An overview on Quantum Machine Learning

Davide Pastorello

Department of Information Engineering and Computer Science
University of Trento

Advanced Topics in Machine Learning and Optimization

Dec 9, 2021

Introduction

Quantum Machine Learning (QML): ML with quantum computers.

First proposals (20 years ago)

Quantum associative memories and pattern recognition.
D. Ventura, T. Martinez, *Information Sciences* ***124*** (2000)
R. Schützhold, *Phys. Rev. A* ***67*** (2003)

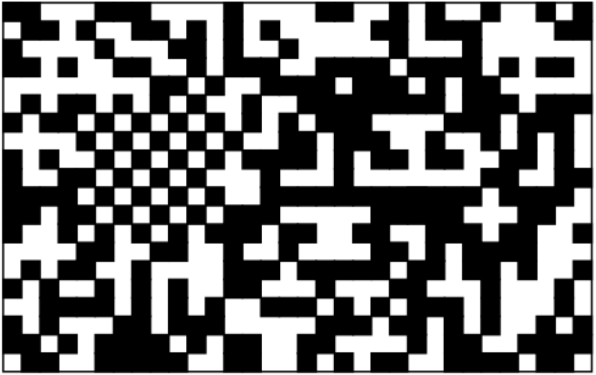
Main results on QML (since 2013)

Algorithm	**Quantum speedup**
K-medians	Quadratic/Exponential
Hierarchical clustering	Quadratic
K-means	Exponential
Principal component analysis	Exponential
Support vector machines	Exponential
Nearest neighbors	Quadratic / Exponential
Neural networks	?

Current challenge

Devising *quantum learning mechanisms* for available or near-term quantum machines.

Introduction



Implementing the Quantum Fourier transform, the chessboard structure can be found **easily** by a quantum computer.

Introduction

Data processing based on quantum effects like *quantum superposition* and *quantum entanglement*.

Examples of quantum advantages

- Calculation of Euclidean distance in $\mathbb{R}^d$:

  Classical time complexity $O(d)$.
  Quantum time complexity $O(\log d)$.

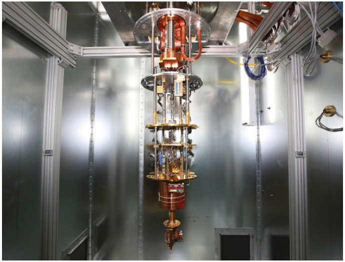
- Search in an unstructured database of N items:

  Query complexity of exhaustive search: $O(N)$.
  Query complexity of Grover's algorithm: $O(\sqrt{N})$.

Qdist and *Grover* are typical subroutine of QML algorithms.

Introduction

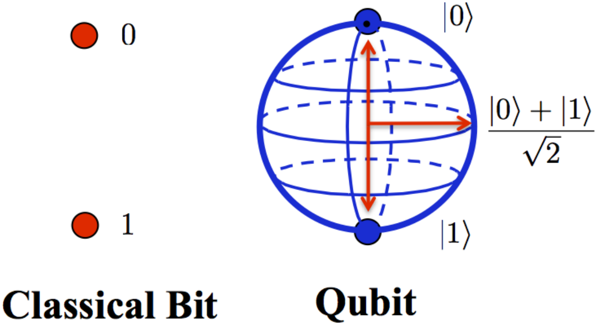
Examples of quantum computers:



D-Wave

IBMQ

Introduction

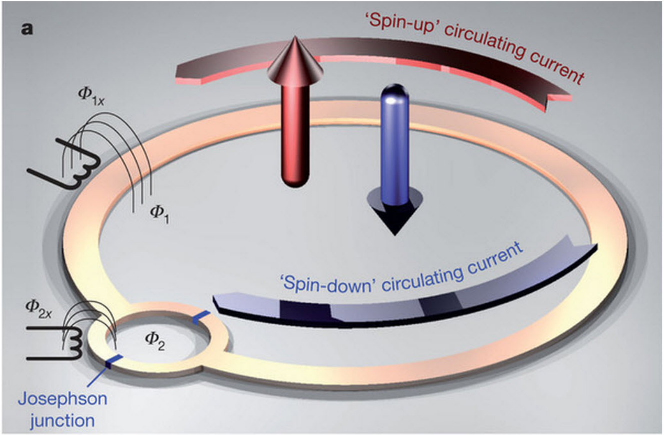


Examples of qubits:

- Particle with spin 1/2;
- Polarized Photons;
- Controlled superconducting circuits.

