieved with photons (Lapkiewicz et al. 2011; Erhard et al. 2018; Luo et al. 2019; Kues et al. 2017; Imany et al. 2019). There the qudit's state is realized by a single photon superposed over d modes which can be spatial, time, frequency or orbital angular momentum ones. With integrated photonics, on-chip generation of two entangled qudits has been reported for $d = 10$ (Kues et al. 2017) and $d = 32$ (Imany et al. 2019) giving good promises for scalability of photonic platforms with d. Superconducting circuits are yet another platform accommodating transmonic qutrits (Fedorov et al. 2012; Kunzhe et al. 2018; Blok et al. 2021) and recently a circuit of five qutrits has been demonstrated (Blok et al. 2021) able to perform teleportation. As the authors state in this work (Blok et al. 2021) these circuits could be possibly extended from qutrits to qudits by leveraging the intrinsic anharmonicity of transmons. Finally, trapped chains of $^{40}Ca^+$ ions with their natural multilevel structure have been tested (Ringbauer et al. 2022) as a platform of qudits with d up to 7.

Circuits made of qubits require entangling operations in order to exploit the exponentially large space. Two-qubits operations define the complexity of a circuit and experimentally form the most challenging part demanding either controllable interaction between qubits or in the case of quantum optical circuits nonlinearities which have weak effect. On the contrary, in this work we build our model on single qubits which only require coherent manipulation achievable in its full range in most of the aforementioned qudit platforms. For transmonic qutrits, single-qutrit gates forming a universal set have been demonstrated (Blok et al. 2021) showing high fidelity and in addition the readout of the states can be performed with high accuracy. Similarly for ion qudits (Ringbauer et al. 2022) full single-qudit control for $d = 3$ and 5 has been achieved. Concerning Rydberg atoms, there is an extensive experimental knowledge on the manipulating of their state via laser fields and one may see for instance (Weggemans et al. 2022) for specific prescriptions leading to full control of a fermionic strontium ^{87}Sr qudit, $d = 10$. Even though the highest d at the moment is achieved in integrated photonics platforms, there the full set of operations that our model demands for is less obvious. However, recent results (Kues et al. 2017) demonstrating gates on frequency modes, achieved using programmable filters and an electro-optic phase modulator, are highly promising, and open new pathways for the applications of qudits.

2.1 The Bloch hyper-sphere of a qudit

As a mathematical entity a qudit state "lives" in the d-dimensional Hilbert space which is spanned by the eigen-

states of the Hamiltonian of the system. Let us denote by $\{|k\rangle\}_{k=0}^{d-1}$ such a set of normalized eigenstates. Then one can express a generic qudit state

$$|\psi\rangle = \sum_{k=0}^{d-1} c_k |k\rangle \tag{1}$$

by d complex amplitudes c over this basis, being constrained by the normalization condition $\sum_{k=0}^{d-1} |c_k|^2 = 1$.

We claim a full $su(d)$ algebra for the system, spanned by $d^2 - 1$ generators $\{\hat{g}_i\}$ that can be chosen to be orthogonal with respect to the Hilbert-Schmidt product such that and $Tr\left(\hat{g}_i^{\dagger}\hat{g}_j\right) = G\delta_{i,j}$ with G a positive constant. For $d = 2$ these generators can be identified with the Pauli operators ($G = 2$) while for $k = 3$ and $G = 2$ with the Gell-Mann operators (see the Appendix). Extending the set $\{\hat{g}_i\}$ by an element $g_0 = \sqrt{\frac{G}{d}}\hat{1}$, the generators of the algebra form a basis in Hilbert-Schmidt space of Hermitian operators so that any observable $\hat{H}$ of the qudit can be written as

$$\begin{aligned}\hat{H} &= \sum_{m=0}^{d^2-1} h_m \hat{g}_m = h_0 \hat{g}_0 + \phi \sum_{m=1}^{d^2-1} n_m \hat{g}_m \\ &= h_0 \hat{g}_0 + \phi\, \vec{n}.\vec{\hat{g}}\end{aligned} \tag{2}$$

with $h_m = Tr\left(\hat{g}_m^{\dagger}\hat{H}\right)/G \in \Re$, $\vec{n} = \{n_1, n_2, \ldots, n_{d^2}\}$ a normalized real vector and ϕ an angle.

The density operator $\hat{\rho}$ of a pure state $|\psi\rangle$, $\hat{\rho} = |\psi\rangle\langle\psi|$, being a positively defined Hermitian matrix with $Tr(\hat{\rho}) = 1$, can also be decomposed on the basis of the generators as

$$\hat{\rho} = \frac{1}{G}\hat{g}_0 + \sum_{m=1}^{d^2-1} r_m \hat{g}_m \tag{3}$$

with $r_m = Tr\left(\hat{g}_m^{\dagger}\hat{\rho}\right)/G$ and $\vec{r} = \{r_1, r_2, \ldots, r_{d^2-1}\}$ proportional by $1/G$ factor to the unit-length Bloch vector, living on the $(d^2 - 2)$-dimensional surface of the so-called Bloch hyper-sphere. For completeness we note here that pure states occupy only a sub-manifold of this surface of dimension $d^2 - d$ while the rest of the surface corresponds to non-positive Hermitian matrices. Mixed states correspond to vectors inside the Bloch hyper-sphere.

Furthermore, since any unitary operation $\hat{U}$ is generated by a Hermitian matrix $\hat{H}$ as $\hat{U} = e^{i\hat{H}}$, in view of the decomposition (2), one can re-write $\hat{U} = e^{i\phi\, \vec{n}.\vec{\hat{g}}}$ up to a phase factor. The latter expression leads (with some extra work) to the interpretation of a unitary operation acting on a pure state

$\hat{U}|\psi\rangle$ (or $\hat{U}\hat{\rho}\hat{U}^{\dagger}$) as a rotation of Bloch vector around the $\vec{n}$-axis for an angle proportional to ϕ. One can also see that the most general unitary operation $U_{\vec{n}}(\phi)$ is parameterized by $d^2 - 1$ real parameters.

Measurable quantities on a qudit are described by Hermitian operators which, again in view of the decomposition presented in Eq. (2), define a direction on the Bloch hyper-sphere. In addition, the d eigenvectors of the observables, corresponding to d real different measurement outcomes, i.e., eigenvalues, are mutually orthogonal to each other and offer a separation of the Bloch hyper-surface into d adjacent segments of equal area, in absence of degeneracies.

3 Employing qudits for supervised classification tasks

Let us consider classical data consisting of n k-dimensional feature vectors $\{\vec{x}\}$, i.e., $\vec{x} = \{x_1, x_2, \ldots x_k\}$. Every data point belongs to one od M classes. A random subset of the data composed of l-elements ($l < n$), $\{\vec{x}\}_l$, is picked as the training set.

3.1 Quantum resources

For this problem, the required resource is a single qudit where $d^2 - 1 \geq k$ and $d^2 - 1 - k$ increasing with the complexity of the task. One should be able to perform the full $SU(d)$ group of operations on the qudit and in addition to measure a single observable $\hat{O}$. For simplicity, we assume the spectrum to be non-degenerate, yielding d distinct measurement outcomes. Since the classification is based on mean values of measurement outcomes, one should be able to perform experiments in identical conditions multiple times.

