A quantum binary classifier based on cosine similarity

Davide Pastorello[a] and Enrico Blanzieri[b]

Department of Information Engineering and Computer Science, University of Trento,
via Sommarive 9, 38123 Povo (Trento), Italy
[a] d.pastorello@unitn.it, [b] enrico.blanzieri@unitn.it

Abstract

We introduce the quantum implementation of a binary classifier based on cosine similarity between data vectors. The proposed quantum algorithm evaluates the classifier on a set of data vectors with time complexity that is logarithmic in the product of the set cardinality and the dimension of the vectors. It is based just on a suitable state preparation like the retrieval from a QRAM, a SWAP test circuit (two Hadamard gates and one Fredkin gate), and a measurement process on a single qubit. Furthermore we present a simple implementation of the considered classifier on the IBM quantum processor ibmq_16_melbourne. Finally we describe the combination of the classifier with the quantum version of a K-nearest neighbors algorithm within a hybrid quantum-classical structure.

1 Introduction

Quantum machine learning is a rapidly emerging research area where quantum computing techniques are applied to machine learning tasks in order to pursue a computational advantage, in terms of time and space, within the present context of ever-growing amounts of data to manage. Within the supervised learning scenario, some interesting quantum classification algorithms have been proposed in recent years [SP18]. In particular in [Re14], the authors present the quantum version of a support vector machine. The quantum SVM is based on advanced quantum subroutines like the *quantum phase estimation* and the *exponentiation of density matrices* that make its realization very hard with the available prototypes of quantum computers. However, there are some recent proposals of quantum classifiers assuming fewer quantum resources, like a distance-based quantum classifier [Sc17] and a quantum implementation of the K-nearest neighbors algorithm [BG20], that may represent more feasible solutions of quantum machine learning. In this paper, we introduce a quantum algorithm for binary classification that implements a model based on cosine similarity to predict the label of a new data instance. The main idea can be summarized as follows: The unclassified instance is put in the quantum superposition of the two possible classifications and then it is compared with all the training points with a single operation by quantum parallelism.

The quantum classifier can be applied to a set of sample vectors, and the set does not, in general, coincide withe whole training set. Given a training set, the choice of the set can be made in different ways, for example it could be the set of support vectors of a support vector machine or it can be the neighborhood of the sample under a suitable metric in an instance-based learning paradigm. Here we propose that the classifier can be used in conjunction with a K-NN quantum implementation that selects the set of neighbors given a sample to classify.

In Section 2, we introduce the quantum algorithm for binary classification that assigns a label to an unclassified data vector evaluating the cosine similarity with all the training vectors. If N is the number of training points and d is the dimension of the feature space where data are represented, the proposed quantum algorithm classifies a new instance in time $O(\log(Nd))$ whereas the classical implementation of the same model presents a time complexity of $O(Nd)$. However we do not define a quantum model that can be trained but an efficient quantum algorithm to implement the binary classification according to a considered cosine similarity-based model. In Section 3, we introduce a low-sized example and perform a test on the IBM quantum processor ibmq_16_melbourne to check the correct classification of a data vector. In Section 4 we develop the idea of the combined use of the our classifier with a quantum implementation of K-NN. In Section 5 we draw the conclusions.

2 The proposed quantum binary classifier

Let $X = \{\mathbf{x}_i, y_i\}_{i=0,...,N-1}$, with $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\} \ \forall i \in \{0, ..., N-1\}$, be a training set of N data instances with two-valued labels that are represented in a real feature space of dimension d. Let $\mathbf{x} \in \mathbb{R}^d$ be a new data instance to be classified as either 1 or -1. The classification model that we consider for the quantum implementation is defined as follows:

$$y(\mathbf{x}) := \text{sgn}\left(\sum_{i=0}^{N-1} y_i \cos(\mathbf{x}_i, \mathbf{x})\right),\qquad(1)$$

where *cosine similarity* is defined by:

$$\cos(\mathbf{x}, \mathbf{y}) := \frac{\mathbf{x} \cdot \mathbf{y}}{\parallel \mathbf{x} \parallel \parallel \mathbf{y} \parallel} \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.\qquad(2)$$

