arXiv:2602.01054v1 [quant-ph] 1 Feb 2026

量子学习动物园

（关于经典概念的量子学习综述）

Sagnik Chatterjee[1*]

[1]理论计算机科学学院（STCS），TIFR 孟买分校。

通讯作者。电子邮箱：chatsagnik@gmail.com;

摘要

本文综述了量子学习理论领域的多项成果，具体聚焦于在可能近似正确（PAC）框架下学习量子编码的经典概念。本文的核心在于强调：在针对不同标记预言机进行带查询访问的学习时，经典学习与量子学习在查询、样本和时间复杂度方面出现的分离。本文旨在汇总上述范畴下该领域所有已知成果，并通过提出 23 个开放问题，让读者认识到我们当前理解的局限。

关键词： 量子计算、PAC 学习、综述、时间与样本复杂度

1 引言

自从计算机科学家借用量子力学原理，提出利用量子力学原理的计算设备这一概念 [1–3], 以来，量子计算领域始终致力于证明量子计算任务相较于经典计算任务的加速优势。Grover [4] 和 Shor [5] 的开创性成果让这一目标达到高潮，为在传统计算机科学的各个领域探索计算加速打开了闸门；此前，人们认为这些领域已触及计算能力的极限。

学习理论 [6–9] 研究算法如何从数据中进行泛化。这个理论计算机科学分支与密码学、编码理论和电路复杂度等领域密切相关，因此是一个极具研究价值的领域。二十世纪末，一系列研究成果

围绕经典不可能性障碍下的高效学习展开，[1] 即使是布尔函数这样极其简单的概念类也不例外。正是在这一关键时刻，Bshouty 和 Jackson [10]率先推动了量子学习理论这一分支的发展，他们利用量子查询算法绕过了已知的经典障碍，从而实现了对析取范式公式的学习。本综述旨在涵盖量子学习理论的大量文献，但会受到若干限制。

1.1 结构与组织

在第 2 节中，我们确立基本符号，并概述量子计算与计算学习理论，随后在第 2.4 节介绍 PAC 学习框架，重点讨论样本复杂度（可学习性）和时间复杂度（高效可学习性）的概念。我们介绍各种查询模型，并在第 3 节讨论它们之间的分离。其中尤为重要的是第 3.2 节和3.3，我们分别阐述经典和量子设定下采样与成员查询之间的分离，并针对量子情形提出一种新的、广义的采样与成员查询分离度量。在第 4 节中，我们综述经典与量子 PAC 学习算法之间的样本复杂度分离。我们的讨论不局限于以往综述（例如 [11, 12],）所考察的传统设定，还涵盖了较新的研究方向，包括针对广义预言机的分离、多分类设定下的分离，以及白盒访问下的分离。在第 5 节中，我们综述经典与量子学习算法之间的计算分离。我们正式引入由隐藏子群问题诱导的概念类这一概念，并说明已有分离可由这一一般概念类中的分离推出。随后，我们将带误差学习（LWE）和带噪声奇偶学习（LPN）中看似神奇的量子学习加速，解释为在强大的新型预言机模型下进行学习的结果。最后，我们综述关于经典布尔概念类（如半空间、决策树、Junta 函数、析取范式和浅层电路）的学习分离及不可能性结果的最新进展。

本综述不涉及的内容

如摘要所述，本文的主要目标是综述经典与量子学习算法在样本复杂度和时间复杂度方面的理论分离，讨论范围聚焦于经典概念类的学习，特别是将其限制在 PAC 学习框架内。本文可视为对 Arunachalam 和 de Wolf 于 2017年发表的早期综述的扩展 [11], ，既概述过去十年中的新量子学习成果，也着重阐释学习分离背后的基本理念，即基于标记查询预言机进行学习。我们会先说明本文综述的范围，再转入所讨论的主要内容。

[1]可学习性和高效可学习性的概念将在后文正式介绍。

本综述未涉及量子学习理论中的许多重要分支，尤其是采用不同学习模型的研究。下面着重介绍其中两类：

1. 面向量子对象的量子学习。量子学习理论的这一子领域契合了费曼最初的愿景，即利用量子计算机来理解量子现象。示例包括学习任意量子态、量子信道和量子过程。关于该领域更全面的介绍，请参阅其他综述 [12, 13]。

2. 在不同学习框架下进行量子学习，例如量子统计学习、量子精确学习等。关于量子统计学习的介绍，请参阅 Nietner 的以下手稿 [14]；关于量子精确学习的更多详情，请参阅以下资料 [11, 15] 。

2 背景与预备知识

2.1 基本记号

我们用 log 表示以 2 为底的对数。对于 $n \in \mathbb{N}$， n 中的任意正多项式函数记作 $\mathrm{poly}(n)$。关于 n 个变量的多元多项式记作 $\mathrm{poly}(x_1, x_2, \ldots\ldots, x_n)$。对于 $n \in \mathbb{N}$，我们用 $[n]$ 表示集合 $\{1, \ldots, n\}$。

阶为 2 的有限域记为 $\mathbb{F}_2 = \{-1, +1\}$ 或 $\mathbb{Z}_2 = \{0, 1\}$，而 $\mathbb{F}_2^n$（分别为 $\mathbb{Z}_2^n$）表示 $\mathbb{F}_2$（分别为 $\mathbb{Z}_2$）上的 n 维向量空间。字符串 $x \in \mathbb{F}_2^n$是 $\mathbb{F}_2^n$中的一个 n 维向量，其中 x_i 表示 x的第 i 位。布尔文字是一个比特 x_i 或其取反 x_i。对于字符串 $x \in \mathbb{F}_2^n$，字符串 $x \in \mathbb{F}_2^n$ 是由 $x_i = 1 \oplus x_i$给出的字符串，其中 $\oplus$是逐位求和运算。对于 $a, b \in \mathbb{F}_2^n$（或 $\mathbb{Z}_2^n$），任意逐位二元运算 $\odot$ 都会得到一个字符串（$a \odot b$） $\in \mathbb{F}_2^n$（或分别为 $\mathbb{Z}_2^n$）。逐位与、或和异或运算分别记为 $a \wedge b$、 $a \vee b$和 $a \oplus b$。 $x \in \mathbb{F}_2^n$ 的汉明重量记为 $|x|$。

