15 October 2002

QUANTUM LEARNING SEMINAR
LECTURE 3: SINGLE QUERY LEARNING

David A. Meyer

Project in Geometry and Physics, Department of Mathematics
University of California/San Diego, La Jolla, CA 92093-0112
`http://math.ucsd.edu/~dmeyer/`; `dmeyer@math.ucsd.edu`

Introduction

As we noted in the first lecture, quantum computation is a relatively young subject, and quantum learning is even younger. The goal of this lecture is to explain that there is a conjecture about quantum learning dating back to '93. That is, 1893!

CONJECTURE (Hadamard [1]): For $N \equiv 0 \bmod 4$ there is a concept class of size N such that quantum learning from membership queries has sample complexity 1.

Concept classes

In the last two lectures we introduced the concept class

$$\mathcal{G}^n = \{g_a : \mathbb{Z}_N \to \mathbb{Z}_2 \mid a \in \mathbb{Z}_N \text{ and } g_a(x) = \delta_{xa}\},$$

and found that Grover's algorithm [3] provides a learning algorithm with sample complexity $O(\sqrt{N})$ for concepts in $\mathcal{G}^n$. Figure 3.1 shows a typical concept in $\mathcal{G}^n$ for the case $N = 8$ $(n = 3)$.

Now consider a different concept class:

$$\mathcal{BV}^n = \{f_a : \mathbb{Z}_2^n \to \mathbb{Z}_2 \mid a \in \mathbb{Z}_2^n \text{ and } f_a(x) = a \cdot x \bmod 2\},$$

(named after Bernstein and Vazirani who first investigated this set in the context of quantum computing [4]). Figure 3.2 shows the first four concepts in $\mathcal{BV}^3$. Notice that except for $a = 0$, $|f_a^{-1}(1)| = N/2$, in contrast to Grover concepts each of which contains exactly 1 element. There are still $N = 2^n$ concepts in $\mathcal{BV}^n$, just as there are in $\mathcal{G}^n$. It is easy to see, however, that the classical complexity of learning from $\mathcal{BV}^n$ is $O(\log N)$. A learning algorithm can exploit the *structure* of $\mathcal{BV}^n$, which is best illustrated not by graphing the concepts as in Figure 3.2, but as in Figure 3.3 where the possible inputs are the vertices of a (hyper)cube. The membership oracle need only be queried about the n basis vectors of $\mathbb{Z}_2^n$. A deterministic algorithm that does this, in the format of Figure 1.2 is

```
**Algorithm D$\mathcal{BV}^n$.**
  **0**. Set $X \leftarrow \mathbb{Z}_2^n$.
  **1**. Set $x \leftarrow 1 = 0\ldots01 \in \mathbb{Z}_2^n$.
  **2**. While $x < 2^n$,
      **2a**. Set hypothesis to $f_x$.
      **2b**. Evaluate $f_x(x)$.
      **2c**. Query the membership oracle about $x$.
      **2d**. If $f_x(x) = \mathrm{MO}(x)$ then adjust $X \leftarrow X \cap f_x^{-1}(1)$ else adjust $X \leftarrow X \cap f_x^{-1}(0)$.
      **2e**. Adjust $x \leftarrow 2x$.
  **3**. Set $\{x\} \leftarrow X$; output $f_x$.
```

This algorithm identifies the bits of a one by one, reducing the set X of possible as by half as each bit is identified. After n iterations of step **2**, X consists of a single element, a, which determines the concept learned, f_a.

Quantum learning

Classically $\mathcal{BV}^n$ is a much easier concept class from which to learn than is $\mathcal{G}^n$. This is also true quantum mechanically: The first step is to prepare the same query state (1.1) as in Grover's algorithm. Submitting it to the membership oracle causes the unitary transformation:

$$\begin{aligned} \sum_x \frac{1}{\sqrt{N}}|x\rangle\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) &\mapsto \sum_x \frac{1}{\sqrt{N}}|x\rangle(-1)^{f_a(x)}\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \\ &= \sum_x \frac{1}{\sqrt{N}}|x\rangle(-1)^{a\cdot x}\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \end{aligned} \tag{3.1}$$