Introduction

Qubit as a superconducting circuit:



Qubits

Quantum state superposition

The state of a qubit is a unit vector in $\mathbb{C}^2$:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad \alpha, \beta \in \mathbb{C} \ , \quad |\alpha|^2 + |\beta|^2 = 1$$

$\{|0\rangle, |1\rangle\}$ computational basis.

A measurement process affects the qubit state

Measurement of a qubit (e.g. we measure the polarization of a photon)

$$\mathbb{P}(0) = |\alpha|^2 \qquad \mathbb{P}(1) = |\beta|^2$$

The qubit state after the measurement is $|i\rangle$ if the outcome is $i = 0, 1$.

Remark:
The vectors $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ with $\theta \in \mathbb{R}$ represent the same *physical state*.

Qubits

Tensor product

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathbb{C}^2 \qquad |\varphi\rangle = \begin{pmatrix} \gamma \\ \delta \end{pmatrix} \in \mathbb{C}^2$$

$$|\psi\rangle \otimes |\varphi\rangle = \begin{pmatrix} \alpha\gamma \\ \alpha\delta \\ \beta\gamma \\ \beta\delta \end{pmatrix} \in \mathbb{C}^4$$

2 qubits (as a composite system) are described in:

$$\mathbb{C}^2 \otimes \mathbb{C}^2 := \text{span}\{|\psi\rangle \otimes |\varphi\rangle \,:\, |\psi\rangle, |\varphi\rangle \in \mathbb{C}^2\} = \mathbb{C}^4.$$

n qubits (as a composite system) are described in:

$$(\mathbb{C}^2)^{\otimes n} = \mathbb{C}^{2^n}.$$

Entangled qubits

Let $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ be the state of a qubit pair.

$|\Psi\rangle$ is said to be:

- **separable** if it has form $|\Psi\rangle = |\psi\rangle \otimes |\varphi\rangle \equiv |\psi\varphi\rangle$;
- **entangled** otherwise.

Example

Entangled state:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Measure the **first qubit**:

- The probability of measuring 0 is $\frac{1}{2}$.
- If the outcome is 0 then the post-measurement state is $|\Psi_0\rangle = |00\rangle$.
- Non-local action on the **second qubit** (quantum correlation).

EPR paradox: inconsistency with QM and Einstein's locality.

A new kind of information

Empirical evidence:
Quantum randomness, state superpositions, entanglement are **physical phenomena** not simply *theoretical interpretations*.

$$\Downarrow$$

Encoding information into qubits (or more general quantum objects) allows completely new kinds of:

information processing, telecommunication, data security...

Quantum encoding

Basis encoding

Elements of $X = \{0,1\}^n$ encoded into the states of n qubits

$$\{0,1\}^n \ni (x_1 \cdots x_n) \mapsto |x_1\rangle \otimes \cdots \otimes |x_n\rangle \equiv |x_1 \cdots x_n\rangle \in \mathbb{C}^{2^n}$$

Observation:

The system (n-qubit register) can be prepared in a superposition of data that can be processed in parallel by linearity.

The register can be initialized in a superposition of all data:

$$|\Psi\rangle = \sum_{x=0}^{2^n-1} |x\rangle$$

Quantum encoding

Amplitude encoding

$\mathbf{x} \in \mathbb{C}^d$ with $\| \mathbf{x} \| := \sqrt{\sum_{i=1}^{d} |x_i|^2} = 1$.

