Proceedings

Supervised Quantum State Discrimination †

Marco Fanizza * , Andrea Mari and Vittorio Giovannetti

NEST, Scuola Normale Superiore, Istituto Nanoscienze-CNR, I-56126 Pisa, Italy

* Correspondence: marco.fanizza@sns.it

† Presented at the 11th Italian Quantum Information Science conference (IQIS2018), Catania, Italy, 17–20 September 2018.

1. Introduction

Combining machine learning and quantum information ideas has been a fruitful line of research in the last few years [1–3]. These new proposals can be divided in three main categories: using machine learning for enhancing numerical methods and experimental techniques, devising quantum algorithms to attack classical machine learning problems, and studying quantum generalisations of learning tasks. In this contribution we address the latter category with an example of a supervised learning with quantum data. For the sake of our discussion we use a very broad characterisation of quantum learning tasks. First of all we consider a genuine quantum information theory setting, where an agent receives a quantum source (of quantum or classical information), and can operate on it with any processing device allowed by the rules of quantum mechanics. Secondly, we refer to tasks in which a machine should be trained to perform a certain quantum operation and this training can be done through quantum processing, that means with quantum training data and quantum operations.

In a classical supervised learning classification problem a machine is given a set of labelled data, and uses this information to produce a classifier which can be used to assign a label to new unlabelled data. In a probabilistic setting the data $x \in X$ and the labels $y \in Y$ are distributed according to the probability distribution $P : (X, Y) \to [0, 1]$ and a classifier is a conditional probability distribution $C(Y|X)$. For each P there exist an optimal classifier which minimises the probability of misclassification, and a good learning algorithm is expected to give a good approximation of the optimal classifier, at least when the training dataset becomes large. Even better, if the agent has some prior information on the possible distributions one can define what is the optimal training and test algorithm, which is the one with the lowest probability on average, averaging over all the possible distributions assuming the prior.

Since classical learning tasks can be studied in a probabilistic formulation, a straightforward generalisation of a learning task can be obtained by reinterpreting the task on quantum states rather then probability distributions.

The analogy is clear and fundamental in quantum information theory, as most of the information theoretic tasks that can be defined using probability distribution, like compression and communication, can be generalised to quantum states.

2. The Model

On the basis of these observation, we consider the following problem: an agent is asked to correctly guess the state of a qubit system X initialised with equal probability in the state ρ_1 or ρ_2, bur ρ_1 and ρ_2 are unknown to the agent. Instead, he receives as a training set for the task a system A

made of n qubits known to be in the state ρ_1 and a system B of n other qubits known to be in the state ρ_2. The agent may have some kind of prior information on ρ_1 and ρ_2, like their purity or their overlap. The question we ask is to find the two-outcome measurement $\hat{\mathcal{M}} \equiv \{\Pi_1, \Pi_2\}$ on the joint input made of the training and test data which minimises the probability of misclassification error, calculated by averaging over all the possible couples ρ_1 and ρ_2 according to the prior information. Since the input state of the machine can be one of two alternatives $\tau_1 = \rho_1^{\otimes n} \otimes \rho_1 \otimes \rho_2^{\otimes n}$ (if X in ρ_1) and $\tau_2 = \rho_1^{\otimes n} \otimes \rho_2 \otimes \rho_2^{\otimes n}$ (if X in ρ_2), the average probability reads

$$P_{err}^{(n)} = \int d\mu(\rho_1, \rho_2) \frac{\mathrm{Tr}[\tau_1 \hat{\Pi}_2] + \mathrm{Tr}[\tau_2 \hat{\Pi}_1]}{2}, \tag{1}$$

where $d\mu(\rho_1, \rho_2)$ is a classical probability distribution on the states ρ_1, ρ_2, and encodes the prior information of the agent. This problem can be translated in a binary state discrimination task for two known effective states

$$\alpha^{(n)} \equiv \int d\mu(\rho_1, \rho_2) \rho_1^{\otimes n} \otimes \rho_1 \otimes \rho_2^{\otimes n}, \qquad \beta^{(n)} \equiv \int d\mu(\rho_1, \rho_2) \rho_1^{\otimes n} \otimes \rho_2 \otimes \rho_2^{\otimes n}. \tag{2}$$

Therefore, defined $\Theta \equiv \alpha^{(n)} - \beta^{(n)}$, our figure of merit can be written as

$$P_{err,min}^{(n)} = \frac{1}{2} - \frac{1}{4} \|\Theta\|_1, \tag{3}$$

with the symbol $\| \cdots \|_1$ indicating the trace norm.

