Machine Learning in a Quantum World

Sébastien Gambs

Joint work with Esma Aïmeur and Gilles Brassard
based on Canadian AI 2006 and work in progress

18 August 2006

Université de Montréal

Introduction

Machine Learning

Machine Learning (ML) is the field that studies techniques to *give to machines the ability to learn from past experience*.

Supervised learning:

- Predict the class of an object (*classification*)
- or some unobserved characteristic (*regression*) based on observations on this object.

Unsupervised learning:

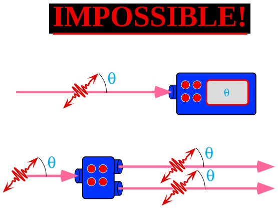
- Discover clusters of similar objects (*clustering*)
- find a meaningful low dimensional representation (*dimensionality reduction*)
- or estimate the distribution of the data (*density estimation*).

Quantum Information Processing

Quantum Information Processing (QIP) is the field that studies *the implication of quantum mechanics for information processing purposes*.

Quantum information is very different from its classical counterpart

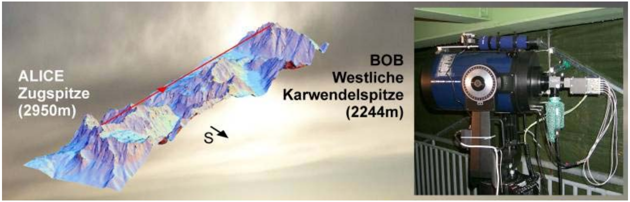
- It can exist in a *superposition* of states
- Quantum states can be *entangled*
- They can be *teleported*
- *They cannot be measured reliably*
- They are *disturbed by observation*
- *They cannot be cloned*
- ...



Joining classical and quantum information

Together, quantum and classical information can realize wonders such as

- Factorize efficiently large integers (Shor 94)
- Search elements in an unstructured database in time $\Theta(\sqrt{n})$ (Grover 96)
- Quantum cryptography (Bennett and Brassard 84)
- ...



Previous encounters of ML with QIP

- Study and comparison of learnability in the classical and quantum settings (Servedio and Gortler 04)
- Quantum neural networks (Ezhov and Berman 03)
- Design of classical clustering algorithms inspired from quantum mechanics (Horn and Gottlieb 01)
- Application of the maximum likelihood principle to quantum channel modelling
(Ziman, Plesch, Bužek and Štelmachovič 05)
- Quantum Bayesian calculus (Kuzmin and Warmuth 06)
- . . .

Novel task: Learning in a quantum world

Initial interrogation: Generally, ML learns from a training set that contains *classical observations* about *classical objects*.
What would happen if the training dataset contained *quantum objects*?

Main motivations:

- For people of the QIP community: Can the machine learning paradigm provide a constructive approach to resolve some quantum detection scenarios?
- For machine learning people: How does the change of physical theory influence the learning process?

Learning in a quantum world

Scenario

A space probe sent in exploration far away in the galaxy has encountered some quantum phenomena and sampled from them. These samples would constitute the training dataset.
What can be learned from this data?

Alternative scenario: A physicist in his laboratory has encountered some quantum phenomena during his experiments.
What can be learned from the observations he made about them?

ML with a classical dataset

A classical training dataset containing n data points:
$D_n = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ where:

- $\mathbf{x}_i$ are *observations* on the *characteristics* of the i^{th} object (or data point)
- and y_i is the *class* of that object.

Typical example: if each object is described using d real-valued attributes then $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$ for binary classification.

Remark: in unsupervised learning, as opposed to supervised learning, the y_i values are unknown.

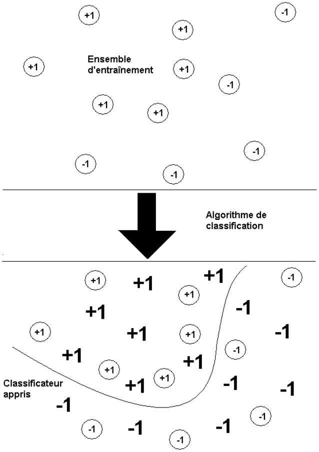
Example 1 of machine learning task: Classification

Classification: from the training dataset, learn a classifier that can be used later to *predict the class of a new object based on observations on this object*.

More formally: learn a function f, called *classifier*, which can associate to a vector of observations $\mathbf{x}$, its corresponding class y.

A few examples : music genre classification, spam detection, recognition of the digital fingerprints or the face of a person, . . .