Consider a quantum register of $\log d$ qubits.

Components of $\mathbf{x}$ encoded into the amplitudes of a quantum state $|\psi_{\mathbf{x}}\rangle$

Let $\{|i\rangle\}_{i=1,\dots,d}$ be the computational basis:

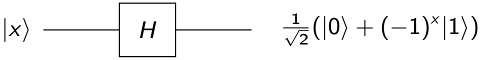
$$|\psi_{\mathbf{x}}\rangle = \sum_{i=1}^{d} x_i |i\rangle.$$

Remark:
Quantum amplitudes cannot be directly observed.
From $|\psi_{\mathbf{x}}\rangle$ we can retrieve only $|x_i|^2$.

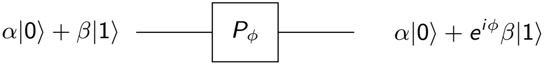
Quantum gates

Let $\{|0\rangle, |1\rangle\}$ be the computational basis

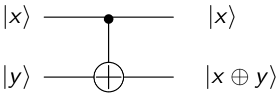
- Hadamard gate:



- Phase gate:



- CNOT gate:



Theorem: $\{H, P_{\pi/4}, CNOT\}$ is a universal set for quantum computation.

A quantum binary classifier

The model

Let $\{(\mathbf{x}_i, y_i)\}_{i=0,\dots,N-1}$ be a training set where:
$x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$ for any $i = 0, \dots, N-1$

$$y(\mathbf{x}) := \operatorname{sgn}\left(\sum_{i=0}^{N-1} y_i \cos(\mathbf{x}_i, \mathbf{x})\right) \qquad \cos(\mathbf{x}, \mathbf{z}) := \frac{\mathbf{x} \cdot \mathbf{z}}{\| \mathbf{x} \| \| \mathbf{z} \|} \qquad \mathbf{x}, \mathbf{z} \in \mathbb{R}^d.$$

Quantum implementation on the IBM ibmq_melbourne

- Quantum superposition of training vectors;
- Test instance in quantum superposition of the two classes;
- Cosine similarities computed by SWAP test.

Classical complexity: $\mathcal{O}(Nd)$
Quantum complexity: $\mathcal{O}(\log(Nd))$

A quantum binary classifier

Training set stored in a n-qubit register within the amplitude encoding.
$\log N$-qubit register, with Hilbert space $\mathcal{H}_{index} \simeq (\mathbb{C}^2)^{\otimes \log N}$, to encode the indexes of training data vectors.

We can construct the state:

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

where $\mathcal{H}_l$ is a 1-qubit register encoding the labels with $b_i = \frac{1-y_i}{2} \in \{0, 1\}$.

We can construct also:

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

where $|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$.

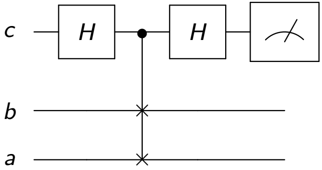
A quantum binary classifier

Add 1 ancillary qubit to the registers and construct:

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

that can be retrieved from the QRAM in time $O(log(Nd))$.

Perform the SWAP test:



where qubit b is prepared in $|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$ and qubit c is prepared in $|0\rangle$.

A quantum binary classifier

The probability of measuring 1 on qubit c is:

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

(Check as an exercise!)