The first tensor factor of the resulting vector is one of $N = 2^n$ different vectors, depending on the concept f_a. For $N = 8$ they are the columns of the matrix

$$\begin{array}{cc} a = & \begin{array}{cccccccc} \scriptstyle 000 & \scriptstyle 001 & \scriptstyle 010 & \scriptstyle 011 & \scriptstyle 100 & \scriptstyle 101 & \scriptstyle 110 & \scriptstyle 111 \end{array} \\ A = \dfrac{1}{\sqrt{8}} & \begin{pmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{pmatrix} \end{array}, \tag{3.2}$$

where $+$ and $-$ denote $+1$ and -1, respectively.

It is easy to check that the columns of A are orthogonal for any N: $A^{\dagger}A = I$. Thus these vectors define an orthonormal basis and hence, according to the definition from the previous lecture, a projective measurement on $\mathbb{C}^N$. Since the first tensor factor on the right hand side of (3.1) is exactly one of these basis vectors (depending on the concept f_a), a projective measurement in this basis will observe a with probability 1. Thus, for concepts in $\mathcal{BV}^n$, the sample complexity of quantum learning from membership queries is 1.

Hadamard matrices

This is clearly equally true for any concept class consisting of concepts that are *orthogonal* in the sense that the matrix corresponding to (3.2) is orthogonal. The first part of the following definition is standard; the second part may not be.

DEFINITION. An $N \times N$ matrix H with elements in $\{\pm 1\}$ that satisfies $H^{\mathrm{T}}H = NI$ (and consequently also $HH^{\mathrm{T}} = NI$) is called a *Hadamard matrix.* Correspondingly, if a concept class $\mathcal{C}$ of maps $X \to \mathbb{Z}_2$ satisfies

$$\sum_{x \in X} (-1)^{c(x)}(-1)^{c'(x)} = 0$$

for all $c \neq c' \in \mathcal{C}$, we will call $\mathcal{C}$ a *Hadamard concept class.*

It is an immediate consequence of our definitions that any Hadamard concept class has sample complexity 1 for quantum learning from a membership oracle. As we will see in a subsequent lecture, the classical sample complexity is $\Omega(\log N)$, as it is for $\mathcal{BV}^n$. These statements are only interesting, of course, if there are Hadamard concept classes other than the family $\mathcal{BV}^n$. These are not so easy to find, but mathematicians have been looking for the corresponding Hadamard matrices since 1867 when Sylvester discovered the family $\left(\begin{smallmatrix} + & + \\ + & - \end{smallmatrix}\right)^{\otimes n}$ [5]—corresponding to $\mathcal{BV}^n$. Some time afterwards Hadamard proved:

THEOREM (Hadamard [2]). *If H is an $N \times N$ Hadamard matrix, then $N \in \{1, 2\}$ or $N \equiv 0$ mod 4.*

Proof. Multiplying any row or column by -1 leaves a Hadamard matrix Hadamard. By doing so appropriately we can change the first row and first column to all +1s; a Hadamard matrix in this form is called *normalized.* Permuting the columns of a Hadamard matrix also leaves it Hadamard. Thus we can arrange the columns so that the first three rows have the form

$$\begin{array}{c}
+\ \text{—————————————}\ + \\
+\ \text{——————}\ +\ -\ \text{——————}\ - \\
\underbrace{+\ \text{—}\ +}_{i\,+\mathrm{s}}\underbrace{-\ \text{—}\ -}_{j\,-\mathrm{s}}\underbrace{+\ \text{—}\ +}_{k\,+\mathrm{s}}\underbrace{-\ \text{—}\ -}_{l\,-\mathrm{s}}
\end{array}$$

Since these rows are mutually orthogonal,

$$\begin{aligned} i + j - k - l &= 0 \\ i - j + k - l &= 0 \\ i - j - k + l &= 0, \end{aligned}$$

which implies $i = j = k = l$. Thus $N = 4k$, if $N \geq 3$. ∎

This result suggested the conjecture with which we started this lecture; the standard phrasing, of course, is:

CONJECTURE (Hadamard [1]). *There is an $N \times N$ Hadamard matrix for all $N \equiv 0 \bmod 4$.*