Illustration of the classification task (a French example :-))



Example 2 of machine learning task: Clustering

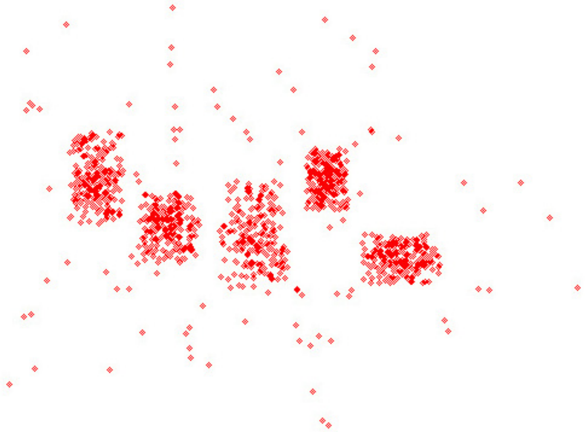
Clustering: try to *discover natural clusters which are hidden inside the data*.

Recall that in the unsupervised setting: $D_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$.

Formal goal: associate to each $\mathbf{x}$, a category (or cluster) $y \in \{1, \ldots, k\}$ such that similar objects are grouped together in the same cluster (*intra-similarity*) and dissimilar objects are put in different clusters (*inter-dissimilarity*).

Examples of applications: find the typical sociological categories existing inside a population, automatically group songs according to their genre, ...

Illustration of the clustering task



The case of ML with a quantum dataset

A quantum training dataset containing n quantum states:
$D_n = \{(|\psi_1\rangle, y_1), \ldots, (|\psi_n\rangle, y_n)\}$ where:

- $|\psi_i\rangle$ is i^{th} quantum state of the training set
- and y_i is the *class* of that quantum state.

Typical example: for a quantum state living in a Hilbert space formed by d qubits, $|\psi_n\rangle \in \mathbb{C}^{2^d}$ and $y_i \in \{-1, +1\}$ for binary classification.

Remark: further generalization would include quantum superposition of classes as well.

Learning classes

Definition: $L_{goal}^{context}$ is a learning class, where *goal* is the learning goal and *context* the form of the training dataset and/or the learner's abilities.

Example of learning classes:

- L_q^c: all descriptions of the quantum states are given classically.
- $L_q^{\otimes k}$: we received k copies of each training quantum state.
- L_c^c: ML in a classical world.
- L_c^q: classical ML goal but with the help of a quantum computer.

Possible learning strategies

Quantum classification: task of predicting the class of an unknown quantum state $|\psi_?\rangle$ *given a single copy of this state*.

If $D_n \in L_q^c$, it is possible to *maximize the probability of a good guess* of the class of $|\psi_?\rangle$ or *minimize the probability of making a wrong guess* (*unambiguous discrimination*).

If $D_n \in L_q^{\otimes k}$, possible strategies include:

(1) Estimation of the training set by making measurements (joint or not) on some of the copies

(2) Classification mechanism using the copies only when the time to classify $|\psi_?\rangle$ comes or

(3) Hybrid strategy based on (1) and (2).

Some facts on the hierarchy of quantum learning classes

- $L_q^{\otimes k} \equiv_\ell L_q^c$ as $k \to \infty$.
  "An infinite number of copies is as a good as a classical description (due to quantum tomography)."
- $L_q^{\otimes 1} \leq_\ell \ldots \leq_\ell L_q^{\otimes k} \leq_\ell L_q^{\otimes k+1} \leq_\ell \ldots \leq_\ell L_q^c$.
  "Adding more copies can never hurt."
- $L_q^{\otimes k} + L_q^{\otimes 1} \leq_\ell L_q^{\otimes k+1}$, where "$+$" denotes a restriction that the first k copies must be measured separately from the the last.
  "Joint measurement can sometimes give more information than separate ones."

Bounds on the training error for classification

Results from quantum detection and estimation theory can give bounds on the best training error we could hope for.

Example:

- If $D_n \in L_q^c$, probability of distinguishing between the two classes is bounded above by $(1 + D(\rho_-, \rho_+))/2$, where $D(\rho_-, \rho_+)$ is a distance measure between ρ_- and ρ_+ (Helstrom 76)[1].
- Recent bounds for *unambiguous discrimination* (Herzog and Bergou 05).

Remark: The goal of a quantum ML algorithm is to give a constructive way to come close (or to achieve) these bounds.

[1] $\rho_- = \frac{1}{m_-} \sum_{i=1}^{n} \frac{1-y_i}{2} |\psi_i\rangle\langle\psi_i|, \rho_+ = \frac{1}{m_+} \sum_{i=1}^{n} \frac{1+y_i}{2} |\psi_i\rangle\langle\psi_i|$

Illustration: Clustering with a quantum dataset

Notion of fidelity

The fidelity $Fid(|\psi\rangle, |\phi\rangle) = |\langle\phi|\psi\rangle|^2$ is a *measure of similarity* between quantum states.