A typical example where cosine similarity is adopted for classification and clustering is text analysis [Ho12, Al16, FA17]. Furthermore, in the case of normalized data vectors, (2) reduces to the dot product and it is directly related to the Euclidean distance by $\parallel \mathbf{x} - \mathbf{y} \parallel = \sqrt{2(1 - \mathbf{x} \cdot \mathbf{y})}$. In the model (1), any training vector contributes to the prediction of the new label and such a contribution is weighted by the cosine similarity with the new instance.

On one hand, the classical calculation of the new label according to (1) requires N cosine similarities each computed in time $O(d)$ then the overall time complexity is $O(Nd)$. On the other hand, the quantum implementation of the model (1) that we introduce in the following is based on the encoding of data vectors into amplitudes of a coherent superposition of quantum states (amplitude encoding), on a suitable state preparation, and on the so-called SWAP test [Bu01]. Assuming $d = 2^n$ without loss of generality, the data vector $\mathbf{x}_i \in \mathbb{R}^d$ is encoded in the amplitudes of a quantum state of n qubits:

$$|\mathbf{x}_i\rangle = \frac{1}{\|\mathbf{x}_i\|} \sum_{j=0}^{d-1} x_{ij}|j\rangle \in \mathsf{H}_n, \tag{3}$$

where $|j\rangle$ is an orthonormal basis of the n-qubit Hilbert space $\mathsf{H}_n \simeq (\mathbb{C}^2)^{\otimes n}$ and x_{ij} is the jth component of $\mathbf{x}_i$. Within the amplitude encoding we have the correspondence $\cos(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}|\mathbf{y}\rangle$. One of the key feature of quantum machine learning is that a quantum random access memory (QRAM) allows to retrieve data in parallel. Assuming that the real components $\{x_{ij}\}_{j=0,...,d-1}$ of $\mathbf{x}_i$ are stored in an array of memory cells as floating point numbers and the norm $\|\mathbf{x}_i\|$ is given separately (or we are considering normalized vectors), the retrieval of the state (3) can be done in $O(\log d)$ steps according to the *bucket brigade architecture* of the QRAM introduced in [GLM08]. Let us consider also a $\log N$-qubit register, with Hilbert space $\mathsf{H}_{index} \simeq (\mathbb{C}^2)^{\otimes \log N}$, to encode the indexes of training data vectors and construct the state:

$$|X\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle|\mathbf{x}_i\rangle|b_i\rangle \in \mathsf{H}_{index} \otimes \mathsf{H}_n \otimes \mathsf{H}_l, \tag{4}$$

where H_l is the Hilbert space of a single qubit used for encoding the values of the labels with $b_i = \frac{1-y_i}{2} \in \{0, 1\}$, then $|b_i\rangle$ is eigenstate of the Pauli matrix σ_z with eigenvalue y_i. The entangled state (4) encodes the training set X as a coherent superposition of its elements with respective labels, note that just one qubit is sufficient for the encoding of all the labels. Moreover, in the same registers we can construct the state:

$$|\psi_{\mathbf{x}}\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle|\mathbf{x}\rangle|-\rangle \in \mathsf{H}_{index} \otimes \mathsf{H}_n \otimes \mathsf{H}_l, \tag{5}$$

where we have the label qubit in the state $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, so the new data vector $\mathbf{x}$ is represented in a quantum superposition of the two possible classifications. Now consider an ancillary qubit, called qubit a, and let us prepare the state:

$$\frac{1}{\sqrt{2}}\left(|X\rangle|0\rangle + |\psi_{\mathbf{x}}\rangle|1\rangle\right) \in \mathsf{H}_{index} \otimes \mathsf{H}_n \otimes \mathsf{H}_l \otimes \mathsf{H}_a, \tag{6}$$

that can be retrieved from the QRAM in time $O(\log(Nd))$. Now we perform the SWAP test (circuit (7)) between a second ancillary qubit, called qubit b, prepared in $|+\rangle =$