事件 E 的指示随机变量记为 $\mathbb{I}_E$。集合 $\mathcal{S}$ 上的均匀分布记为 $\mathcal{U}(\mathcal{S})$。布尔超立方体 $\mathbb{F}_2^n$ 上的分布 $\mathfrak{D}$ 的支撑集记为 $\mathrm{supp}(\mathfrak{D}) = \{z \in \mathbb{F}_2^n \mid \mathfrak{D}(z) \neq 0\}$。 $x \sim \mathfrak{D}$ 表示从概率分布 $\mathfrak{D}$中抽取的字符串，即 $\Pr(\mathbf{X} = x) = \mathfrak{D}(x)$。类似地，$x \sim \mathcal{S} \iff x \sim \mathcal{U}(\mathcal{S})$，其中 $\mathcal{S} \subseteq [n]$。我们用 $\mathop{\mathbb{E}}_{\mathfrak{D}}[\mathbf{X}]$ 表示随机变量 $\mathbf{X}$ 相对于分布 $\mathfrak{D}$的期望，用 $\mathbb{E}[\mathbf{X}]$ 表示$\mathbf{X}$相对于均匀分布的期望。在讨论内积空间时，我们有时将 $\mathop{\mathbb{E}}_{\mathfrak{D}}[\mathbb{X}]$ 记为 $\langle \mathfrak{D}, \mathbb{X} \rangle$。

函数 $f : \mathbb{F}_2^n \mapsto \mathbb{F}_2^m$称为**布尔函数**。回顾一下，阶为 2 的有限域记作 $\mathbb{F}_2 = \{-1, +1\}$或 $\mathbb{Z}_2 = \{0, 1\}$，而 $\mathbb{F}_2^n$（分别是 $\mathbb{Z}_2^n$）表示一个 n维的、定义在 $\mathbb{F}_2$（分别是 $\mathbb{Z}_2$）上的向量空间。本综述关注判定问题，也就是二值函数，因此除非另有说明，我们主要考虑如下形式的布尔函数 $f : \mathbb{F}_2^n \mapsto \mathbb{F}_2$。关于布尔函数表示的更多细节，请参阅B 节 。

2.2 量子计算基础

与 n 量子比特量子态对应的密度矩阵 ρ 是迹为 1 的 $n \times n$ 半正定（PSD）矩阵：$\rho = \sum_{i\in[n]} p_i |\psi_i\rangle \langle\psi_i|$，其中 $\{|\psi_i\rangle\}_{i\in[n]}$ 是纯态，$\{p_i\}_{i\in[n]}$ 构成一个概率分布。与纯态对应的密度矩阵的秩为 1，即 $\rho = |\psi\rangle \langle\psi|$。

定义 2.1 (迹距离)。 两个量子态 ρ和 σ 之间的迹距离记为 $D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$. 当 ρ 和 σ 对应于纯态时 $|\phi\rangle$ 和 $|\psi\rangle$ 迹距离表示为 $D(|\phi\rangle, |\psi\rangle) = \sqrt{1 - |\langle\phi|\psi\rangle|^2}$.

事实 2.1 (区分量子态)。 假设某个算法的任务是区分两个混合态 ρ_0 和 ρ_1。如果算法获得 ρ_b，其中b 是从 $\{0, 1\}$中均匀随机选取的，那么区分 ρ_0和 ρ_1 的最佳成功概率为 $\frac{1}{2} + \frac{1}{2}D(\rho_0, \rho_1) = \frac{1}{2} + \frac{1}{2}\sqrt{1 - |\langle\rho_0|\rho_1\rangle|^2}$。

给定一个布尔函数 f，若量子电路 f 的误差为 $\varepsilon > 0$，且$\forall x \in \mathbb{F}_2^n$，量子电路 Q能够以至少 $f(x)$ 的概率计算 $1 - \varepsilon$。[2]计算 f 的有界错误查询复杂度，是指任何预言机 O_x 被调用的最小次数，使得 Q 能以 $\geq 1 - \varepsilon$的概率计算 f。有界错误时间复杂度还会考虑实现酉算子 V_1、.....、V_k所需的时间。复杂性类 有界错误量子多项式时间是指可由量子计算机高效求解的判定问题类，定义为这样一类语言：$L \subseteq \{0, 1\}^*$ 可由一致量子电路族 $\{Q_n\}_n$在多项式时间内判定，且有界错误概率至多为 $1/3$ 。

注 2.1。 在本综述中，当我们说存在一种能够完成某项特定任务的量子算法时，意思是存在一个能够完成该任务的一致量子查询电路族。

更多关于量子算法的讨论，请参阅A 节。

2.3 计算学习理论概述

令 $\mathcal{X}_n \subseteq \mathbb{F}_2^n$ 为由 n 位实例（或字符串）组成的集合，$\mathcal{Y}$ 为标签集。对于 $n \in \mathbb{N}$，我们用 $\mathcal{C} = \cup_{n>0}\mathcal{C}_n$ 表示，其中 $\mathcal{C}_n = \{c_n \mid c_n : \mathcal{X}_n \mapsto \mathcal{Y}\}$，这是一类将实例映射到标签的函数，称为 **概念类**。类似地，**假设类** $\mathcal{H} = \cup_{n>0}\mathcal{H}_n$（其中 $\mathcal{H}_n = \{h_n \mid h_n : \mathcal{X}_n \mapsto \mathcal{Y}\}$）也是一类将实例映射到标签的函数。当上下文明确时，我们会从定义中省略下标 n 。在本文的大部分内容中，我们假设标签集是二元的。因此，上述讨论的概念和假设都是布尔函数。

[2]这里所说的计算，是指辅助量子比特的测量结果为 $f(x)$。

由于多个布尔电路类可以计算相同的布尔函数（详见 附录 B 中的详细讨论），因此，概念或假设的表示对于计算任务而言变得十分重要。我们假设任何概念类或假设类的表示都是 固定的。

注释 2.2. 在本综述中，所考虑的底层表示是概念类 $\mathcal{C}$ /假设类 $\mathcal{H}$ ，其对应的规模分别记为 规模（$\mathcal{C}$）和规模（$\mathcal{H}$），分别表示 $\mathcal{C}$ 和 $\mathcal{H}$ 的表示所需的最大成本。