It ranges from

- 0 if the states are *orthogonal* (i.e. perfectly distinguishable) to
- 1 if the states are *identical*.

Some properties of fidelity:

- *Symmetry*: $Fid(|\psi\rangle, |\phi\rangle) = Fid(|\phi\rangle, |\psi\rangle)$.
- *Invariance under unitary transformations*: $Fid(|U\psi\rangle, |U\phi\rangle) = Fid(|\psi\rangle, |\phi\rangle)$, where U is any unitary operation.

Remark: the fidelity can be transformed into a metric obeying the triangle inequality by using $Dist(|\psi\rangle, |\phi\rangle) = \arccos Fid(|\psi\rangle, |\phi\rangle)$.

Control-Swap test: Description

Control-Swap (C-Swap) test: quantum operation that can be used to estimate the fidelity between two unknown quantum states $|\psi\rangle$ and $|\phi\rangle$ (BBDEJM[2] 96, BCWW[3] 01).

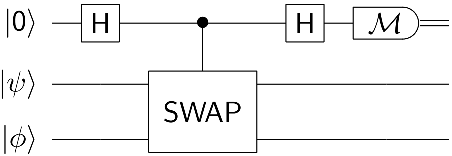


Figure: Circuit of the Control-SWAP test.

[2] Barenco, Berthiaume, Deutsch, Ekert, Jozsa and Macchiavello.
[3] Buhrman, Cleve, Watrous and de Wolf.

Control-Swap test: Fidelity estimator

Result of applying the C-Swap test is:

- $|0\rangle$ with probability 1 if $|\psi\rangle$ and $|\phi\rangle$ are identical.
- Otherwise, $|1\rangle$ with probability $\frac{1}{2} - \frac{1}{2}|\langle\phi|\psi\rangle|^2$.

The C-Swap test provides an *estimator* of the fidelity.

With k copies of $|\psi\rangle$ and $|\phi\rangle$, we can run it k times and estimate $Fid(|\psi\rangle, |\phi\rangle)$ as $1 - 2 \times \# |1\rangle / k$.

Note: A side effect of the C-SWAP test is to irreversibly disturb the input states.

Possible quantum clustering strategies (1)

Quantum clustering: *Group similar quantum states together* and *separate dissimilar quantum states into different clusters*.

If $D_n \in L_q^{\otimes k}$, it is possible to perform a quantum tomography by using all the copies and then reconstruct of the classical description of each training state.

The quality of the reconstruction depends on k, the number of available copies, and $d = 2^n$, the dimension of the Hilbert space for n qubits: $Fid(|\psi_{guess}\rangle, |\psi_{true}\rangle) = \frac{k+1}{k+d}$.

Problem: This requires an exponential number of copies in the number of qubits for good reconstruction.

Possible quantum clustering strategies (2)

More clever approaches:

- Estimate the fidelity between each pair of the training set by using the C-Swap test several times and then run a classical clustering algorithm (such as k-medians) on the data thus obtained.
- Adapt a classical algorithm to the quantum setting.

  Example: An agglomerative algorithm that grows clusters around *quantum seeds* in a adaptive manner.

(Classical) K-medians algorithm

1. Choose randomly k datapoints which will be the initial centers of the clusters.
2. **Do**
   - For each $\mathbf{x}_i$, attach him to the closest center.
   - For each cluster, recompute the center by setting it to the datapoint which is at minimal distance from the other points in the cluster.

   While there is no stabilization of the cluster centers
3. Return the discovered clusters and their centers.

Simulated experimentation as a proof of concept

- Synthetic data of 5 clusters.
- The centre of each cluster is a 13-qubit pure state ($d = 8192$) generated randomly and uniformly according to the Haar measure.
- 20 pure states per cluster obtained by random perturbation of the centre and such that the fidelity with the centre is never below some threshold.

Algorithm:

1. Construction of a *similarity matrix* by estimating the fidelity between each pair of states using the C-Swap test.
2. Running a variant of k-medians on this similarity matrix, which groups the quantum states in 5 clusters.

Clustering quality vs fidelity threshold

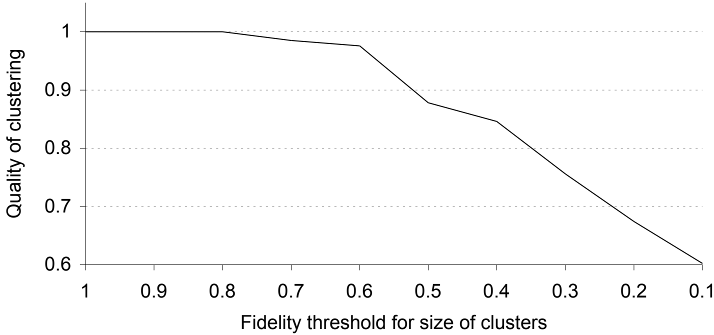


Figure: Evolution of the clustering quality (averaged over 5 trials) as the fidelity threshold decreases. Quality value of 1 = perfect clustering. Random "clustering" results in a quality of approximately 0.37.