3

$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and the qubit a, moreover we need another qubit, say c, prepared in $|0\rangle$ to control the Fredkin gate (i.e. the controlled swap gate):

$$\hspace{12cm} (7)$$

A straightforward calculation[1] shows that after the action of (7) the probability to obtain the outcome 1 measuring the qubit c is:

$$\mathbb{P}(1) = \frac{1}{4}(1 - \langle X|\psi_{\mathbf{x}}\rangle), \tag{8}$$

that is directly related to (1), in fact we have:

$$\langle X|\psi_{\mathbf{x}}\rangle = \frac{1}{N}\sum_{i,k=0}^{N-1}\langle i|k\rangle\langle\mathbf{x}_i|\mathbf{x}\rangle\langle b_i|-\rangle = \frac{1}{N\sqrt{2}}\sum_{i=0}^{N-1}\langle\mathbf{x}_i|\mathbf{x}\rangle(\langle b_i|0\rangle - \langle b_i|1\rangle) \tag{9}$$

$$= \frac{1}{N\sqrt{2}}\sum_{i=0}^{N-1} y_i \cos(\mathbf{x}_i, \mathbf{x}),$$

where we have simply used the identities $\langle i|k\rangle = \delta_{ik}$ and $\langle b_i|0\rangle - \langle b_i|1\rangle = 1 - 2b_i = y_i$ for any $i = 0, ..., N - 1$. Therefore the probability $\mathbb{P}(1)$ is related to the prediction of the label of $\mathbf{x}$, according to the model (1), by means of:

$$y(\mathbf{x}) = \text{sgn}\left[1 - 4\mathbb{P}(1)\right]. \tag{10}$$

After the retrieval of the state (6) from a QRAM in time $O(\log(Nd))$, an appropriate SWAP test, performed in constant time, is sufficient to classify the new insatnce $\mathbf{x}$ according to the model (1) with an exponential speedup w.r.t. the classical calculation. However the procedure *preparation+test* must be repeated several times for sampling the qubit c to estimate $\mathbb{P}(1)$ as the success probability of a Bernoulli trial, so an estimation within an error ϵ requires a number of repetitions growing as $O(\epsilon^{-2})$ as provided by the binomial proportion confidence interval [BCD01]. Thus the overall time complexity of Algorithm 1 is $O(\epsilon^{-2}\log(Nd))$.

---

[1]Appendix A for details.

Input: training set $X = \{\mathbf{x}_i, y_i\}_{i=0,\dots,N-1}$, unclassified instance $\mathbf{x}$.
Result: label y of $\mathbf{x}$.

1 **repeat**
2   initialize the register $\mathsf{H}_{index} \otimes \mathsf{H}_n \otimes \mathsf{H}_l$ and an ancillary qubit a in the state (6);
3   initialize a qubit b in the state $|-\rangle$;
4   perform the SWAP test on a and b with control qubit c prepared in $|0\rangle$; *% circuit (7)*
5   measure qubit c;
6 **until** *desired accuracy on the estimation of* $\mathbb{P}(1)$;
7 Estimate the probability $\mathbb{P}(1)$ as the relative frequency $\hat{\mathbb{P}}$ of the outcome 1;
8 **if** $\hat{\mathbb{P}} > 0.25$ **then**
9   **return** $y = -1$
10 **else**
11   **return** $y = 1$
12 **end**

Algorithm 1: *Quantum implementation of the model (1).*

3   Run on the IBM quantum processor

Let us provide an example of implementation of the introduced quantum binary classifier for $N = 2$ and $d = 2$. Consider a training set of two-dimensional data instances given by $X = \{(\mathbf{x}_0, y_0), (\mathbf{x}_1, y_1)\}$ where $\mathbf{x}_0 = (1, 0)$, $y_0 = 1$ and $\mathbf{x}_1 = (0.718, 0.696)$, $y_1 = -1$. Let $\mathbf{x} = (0.884, 0.468)$ be the unlabelled data instance that we have to classify. In this simple example with normalized data vectors and only two training points, the model (1) predicts the label of $\mathbf{x}$ as a *nearest neighbor* then it returns $y = -1$.

