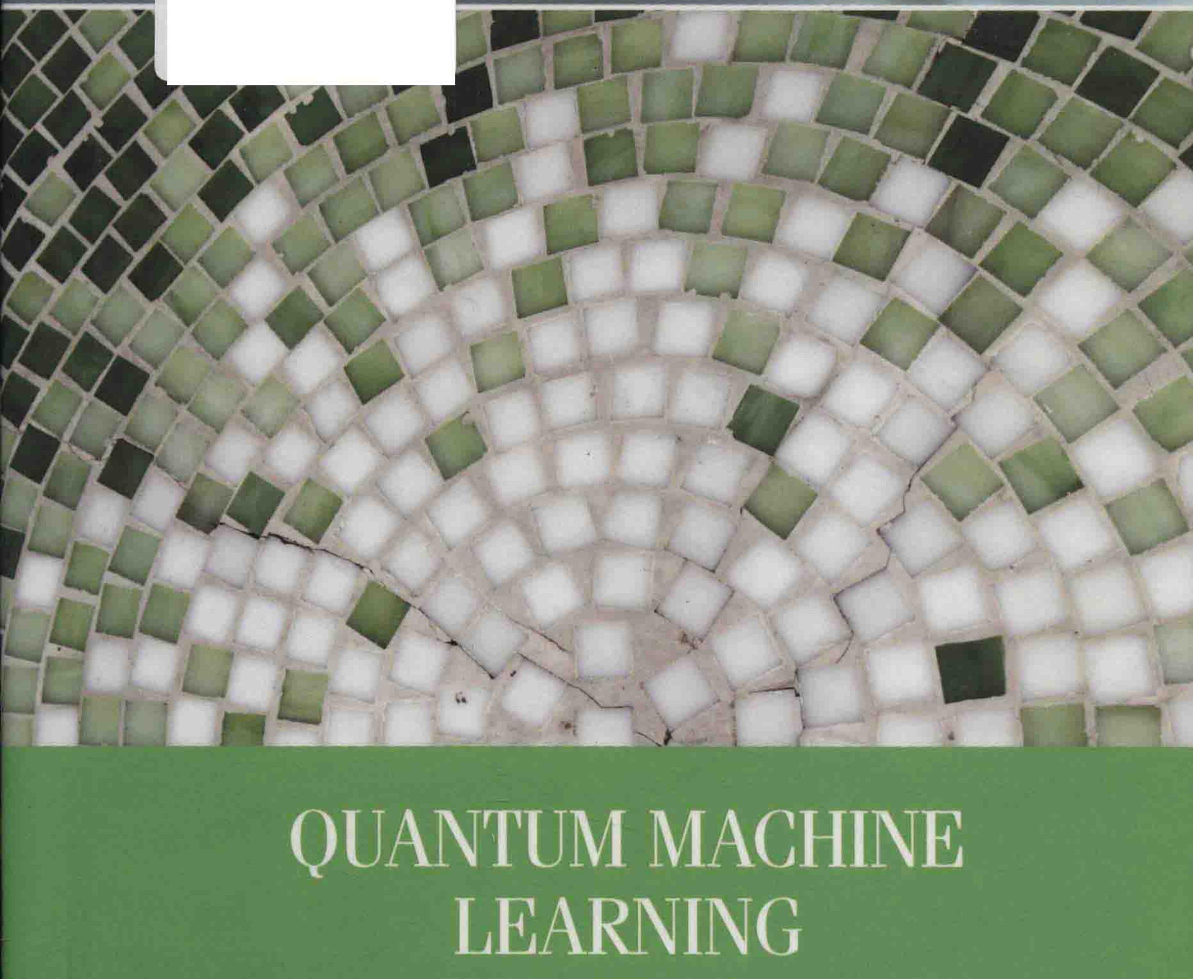
ELSEVIER INSIGHTS



QUANTUM MACHINE LEARNING

WHAT QUANTUM COMPUTING MEANS TO DATA MINING

PETER WITTEK

Quantum Machine Learning

What Quantum Computing Means to Data Mining

Peter Wittek
University of Borås
Sweden

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
32 Jamestown Road, London NW1 7BY, UK

First edition

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-800953-6

For information on all Elsevier publications
visit our website at store.elsevier.com

Quantum Machine Learning

Preface

Machine learning is a fascinating area to work in: from detecting anomalous events in live streams of sensor data to identifying emergent topics involving text collection, exciting problems are never too far away.