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

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

$\langle i|k\rangle = \delta_{ik}$ and $\langle b_i|0\rangle - \langle b_i|1\rangle = 1 - 2b_i = y_i$ for any $i = 0, ..., N-1$

The model:

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

Quantum implementation of the model based on:

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

A quantum binary classifier

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

repeat
  initialize the register $\mathcal{H}_{index} \otimes \mathcal{H}_n \otimes \mathcal{H}_l$ and an ancillary qubit a in the state $|\Psi\rangle$;
  initialize a qubit b in the state $|-\rangle$;
  perform the SWAP test on a and b with control qubit c prepared in $|0\rangle$;
  measure qubit c;
until *desired accuracy on the estimation of* $\mathbb{P}(1)$;
Estimate $\mathbb{P}(1)$ as the relative frequency $\hat{\mathbb{P}}$ of outcome 1;
if $\hat{\mathbb{P}} > 0.25$ **then**
  return $y = -1$
else
  return $y = 1$
end

Overall complexity within an error ϵ in the estimation of $\mathbb{P}(1)$: $O(\epsilon^{-2} \log(Nd))$

Quantum clustering

Qdist is a quantum algorithm based on the SWAP test to calculate Euclidean distance in logarithmic time.

Grover is a quantum search algorithm with quadratic speedup.

Example: K-medians clustering

Input: Data set $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$, number of clusters K
Result: Partition of $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ into K clusters

initialize K centroids $C_1, ..., C_K$ from the elements of the dataset V;
repeat
    foreach $i \leftarrow 1, ..., N$ **do**
        Qdist $(\mathbf{x}_i, C_j)$ $\forall j = 1, ..., K$;
        find $\text{argmin}_j \parallel \mathbf{x}_i - C_j \parallel$ with **Grover**;
    end
    construct the cluster $P_j = \{\mathbf{x}_i : C_j \text{ is the nearest centroid}\}$ for all $j = 1, ..., K$;
    foreach $j \leftarrow 1, ..., K$ **do**
        use **Qdist** and **Grover** for centroid calculation;
    end
until *convergence*;
return $P_1, ..., P_K$

Some QML schemes designed for gate-based quantum computers

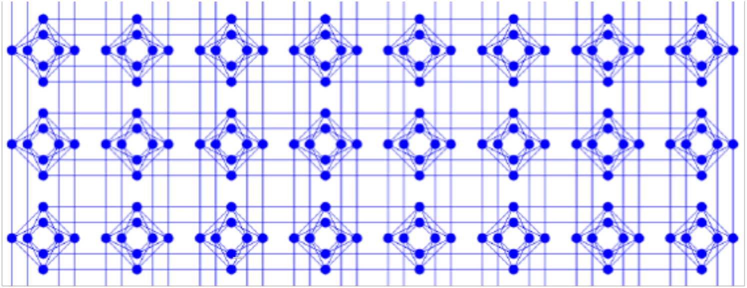
- **Quantum divisive clustering**
  W. Aïmeur et al. *Quantum clustering algorithms* ICML '07: Proceedings of the 24th international conference on Machine learning (2007)
- **Quantum principal component analysis**
  S. Lloyd, et al. *Quantum principal component analysis* Nature Physics 10, 631 (2014)
- **Quantum support vector machine**
  P. Rebentrost et al. *Quantum support vector machine for big data classification* Phys. Rev. Lett. 113, 130503 (2014)
- **Quantum nearest neighbor**
  N. Wiebe et al. *Quantum Algorithms for Nearest-Neighbor Methods for Supervised and Unsupervised Learning* Quantum Information and Computation 15(3,4): 0318- 0358 (2015)
- **Quantum perceptron** M. Schuld *Simulating a perceptron on a quantum computer* Physics Letters A, 379, pp. 660-663 (2015)
- **Quantum meta-learning** M. Wilson et al. *Optimizing quantum heuristic with meta-learning* Quantum Machine Intelligence 3 (2021)

Quantum Annealing

Quantum Annealers

The hardware is a *quantum spin glass*, i.e. a collection of qubits arranged in the vertices of a graph (V, E) where edges represent the interactions between neighbors.

Example: *D-Wave Chimera topology*



Quantum Annealing

Annealing process (annealing time $20\mu s$)

By energy dissipation the quantum system evolves in the **ground state** (the less energetic state) corresponding to the **solution** of a given optimization problem.