In order to run Algorithm 1, according to (4) and (5) we prepare the state:

$$|X\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle|0\rangle + |1\rangle|\mathbf{x}_1\rangle|1\rangle), \tag{11}$$

and the state:

$$|\psi_{\mathbf{x}}\rangle = |+\rangle|\mathbf{x}\rangle|-\rangle. \tag{12}$$

We run the algorithm on a IBM prototype quantum processor that is available within the cloud-based quantum computing service of *IBM Quantum*. The machine is realized by superconducting transmon qubits located in a cryogenic refrigerator at the Thomas J. Watson Research Center, New York [IBM21]. Let us consider the quantum circuit for the construction of the state (11):
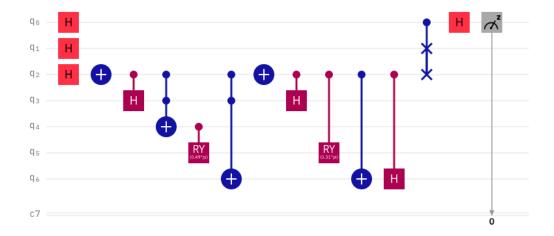

$$\tag{13}$$

where $RY_{(0.49\pi)}$ is a rotation of 0.49π around the y-axis of the Bloch sphere mapping $|0\rangle$ into $|\mathbf{x}_1\rangle = 0.718|0\rangle + 0.696|1\rangle$. The circuit to construct the state (3) is simply:

$$|0\rangle \;\text{---}\; \boxed{H} \;\text{-------} \tag{14}$$

$$|0\rangle \;\text{-}\; \boxed{RY_{(0.31\pi)}} \;\text{---}$$

$$|0\rangle \;\text{-------}\; \oplus \;\text{---}\; \boxed{H}\text{-}$$

where the gate $RY_{(0.31\pi)}$ rotates $|0\rangle$ into $|\mathbf{x}\rangle = 0.884|0\rangle + 0.468|1\rangle$ and we have a bit-flip followed by a Hadamard gate to prepare the label qubit in $|-\rangle$.

In Fig. 1, there is the circuit, represented in the *IBM Quantum Composer*, implementing Algorithm 1 for the example just introduced. The qubits q_0, q_1, q_2 are the ancillas used for the SWAP test, q_3 is the 1-qubit index register, q_4 is an additional ancillary qubit necessary to implement the controls of the gate $RY_{(0.49\pi)}$, q_5 is the 1-qubit register for the amplitude encoding of data and q_6 is the qubit encoding the labels.

Fig.1. Quantum circuit implementing lines 1-7 of Algorithm 1 in the IBM Quantum Composer where $N = 2$, $d = 2$, the training set is $\{((1,0),1)\,,\,((0.718,0.696),-1)\}$ and the unclassified data vector is $\mathbf{x} = (0.884, 0.468)$.

We have run the algorithm on the IBM quantum processor *ibmq_16_melbourne* performing 1024 shots for sampling the qubit q_0 (corresponding to the qubit c of our general algorithm description). The obtained statistic is shown in Fig. 2, the relative frequency of the outcome 1, used to estimate the probability $\mathbb{P}(1)$, is $\hat{\mathbb{P}} = 447/1024 > 0.25$ then the label assigned to $\mathbf{x} = (0.884, 0.468)$ is $y = -1$ as expected.
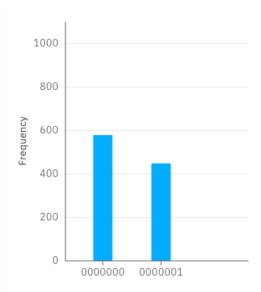
Fig.2. Output statistic of the circuit reported in Fig.1.

