Quantum-inspired Classification Process

Giuseppe Sergioli & **Alophis group**
(Applied Logics, Philosophy and History of Science)
University of Cagliari
[giuseppe.sergioli@gmail.com]

November 3th-4th, Cagliari

Univeristy of Cagliari
Department of Philosophy
Department of Electronic Engineering

Project: "*Modelling the Uncertainty: Quantum Theory at the service of Pattern Recognition*"

List of contents

Training set, Class, Pattern, Feature

Let us consider (as a simple example) two disjoint sets A and B of different objects (say cats and dogs). During the **training set**, we take n objects from the set A and m objects from the set B. Let $C_a \subset A$ and $C_b \subset B$.

We can measure two (or more) **features** of each object $a_i \in C_a$ and $b_i \in C_b$ (for istance the weight and the lenght of the tail).

We say that C_a and C_b are **classes** and the objects a_i and b_i are **patterns** that are characterized by their features.
We write, for example, $a_i = \{x_1, x_2\}$, where x_1 and x_2 are the weight and the lenght of the tail of the cat a_i, respectively.

Nearest Mean Classifier (NMC)

Let us consider the classes $C_a = \{a_1, ..., a_n\}$ and $C_b = \{b_1, ..., b_m\}$, with a_i and b_i belonging to the training set and an arbitrary pattern $c_i = \{x_1, x_2\}$ belonging to the **test set.** The goal is to establish whether is more probably that $c_i \in A$ or $c_i \in B$.

We - only - consider the **centroids** a^* and b^* of C_a and C_b and the **euclidean distances** $Ed(c_i, a^*)$ and $Ed(c_i, b^*)$.

Hence, if $Ed(c_i, a^*) \geq Ed(c_i, b^*)$ then (is more probabily that) $c_i \in B$; otherwise $c_i \in A$.

The notions of "Pattern" and "Classification" are very general and are naturally connected to our common processes of acquiring knowledge.

All we need in order to provide a Quantum representation of NMC are:

- a sutable encoding from patterns to quantum objects
- a quantum counterpart of the centroid
- a quantum counterpart of the Euclidean distance

An Example: Stereographic encoding

It is possible to map the pattern $a = (x, y)$ onto the surface of a radius one sphere by the *stereographic projection*:

$$(x, y) \to \left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}\right).$$

By placing the Bloch components:
$r_1 = \frac{2x}{x^2+y^2+1}$; $r_2 = \frac{2y}{x^2+y^2+1}$; $r_3 = \frac{x^2+y^2-1}{x^2+y^2+1}$ we obtain:

$$\rho_a = \frac{1}{2}\begin{pmatrix} 1 + r_3 & r_1 - ir_2 \\ r_1 + ir_2 & 1 - r_3 \end{pmatrix} = \frac{1}{x^2 + y^2 + 1}\begin{pmatrix} x^2 + y^2 & x - iy \\ x + iy & 1 \end{pmatrix}.$$

Example

Let us consider the pattern $a = \{1, 3\}$. Its corresponding Density pattern ρ_a, is:

$$\rho_a = \frac{1}{11}\begin{pmatrix} 10 & 1-3i \\ 1+3i & 1 \end{pmatrix}$$

We call ρ_a *Density Pattern*.

Moon Dataset

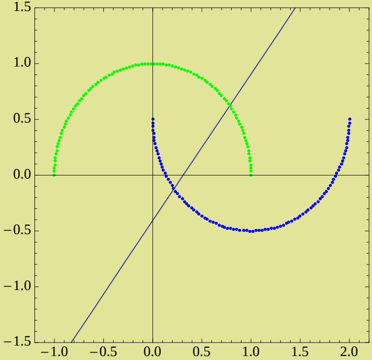


Figure : Classical Patterns

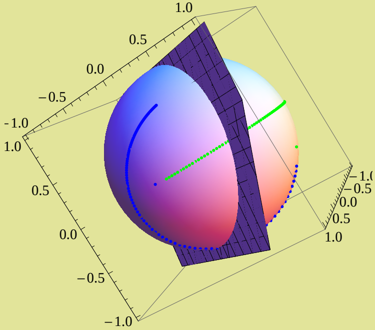


Figure : Density Patterns

Another Example: Projective encoding

$$v \equiv (x, y) \to \left(\frac{x}{||v||}, \frac{y}{||v||}\right) \equiv (\bar{x}, \bar{y})$$

$$|\psi_v\rangle = \bar{x}|0\rangle + \bar{y}|1\rangle$$

$$\rho_v = |\psi_v\rangle\langle\psi_v|$$

...and many others.