3.2 Encoding classical data

Let us now introduce $P = S + W$ adjustable weights that we separate into two groups: $\vec{s} = \{s_1, \ldots s_S\}$ and $\vec{w} = \{w_1, \ldots w_W\}$, with $S, W \leq d^2 - 1$.

In the first part, there is the encoding phase where the classical data, i.e., the elements of the vector $\vec{x}_i$, together with the adjustable weights $\vec{s}$, are "uploaded" on the qudit that is initially in its ground state:

$$\left|\psi_{\vec{x},\vec{s}}\right\rangle = \exp\left[ix_1\sum_{j\in A_1} s_j\hat{g}_j + \ldots ix_k\sum_{j\in A_k} s_j\hat{g}_j\right]|0\rangle\,. \tag{4}$$

where A_j implies different grouping of the generators with $A_j \cap A_k = 0$ being a suggestive condition. With $|0\rangle$ we

denote the ground state of the qudit. Overall, the angles and axis of rotation of the initial vector $|0\rangle$ are related to both classical data and adjustable weights $\vec{s}$ in an intrigued way, and the result of such encoding is a map from the Cartesian space, where the inputs are initially described (k-dimensional real vector $\vec{x}$), onto the surface of the $d^2 - 1$ dimensional hyper-Bloch sphere. Given the requirement $k \leq d^2 - 1$, we actually map the data onto a higher dimensional feature space characterized by the kernel

$$\left| \left\langle \psi_{\vec{x}_2, \vec{s}} \mid \psi_{\vec{x}_1, \vec{s}} \right\rangle \right|^2 \quad . \tag{5}$$

In the Appendix, we provide an explicit expression of the simplest kernel employed in this work, namely the qubit model A (see Table 1). Contrary to the usual rotation encoding consisting of successive rotations around orthogonal directions of the Bloch sphere, resulting in cosine kernels, the "combined" encoding of Eq. (4) results in more intrigued kernels. Naturally, the complexity of these kernels is increasing with d and k.

3.3 Rotating and measuring

After mapping the data onto the hyper-sphere, it is separated into M groups. A projective measurement of $\hat{O}$ observable, provides, with some probability which depends on the state Eq. (4), an outcome from the d values of its spectrum $\{o_1, \ldots, o_d\}$ (arranged in increasing order).

We take the habitual assumption that the whole procedure can be repeated many times in an identical way and use the mean value of $\left\langle \hat{O} \right\rangle$ that lies in the interval $[o_1, o_d]$ to divide the interval (equally or unequally) into M segments, classifying the data, i.e., $[o_1, y_1]$, $[y_1, y_2] \ldots [y_{M-1}, o_d]$. To get optimum results though one should be able to rotate $\hat{O}$ in order to "match" its orientation with the one of the mapped data on the hyper-surface. Alternatively, one can keep $\hat{O}$ intact and rotate $|\psi_{\vec{x}_l, \vec{s}}\rangle$. So, in this stage, one applies arbitrary rotations to the state vector carrying the classical information, yielding

$$\left| \psi_{\vec{x}_l, \vec{s}, \vec{w}} \right\rangle = \exp \left[i \sum_{j=1}^{d^2-1} w_j \hat{g}_j \right] \left| \psi_{\vec{x}_l, \vec{s}} \right\rangle, \tag{6}$$

and measures $\hat{O}$. Let us note that it is not always profitable in terms of capacity to keep all the weights w_j in Eq. (6) and some should be ignored or set zero so that $W \approx S$.

The whole "encode-rotate-measure" scheme is repeated many times in identical conditions until mean value for the measurement

$$\left\langle \hat{O} \right\rangle_{\vec{x}_l, \vec{s}, \vec{w}} = \left\langle \psi_{\vec{x}_l, \vec{s}, \vec{w}} \right| \hat{O} \left| \psi_{\vec{x}_l, \vec{s}, \vec{w}} \right\rangle \tag{7}$$

is obtained that classifies the data point $\vec{x}_l$ according to the choice of segmentation $\{y_i\}$ of the interval $[o_1, o_d]$ of mean values. The values y_i can be also adjustable in the same way the threshold values of perceptrons in neural networks can be variable and optimizable.

One may summarize the total scheme in the following diagram:

Table 1 Qubit models

Model	Encoding-Rotation	k	P=S+W	$\tilde{D}_{LM}$
A	$\exp\left[is_1\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_1\hat{g}_1\right]$	2	2	4
B	$\exp\left[is_1\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_1\hat{g}_1 + iw_2\hat{g}_2 + iw_3\hat{g}_3\right]$	2	4	7
C	$\exp\left[i\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_1\hat{g}_1 + iw_2\hat{g}_2 + iw_3\hat{g}_3\right]$	2	3	3
D	$\exp\left[is_1 x_1\hat{g}_1 + is_2 x_2\hat{g}_2\right]\exp\left[iw_1\hat{g}_1 + iw_2\hat{g}_2 + iw_3\hat{g}_3\right]$	2	5	6
E	$\exp\left[is_1\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_1\hat{g}_1\right]$ $\exp\left[is_2\left(x_1\hat{g}_2 + ix_2\hat{g}_3\right)\right]\exp\left[iw_2\hat{g}_2\right]$	2	4	8
F	$\exp\left[is_1\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_1\hat{g}_1\right]$ $\exp\left[is_2\left(x_1\hat{g}_2 + ix_2\hat{g}_3\right)\right]\exp\left[iw_2\hat{g}_2\right]$ $\exp\left[is_3\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_3\hat{g}_1\right]$	2	6	9
G	$\exp\left[is_1\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_1\hat{g}_1\right]$ $\exp\left[is_2\left(x_1\hat{g}_2 + ix_2\hat{g}_3\right)\right]\exp\left[iw_2\hat{g}_2\right]$ $\exp\left[is_3\left(x_1\hat{g}_1 + ix_2\hat{g}_2\right)\right]\exp\left[iw_3\hat{g}_1\right]$ $\exp\left[is_4\left(x_1\hat{g}_2 + ix_2\hat{g}_3\right)\right]\exp\left[iw_4\hat{g}_2\right]$	2	8	11
H	$\exp\left[is_1\left(x_1\hat{g}_1 + ix_2\hat{g}_2 + ix_3\hat{g}_3\right)\right]\exp\left[iw_1\hat{g}_1 + iw_2\hat{g}_2\right]$	3	3	6
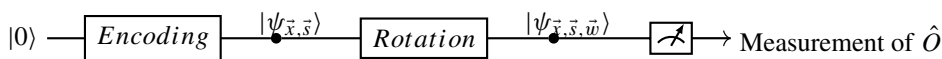
Finally, while the full scheme could be written as

$$\left| \psi_{\vec{x},\vec{s},\vec{w}} \right\rangle = \exp\left[i\hat{H} \right] |0\rangle \,, \tag{8}$$

it is important to note that $\hat{H}$ is highly nonlinear in the input $\vec{x}$ due to BCH formula. Our scheme, in contrast, relies on a simple linear encoding of Eq. (4).

3.4 Training

To perform the training, we define a loss function that penalizes misclassified data of the training set

$$E = \sum_{i \in T} \left(\left\langle \hat{O} \right\rangle_i - Y_i \right)^2 \,, \tag{9}$$

while correctly classified data do not contribute to its value. Here, T is the set of misclassified data of the training set, and Y_i is the upper or lower value of the spectral segment that characterizes correct class for the ith point. In Sect. 6.3 we use for convenience the cross entropy loss function.