学习算法（或学习器） A 是一个随机映射，它将实例集合 $\mathcal{S}$ 中依照未知分布 $\mathfrak{D}$ 独立同分布（ i.i.d. ）采样的实例 [3]，以及由未知目标概念 $c \in \mathcal{C}$ 标记的实例，映射为假设函数 $h \in \mathcal{H}$。[4] 由于标签集是二元的，本综述中讨论的大多数算法都执行二分类。这里需要指出，学习器 A 可以是经典随机算法，也可以是量子算法。在本综述的大部分内容中，我们着重采用一种度量来量化输出假设的性能，即假设误差。误差这一概念从语义上刻画了假设 h 相对于概念 c 在不同形式下所犯的错误数量。

定义 2.2 （误差）。 假设 $h \in \mathcal{H}$： $\mathbb{F}_2^n \mapsto \mathbb{F}_2$ 相对于未知概念 $c \in \mathcal{C}$： $\mathbb{F}_2^n \mapsto \{-1, +1\}$ 以及未知的底层分布$\mathfrak{D}$ 的误差定义如下：

$$\mathrm{err}_{\mathfrak{D},c}(h) := \mathop{\mathbb{E}}_{x\sim\mathfrak{D}} \left[\mathbb{I}_{(h(x)\neq c(x))} \right] = \Pr_{x\sim\mathfrak{D}} \left[h(x) \neq c(x) \right]. \tag{1}$$

假设类 $\mathcal{H}$ 相对于未知概念$c \in \mathcal{C}$ 以及底层分布 $\mathfrak{D}$ 的最优误差定义为 $\mathrm{opterr}_{\mathfrak{D},c}(\mathcal{H}) := \min_{h\in\mathcal{H}} \mathrm{err}_{\mathfrak{D},c}(h)$。

注 2.3。 满足 $\mathrm{err}(h) = 0$ 的假设称为理想假设。

2.4 PAC 学习框架

本节将使用由 Valiant [9]提出的可能近似正确（PAC）学习模型，量化学习算法 A 的性能。

定义 2.3 ((ε, δ)-PAC 可学习性). 学习算法 A (ε, δ)-*PAC*学习一个未知概念类 $\mathcal{C}$ （定义在 n比特实例上），并使用假设类 $\mathcal{H}_A$，如果对于 $\mathcal{X}$上的每一个分布 $\mathfrak{D}$ 以及每一个目标概念 $c \in \mathcal{C}$，都存在一个函数$m_0 : (0, 1)^2 \mapsto \mathbb{N}$，使得对于每一个 ε、 $\delta \in [0, 1]$， A 以训练集 $\mathcal{S}$作为输入，该训练集由 $m > m_0(\varepsilon, \delta)$ 组成，这些样本按 $\mathfrak{D}$ 独立同分布抽取并由 c 标注，并输出

[3]集合 $\mathcal{S}$ 称为 A的训练集。[4]我们假设假设类 $\mathcal{H}$ 和概念类 $\mathcal{C}$ 均已知，但具体的目标概念 c 未知。

$h \in \mathcal{H}_A$ 使得

$$\Pr\left[\mathrm{err}_{\mathfrak{D},c}(h) \leq \varepsilon\right] \geq 1-\delta. \tag{2}$$

式（2）中的概率是针对训练集 $\mathcal{S}$ 的选择以及学习器 A的任何内部随机化而言的。我们还注意到，在 定义 2.3中，隐含着可实现性假设，即对于 $\mathcal{S}$，以概率 1 存在 $h_{\mathrm{opt}} \in \mathcal{H}$ ，使得 $\mathrm{err}_{\mathfrak{D},c}(h_{\mathrm{opt}}) = 0$。

2.4.1 样本复杂度与可学习性

在定义 2.3 中， m 是学习器 A学习 $\mathcal{C}$时的样本复杂度。如果$m = \mathrm{poly}\left(1/\varepsilon, 1/\delta, n, \mathrm{size}\left(\mathcal{H}_A\right)\right)$，那么概念类 $\mathcal{C}$就称为可由学习器 A 使用假设类 $\mathcal{H}_A$进行 PAC 学习。若存在某些 ε、δ、m_0，使得式（2）成立，则概念类 $\mathcal{C}$ 可使用 $\mathcal{H}_A$ 进行学习，或称为 PAC 可学习。

注 2.4。定义 *2.3*遵循 *Shalev-Shwartz* 和 *Ben-David* *[16]*,将样本复杂度的概念与 *PAC* 学习的定义分开。这是为了强调可学习性与高效可学习性之间的区别。在其他标准教材和资料中，通常会发现$m = \mathrm{poly}\left(1/\varepsilon, 1/\delta, n, \mathrm{size}\left(\mathcal{H}_A\right)\right)$ 直接出现在定义 *2.3*本身中。

2.4.2 时间复杂度与高效可学习性

如果 A (ε， δ)-PAC 能在多项式时间内学习 $\mathcal{C}$，即运行时间为 $\mathrm{poly}(1/\varepsilon, 1/\delta, n, \mathrm{size}\left(\mathcal{H}_A\right))$，那么我们称概念类 $\mathcal{C}$ 可由 A使用假设类 $\mathcal{H}_A$**高效 PAC 学习**（或简称**高效学习**）。对于高效学习器，$\mathrm{size}\left(\mathcal{H}_A\right) = \mathrm{poly}\left(1/\varepsilon, 1/\delta, n, \mathrm{size}\left(\mathcal{C}\right)\right)$。这里需要指出，并非所有相对于学习器 A 可学习的概念类，相对于 A都可高效学习。

注 2.5。如果潜在概念类的表示方案表达能力极强／冗长，例如该概念类使用真值表描述，那么所有布尔函数都可以高效地进行 *PAC* 学习。有趣的情况是，潜在概念使用布尔函数表示，而人们认为布尔函数的表达能力有限，例如决策树，因为极其简单的布尔电路就能计算它们。关于与学习算法相关的假设类，更多讨论请参阅*C.1* 节。

尽管对PAC学习算法进行分类的方式有很多，但在本综述中，我们重点关注PAC学习器在各种噪声模型下、以及针对各种预言机查询的样本复杂度和时间复杂度（具体而言，是经典PAC学习器与量子PAC学习器之间的分离）。

