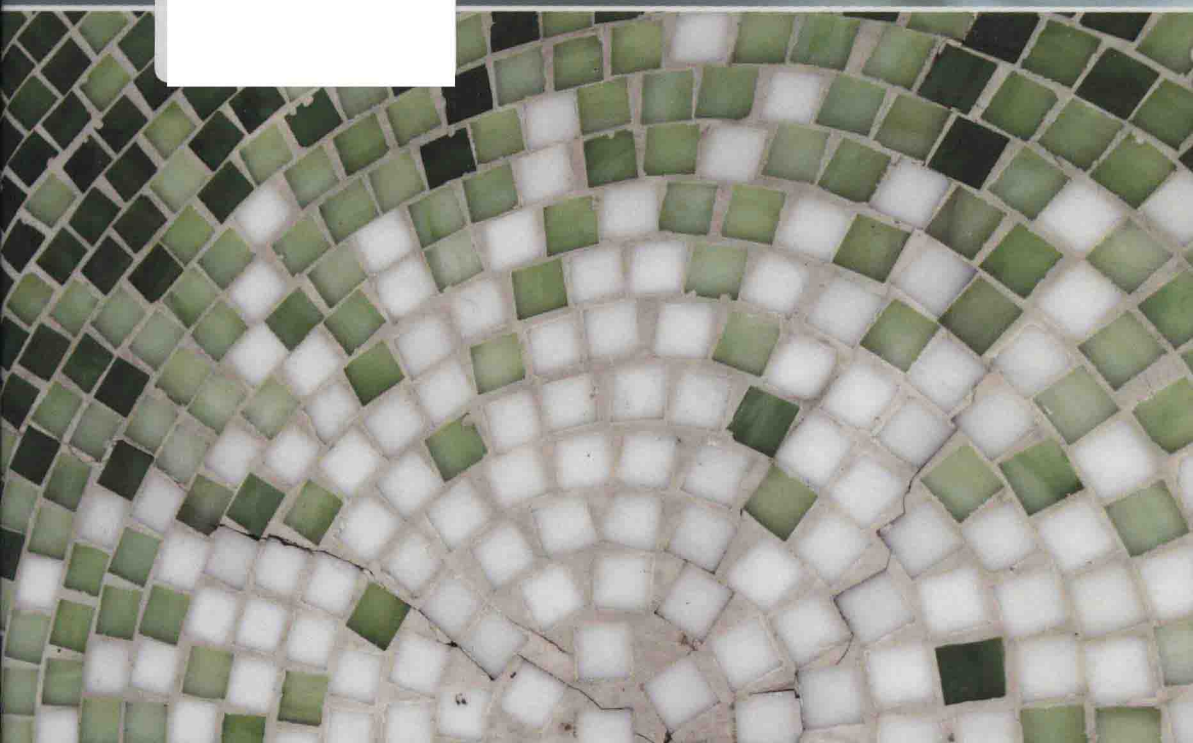




ELSEVIER INSIGHTS



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

Copyright © 2014 by Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangement with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notice

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

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



Working together
to grow libraries in
developing countries

www.elsevier.com • www.bookaid.org

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
\dagger	Hermitian conjugate
$\ \cdot\ $	norm of a vector
$[\cdot, \cdot]$	commutator of two operators
\otimes	tensor product
\oplus	XOR operation or direct sum of subspaces



Contents

Preface	ix
Notations	xi
Part One Fundamental Concepts	1
1 Introduction	3
1.1 Learning Theory and Data Mining	5
1.2 Why Quantum Computers?	6
1.3 A Heterogeneous Model	7
1.4 An Overview of Quantum Machine Learning Algorithms	7
1.5 Quantum-Like Learning on Classical Computers	9
2 Machine Learning	11
2.1 Data-Driven Models	12
2.2 Feature Space	12
2.3 Supervised and Unsupervised Learning	15
2.4 Generalization Performance	18
2.5 Model Complexity	20
2.6 Ensembles	22
2.7 Data Dependencies and Computational Complexity	23
3 Quantum Mechanics	25
3.1 States and Superposition	26
3.2 Density Matrix Representation and Mixed States	27
3.3 Composite Systems and Entanglement	29
3.4 Evolution	32
3.5 Measurement	34
3.6 Uncertainty Relations	36
3.7 Tunneling	37
3.8 Adiabatic Theorem	37
3.9 No-Cloning Theorem	38
4 Quantum Computing	41
4.1 Qubits and the Bloch Sphere	41
4.2 Quantum Circuits	44
4.3 Adiabatic Quantum Computing	48
4.4 Quantum Parallelism	49