The optimization of parameters implies a minimization of E, which is achieved (in all analysis apart from the examples in Sect. 6.3) by gradient descent. The landscape of E though contains a number of local minima, and, when starting from a random initial point in the space of parameters $\vec{s} \vee \vec{w}$, the procedure might get trapped in one of those. To improve minimization, we use a sample of $l = 50$ initial points and we pick the best result among all runs of gradient descent. When dealing with real-world data using a qutrit (in Sect. 6.3) and comparing its outcome the one obtained with classical models, a more advanced stochastic gradient descent is applied.

4 Lossless memory dimension of different encoding-rotation models

In this section, we compare different models of encoding using a measure of capacity with clear theoretical meaning that is also suitable for numerical evaluation. Our aim is not to accurately compare with the capacity of classical neural networks (Wright and McMahon 2020), but to identify optimum way for introducing the trainable parameters in encoding and rotating stages, Eqs. (4) and (6), of the proposed scheme. Due to limited computational capacity our numerical tests are not "exhaustive" but indicative.

We are employing a recently suggested measure (Friedland and Krell 2018), which has been constructed for evaluating the informational/memory capacity of multi-layered classical neural networks, the so-called LM dimension. This is a generalization of the Vapnik-Chervonenkis (VC) dimension (Vapnik and Chervonenkis 1971) that is based on the work of MacKey (Mackay 2003), embedding the memory capacity into the Shannon communication model. The definition of LM dimension (Friedland and Krell 2018) is the following:

- *The LM dimension D_{LM} is the maximum integer number D_{LM} such that for any dataset with cardinality $n \leq D_{LM}$ and points in random position, all possible labelings of this dataset can be represented with a function in the hypothesis space.*
- A set of points $\{x_n\}$ in K-dimensional space is in random position, if and only if from any subset of size $< n$ it is not possible to infer anything about the positions of the remaining points.

For this measure, the authors showed analytically that the upper limit of LM dimension scales linearly with the number of parameters in a classical neural network with a factor of proportionality that is the unity. In practice, a training method cannot be perfect, therefore this linear dependence persists with a lower factor of proportionality, For more details and the informational meaning of this measure we refer the interested reader to the original work of Friedland and Krell (2018).

For quantum models where analytical calculations are not available, we proceed with the numerical evaluation of LM dimension, which we denote as $\tilde{D}_{LM}$. Naturally $\tilde{D}_{LM}$ lower bounds D_{LM} and this can be understood from the procedure that we follow for each encoding-rotation model under test:

- We set the k-dimension of the inputs of the model. According to our general model for a qudit, we have $k \leq d^2 - 1$. We generate a set of $\{x_n\}$ points in random position, which we call *random pattern*, by selecting each of k coordinates from a uniform distribution in the interval $[-0.5, 0.5]$. We start with a $n \leq P$ where $P = S + W$ the total number of parameters.
- According to the definition of LM dimension, we treat only binary classification tasks, and we attribute labels randomly to the vectors of the random pattern to two groups. For a given random pattern, one should test all

2^n different labelings, but not having the computational capacity for this, when $n > 6$ we perform our estimate by taking a sample of 50 different random labelings.

- If the training of parameters via gradient descent, with 50 different starting points, does not lead to classification of the vectors of the random pattern into two groups with 100% success ratio, we repeat for other random patterns $\{x_n\}$ until we find a pattern that is successfully classified for all possible labelings. However, we do not exceed the number of 10 different random patterns under test.
- The number n is step-wised increased up to the point where the classification is no longer successful for any tested random pattern. The empirical LM dimension, $\tilde{D}_{LM}$ is the highest n where the classification is achieved for at least one random pattern (all possible labelings).

In Table 1 we present the results on the empirical estimation of LM dimension for a qubit. The generators of the algebra, $\hat{g}_i$ for a qubit system, are identified as the Pauli operators: $\hat{g}_1 = \hat{\sigma}_x$, $\hat{g}_2 = \hat{\sigma}_y$, $\hat{g}_3 = \hat{\sigma}_z$. For all qubit schemes under study the classification is performed by measuring the operator $\hat{g}_3$ with eigenvalues $\{-1, 1\}$ and corresponding eigenvectors $\{|1\rangle, |0\rangle\}$. The two groups of data are separated according to $\langle \hat{g}_3 \rangle \gtrless 0$. For comparison with our *single-layer* model, we have also included models ($E-F$ Table 1) which implement re-uploading of input data (Pérez-Salinas et al. 2020; Wach et al. 2023).

We proceed with the estimation of $\tilde{D}_{LM}$ for a qutrit with results presented in Table 2. The generators $\hat{g}_i$ of the $SU(3)$ group can be chosen to be the Gell-Mann operators, $\hat{\lambda}_i$ with $i = 1, \ldots, 8$, which are provided in matrix form in the Appendix. According to Sect. 3, during the encoding phase the classical data are mapped onto the Bloch hyper-sphere of a qutrit embedded in the 8-dimensional space which cannot be visualized. To obtain a partial visualization, as, for example, in Sect. 6.1.1, we use the Bloch ball representation offered by the $su(2)$ subalgebra of $su(3)$, spanned by the generators $\left\{\hat{L}_x = \hat{\lambda}_1 + \hat{\lambda}_6,\ \hat{L}_y = \hat{\lambda}_2 + \hat{\lambda}_7, \hat{L}_z = \hat{\lambda}_3 + \sqrt{3}\hat{\lambda}_8\right\}$.

For all schemes, we choose to measure the operator $\hat{L}_z = \hat{\lambda}_3 + \sqrt{3}\hat{\lambda}_8$ that is diagonal in the computational basis and with uniform spectrum $\{-2, 0, 2\}$. For binary classification results, as shown in Table 2, we separate the two groups according to the sign of $\langle \hat{L}_z \rangle$.

With regard to efficiency, the single-layer schemes that achieve $\tilde{D}_{LM} = 2P$ can be considered as the most successful ones, i.e., qubit: A, H, qutrit: $D2$. From the qubit models C, D and qutrit model $D1$ we may conclude that both the absence and the excessive input of parameters in the encoding phase are not recommended. We also observe that the most successful single-layer models are the ones where $k \approx d^2 - 1$.

For classical neural networks, for a fixed input dimension k, one can linearly augment LM dimension with the number of parameters by adding hidden layers (Friedland and Krell 2018). For the model presented here, this becomes possible by using a qudit system where $k < d^2 - 1$. The scaling $\tilde{D}_{LM} = 2P$ is not maintained but one rather achieves $\tilde{D}_{LM} \approx P$ as it is shown with $k = 2$ with qutrit model B. An alternative way to increase LM dimension is to use re-uploading, see qubit models $E - G$, but there the scaling $\tilde{D}_{LM} = 2P$ also is not achieved but rather $\tilde{D}_{LM} = P + L$, with L being a constant. This analysis confirms the findings in Wright and McMahon (2020) and underlines the need for more research in identifying quantum models which exceed the classical limits.