Preservation of the Order

Let $a = \{x_a, y_a\}$ $b = \{x_b, y_b\}$ and $c = \{x_c, y_c\}$ be three arbitrary patterns and let ρ_i be the density pattern associated to the pattern i.

If $Ed(a, b) \leq Ed(b, c)$ (where Ed is the Euclidian distance), is it possible to define a *Quantum distance* such that $Qd(\rho_a, \rho_b) \leq Qd(\rho_b, \rho_c)$?

Normalized Trace Distance

Let us consider two patterns $a = \{x_a, y_a\}$ and $b = \{x_b, y_b\}$.

Let $\rho_a = \frac{1}{2}\begin{pmatrix} 1 + r_{a_3} & r_{a_1} - ir_{a_2} \\ r_{a_1} + ir_{a_2} & 1 - r_{a_3} \end{pmatrix}$ the density pattern associated to a; similarly for b.

Let place $K = \frac{2}{\sqrt{(1-r_{a_3})(1-r_{b_3})}}$ and let we define the **normalized trace distance** as: $K\,Td(\rho_a, \rho_b)$, where Td is the usual *Trace distance*.
It is straightforward to show that

$$Ed(a, b) = K\,Td(\rho_a, \rho_b).$$

Classification

Hence, given a and b as the centroids of C_a and C_b respectively, if

$$K\,Td(\rho_x, \rho_a) \geq K\,Td(\rho_x, \rho_b)$$

then $x \in B$; otherwise $x \in A$. Similarly to the classical case.

Convenience on a Quantum Computer

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

"*Estimating distances between vectors in N-dimensional vector spaces then takes time $O(logN)$ on a quantum computer. By contrast, sampling and estimating distances between vectors on a classical computer is apparently exponentially hard. Quantum machine learning provides an exponential speed-ups over all known classical algorithms for problems involving evaluating distances between large vectors.*"

But it turns out to be convenient mostly on a Classical Computer...

Quantum Centroid

Given a dataset $\{P1, ..., P_n\}$, let us consider the respective set of density patterns $\{\rho_1, ..., \rho_n\}$.

The Quantum Centroid is defined as:

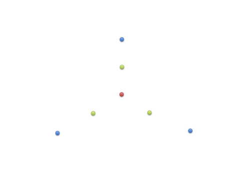
$$\rho_{QC} = \frac{1}{n} \sum_{i=1}^{n} \rho_i.$$

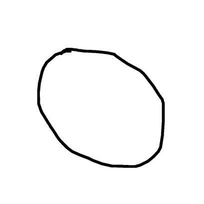
S. Gambs, *Quantum classification*, arXiv:0809.0444v2.

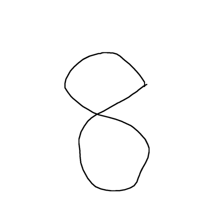
Observations

Some observation:

- The QC ρ_{QC} is not a pure state and it has not any counterpart in the set of classical pattern in $\mathbb{R}^n$;
- In contrast to the Classical Centroid, the QC is "sensitive" to the **distribution** of the patterns.







We provide a comparison between the NMC and the "quantum" classification process based on Density Patterns and Quantum Centroids by involving different kinds of standard datasets on a **Classical Computer**.

We compare the Error E and the reliability (in terms of the Cohen's constant k) for both classifiers.

At a first glance - and in order to provide a clear *visual* representation - we consider that the training and the test sets are the same.

Gaussian Dataset

Gaussian Dataset: 200 Patterns allocated in two Classes.

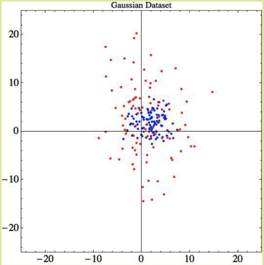


Figure : Gaussian Dataset

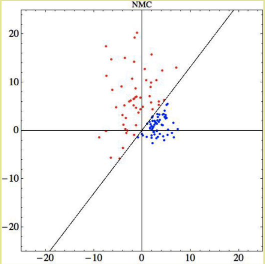


Figure : NMC

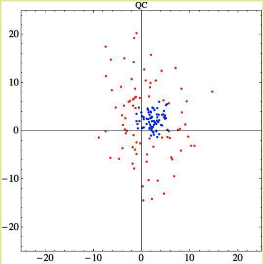


Figure : Quantum Classifier

Table : Gaussian Dataset

	E	E1	E2	Pr	k	TPR	FPR
NMC	**0.445**	0.41	0.48	**0.555**	**0.11**	0.555	0.445
QC	**0.24**	0.28	0.2	**0.762**	**0.52**	0.76	0.24

By randomly dividing the dataset in a training set (80%) and in a test set (20%), the average over 100 experiments gives: $NMC - Error = 44.35 \pm 6.79$; $Q - Error = 23.68 \pm 6.09$.