4.5	Grover's Algorithm	49
4.6	Complexity Classes	51
4.7	Quantum Information Theory	52
Part Two Classical Learning Algorithms		55
5	Unsupervised Learning	57
5.1	Principal Component Analysis	57
5.2	Manifold Embedding	58
5.3	K -Means and K -Medians Clustering	59
5.4	Hierarchical Clustering	60
5.5	Density-Based Clustering	61
6	Pattern Recognition and Neural Networks	63
6.1	The Perceptron	63
6.2	Hopfield Networks	65
6.3	Feedforward Networks	67
6.4	Deep Learning	69
6.5	Computational Complexity	70
7	Supervised Learning and Support Vector Machines	73
7.1	K -Nearest Neighbors	74
7.2	Optimal Margin Classifiers	74
7.3	Soft Margins	76
7.4	Nonlinearity and Kernel Functions	77
7.5	Least-Squares Formulation	80
7.6	Generalization Performance	81
7.7	Multiclass Problems	81
7.8	Loss Functions	83
7.9	Computational Complexity	83
8	Regression Analysis	85
8.1	Linear Least Squares	85
8.2	Nonlinear Regression	86
8.3	Nonparametric Regression	87
8.4	Computational Complexity	87
9	Boosting	89
9.1	Weak Classifiers	89
9.2	AdaBoost	90
9.3	A Family of Convex Boosters	92
9.4	Nonconvex Loss Functions	94

Part Three	Quantum Computing and Machine Learning	97
10	Clustering Structure and Quantum Computing	99
	10.1 Quantum Random Access Memory	99
	10.2 Calculating Dot Products	100
	10.3 Quantum Principal Component Analysis	102
	10.4 Toward Quantum Manifold Embedding	104
	10.5 Quantum K -Means	104
	10.6 Quantum K -Medians	105
	10.7 Quantum Hierarchical Clustering	106
	10.8 Computational Complexity	107
11	Quantum Pattern Recognition	109
	11.1 Quantum Associative Memory	109
	11.2 The Quantum Perceptron	114
	11.3 Quantum Neural Networks	115
	11.4 Physical Realizations	116
	11.5 Computational Complexity	118
12	Quantum Classification	119
	12.1 Nearest Neighbors	119
	12.2 Support Vector Machines with Grover's Search	121
	12.3 Support Vector Machines with Exponential Speedup	122
	12.4 Computational Complexity	123
13	Quantum Process Tomography and Regression	125
	13.1 Channel-State Duality	126
	13.2 Quantum Process Tomography	127
	13.3 Groups, Compact Lie Groups, and the Unitary Group	128
	13.4 Representation Theory	130
	13.5 Parallel Application and Storage of the Unitary	133
	13.6 Optimal State for Learning	134
	13.7 Applying the Unitary and Finding the Parameter for the Input State	136
14	Boosting and Adiabatic Quantum Computing	139
	14.1 Quantum Annealing	140
	14.2 Quadratic Unconstrained Binary Optimization	141
	14.3 Ising Model	142
	14.4 QBoost	143
	14.5 Nonconvexity	143
	14.6 Sparsity, Bit Depth, and Generalization Performance	145
	14.7 Mapping to Hardware	147
	14.8 Computational Complexity	151
	Bibliography	153



Part One

Fundamental Concepts

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

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 (Serfedio and Gortler, 2004). On the other hand, there are apparent computational advantages: certain concept classes