Finally, for single-layer models and $k \approx d^2 - 1$, we see that $\tilde{D}_{LM}$ is higher than the one for the classical neural network. It would be interesting to see whether more exotic classical perceptron models such as product-units (Durbin and Rumelhart 1989; Dellen et al. 2019) or complex-valued perceptron (Dellen and Jaekel 2021) exhibit similar augmentation of LM dimension. In addition we underline the fact that the LM dimension only captures a specific aspect of the model. For a complete evaluation of the quantum model for supervised learning task, other aspects (Abbas et al. 2021) would have to be taken into account, e.g., difficulty in training (barren plateaus problem) and presence of noise in implementation.

Table 2 Qutrit models

Model	Encoding-Rotation	k	P=S+W	$\tilde{D}_{LM}$
A	$\exp\left[ix_1s_1\hat{g}_6 + ix_2s_2\hat{g}_7\right]\exp\left[iw_1\hat{g}_1 + iw_2\hat{g}_4\right]$	2	4	6
B	$\exp\left[ix_1(s_1\hat{g}_1 + s_2\hat{g}_2) + ix_2(s_3\hat{g}_3 + s_4\hat{g}_4)\right]$			
	$\exp\left[i\sum_{j=1}^4 w_j\hat{g}_j\right]$	2	8	8
C	$\exp\left[ix_1(s_1\hat{g}_5 + s_2\hat{g}_6) + ix_2(s_3\hat{g}_7 + s_4\hat{g}_8)\right]$			
	$\exp\left[i\sum_{j=1}^4 w_j\hat{g}_j\right]$	2	8	7
D1	$\exp\left[is_1\sum_{j=1}^8 x_j\hat{g}_j\right]\exp\left[i\sum_{j=1}^7 w_j\hat{g}_j\right]$	8	8	13
D2	$\exp\left[i\sum_{j=1}^8 x_js_{1+j\bmod 4}\hat{g}_j\right]\exp\left[i\sum_{j=1}^4 w_j(\hat{g}_j + \hat{g}_{j+4})\right]$	8	8	16
D3	$\exp\left[i\sum_{j=1}^8 s_jx_j\hat{g}_j\right]\exp\left[iw_1\hat{g}_1\right]$	8	9	17

The conclusions of the numerical studies on LM dimension are illustrated with examples in next sections. Since LM dimension only concerns capacity of binary classification tasks, we address classification problems with $M > 2$ classes as well.

5 Classification problems treated with a qubit

We start the illustration of the suggested method by addressing two typical classification problems with a qubit ($d = 2$). Even though the power of a qubit has been extensively studied in the literature, this is the first example showing that a qubit can be logically complete, i.e., it is able to implement all binary logical functions. This is achieved with the model A, Table 1, which contains two real parameters. This outcome does not come as surprise since model A has LM dimension $\tilde{D}_{LM} = 4$ for $k = 2$, or, in other words, can shatter all possible ways four 2-dimensional vectors in random positions.

5.1 Binary logical functions

Let us consider four datasets on a plane ($k = 2$), as shown in Fig. 1(a). The logical functions for these noisy data correspond to different attributions of each dataset to one of two groups, A and B. For instance, the XOR function requires a classification of the datasets as in Fig. 1(a).

To implement classification according to the logical functions, we first map the data onto the 2-dimensional surface of the Bloch sphere. Even if the feature space has the same dimension as the initial space, the change in topology proves to be helpful. Numerical tests show that all logical functions can be implemented this way with 2 real weights ($S = 1$ and $W = 1$). In more details, we use the encoding and rotation as in model A for a qubit, see Table 1, and the classification is conducted using the sign of $\langle \hat{g}_3 \rangle$.

We successfully solved classification problems for all logical functions (AND, OR, XOR); however, we present in Fig. 1(b) only the results about XOR, which is the most challenging task, since it is a nonlinearly separable problem. The total number of data is 2000 and we use 4% of them for the training. A success ratio of classification of 100% was readily achieved.

It is important to note here that all binary logical functions can be solved with 2 real parameters also by the complex perceptron model presented in Dellen and Jaekel (2021). We proceed with an example that it is not solvable with any single-layer classical perceptron model up to our knowledge.
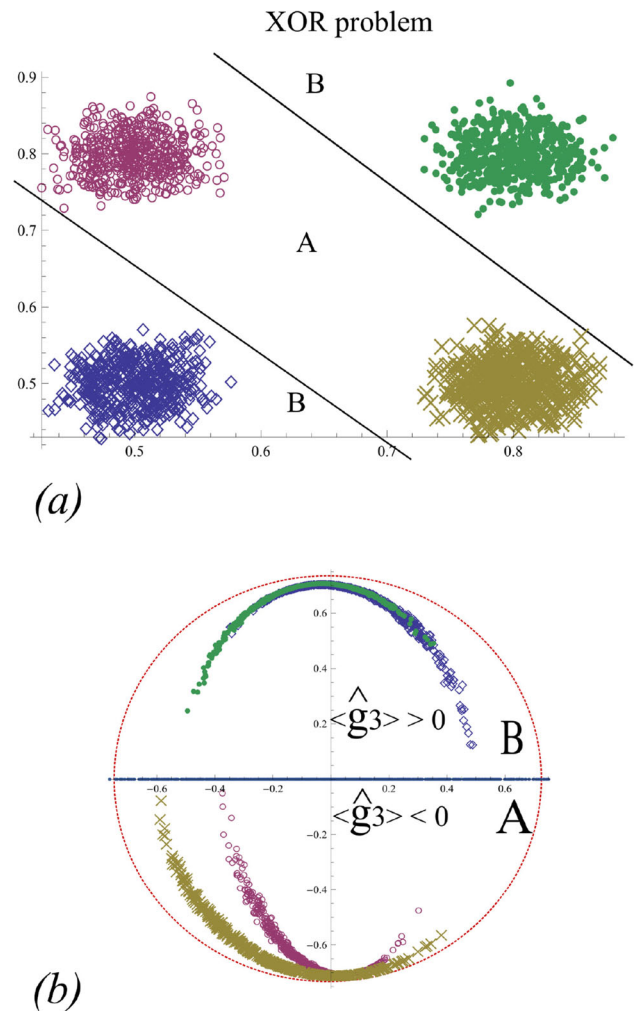

(a)

(b)

Fig. 1 *(a)* Data to be classified according to XOR logical function, into groups A and B. *(b)* The classified data mapped on the surface of Bloch sphere (projection on the x-z plane) after training on the 2 weights has been performed

5.2 Classification for circular boundaries

We proceed with a more complex classification problem and show that it can still be tackled with a single qubit. For this purpose, we employ model B, Table 1, because it achieves a higher LM dimension than model A.

The problem consists of classifying the data (1000 2-dimensional vectors) in Fig. 2(a) into two groups. In Fig. 2(b), we present the classification achieved on the Bloch sphere after the weights s_1, w_1, w_2, w_3 have been optimized. The classification ratio achieved is 100% using 10% of total dataset as training dataset.