注释 2.6。本综述未涵盖的一大类分类体系，涉及学习器生成的输出假设的表示和准确性。关于上述分类体系的简要讨论，请参阅*C.2* 节。

3 查询复杂度分离

3.1 带查询的 PAC 学习

学习算法 A 在定义 2.3(ε, δ)-PAC 学习 $\mathcal{C}$ 随机样本（以训练集 S的形式提供）。或者，我们也可以假设学习器 A 可以进行黑盒查询，即学习器可以访问预言机 $\mathcal{O}$。针对查询进行学习主要用于研究在各种形式的标记噪声下，PAC 学习设置的推广或限制。关于 PAC 框架中标签噪声的详细讨论，请参阅C.4 节。在这个广义模型中，我们将 PAC 学习的概念重新定义如下:

定义 3.1（带预言机访问的 PAC 学习）. 学习器 A (ε, δ)-*PAC*学习概念类 $\mathcal{C}$（定义在 n比特实例上），并具有对预言机 $\mathcal{O}$ 的查询访问，使用假设类$\mathcal{H}_A$，如果对每个 ε、$\delta \in (0,1)$，对 $\mathcal{X}$ 上的每个分布 $\mathfrak{D}$，以及每个目标概念$c \in \mathcal{C}$，A 以训练集 S 作为输入，该训练集由 $m = \text{poly}\left(1/\varepsilon, 1/\delta, n,\text{规模}\left(\mathcal{H}_A\right)\right)$ 个依照 $\mathfrak{D}$ 独立同分布采样并由 c标注的样本组成，并且对 $\mathcal{O}$ 至多进行 m 次查询，并输出 h，使得 h 的误差至多为 ε 的概率至少为 $1-\delta$。

3.1.1 带采样预言机的 PAC 学习

采样预言机也称为样本查询预言机。

3.1.2 经典样本查询预言机

样本查询（EX）预言机：查询经典随机**样本预言机**EX $(\mathfrak{D}, c)$，会生成形式为 $(x, c(x))$ 的带标记样本，其中 $x \sim \mathfrak{D}$。利用 EX 预言机，我们可以重新定义使用 A 进行概念 $c \in \mathcal{C}$ 的 PAC 学习的样本复杂度和时间复杂度，改用 A 对 EX $(\mathfrak{D}, c)$ 发起的查询次数来表示，具体如下。

定义 *3.2*（访问 *EX* 预言机的 (ε, δ)-*PAC* 学习）。如果对每个 ε、$\delta \in (0,1)$，对每个定义在 $\mathcal{X}$上的分布 $\mathfrak{D}$，以及每个目标概念 $c \in \mathcal{C}$，学习器 A (ε, δ) 都至多向 $EX(\mathfrak{D}, c)$ 发起 $m = poly\left(1/\varepsilon, 1/\delta, n, size\left(\mathcal{H}_A\right)\right)$ 次查询，并输出满足以下条件的 h，则称其在 n-位实例上、通过查询访问 $EX(\mathfrak{D}, c)$，使用假设类 $\mathcal{H}_A$，对概念类 $\mathcal{C}$ 进行 (ε, δ)-*PAC* 学习：$Pr\left[err_{\mathfrak{D}, c}(h) \leq \varepsilon\right] \geq 1-\delta$。

随机噪声样本查询（REX）预言机：对 $\text{REX}^{\eta}(\mathfrak{D}, c)$ 预言机进行查询会生成形式为 $\{(x_i, y_i')\}_{i \in [m]}$ 的带标记样本，其中 $y_i' = y_i$ 以概率 η 且$y_i' = y_i$ 以概率 $1-\eta$。因此，随机分类噪声（RCN）设定下的 PAC 学习可以视为学习器 A 对 REX 预言机进行查询。

无关示例查询（AEX）预言机：与 RCN 设定类似，不可知设定下的 PAC 学习可以理解为学习器 A向 AEX 预言机发起查询；调用 AEX（$\mathfrak{D}$）预言机会根据某个未知的联合分布 $\mathfrak{D}$ 在 $\mathcal{X} \times \mathcal{Y}$ 上生成样本（x，y）$\sim \mathfrak{D}$。不可知情形下的可学习性定义如下 [17]。

定义 *3.3*（β-最优（ε, δ）-不可知 *PAC* 学习）。如果对于每个 $\varepsilon, \delta \in (0, 1)$，存在某个 $\beta \in [0, 1/2)$，对于每个在 $\mathcal{X} \times \mathcal{Y}$ 上的分布 $\mathfrak{D}$，学习器 A β 至多使用 *poly*（$1/\varepsilon, 1/\delta, n$, *size*（$\mathcal{H}_A$））次查询访问 *AEX*（$\mathfrak{D}$），并输出 h，使得 *Pr*$[err_{\mathfrak{D}}(h) \leq opterr_{\mathfrak{D}}(\mathcal{C}) + \beta + \varepsilon] \geq 1 - \delta$ 对某个 $\beta \in [0, 1/2$ 成立，则称其使用假设类 $\mathcal{H}_A$，对 n 位实例上的基准概念类 $\mathcal{C}$ 进行 ε-最优（δ）-不可知 *PAC* 学习。

可学习性（以及高效可学习性）在定义 3.3 意味着 A 对 AEX ($\mathfrak{D}$) 发出的查询数量（以及 A 输出 h 所需的时间）受到多项式上界的限制，这与定义 3.2类似。在定义 3.3中，如果存在一个概念 $c \in \mathcal{C}$，使得 $\mathrm{err}_{\mathfrak{D}}(c) := \Pr_{(x,y)\sim\mathfrak{D}}[c(x) \neq y] = 0$，那么 $\mathrm{opterr}_{\mathfrak{D}}(\mathcal{C}) = \beta$，此时我们几乎回到了无噪声 PAC 学习设定。

观察 3.1。概念类 $\mathcal{C}$ 在不可知设定中充当学习器 A 的基准。如果我们选择一个很差的基准概念类，即$\mathrm{opterr}_{\mathfrak{D}}(\mathcal{C}) > 1/2$，那么最好的不可知学习器所生成假设的准确率将低于 $1/2$。