4 Combination with the quantum K-NN

The proposed binary classifier can be combined with a quantum version of the *K-nearest neighbors algorithm* presented in [BG20]. In fact, given the training set $X = \{\mathbf{x}_i, y_i\}_{i=0,\dots,N-1}$ and the unclassified instance $\mathbf{x}$ we can assume to select the K closest training points to $\mathbf{x}$ (w.r.t. cosine similarity) and then performing the quantum binary classification described above. So we do not consider the whole training set for the label assignment but just the training points in a neighborhood of the new data vector. We focus on vectors with positive entries, this is not a sever restriction if we take into account that data vectors are so in text analysis that is the most relevant framework where cosine similarity is used for classification and clustering.

The selection of the K nearest neighbors can be done implementing a SWAP test as well. However the *controlled swap* is not a standard Fredkin gate that conditionally swaps only two qubits but we must swap two n-qubit registers. It is straightforward to show that the controlled swap of two n-qubit registers can be implemented with n Fredkin gates controlled by the same qubit.

Let us consider two n-qubit register (i.e. two copies of the main register where data are encoded), the index register, and an ancillary qubit prepared in the state:

$$\frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle |\mathbf{x}_i\rangle |\mathbf{x}\rangle |0\rangle \in \mathsf{H}_{index} \otimes \mathsf{H}_n \otimes \mathsf{H}_n \otimes \mathsf{H}_a. \tag{15}$$

In (15), we have the superposition of the training points and the new data instance stored in two different registers, instead in the classification algorithm described above

these data are superposed in the same register. Now perform the SWAP test on the two n-qubit registers controlled by the ancillary qubit obtaining:

$$|\Psi\rangle = \frac{1}{2\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \left[(|\mathbf{x}_i\rangle|\mathbf{x}\rangle + |\mathbf{x}\rangle|\mathbf{x}_i\rangle)|0\rangle + (|\mathbf{x}_i\rangle|\mathbf{x}\rangle - |\mathbf{x}\rangle|\mathbf{x}_i\rangle)|1\rangle \right]. \tag{16}$$

The probability to obtain outcome $\alpha \in \{0, 1\}$ by a measurement process on the ancillary qubit is:

$$\mathbb{P}(\alpha) = \frac{1}{2} + (-1)^\alpha \frac{1}{2N} \sum_{i=0}^{N-1} |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2, \tag{17}$$

and the corresponding post-measurement state of the registers is:

$$|\Psi_\alpha\rangle = \frac{\sum_{i=0}^{N-1} |i\rangle(|\mathbf{x}_i\rangle|\mathbf{x}\rangle + (-1)^\alpha|\mathbf{x}\rangle|\mathbf{x}_i\rangle)}{\sqrt{2\left(N + (-1)^\alpha \sum_{i=0}^{N-1} |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2\right)}} |\alpha\rangle. \tag{18}$$

Once obtained the outcome α measuring the ancillary qubit, we can perform a measurement on the index register, the probability[2] to obtain the outcome i is:

$$\mathbb{P}(i|\alpha) = \frac{1 + (-1)^\alpha |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2}{N + (-1)^\alpha \sum_{i=0}^{N-1} |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2}. \tag{19}$$

We have:

$$\mathbb{P}(i|0) - \mathbb{P}(i|1) = \frac{2(|\langle \mathbf{x}|\mathbf{x}_i\rangle|^2 - C)}{N(1 - C^2)}, \tag{20}$$

where $C = \frac{1}{N} \sum_i |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2$ is a constant, so (20) is proportional to the square cosine similarity between $\mathbf{x}_i$ and $\mathbf{x}$ that is the *fidelity* between the quantum states $|\mathbf{x}\rangle$ and $|\mathbf{x}_i\rangle$. Since we are considering only vectors with positive entries, sampling the index register we obtain the most likely indexes corresponding to the closest vectors to $\mathbf{x}$. Then we can run the binary classifier defined above without evaluating any training point but only those in a neighborhood of the new data instance.