This conjecture is sometimes attributed to Paley [6], but in 1893 Hadamard wrote: "*J'ai formé des déterminants réels pour $n = 12$ et $n = 20$, sans avoir pu néanmoins reconnaître d'une façon certaine s'il en existe chaque fois que n est divisible par* 4." [1], so it is clear that he thought it might be true. As of 2002, the smallest multiple of 4 for which no Hadamard matrix is known is $N = 428$ [7]. Since the Sylvester matrix $\left(\begin{smallmatrix} + & + \\ + & - \end{smallmatrix}\right)^{\otimes n}$ has dimension 2^n, this implies that there must be other ways to contruct Hadamard matrices. We describe the simplest, which produces the $N = 12$ matrix found by Hadamard (I haven't checked his example for $N = 20$.), as the smallest of an infinite family of matrices different from the Sylvester matrices, next.

Paley's construction

We begin with some number theoretic preliminaries. Let $p > 2$ be a prime number.

DEFINITION. Let $0 < a \in \mathbb{Z}$. The elements of $\{a^2 \bmod p\}$ are called *quadratic residues* mod p.

To compute the number of quadratic residues, note that since

$$(a + p)^2 = a^2 + 2ap + p^2 \equiv a^2 \bmod p,$$

we need only consider $0 \leq a < p$. Furthermore, since

$$(p - a)^2 = p^2 - 2pa + a^2 \equiv a^2 \bmod p,$$

we need only consider $0 \leq a \leq \frac{p-1}{2}$. But $1^2, 2^2, \ldots, \left(\frac{p-1}{2}\right)^2$ mod p are all distinct since

$$\begin{aligned} a^2 \equiv b^2 \bmod p &\implies p | a^2 - b^2 = (a + b)(a - b) \\ &\implies a = b. \end{aligned}$$

Therefore there are $\frac{p-1}{2}$ quadratic residues mod p. The other $\frac{p-1}{2}$ nonzero elements of $\mathbb{Z}_p$ are called *nonresidues*. 0 is neither a residue nor a nonresidue.

Definition. The *Legendre symbol* is defined to be

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & \text{if } a \text{ is a residue;} \\ -1 & \text{if } a \text{ is a nonresidue;} \\ 0 & \text{if } a = 0. \end{cases}$$

Definition. For $p \equiv -1 \bmod 4$, the *Jacobsthal matrix* is the $p \times p$ matrix Q with elements $q_{ij} = \left(\frac{j-i}{p}\right)$, for $i, j \in \{0, \ldots, p-1\}$.

Example. For $p = 11$,

$$Q = \begin{pmatrix}
0 & + & - & + & + & + & - & - & - & + & - & - \\
- & 0 & + & - & + & + & + & - & - & - & + & - \\
- & - & 0 & + & - & + & + & + & - & - & - & + \\
+ & - & - & 0 & + & - & + & + & + & - & - & - \\
- & + & - & - & 0 & + & - & + & + & + & - & - \\
- & - & + & - & - & 0 & + & - & + & + & + & - \\
- & - & - & + & - & - & 0 & + & - & + & + & + \\
+ & - & - & - & + & - & - & 0 & + & - & + & + \\
+ & + & - & - & - & + & - & - & 0 & + & - & + \\
+ & + & + & - & - & - & + & - & - & 0 & + & - \\
- & + & + & + & - & - & - & + & - & - & 0 & + \\
+ & - & + & + & + & - & - & - & + & - & - & 0
\end{pmatrix}.$$

Notice that $Q^{\mathrm{T}} = -Q$. This is true for every Jacobsthal matrix; it is a consequence of the Legendre symbol being a character. Another consequence is that

$$QQ^{\mathrm{T}} = -Q^2 = pI - \mathbf{1}\mathbf{1}^T,$$

where $\mathbf{1}$ is the $p \times 1$ vector of all +1s.