观察 3.2。在经典采样预言机下进行学习时，学习器 A 会获得凸组合的副本，其中权重为概率质量。

$$\sum_{x\in\mathcal{X}} \mathfrak{D}_x \ (x, c(x)), \quad \sum_{x\in\mathcal{X}} \mathfrak{D}_x = 1, \quad \mathfrak{D} : \mathcal{X} \to [0, 1]. \ \textit{(Noiseless setting.)}$$

$$\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \mathfrak{D}_{x,y} \ (x, y), \quad \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \mathfrak{D}_{x,y} = 1, \quad \mathfrak{D} : \mathcal{X} \times \mathcal{Y} \to [0, 1]. \ \textit{(Agnostic setting.)}$$

3.1.3 量子样本查询预言机

量子样本查询（QEX）预言机： Bshouty 和 Jackson [10] 将(ε, δ)-PAC 学习通过访问 EX 预言机的概念推广到了量子设定，引入了 QEX 预言机（参数化为 QEX($\mathfrak{D}, c$)），并将通过访问 QEX 预言机进行量子 PAC 学习定义如下。

$$\mathrm{QEX}(\mathfrak{D}, c) : |0, 0\rangle \mapsto \sum_{x\in\mathbb{F}_2^n} \sqrt{\mathfrak{D}(x)}\,|x, c(x)\rangle\,. \tag{3}$$

量子 RCN 样本查询（QREX）预言机： 我们可以如下定义一个能够刻画 RCN设定的量子样本预言机。

$$\mathrm{QREX}^p(\mathfrak{D}, c) : |0, 0\rangle \mapsto \sum_{x\in\mathbb{F}_2^n} \sqrt{(1-p)\cdot\mathfrak{D}(x)}\,|x, f(x)\rangle + \sum_{x\in\mathbb{F}_2^n} \sqrt{p\cdot\mathfrak{D}(x)}\,\left|x, \overline{f(x)}\right\rangle. \tag{4}$$

定义 3.4（使用 QEX（分别为 QREX）预言机进行量子 PAC 学习）。如果量子算法 A（ε, δ）使用假设类 $\mathcal{H}_A$，并且对任意$\varepsilon, \delta \in (0、1)$，对任意定义在 $\mathcal{X}$上的分布 $\mathfrak{D}$ 以及任意目标概念 $c \in \mathcal{C}$，A 满足以下条件，则称其在 n-位实例上，通过查询访问 QEX（$\mathfrak{D}, c$）（分别为 QREX^p（$\mathfrak{D}, c$），以 $\mathcal{C}$ 为参数量子 *PAC* 学习概念类 $\mathcal{C}$：

对 QEX（$\mathfrak{D}, c$）（分别为 QREX^p（$\mathfrak{D}, c$）的查询次数至多为 $m = \mathrm{poly}(1/\varepsilon, 1/\delta, n, \mathrm{size}(\mathcal{H}_A)$，并输出 h，使得 $\Pr[\text{错误}_{\mathfrak{D},\, c}(h) \le \varepsilon] \ge 1 - \delta$。

量子不可知示例查询（QAEX）预言机： Arunachalam 和 de Wolf [11] 将不可知例子查询（AEX）预言机和β-最优（ε，δ）-不可知PAC学习推广到量子设定，如下所示：

$$\mathrm{QAEX}(\mathfrak{D}) : |0,0\rangle \mapsto \sum_{(x,y)\in\mathbb{F}_2^{n+1}} \sqrt{\mathfrak{D}(x,y)}\,|x,y\rangle\,. \tag{5}$$

定义 *3.5*（使用 *QAEX* 预言机的量子不可知 *PAC* 学习）。若量子学习器 A β 能够以 ε 最优的方式，对定义在 n 位实例上的基准概念类 $\mathcal{C}$ 进行 *(ε, δ)-PAC* 学习，且可查询 *QAEX*（$\mathfrak{D}$），并使用假设类 $\mathcal{H}_A$，则对于任意 $\varepsilon, \delta \in$ *(0, 1)*，存在某个 $\beta \in [0, 1/2)$，对于定义在 $\mathcal{X} \times \mathcal{Y}$ 上的任意分布 $\mathfrak{D}$，A 对 *QAEX*（$\mathfrak{D}$）的查询次数至多为 $m = \mathit{poly}(1/\varepsilon, 1/\delta, n, \mathit{size}(\mathcal{H}_A))$，并输出 h，使得 $\mathit{Pr}[\mathit{err}_{\mathfrak{D}}(h) \le \mathit{opterr}_{\mathfrak{D}}(\mathcal{C}) + \beta + \varepsilon] \ge 1 - \delta$，且此结论对某个 $\beta \in [0, 1/2$ 成立。

观察 3.3。如果学习器是经典的，但查询预言机是量子的，那么学习器唯一能做的就是测量方程 (3) 至(5) 中的状态，从而获得一个带标签样本。因此，$\mathrm{EX}(\mathfrak{D}, c)$ 和 $\mathrm{QEX}(\mathfrak{D}, c)$ 预言机（以及它们各自的含噪版本）等价于经典学习算法。

备注 3.1。在 定义 *3.4* 和 *3.5*中，我们说量子学习器输出一个假设，意思是量子学习器执行 *POVM* 测量，使测量的每个结果都与一个假设相关联。

备注 3.2。量子 *PAC* 设定中的可学习性和高效可学习性概念与经典设定中的类似。在量子 *PAC* 设定中，我们关心的复杂度概念同样是样本（和/或查询）复杂度以及学习器的时间复杂度。

注释 3.3。我们用 $A^{\mathcal{O}}$ 表示学习器 A 向预言机 $\mathcal{O}$进行查询。

观察 3.4。在量子采样预言机下进行学习时，学习器 A会获得线性组合的副本，其中权重为概率幅。

$$\sum_{x\in\mathcal{X}} \sqrt{\mathfrak{D}_x}\ (x, c(x)),\quad \sum_{x\in\mathcal{X}} \mathfrak{D}_x = 1,\quad \mathfrak{D} : \mathcal{X} \to [0,1].\ \textit{(Noiseless setting.)}$$

$$\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \sqrt{\mathfrak{D}_{x,y}}\ (x, y),\quad \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \mathfrak{D}_{x,y} = 1,\quad \mathfrak{D} : \mathcal{X}\times\mathcal{Y} \to [0,1].\ \textit{(Agnostic setting.)}$$