Algorithm 2 is a hybrid quantum-classical algorithm for binary classification providing the quantum selection of the K closest training points to the unlabelled data vector and the following assignment of the label running Algorithm 1. In the algorithm there are classical operations to read the output of the K-NN and initialize the quantum classification routine in a neighborhood of the unclassified vector.

---

[2]The probabilities (17) and (19) and the state (18) are explicitly calculated in Appendix B.

Input: training set $X = \{\mathbf{x}_i, y_i\}_{i=0,...,N-1}$, unclassified instance $\mathbf{x}$.
Result: label y of $\mathbf{x}$.

1 initialize the register $\mathsf{H}_{index} \otimes \mathsf{H}_n \otimes \mathsf{H}_n$ and an ancillary qubit a in the state (15);
2 perform the SWAP test on the n-qubit registers controlled by the qubit a;
3 measure the qubit a;
4 compute (20) sampling the index register;
5 select the K most likely indexes $\{i_1, ..., i_K\}$;
6 $X \leftarrow \{\mathbf{x}_{i_k}, y_{i_k}\}_{k=1,...,K}$;
7 $y \leftarrow Algorithm\,1(X, \mathbf{x})$;
8 **return** y;

Algorithm 2: *Hybrid quantum-classical algorithm for binary classification.*

5 Conclusions

In this paper we have presented a quantum implementation of an elementary binary classifier based on cosine similarity with an exponential speedup w.r.t. to any classical implementation of the same model. Then we have tested the proposed quantum algorithm on a IBM quantum processor to show that it can be easily run on available quantum machines. The algorithm shares some characteristics with known quantum machine learning algorithms based on the SWAP test [Ll13, Re14, BG20]. The peculiarity of our proposal is the quantum calculation of the cosine similarity between the unclassified vector and all the training vectors weighted with the labels without parameters that must be learned in a training phase. A natural application framework of this classification model may be text analysis where documents are typically compared by means of cosine similarity.

We also provided and example of use of our classifier in conjunction with a quantum implementation of the K-nearest neighbors. The classifier is applied to the K vectors outputted by the K-NN. Although the combination is hybrid quantum-classical in nature and further work is needed in order to integrate the two in a completely quantum version, it illustrates the potentiality of our proposal to be a component of novel quantum machine learning algorithms.

The exponential speedup of the quantum classifier is due to the efficient preparation of quantum states in logarithmic time and to the classification itself performed in constant time (that depends on the desired accuracy). However the efficient retrieval is a common bottleneck in quantum machine learning as the QRAM architecture is extremely expansive in terms of space resources. Thus we conclude that new effective procedures for representing classical data into quantum hardware are a crucial direction of investigation for quantum machine learning. Moreover a goal of the pivotal proposal of this paper is the suggestion of a general approach in devising quantum machine learning algorithms that should have the requirement of being easily and efficiently implemented on the available or near-term quantum machines of the NISQ era.

Acknowledgements

Appendix A: Derivation of (8)

Consider the circuit (7) for the SWAP test: If the input state is $|0\rangle_c|\varphi\rangle_b|\psi\rangle_a$, we have the following processing:

$$|0\rangle_c|\varphi\rangle_b|\psi\rangle_a \xrightarrow{H\otimes\mathbb{I}_b\otimes\mathbb{I}_a} \frac{1}{\sqrt{2}}(|0\rangle_c + |1\rangle_c)|\varphi\rangle_b|\psi\rangle_a \xrightarrow{Fredkin\ gate} \frac{1}{\sqrt{2}}(|0\ \varphi\ \psi\rangle_{cba} + |1\ \psi\ \varphi\rangle_{cba}) \quad (21)$$