Following the same encoding-rotation scenario (model B), we are able to treat elliptical data (not presented here), but with a lower final classification ratio ($\approx 90\%$).
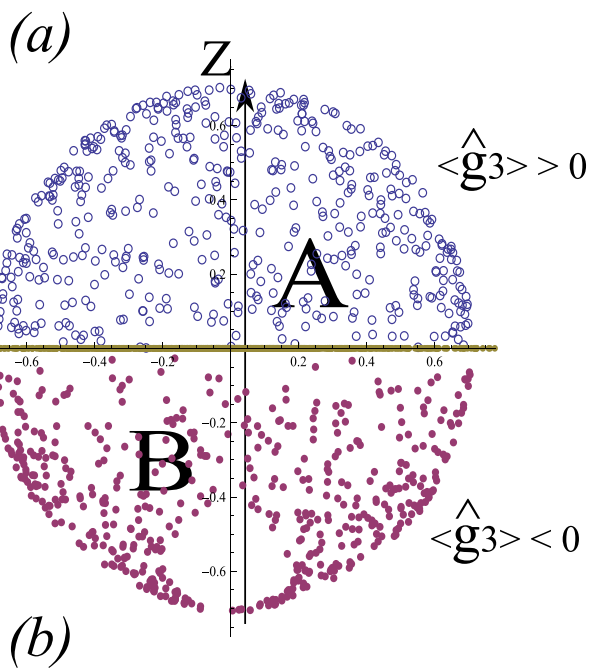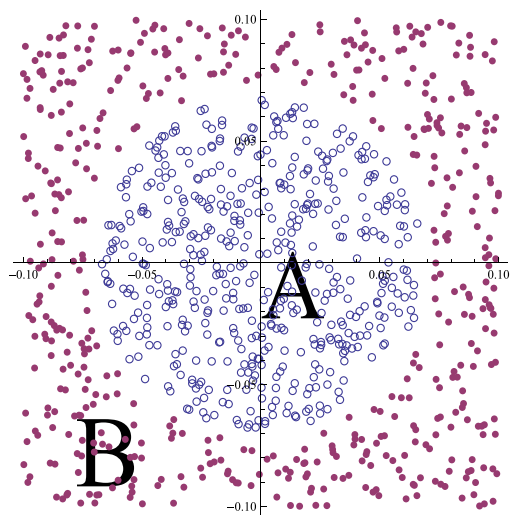
Circular boundaries problem

(a)

(b)

Fig. 2 *(a)* The initial data (1000 points) to be classified into groups A and B. *(b)* The data are mapped on the Bloch surface and perfectly classified

6 Examples solved with a qutrit

Even though we have been able to solve a couple of basic classification problems with one qubit, it is obvious that one needs a higher dimensional space d to resolve more complicated problems since one qubit can accommodate at most 5 parameters/weights according to our single-layer model. As shown in Sect. 4 qutrits may accommodate more parameters and therefore achieve higher LM dimension. In addition,

tests have shown that qutrit models perform better than qubit models for classification tasks into $M > 2$ groups. This is not obvious studying LM dimension alone.

6.1 Noisy XOR

We first investigate the binary classification task presented in Fig. 3(a) for which all qubit models exhibited low performance but where qutrit's model B, see Table 2, gives adequate results. More specifically, we use 1% of the total data (2000 points) for training and achieve a success classification ratio of 96%.

6.1.1 Classification into three groups and a geometric picture

We increase the difficulty of the previous problem by demanding classification in 3 groups of data and reducing the margins between sets, as shown in Fig. 4(b). We use a comparable number of weights (9), but now the encoding-rotation model is:

- Encoding via

$$\left|\psi_{\vec{x},\vec{s}}\right\rangle = \exp\left[ix_1(s_1\hat{g}_3 + s_2\hat{g}_5 + s_3\hat{g}_7)\right.$$
$$\left. + ix_2(s_4\hat{g}_4 + s_5\hat{g}_6 + s_6\hat{g}_8)\right]|0\rangle \quad . \quad (10)$$

- Rotation via

$$\left|\psi_{\vec{x}_l,\vec{s},\vec{w}}\right\rangle = \exp\left[i\sum_{j=1}^{3} w_j\hat{L}_j\right]\left|\psi_{\vec{x}_l,\vec{s}}\right\rangle \quad (11)$$

where $\hat{L}_1 = \hat{L}_x$, $\hat{L}_2 = \hat{L}_y$, $\hat{L}_3 = \hat{L}_z$.
- Measurement of $\hat{L}_z$ and classification by comparing the value of $\left\langle\hat{L}_z\right\rangle$ with $A : [-2, -2 + 4/3]$, $B : [-2 + 4/3, 2 - 4/3]$, $C : [2 - 4/3, 2]$.

Using 4% of the total data (2000 points) for training, a success ratio of classification 87% is achieved for the rest of data. In Fig. 4, we depict the mapping (with optimized parameters) of the data into the $SU(2)$ Bloch ball generated by $\left\{\hat{L}_x, \hat{L}_y\hat{L}_z\right\}$ operators. The classification "intervals" for $\left\langle\hat{L}_z\right\rangle$ are also presented in the picture as horizontal lines. This "local" picture offered by the subgroup is equivalent to the picture one would obtain by inspecting the local density matrix of an entangled system. One can thus claim by borrowing terms by the notion of generalized entanglement (Barnum et al. 2004) that the self-entanglement of a qutrit has the same use in the classification procedure as physical entanglement between subsystems, i.e., this extends the mapping from the
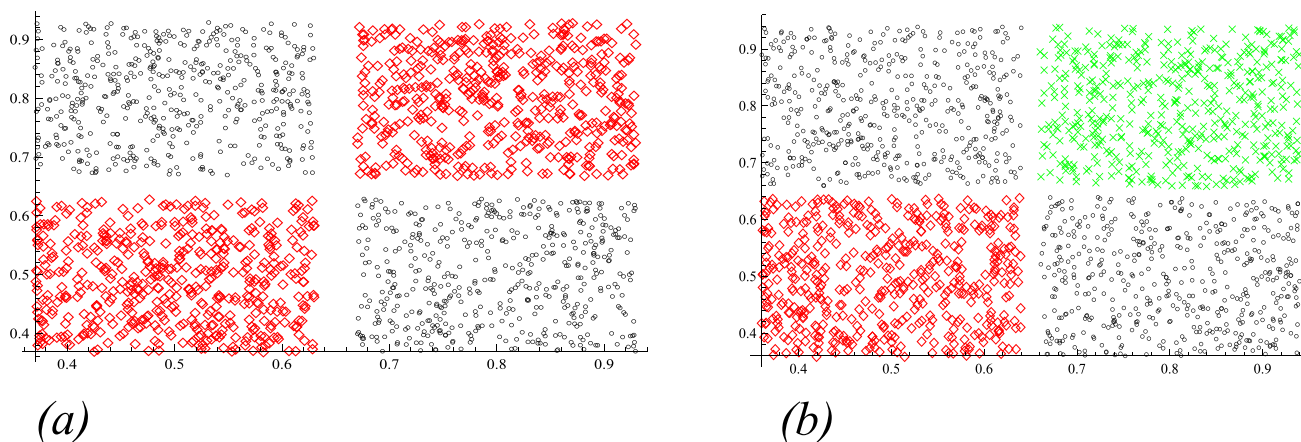
Fig. 3 Qutrit model: *(a)* classification into 2 groups employing 8 weights, *(b)* classification into 3 groups employing 9 weights

surface to the inside area of the Bloch hyper-sphere of a sub-system. The generation of self-entanglement in a qudit does require the ability to fully operate the system but in practice this is less demanding than the entangling interaction between subsystems.

6.2 Classifying moon sets with a qutrit

Finally, by using qutrit model C of Table 2, we attempt a common classification task, the one of moon sets. By optimizing the 8 parameters of the model, we achieve a classification ratio of 90% using 10% of 800 total data points. In Fig. 5,

we present $\langle \hat{L}_z \rangle$ for the optimized set of parameters, together with the datasets.