注释 3.4。 最近，*Arunachalam* 等人 *[18]* 提出了一种更强的量子不可知学习模型，并在该模型下获得了与第 *5* 节所述结果类似的最先进可学习性结果。在他们的模型中，*Arunachalam* 等人 *[18]*考虑一种基准量子相位态，该态编码了有关目标概念类的信息。输入

3.1.4 使用成员查询预言机进行 PAC 学习

PAC 学习中最重要的预言机之一是成员查询预言机 MEM_c。向 MEM_c 预言机查询一个样例 $x \in \mathbb{F}_2^n$，会直接返回该样例的标签 $c(x)$。正如我们所见，成员查询（MQ）预言机允许我们进行非常强大的分布外 查询。

3.2 经典成员查询的威力

使用成员查询进行学习是一种主动学习模型，学习器可以选择查询其希望查询的任意实例—标签对；而在被动学习模型中，学习器只能获取根据未知分布抽样得到的实例—标签对。下面陈述两个结果，以帮助理解这两种学习查询模型的相对能力。
Bshouty [19]证明了以下结果。

引理3.1（布尔函数的可学习性）。 任意 n 位布尔函数都可以通过成员查询在 n 和其析取范式（*DNF*）规模或决策树规模的多项式时间内进行 *PAC* 学习。

回想一下，$(\varepsilon,\ \delta)$-PAC 学习通过访问 EX 预言机 不过是对普通 PAC 学习的重新表述，改为通过 EX 预言机进行查询访问。因此，引理3.1 立即表明，对于 DNF，使用 EX 查询和 MQ 查询进行 PAC 学习之间可能存在学习分离；而 DNF 是否能通过 EX 查询实现 PAC 学习尚不明确。以下命题由 的类似论证得出，它无条件地（尽管是多项式级别地）区分了 MQ 预言机与 EX 预言机。

命题3.1 (MQ 与 EX 预言机之比较)。假设某个（二者均已知的）实例集合$\mathcal{S}$ 是二元概念类 $\mathcal{C}$所能打散的最大集合，并设 $\mathfrak{D}:\ \mathcal{S} \mapsto [0,1]$ 是一个分布，其大部分概率质量集中在某个特定实例上，如下所示：

$$\mathfrak{D}(x) = \begin{cases} 1-3\varepsilon & ,x = x', \varepsilon = 1/\mathrm{poly}(|\mathcal{S}|) \\ \frac{3\varepsilon}{|\mathcal{S}|}, & ,x \in \mathcal{S} \setminus \{x'\}. \end{cases}$$

那么，任何针对 c 、相对于 $\mathfrak{D}$ 的(ε,δ)*PAC* 学习器，若通过MEM_c 预言机进行查询，其查询复杂度都比通过$\mathrm{EX}(\mathfrak{D},c)$ 预言机进行查询的学习器小一个多项式量级。

Proof **相对于成员查询预言机进行学习：**由于 $\mathcal{C}$ 的VC维数为 $|\mathcal{S}|$，存在一个概念 $c \in \mathcal{C}$，使得 $c(x') = 0$，并为实例空间中的其他元素赋予 $2^{|\mathcal{S}|-1}$ 中的某种标注。设 A 是一个针对 c、相对于 $\mathfrak{D}$ 的 PAC 学习器，它向 MEM_c 预言机进行查询的次数至多为 Q_{MQ}。A 需要 $Q = \mathrm{O}\left(\frac{2}{3}|\mathcal{S}|\right)$ 次查询，才能针对 $\mathfrak{D}$ PAC学习 c，因为我们需要学习至多 $\frac{2}{3}|\mathcal{S}|$ 个实例的标签，才能将 c 近似到 ε。

关于 EX 预言机的学习：根据 Arunachalam 和 de Wolf 的引理12 [20], ，我们可以直接得到 $Q_{\mathrm{EX}} = \mathrm{poly}(|\mathcal{S}|)$ 次查询。这比 MQ 情况多了一个多项式因子。□

引理3.1和命题3.1都说明了，使用 MQ 预言机的主动学习为何比依赖采样查询进行学习更强大。在下一节中，我们将说明，量子主动查询与量子采样查询之间也存在一种分离。尽管 MQ 预言机有诸多优势，但也有人对其提出批评，详见C.3节。

3.3 量子成员查询预言机的能力

量子 MQ 预言机 $\mathrm{QMEM}(f)$：$|x\rangle \mapsto |x\rangle|f(x)\rangle$ 本质上是一个函数预言机（见图 A1），在设计量子算法时十分常见，例如算法1。现在我们来介绍有偏量子成员查询预言机的命名体系。

定义 3.6 (ε-有偏预言机)。对于某个 $0 < \varepsilon < \frac{1}{2}$，当满足以下条件时，与布尔函数 f 对应的量子样本查询预言机 $\mathcal{O}_f^{\varepsilon}$ 称为 ε-有偏预言机：

$$\mathcal{O}_f^{\varepsilon}: |0\rangle^{\otimes n}|0\rangle \mapsto \sum_{x\in\mathbb{F}_2^n} \alpha|x\rangle|f(x)\rangle + \sqrt{1-\alpha^2}|x\rangle\left|\overline{f(x)}\right\rangle, \quad \alpha > \frac{1}{2} + \varepsilon.$$

QREX 预言机是 ε-有偏预言机。回想一下，无偏预言机是 $\alpha = 1$ 时的 ε-有偏预言机。Iwama等人 [22] 证明了以下引理:

引理3.2 (使用 ε-有偏预言机模拟无偏预言机)。对于任何使用对无偏预言机的查询访问、以高概率解决某个问题的 $\mathrm{O}(T)$查询量子算法，都存在一个 $\mathrm{O}\left(\frac{T}{\varepsilon}\right)$ 查询量子算法，使用对某个 ε-有偏预言机的查询访问，以高概率解决同一问题。

后来，ε-有偏预言机 的概念由 Chatterjee等人 [23], 进一步推广，他们引入了以下预言机模型。

定义3.7 （强偏置预言机）。与布尔函数 f 相对应的量子样本查询预言机 $\mathcal{O}_f^{\varepsilon_x}$ *cor-* 被称为某个 $0 < \varepsilon_x < \frac{1}{2}, \forall x \in \mathbb{F}_2^n$ 的强偏置预言机，当

$$\mathcal{O}_f^{\varepsilon_x}: |0\rangle^{\otimes n}|0\rangle \mapsto \sum_{x\in\mathbb{F}_2^n} \alpha_x|x\rangle|f(x)\rangle + \sqrt{1-\alpha_x^2}|x\rangle\left|\overline{f(x)}\right\rangle, \quad \alpha_x > \frac{1}{2} + \varepsilon_x, \forall x \in \mathbb{F}_2^n.$$