$$\xrightarrow{H\otimes\mathbb{I}_b\otimes\mathbb{I}_a} \frac{1}{2}|0\rangle_c(|\varphi\ \psi\rangle_{ba} + |\psi\ \varphi\rangle_{ba}) + \frac{1}{2}|1\rangle_c(|\varphi\ \psi\rangle_{ba} - |\psi\ \varphi\rangle_{ba}),$$

where $\mathbb{I}_{a,b}$ are the identity operators acting on the qubits a and b. Therefore the probability of measuring 1 on the qubit c is:

$$\mathbb{P}(1) = \frac{1}{4}(\langle\varphi\ \psi|_{ba} - \langle\psi\ \varphi|_{ba})(|\varphi\ \psi\rangle_{ba} - |\psi\ \varphi\rangle_{ba}) = \frac{1}{2}(1 - |\langle\varphi|\psi\rangle|^2), \quad (22)$$

so the SWAP test is a useful and well-known tool to estimate $|\langle\varphi|\psi\rangle|^2$ given two unknown pure states $|\psi\rangle$ and $|\varphi\rangle$.

In our case the qubit b is in $|+\rangle$ and the qubit a is entangled with the n-qubit register encoding the data as provided by the state (6). Then the probability to obtain 1 by a measurement on the qubit c is $\mathbb{P}(1) = \frac{1}{2}(1 - \langle\Phi|\Phi\rangle)$ where Φ is the non-normalized vector obtained taking the inner product of $|+\rangle$ and the ancillary tensor factor of the state (6):

$$\Phi = \frac{1}{\sqrt{2}}(\langle+|0\rangle|X\rangle + \langle+|1\rangle|\psi_{\mathbf{x}}\rangle) = \frac{1}{2}(|X\rangle + |\psi_{\mathbf{x}}\rangle). \quad (23)$$

Therefore:

$$\mathbb{P}(1) = \frac{1}{2}(1 - \langle\Phi|\Phi\rangle) = \frac{1}{2}\left[1 - \frac{1}{4}(2 + 2\langle X|\psi_{\mathbf{x}}\rangle)\right] = \frac{1}{4}(1 - \langle X|\psi_{\mathbf{x}}\rangle). \quad (24)$$

Appendix B: Derivation of (17), (18), (19)

The probability to obtain the outcome 0 by a measurement process on the ancillary qubit in the state (16) is:

$$\mathbb{P}(0) = \langle\Psi|P_0\Psi\rangle \quad \text{with} \quad P_0 = \mathbb{I}_{index,n} \otimes |0\rangle\langle0|, \quad (25)$$

where $\mathbb{I}_{index,n}$ is the identity operator on the Hilbert space $\mathsf{H}_{index} \otimes \mathsf{H}_n^{\otimes 2}$ and $|0\rangle\langle 0|$ is the projector onto the 1-dimensional subspace of H_a spanned by $|0\rangle$, as provided by the postulates of quantum mechanics. Therefore:

$$\mathbb{P}(0) = \frac{1}{4N} \sum_{i,l=0}^{N-1} \langle i|l\rangle ((\langle \mathbf{x}_i|\langle \mathbf{x}| + \langle \mathbf{x}|\langle \mathbf{x}_i|)(|\mathbf{x}_l\rangle|\mathbf{x}\rangle + |\mathbf{x}\rangle|\mathbf{x}_l\rangle)) = \tag{26}$$

$$= \frac{1}{4N} \sum_{i=0}^{N-1} (2 + 2\langle \mathbf{x}_i|\mathbf{x}\rangle\langle \mathbf{x}|\mathbf{x}_i\rangle) = \frac{1}{2} + \frac{1}{2N} \sum_{i=0}^{N-1} |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2.$$

A similar calculation, considering the projector $P_1 = \mathbb{I} \otimes |1\rangle\langle 1|$, yields $\mathbb{P}(1) = \langle \Psi|P_1\Psi\rangle = \frac{1}{2} - \frac{1}{2N} \sum_{i=0}^{N-1} |\langle \mathbf{x}_i|\mathbf{x}\rangle|^2$. Thus we have (17).