Quantum information theory also teems with excitement. By manipulating particles at a subatomic level, we are able to perform Fourier transformation exponentially faster, or search in a database quadratically faster than the classical limit. Superdense coding transmits two classical bits using just one qubit. Quantum encryption is unbreakable—at least in theory.

The fundamental question of this monograph is simple: What can quantum computing contribute to machine learning? We naturally expect a speedup from quantum methods, but what kind of speedup? Quadratic? Or is exponential speedup possible? It is natural to treat any form of reduced computational complexity with suspicion. Are there tradeoffs in reducing the complexity?

Execution time is just one concern of learning algorithms. Can we achieve higher generalization performance by turning to quantum computing? After all, training error is not that difficult to keep in check with classical algorithms either: the real problem is finding algorithms that also perform well on previously unseen instances. Adiabatic quantum optimization is capable of finding the global optimum of nonconvex objective functions. Grover's algorithm finds the global minimum in a discrete search space. Quantum process tomography relies on a double optimization process that resembles active learning and transduction. How do we rephrase learning problems to fit these paradigms?

Storage capacity is also of interest. Quantum associative memories, the quantum variants of Hopfield networks, store exponentially more patterns than their classical counterparts. How do we exploit such capacity efficiently?

These and similar questions motivated the writing of this book. The literature on the subject is expanding, but the target audience of the articles is seldom the academics working on machine learning, not to mention practitioners. Coming from the other direction, quantum information scientists who work in this area do not necessarily aim at a deep understanding of learning theory when devising new algorithms.

This book addresses both of these communities: theorists of quantum computing and quantum information processing who wish to keep up to date with the wider context of their work, and researchers in machine learning who wish to benefit from cutting-edge insights into quantum computing.

I am indebted to Stephanie Wehner for hosting me at the Centre for Quantum Technologies for most of the time while I was writing this book. I also thank Antonio Acín for inviting me to the Institute for Photonic Sciences while I was finalizing the manuscript. I am grateful to Sándor Darányi for proofreading several chapters.

Peter Wittek
Castelldefels, May 30, 2014

Notations

$\mathbb{1}$	indicator function
$\mathbb{C}$	set of complex numbers
d	number of dimensions in the feature space
E	error
$\mathbb{E}$	expectation value
$\mathbf{G}$	group
H	Hamiltonian
$\mathcal{H}$	Hilbert space
I	identity matrix or identity operator
K	number of weak classifiers or clusters, nodes in a neural net
N	number of training instances
P_i	measurement: projective or POVM
$\mathbf{P}$	probability measure
$\mathbb{R}$	set of real numbers
ρ	density matrix
$\sigma_x, \sigma_y, \sigma_z$	Pauli matrices
tr	trace of a matrix
U	unitary time evolution operator
$\mathbf{w}$	weight vector
$\mathbf{x}, \mathbf{x}_i$	data instance
X	matrix of data instances
y, y_i	label
$^\top$	transpose
†	Hermitian conjugate
$\lVert . \rVert$	norm of a vector
$[., .]$	commutator of two operators
$\otimes$	tensor product
$\oplus$	XOR operation or direct sum of subspaces

Contents

Part One

Fundamental Concepts

Introduction

1

The quest of machine learning is ambitious: the discipline seeks to understand what learning is, and studies how algorithms approximate learning. Quantum machine learning takes these ambitions a step further: quantum computing enrolls the help of nature at a subatomic level to aid the learning process.

Machine learning is based on minimizing a constrained multivariate function, and these algorithms are at the core of data mining and data visualization techniques. The result of the optimization is a decision function that maps input points to output points. While this view on machine learning is simplistic, and exceptions are countless, some form of optimization is always central to learning theory.