请注意，在采样预言机 QAEX 下学习，是比在强偏置预言机下学习更强的模型。为了具体证明，即使在量子设定下，采样与主动查询之间也存在分离，我们首先重新证明 Bshouty 和 Jackson 的一个结果 [10]，该结果表明，关于均匀分布，QEX 预言机无法精确模拟QMEM 预言机。首先，我们定义精确模拟性的概念，然后将其扩展为近似模拟性的概念。接着，我们推广 Bshouty 和 Jackson 的结果[10]，证明关于均匀分布，QEX 预言机甚至无法近似模拟QMEM 预言机。

定义 3.8（QMEM 的精确模拟性）。如果存在一个有界错误量子多项式时间（分别为有界错误概率多项式时间）算法 A，使得对于所有 $f \in \mathcal{F}$ 和所有 $x \in \mathbb{F}_2^n$，运行 A 并访问 $\mathcal{O}(\mathfrak{D}, f)$ 时，能够在 x 上生成 $f(x)$，那么量子（分别为经典）样例预言机类 $\mathcal{O}$ 就能针对布尔概念类 $\mathcal{F}$ 和分布 $\mathcal{D}$ 精确模拟 QMEM（分别为 MQ）。

定理 3.1（通过 QEX 精确模拟，Bshouty 和 Jackson [10]）。*QEX* 无法针对 QMEM 相对于 $\mathcal{U}(\mathbb{F}_2^n)$ 进行精确模拟。

*Proof*考虑两个布尔函数 f 和 g，它们仅在一个字符串 $x' \in \mathbb{F}_2^n$ 上不同，例如，令 $f(x) = 0$，对于所有 $x \in \mathbb{F}_2^n$，并且

$$g(x) = \begin{cases} 1 & , x = x', \\ 0 & , \text{otherwise.} \end{cases}$$

考虑一个布尔函数 h，并保证它要么是 f，要么是 g。假设量子学习器 A 向 QEX 预言机 $\mathrm{QEX}(\mathcal{U}, h)$ 发起 Q 次查询，并想要确定它查询的是哪个预言机。此任务等同于尝试区分以下一对量子态：

$$|\psi_f\rangle = \left(\sum_{x \in \mathbb{F}_2^n} \sqrt{\mathcal{U}(x)}\, |x, f(x)\rangle \right)^{\otimes Q}, \quad |\psi_g\rangle = \left(\sum_{x \in \mathbb{F}_2^n} \sqrt{\mathcal{U}(x)}\, |x, g(x)\rangle \right)^{\otimes Q}. \tag{6}$$

根据事实 2.1，我们知道区分 $|\psi_f\rangle$ 和 $|\psi_g\rangle$ 的最佳成功概率为 $\frac{1}{2} + \frac{1}{2} D\left(|\psi_f\rangle, |\psi_g\rangle\right)$。如果对于任意 $\delta \in$ ，该概率都大于 $1 - \delta\,(0, 1/2)$，那么 $\langle \psi_f | \psi_g \rangle \leq 2\sqrt{\delta \cdot (1-\delta)}$。因此，

$$\langle \psi_f | \psi_g \rangle \leq 2\sqrt{\delta \cdot (1-\delta)} \tag{7}$$

$$\implies \left(\sum_{x \neq x'} \sqrt{\mathcal{U}(x)} \right)^{Q} \leq 2\sqrt{\delta \cdot (1-\delta)} \tag{8}$$

$$\implies \left(1 - \mathcal{U}(x')\right)^{Q/2} \leq 2\sqrt{\delta \cdot (1-\delta)}. \tag{9}$$

令 $\varepsilon = \mathcal{U}(x') = \frac{1}{2^{n\circ}}$ 那么，我们有：

$$(1-\varepsilon)^{Q/2} \leq 2\sqrt{\delta \cdot (1-\delta)} \tag{10}$$

$$\implies (1-\varepsilon)^{Q} \leq 4\delta \tag{11}$$

$$\implies Q \log(1-\varepsilon) \leq \log(4\delta) \tag{12}$$

$$\implies \frac{-Q\varepsilon}{1-\varepsilon} \leq \log(4\delta). \tag{13}$$

式（13）中的不等式可由对$1+x$）$\geq x/1+x$，$\forall x > -1$ 应用不等式 log(，其中左端为式（12）。因此，我们有 $Q = \Omega\left(\frac{1}{\varepsilon}\log\left(\frac{1}{\delta}\right)\right)$，即查询数量是超多项式级别的。因此，不存在查询高效学习器 A^{QEX} 能够区分$|\psi_f\rangle$与 $|\psi_g\rangle$，且成功概率为 $\geq 1-\delta$。□

备注 3.5（QMEM 与 QEX 关于均匀分布的无条件分离）。尽管关于均匀分布，*QMEM*无法由*QEX*精确模拟，但我们可以通过向QMEM 预言机查询1 次来模拟*QEX*预言机，如下所示：

$$\mathrm{QMEM}(f)H^{\otimes n}|0\rangle^{\otimes n}|0\rangle = \sum_{x\in\mathbb{F}_2^n}\frac{1}{2^{n/2}}|x, f(x)\rangle \tag{14}$$

这表明，至少对于均匀分布而言，QMEM 查询明确强于*QEX*查询，因此也强于其带噪声的对应形式，即*QREX*和*QAEX*查询。

定理 3.1 的证明表明，如果存在两个仅在一个坐标上不同的函数，那么任何多项式查询量子算法都无法区分相应的量子态。现在，我们将这一概念形式化，以定义近似模拟性的概念。

定义 3.9（ε-远函数）。如果一对布尔函数 f、g 仅在定义域中占比为 ε的一部分上 $f \neq g$，则对于任意 $\varepsilon > 0$，它们是 ε-远的。

我们现在关注以下问题：对于任意一对 ε-远的 f，$g \in \mathcal{F}$，拥有（承诺）采样预言机访问权限的学习算法能否以高概率高效地区分 f 和 g？