A remark

Even if the error of the Quantum Classifier is lower than the Error of the NMC, there are some patterns that are correctly classified by the NMC but not by the Quantum Classifier. Hence, it makes sense to consider a "merging" of the NMC and the Quantum Classifier.

Gaussian Dataset

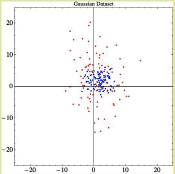


Figure : Gaussian Dataset

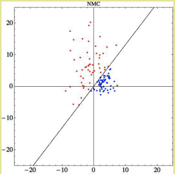


Figure : NMC

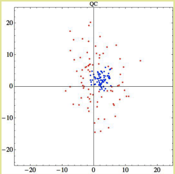


Figure : Quantum Classifier

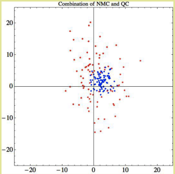


Figure : NMC & Quantum Classifier

Table : Gaussian Dataset

	E	E1	E2	Pr	k	TPR	FPR
NMC	**0.445**	0.41	0.48	**0.555**	**0.11**	0.555	0.445
QC	**0.24**	0.28	0.2	**0.762**	**0.52**	0.76	0.24
NMC-QC	**0.13**	0.14	0.12	**0.87**	**0.74**	0.87	0.13

Moon Dataset

Moon Dataset: 200 patterns allocated in two Classes.

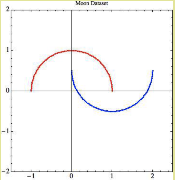


Figure : Moon Dataset

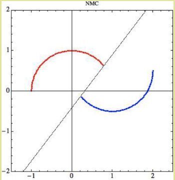


Figure : NMC

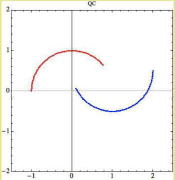


Figure : Quantum Classifier

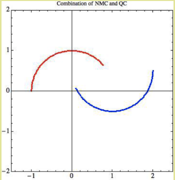


Figure : NMC & Quantum Classifier

Table : Moon Dataset

	E	E1	E2	**Pr**	**k**	TPR	FPR
NMC	**0.22**	0.22	0.22	**0.78**	**0.56**	0.78	0.22
QC	**0.18**	0.14	0.22	**0.822**	**0.64**	0.82	0.18

By randomly dividing the dataset in a training set (80%) and in a test set (20%), the average over 100 experiments gives: $NMC - Error = 22.32 \pm 6.32$; $Q - Error = 17.85 \pm 5.46$.

Banana Dataset

Banana Dataset: 5300 patterns; 2376 belonging to the first Class and 2924 to the second Class.

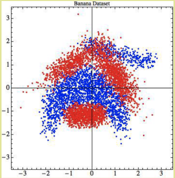


Figure : Banana Dataset

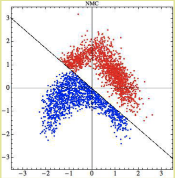


Figure : NMC

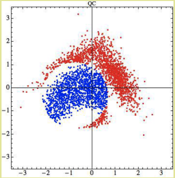


Figure : Quantum Classifier

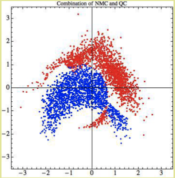


Figure : NMC & Quantum Classifier

Table : Banana Dataset

	E	E1	E2	**Pr**	**k**	TPR	FPR
NMC	**0.447**	0.423	0.468	**0.554**	**0.108**	0.555	0.445
QC	**0.418**	0.382	0.447	**0.585**	**0.168**	0.585	0.415
NMC-QC	**0.345**	0.271	0.406	**0.661**	**0.317**	0.662	0.338

By randomly dividing the dataset in a training set (80%) and in a test set (20%), the average over 100 experiments gives: $NMC - Error = 44.88 \pm 1.74$; $Q - Error = 41.57 \pm 1.21$.

3Gaussian Dataset

3Gaussian Dataset: 450 Patterns allocated in three Classes.