Work in progress: Towards a quantum analogue of boosting

Ensemble methods, boosting and AdaBoost

Illustration of the motto "Unity and diversity make strength".

Principle of ensemble methods: Several classifiers are trained and then combined into a single efficient mechanism (generaly using a voting mechanism).

Philosophy of boosting: construction of an efficient classifier by iteratively adding several *weak classifiers* whose predictions need only to be a little bit better than a random guess.

AdaBoost (*Ada*ptive *Boost*ing): state-of-the-art boosting algorithm (Freund and Schapire 97).

AdaBoost algorithm

Iterative boosting algorithm that stops after T iterations. Each datapoint has a weight which changes at each iteration. This weight reflects *how hard it is to classify a datapoint*.

Details of one iteration:

1. Find the weak classifier that *minimizes the weighted error* on the datapoints.
2. Compute the coefficient of this weak classifier from its weighted error.
3. Reweight the datapoints.
   For each datapoint, if it is correctly classified by the weak classifier of the current iteration decrease its weight, otherwise increase it.

Weak measurement

Weak measurement: measurement which learns only a small amount of information about an unknown quantum state $|\psi_?\rangle$, thus disturbing it only a little.

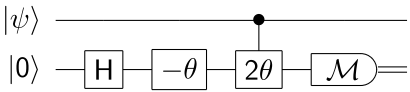


Figure: Example of a weak measurement circuit. θ is a unitary operation (here a small rotation) close to the identity and whose strength can be tune to change the amount of information acquired.

Resemblance between a weak classifier and a weak measurement.
Currently finishing to develop a quantum version of AdaBoost.

Conclusions and open problems

Conclusions

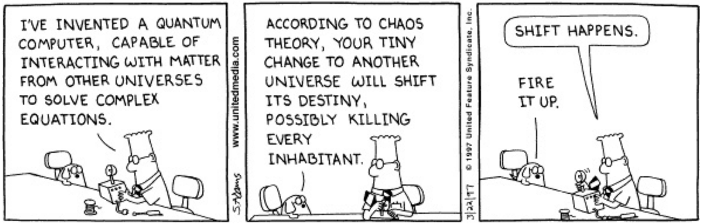
- Novel learning task and new model of ML.
- Using quantum information has a great impact on the learning process.
- Offers a lot of interesting questions and perspectives whose study could lead to insights both in ML and QIP.

Open problems

- Define analogues of classical notions of ML in the quantum setting (such as the *generalization error* or the *margin*).
- Study the effect of noise (both classical and quantum).
- Extend or generalize classical ML algorithms to the quantum world.
  Currently in development: Quantum analogue of boosting and ID3.
- Devise brand new ML algorithms.
- ...
- Suggestions?

This is the end!!!

Thank you for your attention. Questions?



References: Some encounters between ML and QIP

- Aimeur, E., Brassard, G. and S. Gambs: *Machine learning in a quantum world*. Proceedings of Canadian AI 2006, pages 433-444, 2006.
- Ezhov, A.A. and G.P. Berman: *Introduction to quantum neural technologies*. Rinton Press, 2003.
- Horn, D. and A. Gottlieb: *The method of quantum clustering*. Proceedings of the Neural Information Processing Systems (NIPS'01), pages 769-776, 2001.
- Servedio, R. and A. Gortler: *Equivalences and separations between quantum and classical learnability*. SIAM Journal of Computing, 33(5):1067-1092, 2004.
- Ziman, M., M. Plesch, V. Bužek and P. Štelmachovič: *Process reconstruction: From unphysical to physical maps via maximum likelihood*. Physical Reviews A **72**, article 022106, 2005.

This is the end (bis)!!!

Many thanks to the organizer!!!

Bonus slide: Towards a quantum analogue of ID3

- ID3 (Quinlan 86) is a ML algorithm that produce a decision tree and uses Shannon entropy as a splitting criterion.
- The von Neumann entropy is a measure of the uncertainty we have about a quantum state.
- Strong relationship between Shannon and von Neumann entropies: $S(\rho) = H(\lambda_i)$ if λ_i and $|\psi_i\rangle$ are respectively the *eigenvalues* and *eigenvectors* of ρ.
- Idea: construct a quantized version of ID3 that uses von Neumann entropy as a splitting criterion.