Quantum Annealer task

Minimization (w.r.t. $\mathbf{z}$) of the cost function (Ising model):

$$E(\theta, \mathbf{z}) = \theta_0 + \sum_{i \in V} \theta_i z_i + \sum_{(i,j) \in E} \theta_{ij} z_i z_j \qquad \mathbf{z} \in \{-1, 1\}^{|V|}\ ,\ \theta_0, \theta_i, \theta_{ij} \in \mathbb{R}$$

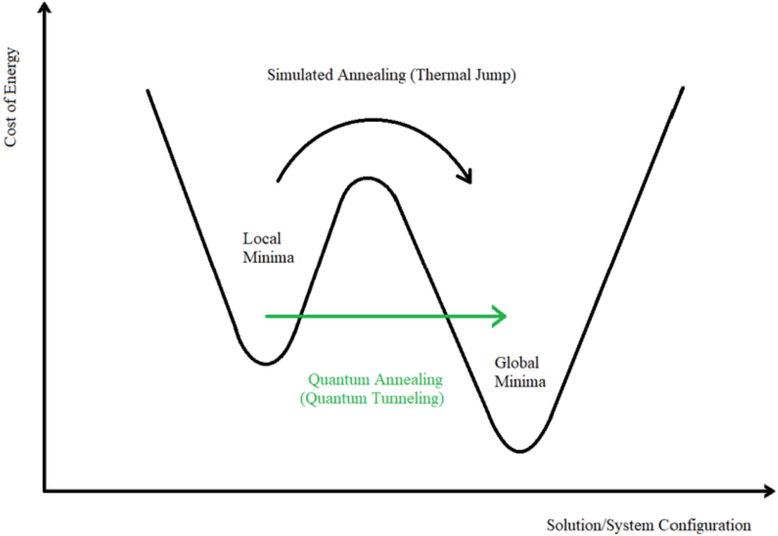
Initialization of the machine

- Assignment of the weights θ;
- Mapping the binary variables into the qubits.

QA vs Simulated Annealing

QA employs the tunnel effect to escape from local minima.
SA employs thermal hill-climbing.

Quantum Annealing



ML with quantum annealers

- **Boltzmann machine implementation**
  M.H. Amin *Quantum Boltzmann machine* Phys. Rev. X 8, 021050 (2018)
- **Classification**
  N. T. Nguyen et al. *Image classification using quantum inference on the D-wave 2X*. In: 2018 IEEE International Conference on Rebooting Computing (ICRC), pp. 1-7. IEEE (2018)
- **Clustering**
  V. Kumar et al. *Quantum Annealing for Combinatorial Clustering* Quantum Inf Process (2018) 17: 39
- **Training of a SVM**
  D. Willsch et al. *Support vector machines on the D-wave quantum annealer* Computer Physics Communications 248, 107006 (2020)

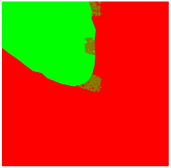
Quantum-inspired ML

Using quantum formalism to devise classical ML algorithms

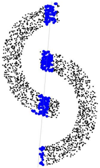
Quantum-inspired classifiers:

- Data encoding into density operators (quantum states);
- Construction of the centroids in the space of quantum states;
- Application of *discrimination of quantum states* to attach a new data instance to the *most similar* centroid.

Example:



Quantum-inspired classifier



SVM with linear kernel

Ref.: R. Leporini and D. P.. *Support vector machines with quantum state discrimination*. Quantum Reports vol. 3, n. 3 (2021)

Conclusions

Goals of the scientific research on QML:

- Exploiting exponential improvements in time and space of quantum techniques to find new strategies to deal with big data.
- Finding promising commercial applications of quantum computing. (Nowadays attraction of investments has strong impact also on research in the academy... like it or not!)
- Increasing our knowledge about the connection between the abstract concept of *learning* and the quantum nature of the World.
- Devising *learning mechanisms* of new quantum algorithms, since providing efficient quantum algorithms is an extremely difficult task.

In other words, quantum machines must learn how to solve problems on their own...