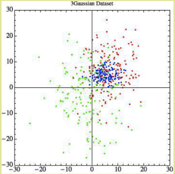


Figure : 3Gaussian Dataset

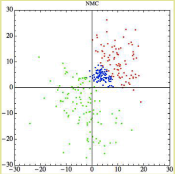


Figure : NMC

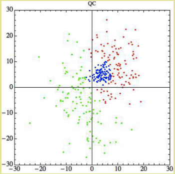


Figure : Quantum Classifier

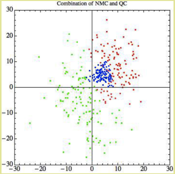


Figure : NMC & Quantum Classifier

Here we randomly divide the dataset in a Training set (80% of the patterns) and a Test set (20% of the patterns). We calculate the average over 100 runs for each experiments.

A full comparison

Datasets		NMC			QNMC		
Name	Tr/Test(d)	ACC	TPR	TNR	ACC	TPR	TNR
Banana	4240/1060(2)	55.0 ± 1.8	57.1 ± 2.5	53.4 ± 2.2	**71.0 ± 1.2**	**70.5 ± 1.6**	**71.4 ± 1.8**
Gaussian	160/40(2)	55.5 ± 7.7	61.4 ± 10.7	49.7 ± 11.4	**76.2 ± 5.6**	**70.9 ± 8.1**	**81.8 ± 8.7**
Moon	160/40(2)	77.9±5.7	78.7 ± 9.2	76.9 ± 9.2	**88.9 ± 4.4**	**94.3 ± 5.8**	**83.4 ± 7.9**
Diabetes	614/154(8)	63.4 ±3.9	**73.5 ± 4.6**	44.9 ± 7.2	**68.7 ± 3.2**	69.5 ± 4.1	**67.1 ± 5.1**
Cancer	546/137 (10)	**96.4 ± 1.4**	**97.9 ± 1.3**	93.5 ± 3.0	93.7 ± 1.9	90.4 ± 2.9	**100.0 ± 0.0**
Liver	463/116(10)	53.8 ± 4.2	39.3 ± 5.9	**89.3 ± 5.6**	**64.3 ± 3.5**	**59.6 ± 4.2**	75.9 ± 6.9
Ionosphere	280/71(34)	72.9 ± 4.5	**74.0 ± 7.6**	72.3 ± 6.4	**83.7 ± 4.3**	58.6 ± 9.9	**98.4 ± 1.6**

The Quantum Centroid is not invariant under rescaling $\rightarrow$ the "Quantum" Classifier is not invariant under rescaling!

Is it an Embarrasment or is it an Asset?

The Error is dependent on both the rescaling of the Patterns and the different encoding.

We show how the Error changes by changing the rescaling and for two different encodings.

Rescaling

How the Error of the Quantum Classifier changes by ranging the value of the rescaling.

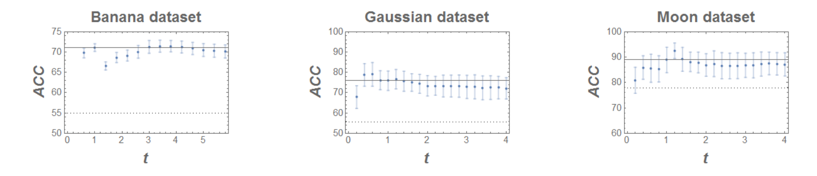


Figure : Accuracy and Rescaling

Further developments

As further developments, it will be checked whether the Quantum Classifier could bring some benefit for practical implementations, such as

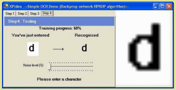


Figure : Handwriting

Figure : Fingerprint Recognition

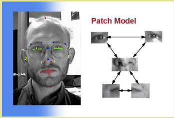


Figure : Face Recognition

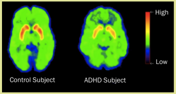


Figure : Biomedical Imaging

Observations and Open Problems

- The choise of the "best" Encoding (and/or the best Rescaling) is mostly empirical and it is strictly dependent on the Database (*No Free Lunch Theorem*).
- A comparison with more performant classifiers (Linear Discriminant Analysis, Quadratic Discriminant Analysis ...) could be investigated. The NMC and the Quantum Classifier are based on the concepts of **centroid** and **distance** only.

Suggestions are wellcome!

G. Sergioli, E. Santucci, L. Didaci, J.A. Miskczak, R. Giuntini, *Pattern Recognition on the Bloch Sphere*, arXiv:1603.00173 (2016).