In the limit of $n \to \infty$ one expects that this quantity converges to the averages of the probability of error for the Helstrom measurement for known ρ_1 and ρ_2, since the classical description of the template states can be recovered exactly, for example using tomography. In fact this limit is always a lower bound to the probability of error at finite n. We are interested in the finite size correction to this value.

3. Results

Our contribution consists in calculating $P_{err,min}^{(n)}$ for a number of priors, generalising previuos results [4,5]. As detailed in the preprint [6] we extensively use the symmetric and covariant properties of Θ in order to reduce the problem to a simple analytic diagonalisation of 2×2 matrices. Moreover, we perform asymptotic expansions of the sum of eigenvalues arising from Equation (3) in order to obtain finite size corrections to the asymptotic limit. In particular we consider these cases:

(i) ρ_1, ρ_2 have assigned purities—the moduli of their Bloch vectors being respectively r_1 and r_2; uniform prior on the Bloch vector's directions,

$$
\begin{aligned}
P_{err,min}^{(n \gg 1)} = \frac{1}{2} - \frac{1}{24} \frac{(r_1 + r_2)^3 - |r_1 - r_2|^3}{r_1 r_2} \\
+ \frac{5}{24\,n} \frac{(r_1 + r_2)^3 + |r_1 - r_2|^3}{r_1^2 r_2^2} - \frac{1}{24\,n} \frac{(r_1 + r_2)^5 - |r_1 - r_2|^5}{r_1^3 r_2^3} + o\left(\frac{1}{n}\right).
\end{aligned}
$$

where the notation $o\left(\frac{1}{n}\right)$ indicate terms that goes to zero faster than $\frac{1}{n}$.

(ii) ρ_1 and ρ_2 are generically mixed qubit states, with a constant density Bloch sphere prior,

$$P_{err,min}^{(n \gg 1)} = \frac{17}{70} + \frac{18}{35n} + o\left(\frac{1}{n}\right). \tag{4}$$

(iii) ρ_1 and ρ_2 are pure and they have a fixed overlap $\text{Tr}[\rho_1\rho_2] = \sin^2 \frac{\theta}{2}$; uniform prior on the global orientation,

$$P_{err,min}^{(n \gg 1)} = \frac{1}{2}\left(1 - |\cos \tfrac{\theta}{2}|\right) + \frac{3+\cos\theta}{8\sqrt{2}\sqrt{1+\cos\theta}}\frac{1}{n} + \frac{1-60\cos\theta-5\cos2\theta}{128\sqrt{2}(1+\cos\theta)^{3/2}}\frac{1}{n^2} + o\left(\frac{1}{n^2}\right).$$ (5)

References

1. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202.
2. Dunjko, V.; Briegel, H.J. Machine learning & artificial intelligence in thequantum domain: A review of recent progress. *Rep. Prog. Phys.* **2018**, *81*, 074001.
3. Schuld, M.; Petruccione, F. *Supervised Learning with Quantum Computers*; Springer: Cham, Switzerland 2018.
4. Hayashi, A.; Horibe, M.; Hashimoto, T. Quantum pure-state identification. *Phys. Rev. A* **2005**, *72*, 052306.
5. Sentís, G.; Bagan, E.; Calsamiglia, J.; Munoz-Tapia, R. Multicopy programmable discrimination of general qubit states. *Phys. Rev. A* **2010**, *82*, 042312.
6. Fanizza, M.; Mari, A.; Giovannetti, V. Optimal universal learning machines for quantum state discrimination. *arXiv* **2018**, arXiv:1805.03477.