6.3 Real-world data multi-class classification

We will now turn to multi-class classification tasks using real-world data and more advanced methods of training. We use datasets from the UCI Machine Learning Repository, a widely used and publicly available repository (Dua and Graff 2017), maintained by the University of California, Irvine. Our aim is to explore the feasibility of using a single qutrit to accurately distinguish between three classes in datasets with more than two dimensions, such as the Iris and Wine
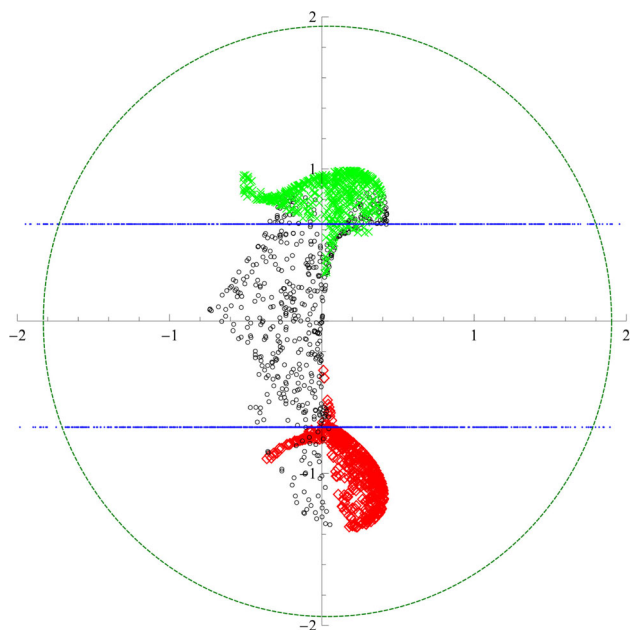

Fig. 4 The classification of data of Fig. 4(b) into 3 groups as perceived in the $SU(2)$ Bloch sphere representation provided by the operators $\left\{ \hat{L}_x, \hat{L}_y, \hat{L}_z \right\}$
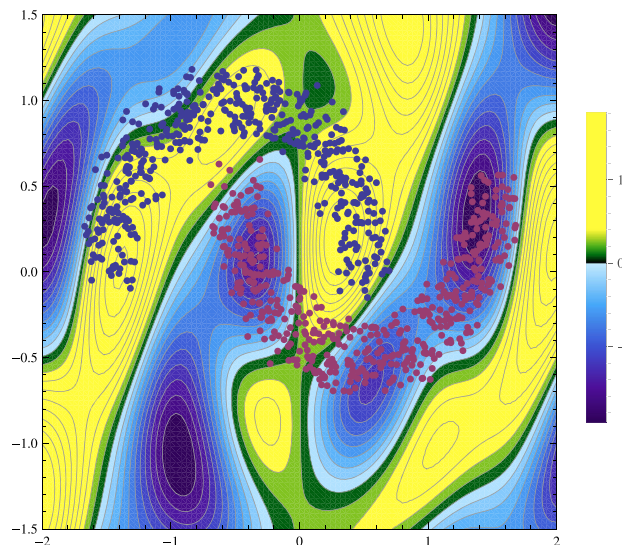

Fig. 5 Classification of moon sets with the qutrit model C using 8 weights. A contour plot of $\langle \hat{L}_z \rangle$ is depicted together with the moon data after optimization has been performed

Table 3 Real-world data treated with a qutrit

Dataset	Features	Classes	Samples	Pre-processing
IRIS	4	3	150	none
WINE	13	3	178	PCA

datasets. Our results illustrate that supervised learning in the context of less structured data is achievable.

The Iris dataset consists of 150 samples of iris flowers, with measurements of four features: sepal length, sepal width, petal length, and petal width. Each sample is labeled with one of three possible iris species. The Wine Cultivars dataset consists of measurements of thirteen chemical constituents found in three different wine cultivars. The objective is to classify the cultivar of the wine based on the chemical composition measurements.

Our aim is to use the same encoding and number of parameters for both datasets. Thus for the Wine Cultivars dataset which possesses 13 different features we employ Principal Component Analysis (PCA) (Hotelling 1933) in order to reduce the number of features to four (Table 3).

The encoding and rotating scheme that we follow is

- Encoding via

$$\left|\psi_{\vec{x}_l,\vec{s}}\right\rangle = \exp\left[is\sum_{j=1}^{4} x_j\,\hat{g}_j\right]|0\rangle \quad . \tag{12}$$

- Rotation via

$$\left|\psi_{\vec{x}_l,\vec{s},\vec{w}}\right\rangle = \exp\left[i\sum_{j=5}^{8} w_{j-4}\,\hat{g}_8\right]\left|\psi_{\vec{x}_l,\vec{s}}\right\rangle \tag{13}$$

where the variational weights $\vec{w} = (s, w_1, w_2, w_3, w_4)$ are the parameters to be optimized.

We ensured an equal representation of each class. For reproducibility, we used the same seed to split the data into train and test sets. To avoid overfitting, early stopping was employed and different gradient-based methods were trialed to combat the barren plateaus problem before settling to stochastic gradient descent (SGD) (Schuld et al. 2019) using the parameter-shift rule. This method reduces the number of measurements needed during implementation compared to the standard method, making it more efficient and practical for quantum machine learning. SGD is a variant of gradient descent that randomly selects a subset of data points, called mini-batch, to calculate the gradient of the cost function at each iteration. Since these are multi-class problems categorical cross entropy loss was used as a cost function. Using

this approach, we achieved competitive scores with a single qutrit as can be seen in Table 4.

Since these are multi-class problems categorical cross entropy was used as cost function, which combines the softmax activation and the negative log likelihood loss as follows:

$$\text{loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} t_{ij}\,\log(p_{ij}) \ . \tag{14}$$

Here N is the number of samples, C is the number of classes, t_{ij} represents the true label for sample i and class j, and p_{ij} represents the predicted probability.

Using this approach, we achieved competitive scores with a single qutrit as can be seen in Table 4. In these benchmarks, we present the results of the single-qutrit model against a classical machine learning model using Support Vector Machines (SVM) and a Variational Quantum Classifier (VQC) model with entangled qubits in Qiskit. These tests were conducted using four qubits and the popular ZZ feature map with twelve parameters, utilizing the Limited-memory Broyden-Fletcher-Goldfarb-Shanno Bound (L-BFGS-B) optimizer to minimize sensitivity to local minima and the barren plateau issue (Liu and Nocedal 1989).

These results showcase that even a single-qudit classifier is capable of multi-class classification for multi-dimensional real-world data. Although on the Iris dataset the four-qubits model outperformed the single-qutrit model, the single-qutrit model produced better results even with five parameters compared to the twelve used by the ZZ feature map on the Wine dataset. Increasing the encoding layers could further enhance the classifier's performance, but since the aim of our study was to demonstrate that a single-layer qudit classifier can accurately distinguish between multiple classes, it is not further investigated here.

7 Discussion

Qudits are extensions of qubit units to higher dimensions, which can enhance the performance in quantum computing (Gokhale et al. 2019; Pavlidis and Floratos 2021; Gedik et al.