The idea of using quantum mechanics for computations stems from simulating such systems. Feynman (1982) noted that simulating quantum systems on classical computers becomes unfeasible as soon as the system size increases, whereas quantum particles would not suffer from similar constraints. Deutsch (1985) generalized the idea. He noted that quantum computers are universal Turing machines, and that quantum parallelism implies that certain probabilistic tasks can be performed faster than by any classical means.

Today, quantum information has three main specializations: quantum computing, quantum information theory, and quantum cryptography (Fuchs, 2002, p. 49). We are not concerned with quantum cryptography, which primarily deals with secure exchange of information. Quantum information theory studies the storage and transmission of information encoded in quantum states; we rely on some concepts such as quantum channels and quantum process tomography. Our primary focus, however, is quantum computing, the field of inquiry that uses quantum phenomena such as superposition, entanglement, and interference to operate on data represented by quantum states.

Algorithms of importance emerged a decade after the first proposals of quantum computing appeared. Shor (1997) introduced a method to factorize integers exponentially faster, and Grover (1996) presented an algorithm to find an element in an unordered data set quadratically faster than the classical limit. One would have expected a slew of new quantum algorithms after these pioneering articles, but the task proved hard (Bacon and van Dam, 2010). Part of the reason is that now we expect that a quantum algorithm should be faster—we see no value in a quantum algorithm with the same computational complexity as a known classical one. Furthermore, even

Quantum Machine Learning. http://dx.doi.org/10.1016/B978-0-12-800953-6.00001-3

with the spectacular speedups, the class NP cannot be solved on a quantum computer in subexponential time (Bennett et al., 1997).

While universal quantum computers remain out of reach, small-scale experiments implementing a few qubits are operational. In addition, quantum computers restricted to domain problems are becoming feasible. For instance, experimental validation of combinatorial optimization on over 500 binary variables on an adiabatic quantum computer showed considerable speedup over optimized classical implementations (McGeoch and Wang, 2013). The result is controversial, however (Rønnow et al., 2014).

Recent advances in quantum information theory indicate that machine learning may benefit from various paradigms of the field. For instance, adiabatic quantum computing finds the minimum of a multivariate function by a controlled physical process using the adiabatic theorem (Farhi et al., 2000). The function is translated to a physical description, the Hamiltonian operator of a quantum system. Then, a system with a simple Hamiltonian is prepared and initialized to the ground state, the lowest energy state a quantum system can occupy. Finally, the simple Hamiltonian is evolved to the target Hamiltonian, and, by the adiabatic theorem, the system remains in the ground state. At the end of the process, the solution is read out from the system, and we obtain the global optimum for the function in question.

While more and more articles that explore the intersection of quantum computing and machine learning are being published, the field is fragmented, as was already noted over a decade ago (Bonner and Freivalds, 2002). This should not come as a surprise: machine learning itself is a diverse and fragmented field of inquiry. We attempt to identify common algorithms and trends, and observe the subtle interplay between faster execution and improved performance in machine learning by quantum computing.

As an example of this interplay, consider convexity: it is often considered a virtue in machine learning. Convex optimization problems do not get stuck in local extrema, they reach a global optimum, and they are not sensitive to initial conditions. Furthermore, convex methods have easy-to-understand analytical characteristics, and theoretical bounds on convergence and other properties are easier to derive. Nonconvex optimization, on the other hand, is a forte of quantum methods. Algorithms on classical hardware use gradient descent or similar iterative methods to arrive at the global optimum. Quantum algorithms approach the optimum through an entirely different, more physical process, and they are not bound by convexity restrictions. Nonconvexity, in turn, has great advantages for learning: sparser models ensure better generalization performance, and nonconvex objective functions are less sensitive to noise and outliers. For this reason, numerous approaches and heuristics exist for nonconvex optimization on classical hardware, which might prove easier and faster to solve by quantum computing.

As in the case of computational complexity, we can establish limits on the performance of quantum learning compared with the classical flavor. Quantum learning is not more powerful than classical learning—at least from an information-theoretic perspective, up to polynomial factors (Servedio and Gortler, 2004). On the other hand, there are apparent computational advantages: certain concept classes