The post-measurement state (18) is obtained normalizing the projected vector $P_\alpha|\Psi\rangle$.

The probability (19) is given by:

$$\mathbb{P}(i|\alpha) = \langle \Psi_\alpha|Q_i\Psi_\alpha\rangle \quad \text{with} \quad Q_i = |i\rangle\langle i| \otimes \mathbb{I}_{n,a}, \tag{27}$$

where $\mathbb{I}_{n,a}$ is the identity operator on the Hilbert space $\mathsf{H}_n^{\otimes 2} \otimes \mathsf{H}_a$ and $|i\rangle\langle i|$ is the projector onto the subspace of H_{index} spanned by $|i\rangle$. Therefore:

$$\mathbb{P}(i|\alpha) = \frac{((\langle \mathbf{x}_i|\langle \mathbf{x}| + (-1)^\alpha \langle \mathbf{x}|\langle \mathbf{x}_i|)(|\mathbf{x}_i\rangle|\mathbf{x}\rangle + (-1)^\alpha |\mathbf{x}\rangle|\mathbf{x}_i\rangle))}{2(N + (-1)^\alpha \sum_{i=0}^{N-1} |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2)} = \tag{28}$$

$$= \frac{2 + 2(-1)^\alpha \langle \mathbf{x}_i|\mathbf{x}\rangle\langle \mathbf{x}|\mathbf{x}_i\rangle}{2(N + (-1)^\alpha \sum_{i=0}^{N-1} |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2)} = \frac{1 + (-1)^\alpha |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2}{N + (-1)^\alpha \sum_{i=0}^{N-1} |\langle \mathbf{x}|\mathbf{x}_i\rangle|^2}.$$

References

[Al16] S. Al-Ananzi, H. AlMahmoud, I. Al-Turaiki *Finding similar documents using different clustering techniques* Procedia Computer Science 82 (2016)

[BG20] A. A. Basheer, S. K. Goyal *Quantum k-nearest neighbor machine learning algorithm* preprint (2020)

[BCD01] L. D. Brown, T. T. Cai, A. DasGupta *Interval Estimation for a Binomial Proportion* Statist. Sci. 16 (2) 101 - 133 (2001)

[Bu01] H. Buhrman, R. Cleve, J. Watrous, R. de Wolf *Quantum Fingerprinting* Phys. Rev. Lett. 87, 167902 (2001)

[DM01] I. S. Dhillon, D. S. Moda *Concept decomposition for large sparse text data using clustering* Machine Learning, 42, 143-175 (2001)

[FA17] C. Focil-Arias, J. Ziiniga, G. Sidorov, I. Batyrshin, A. Gelbukh *A tweets classifier based on cosine similarity* CEUR Workshop Proceedings (2017)

[Ho12] K. Hornik, I. Feinerer, M. Kober, C. Buchta *Spherial k-means clustering* Journal of Statistical Software vol. 50, n. 10 (2012)

[IBM21] https://quantum-computing.ibm.com

[GLM08] V. Giovannetti, S. Lloyd, L. Maccone *Quantum random access memory* Phys. Rev. Lett. 100, 160501 (2008)

[Ll13] S. Lloyd, Mohseni, M., Rebentrost, P. *Quantum algorithms for supervised and unsupervised machine learning.* Preprint at https://arxiv.org/abs/1307.0411 (2013)

[Re14] P. Rebentrost, M. Mohseni, S. Lloyd *Quantum support vector machine for big data classification* Phys. Rev. Lett. 113, 130503 (2014)

[Sc17] M. Schuld et al. *Implementing a distance-based classifier with a quantum interference circuit* EPL 119 60002 (2017)

[SP18] M. Schuld, F. Petruccione *Supervised learning with quantum computers* Springer International Publishing 2018