Now let H be the $p+1 \times p+1$ matrix:

$$H = \begin{pmatrix} 1 & \mathbf{1}^T \\ \mathbf{1} & Q - I \end{pmatrix}. \tag{3.3}$$

This is Paley's (Type I) construction [6], which was apparently discovered by Gilman a few years earlier [8]. It produces a Hadamard matrix, since

$$HH^{\mathrm{T}} = \begin{pmatrix} p+1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{1}\mathbf{1}^T + (Q-I)(Q-I)^{\mathrm{T}} \end{pmatrix},$$

where $\mathbf{0}$ is the $p \times 1$ zero vector. The lower right block is

$$\mathbf{1}\mathbf{1}^T + QQ^{\mathrm{T}} - Q^{\mathrm{T}} - Q + I = \mathbf{1}\mathbf{1}^T + pI - \mathbf{1}\mathbf{1}^T + I = (p+1)I,$$

where we have used the two facts about Jacobsthal matrices noted in the Example above. Thus $HH^{\mathrm{T}} = (p+1)I$, which means H is a Hadamard matrix.

The $p = 11$ Jacobsthal matrix of the Example above produces the unique 12×12 Hadamard matrix by this construction. For $p = 3$ and $p = 7$, Paley's/Gilman's construction reproduces the 4×4 and 8×8 Sylvester matrices. For $p = 31$, however, it produces a 32×32 Hadamard matrix that is distinct from the 32×32 Sylvester matrix. The number of distinct Hadamard matrices is known currently only up to $N = 28$ [9]:

$N =$	1	2	4	8	12	16	20	24	28
$\# =$	1	1	1	1	1	5	3	60	487

Each distinct Hadamard matrix defines a Hadamard concept class in which quantum learning with access to a membership oracle succeeds with sample complexity 1. The concept class $\mathcal{BV}^n$ corresponding to the $2^n \times 2^n$ Sylvester matrices consists of concepts of the form "numbers x such that $x \cdot a = 1 \bmod 2$". If the rows and columns of (3.3) are labelled from 0 to p, the corresponding concept class is "numbers x such that $x \neq 0$ and $x - a$ is a square mod p for $a \neq 0$". In a paper that inspired this lecture, van Dam has combined the Paley/Gilman construction and Paley's (Type II) construction for primes $p \equiv 1 \bmod 4$ into the *Shifted Legendre Symbol* problem and proved that it has sample complexity 2 [10].

References

[1] M. Hadamard, "*Sur le module maximum que puisse atteindre un déterminant*", *C. R. Acad. Sci. Paris* **116** (1893) 1500.

[2] M. J. Hadamard, "*Résolution d'une question relative aux déterminants*", *Bull. des Sciences Mathematiques* **17** (1893) 240–246.

[3] L. K. Grover, "A fast quantum mechanical algorithm for database search", in *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, Philadelphia, PA, 22–24 May 1996 (New York: ACM 1996) 212–219;
L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack", *Phys. Rev. Lett.* **79** (1997) 325–328.

[4] E. Bernstein and U. Vazirani, "Quantum complexity theory", in *Proceedings of the 25th ACM Symposium on Theory of Computing*, San Diego, CA, 16–18 May 1993 (New York: ACM Press 1993) 11–20;
E. Bernstein and U. Vazirani, "Quantum complexity theory", *SIAM J. Comput.* **26** (1997) 1411–1473.

[5] J. J. Sylvester, "Thoughts on inverse orthogonal matrices, simultaneous sign-successions, and tesselated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers", *Phil. Mag.* ser. IV **34** (1867) 461–475.

[6] R. E. A. C. Paley, "On orthogonal matrices", *J. Math. and Phys.* **12** (1933) 311–320.

[7] S. Georgiou, C. Koukouvinos and J. Seberry, "Hadamard matrices, orthogonal designs and construction algorithms", in W. D. Walis, ed., *Designs 2002: Further Combinatorial and Constructive Design Theory* (Kluwer, in press).

[8] R. E. Gilman, "On the Hadamard determinant theorem and orthogonal determinants", *Bull. Amer. Math. Soc.* **37** (1931) 30–31.

[9] N. J. A. Sloane, "A library of Hadamard matrices", `http://www.research.att.com/~njas/hadamard/`.

[10] W. van Dam, "Quantum algorithms for weighing matrices and quadratic residues", *Algorithmica* (2002) OF1–OF16.