Table 4 Comparative numerical studies for classification of Iris and Wine data

	Classical model	Entangled qubits	Single qutrit
Iris TrSA	97.5%	85%	86.67%
Iris TeSA	100%	86.6%	84.44%
Wine TrSA	97.9%	65.5%	77.42%
Wine TeSA	91.7%	69.4%	85.19%

The train set accuracy (TrSA) and test set accuracy (TeSA) reached with different methods

2015; Kunzhe et al. 2018; Wang et al. 2020), gate decomposition (Fedorov et al. 2012), error correction (Rosenblum 2018), communication (Cozzolino et al. 2019; Sheridan and Scarani 2010; Amblard and Arnault 2015) and variational algorithms (Deller et al. 2023; Weggemans et al. 2022). These are experimentally realizable with different physical models and recent proposals also use them in quantum machine learning (Wach et al. 2023; Useche et al. 2022). In this work, we have described a model for data classification using a single qudit. The parametrization is introduced according to geometric intuition, partially for controlling the mapping on the Bloch hyper-surface and partially for adjusting the projective measurement to the dataset's orientation on the Bloch hyper-sphere. Entangling or adding more layers can certainly enhance the quantum classifier, similar to how classical neural networks yield better results with increased depth. Nonetheless, given the expense and error-prone nature of entangling in near-term quantum hardware, our results indicate that even a low-depth single-qudit classifier holds a promise for quantum machine learning, if it is thoughtfully employed with a balanced distribution of parameters in the encoding and rotating steps.

The simple model that we present shares obvious similarities and borrows ideas from previous works (Pérez-Salinas et al. 2020; Farhi and Neven 2018; Havlíček et al. 2019; Wach et al. 2023). Being though only in the mid-way of exploration of the potential role of quantum systems for ML tasks, this geometrically dressed entanglement-free proposal gives its own contribution, connecting current efforts with the geometry of Hilbert-Schmidt space and underlying the equivalence of self-entanglement (Barnum et al. 2004) with physical one in practice. In addition, with the help of empirical estimation of LM dimension for a qubit and a qutrit, we have been able to demonstrate that the "capacity" of single-layer quantum systems can be higher than for classical neural network systems

bearing the same number of training parameters. It remains an open question for future work to investigate and compare the capacity of the quantum model with more intrigued single-layer classical perceptron models but also to investigate whether quantum multi-layer structures can exist which can keep the advantage in LM dimension over classical NN.

Appendix A: Qubit kernel

There is the common belief that the encoding of classical data via rotation angles results in a simple cosine kernel, that is easily classically reproducible. However, cosine kernels only emerge when the rotation encoding is successive as

$$\exp\left[is x_1 \hat{g}_1\right] \exp\left[is x_2 \hat{g}_2\right] \ldots |0\rangle \quad , \tag{15}$$

or concerns a setting of non-interacting qubits,

$$\exp\left[is x_1 \hat{g}_1\right] |0\rangle \otimes \exp\left[is x_2 \hat{g}_2\right] |0\rangle \ldots \quad . \tag{16}$$

In this work we use rotation encoding that looks very similar

$$\left|\psi_{\vec{x},s}\right\rangle = \exp\left[is(x_1 \hat{g}_1 + x_2 \hat{g}_2)\right] |0\rangle \tag{17}$$

staying linear in the input $\vec{x}$ but whose corresponding kernel is more intrigued than cosine one due to BCH formula. Straightforward calculations show that for qubit model A the kernel writes as

$$
\begin{aligned}
\left|\langle \psi_{\vec{x},s} | \psi_{\vec{y},s}\rangle\right|^2 = \frac{1}{8x^2y^2} \Big(& 2x^2y^2 \cos(2sy) + x^2y^2 \cos(2s(x+y)) \\
& + \left(x^2\left(y^2 - y_2^2\right) - x_2^2 y^2 + 2x_1 x y y_1 + (x_1 y_1 + x_2 y_2)^2\right) \cos(2s(x-y)) \\
& + 2\left(x^2\left(y^2 + y_2^2\right) - x_2^2 y^2 - (x_1 y_1 + x_2 y_2)^2\right) \cos(2sx) - 2x^2 y_2^2 \cos(2sy) - x^2 y_2^2 \cos(2s(x+y)) \\
& + 2x_2^2 y^2 \cos(2sy) - x_2^2 y^2 \cos(2s(x+y)) - 2x x_1 y_1 y \cos(2s(x+y)) - 2x_1^2 y_1^2 \cos(2sy) \\
& + x_1^2 y_1^2 \cos(2s(x+y)) - 2x_2^2 y_2^2 \cos(2sy) + x_2^2 y_2^2 \cos(2s(x+y)) - 4x_1 x_2 y_1 y_2 \cos(2sy) \\
& + 2x_1 x_2 y_1 y_2 \cos(2s(x+y)) + 2x^2 y^2 + 2x^2 y_2^2 + 2x_2^2 y^2 + 2x_1^2 y_1^2 + 2x_2^2 y_2^2 + 4x_1 x_2 y_1 y_2 \Big) \tag{18}
\end{aligned}
$$

where $x = \sqrt{x_1^2 + x_2^2}$ and $y = \sqrt{y_1^2 + y_2^2}$. Naturally, the intricacy of kernels emerging in this work is increasing with the dimension of the input k and dimension d.

Appendix B: Gell-Mann operators

Here we list the generators of $su(3)$ algebra, the so-called Gell-Mann operators, as matrices in the computational basis of a qutrit, i.e., $\{|0\rangle, \ |1\rangle, \ |2\rangle\}$,

$$\hat{\lambda}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \hat{\lambda}_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \hat{\lambda}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\hat{\lambda}_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \hat{\lambda}_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \hat{\lambda}_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\hat{\lambda}_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \hat{\lambda}_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

For the extended set one should add $\hat{g}_0 = \sqrt{\frac{2}{3}}\hat{1}$.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abbas A, Sutter D, Zoufal C et al (2021) The power of quantum neural networks. Nat Comput Sci 1:403–409. https://doi.org/10.1038/s43588-021-00084-1

Amblard Z, Arnault F (2015) A Quantum Key Distribution Protocol for qudits with better noise resistance. https://doi.org/10.48550/arXiv.1504.08161

Barnum H, Knill E, Ortiz G, Somma R, Viola L (2004) A subsystem-independent generalization of entanglement. Phys Rev Lett 92(7):107902. https://doi.org/10.1103/PhysRevLett.92.107902

Biamonte J, Wittek P, Pancotti N et al (2017) Quantum machine learning. Nature 549(7671):195–202. https://doi.org/10.1038/nature23474

Blok MS et al (2021) Quantum information scrambling on a superconducting qutrit processor. Phys Rev X 11:021010. https://doi.org/10.1103/PhysRevX.11.021010

Cao Y, Guerreschi G, Aspuru-Guzik A (2017) Quantum Neuron: an elementary building block for machine learning on quantum computers. https://doi.org/10.48550/arXiv.1711.11240

Cerezo M, Arrasmith A, Babbush R et al (2021) Variational quantum algorithms. Nat Rev Phys 3:625–644. https://doi.org/10.1038/s42254-021-00348-9

Cozzolino D, Da Lio B, Bacco D (2019) High-dimensional quantum communication: benefits, progress, and future challenges. Adv Quantum Technol 2:1900038. https://doi.org/10.1002/qute.201900038

Dellen B, Jaekel U (2021) Solving nonlinear classification and regression problems by wave interference with a single-layer complex-valued perceptron. 2021 IEEE Symposium series on computational intelligence (SSCI), Orlando, FL, USA, pp 01–08. https://doi.org/10.1109/SSCI50451.2021.9659865.

Dellen B, Jaekel U, Wolnitza M (2019) Function and pattern extrapolation with product-unit networks. Computational science – ICCS 2019, Springer, pp 174–188 isbn 978-3-030-22741-8

Deller Y, Schmitt S, Lewenstein M et al (2023) Quantum approximate optimization algorithm for qudit systems. Phys Rev A 107:062410. https://doi.org/10.1103/PhysRevA.107.062410

Dua D, Graff C, (2017) UCI Machine learning repository. http://archive.ics.uci.edu/ml

Dunjko V, Briegel HJ (2018) Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Rep Prog Phys 81(7):074001–67. https://doi.org/10.1088/1361-6633/aab406

Durbin R, Rumelhart DE (1989) Product units: a computationally powerful and biologically plausible extension to backpropagation networks. Neural Comput 1(1):133–142. https://doi.org/10.1162/neco.1989.1.1.133

Erhard M, Malik M, Krenn M et al (2018) Experimental Greenberger-Horne-Zeilinger entanglement beyond qubits. Nature Photon 12:759–764. https://doi.org/10.1038/s41566-018-0257-6

Farhi E, Neven H (2018) Classification with quantum neural networks on near term processors. https://doi.org/10.48550/arXiv.1802.06002

Fedorov A, Steffen L, Baur M et al (2012) Implementation of a Toffoli Gate with Superconducting circuits. Nature 481:170–172. https://doi.org/10.1038/nature10713

Friedland G, Krell M (2018) A capacity scaling law for artificial neural networks. https://doi.org/10.48550/arXiv.1708.06019

Gedik Z, Silva I, Çakmak B et al (2015) Computational speed-up with a single qudit. Sci Rep 5:14671. https://doi.org/10.1038/srep14671

Gokhale P, Baker JM, Duckering C, Brown NC, Brown KR, Chong FT (2019) Asymptotic improvements to quantum circuits via qutrits. ISCA '19: Proceedings of the 46th international symposium on computer architecture, pp 554–566. https://doi.org/10.1145/3307650.3322253

Havlíček V, Córcoles AD, Temme K et al (2019) Supervised learning with quantum enhanced feature spaces. Nat 567:209. https://doi.org/10.1038/s41586-019-0980-2

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:7. https://doi.org/10.2307/233395

Imany P, Jaramillo-Villegas JA, Alshaykh MS et al (2019) High-dimensional optical quantum logic in large operational spaces. npj Quantum Inf 5:59. https://doi.org/10.1038/s41534-019-0173-8

Kues M, Reimer C, Roztocki P et al (2017) On-chip generation of high-dimensional entangled quantum states and their coherent control. Nature 546:622–626. https://doi.org/10.1038/nature22986

Kunzhe D et al (2018) Demonstration of quantum permutation parity determine algorithm in a superconducting qutrit. Chin Phys B 27:060305. https://doi.org/10.1088/1674-1056/27/6/060305

Lapkiewicz R, Li P, Schaeff C et al (2011) Experimental non-classicality of an indivisible quantum system. Nature 474:490–493. https://doi.org/10.1038/nature10119

Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. Math Program 45:503. https://doi.org/10.1007/BF01589116

Luo YH et al (2019) Quantum teleportation in high dimensions. Phys Rev Lett 123:070505. https://doi.org/10.1103/PhysRevLett.123.070505

Mackay DJC (2003) Information theory, inference, and learning algorithms. Cambridge University Press, New York, USA

Pavlidis A, Floratos E (2021) Quantum-Fourier-transform-based quantum arithmetic with qudits. Phys Rev A 103:032417. https://doi.org/10.1103/PhysRevA.103.032417

Pérez-Salinas A, Cervera-Lierta A, Gil-Fuster E, Latorre JI (2020) Data re-uploading for a universal quantum classifier. Quantum 4:226. https://doi.org/10.22331/q-2020-02-06-226

Ringbauer M, Meth M, Postler L et al (2022) A universal qudit quantum processor with trapped ions. Nat Phys 18:1053–1057. https://doi.org/10.1038/s41567-022-01658-0

Rosenblum S et al (2018) Fault-tolerant detection of a quantum error. Science 361:266. https://doi.org/10.1126/science.aat3996

Schuld M (2021) Supervised quantum machine learning models are kernel methods. https://doi.org/10.48550/arXiv.2101.11020

Schuld M, Killoran N (2019) Quantum machine learning in feature hilbert spaces. Phys Rev Lett 122:040504. https://doi.org/10.1103/PhysRevLett.122.040504

Schuld M, Sinayskiy I, Petruccione F (2014) The quest for a Quantum Neural Network. Quantum Inf Process 13:2567. https://doi.org/10.1007/s11128-014-0809-8

Schuld M, Bocharov A, Svore KM, Wiebe N (2020) Circuit-centric quantum classifiers. Phys Rev A 101(3):032308. https://doi.org/10.1103/PhysRevA.101.032308

Schuld M, Bergholm V, Gogolin C, Izaac J, Killoran N (2019) Evaluating analytic gradients on quantum hardware. Phys Rev A 99:032331. https://doi.org/10.48550/arXiv.1811.11184

Sheridan L, Scarani V (2010) Security proof for quantum key distribution using qudit systems. Phys Rev A 82:030301(R). https://doi.org/10.1103/PhysRevA.82.030301

Tacchino F, Macchiavello C, Gerace D et al (2019) An artificial neuron implemented on an actual quantum processor. npj Quantum Inf 5:26. https://doi.org/10.1038/s41534-019-0140-4

Torrontegui E, García-Ripoll JJ (2019) Unitary quantum perceptron as efficient universal approximator. Europhys Lett 125:30004. https://doi.org/10.1209/0295-5075/125/30004

Useche DH, Giraldo-Carvajal A, Zuluaga-Bucheli HM et al (2022) Quantum measurement classification with qudits. Quantum Inf Process 21:12. https://doi.org/10.1007/s11128-021-03363-y

Vapnik VN, Chervonenkis AY (1971) On the uniform convergence of relative frequencies of events to their probabilities. Theory Probab Appl 16(2):264–280

Verdon G, Broughton M, Biamonte J (2017) A quantum algorithm to train neural networks using low-depth circuits. https://doi.org/10.48550/arXiv.1712.05304

Wach NL, Rudolph MS, Jendrzejewski F, Schmitt S (2023) Data re-uploading with a single qudit. https://doi.org/10.48550/arXiv.2302.13932

Wan KH, Dahlsten O, Kristjánsson H et al (2017) Quantum generalisation of feedforward neural networks. npj Quantum Inf 3:36. https://doi.org/10.1038/s41534-017-0032-4

Wang Y, Hu Z, Sanders BC, Kais S (2020) Qudits and high-dimensional quantum computing. Front. Phys. 8. https://doi.org/10.3389/fphy.2020.589504

Weggemans JR, Urech A, Rausch A et al (2022) Solving correlation clustering with QAOA and a Rydberg qudit system: a full-stack approach. Quantum 6:687. https://doi.org/10.22331/q-2022-04-13-687

Wright LG, McMahon PL, (2020) The Capacity of Quantum Neural Networks. (2020) Conference on Lasers and Electro-Optics (CLEO). San Jose, CA, USA, pp 1–2