定义 3.10（QMEM 的近似模拟性）。考虑任意一对彼此 ε-远的布尔函数 $\{f, g\}$，其中 $\varepsilon > 0$，并令 $h \in \{f, g\}$。对于任意 $\delta > 0$，如果存在一个有界错误量子多项式时间（分别为有界错误概率多项式时间）算法 A，它能够查询访问任意 $\mathcal{O}$（$\mathfrak{D}$、h）A，并以概率至少 $\geq 1-\delta$ 判断 $h = f$ 是否成立，或者 $h = g$ 是否成立，则一类量子（分别为经典）样例预言机 $\mathcal{O}$ 能够相对于分布 $\mathfrak{D}$为布尔概念类 $\mathcal{F}$（ε、δ）模拟（或近似模拟）QMEM（分别为成员查询）。

定理 3.2。*QEX* 无法相对于 $\mathcal{U}$（$\{0,1\}^n$）以 ε、δ）模拟 QMEM。

Proof 假设 f 和 g 是布尔函数，它们在一个多项式规模子集 $\mathcal{S} \subset \mathbb{F}_2^n$上不同，即 $|\mathcal{S}| = \mathrm{poly}(n)$。因此，$\|f-g\|_1 = \frac{1}{2^n}\sum_x |f(x)-g(x)| = \frac{|\mathcal{S}|}{2^n} = \mathrm{poly}(n)/2^n = \varepsilon$。假设量子学习器$A$向 QEX 预言机 QEX$(\mathcal{U}, h)$发起 Q 次查询，并希望区分式 (6) 中描述的量子态 $|\psi_f\rangle$和 $|\psi_g\rangle$。根据事实 2.1，我们知道区分$|\psi_f\rangle$ $|\psi_g\rangle$ $\frac{1}{2}$ +和 的最佳$\frac{1}{2}D(|\psi_f\rangle$ $|\psi_g\rangle)$ $1-\delta$ $\delta \in$ 0 1/2 $\langle\psi_f|\psi_g\rangle \leq$ −成功概率为，。如果对于任意（，），该概率都大于，那么

$2\sqrt{\delta\cdot(1-\delta)}$。因此，

$$\langle\psi_f|\psi_g\rangle \le 2\sqrt{\delta\cdot(1-\delta)} \tag{15}$$

$$\implies \left(\sum_{x\notin\mathcal{S}}\sqrt{\mathcal{U}(x)}\right)^{Q} \le 2\sqrt{\delta\cdot(1-\delta)} \tag{16}$$

$$\implies \left(1-\sum_{x\in\mathcal{S}}\sqrt{\mathcal{U}(x)}\right)^{Q} \le 2\sqrt{\delta\cdot(1-\delta)} \tag{17}$$

$$\implies (1-\varepsilon)^{Q} \le 2\sqrt{\delta\cdot(1-\delta)} \tag{18}$$

$$\implies (1-\varepsilon)^{2Q} \le 4\delta \tag{19}$$

$$\implies 2Q\log(1-\varepsilon) \le \log(4\delta) \tag{20}$$

$$\implies \frac{-2Q\varepsilon}{1-\varepsilon} \le \log(4\delta)\,. \tag{21}$$

公式 (21) 中的不等式公式 (21) 由对公式 (20) 的左端应用不等式 $\log(1+x) \ge x/1+x$，$\forall x > -1$得到。代入 ε 的值可知，$Q = \Omega\left(2^n \operatorname{poly}(n)\log\left(\frac{1}{\delta}\right)\right)$。因此，不存在任何查询高效量子学习器 A^{QEX}，能够以概率 $\ge 1-\delta$ 区分 $|\psi_f\rangle$ 和 $|\psi_g\rangle$。□

表 1 经典与量子 PAC 学习预言机分类。经典**采样**预言机会生成随机样本，而量子**采样**预言机会生成量子态。**主动**预言机允许任意查询。表示列对比了经典比特对与量子叠加态。

Name	范式描述	获得的样本/量子态
经典预言机		
EX	采样	标准样本 $(x, c(x))$，其中 $x\sim\mathfrak{D}$
REX	采样	随机分类噪声 (x, y')，其中 y' 以概率 η翻转
AEX	采样	无关示例 $(x,\ y)\sim\mathfrak{D}_{\mathcal{X}\times\mathcal{Y}}$（联合分布）
MEM	主动	成员查询输入 $x\to$ 返回 $c(x)$
量子谕言机		
QEX	采样	量子示例 $\sum_x \sqrt{\mathfrak{D}(x)}\,\lvert x,\ c(x)\rangle$
QREX	采样	量子 RCN $\sum_x\left(\sqrt{(1-p)\mathfrak{D}(x)}\,\lvert x, c(x)\rangle + \sqrt{p\mathfrak{D}(x)}\,\lvert x, c(\bar{x})\rangle\right)$
QAEX	采样	量子不可知 $\sum_{(x,y)} \sqrt{\mathfrak{D}(x,\ y)}\,\lvert x,\ y\rangle$
QMEM	主动	量子记忆 $\lvert x\rangle\lvert 0\rangle \mapsto \lvert x\rangle\lvert c(x)\rangle$（叠加态）

3.4 其他查询模型中的分离

为回应人们对学习器向 MQ 预言机发起查询时不受限制的性质所提出的担忧，Awasthi 等人 [24] 引入了局部 MQ 预言机模型，在该模型中

学习器只能在[5] 提供给学习器的训练集 S的某个邻域内向 MQ 预言机发起查询。不幸的是，Bary-Weisberg 等人 [25] 证明，尽管使用 q-局部 MQ 进行学习，对于任意常数 $q > 0$ 而言，其能力强于普通的 PAC 学习，但在分布无关设定下，仍不足以学习诸如 juntas 或决策树等相对简单的布尔函数类。另一种类似的预言机模型是 Bshouty 等人提出的随机游走查询（RWQ）预言机模型， $\mathbb{F}_2^n$其中，提供给学习器的第一个点是从 $\mathbb{F}_2^n$中均匀采样得到的，而后续每个点都是通过在上进行均匀随机游走获得的。

备注 3.6. 这里指出，局部 *MQ* 和 *RWQ* 预言机可用于针对均匀分布 *[24, 26]*高效地学习析取范式的 *PAC* 学习器。

Childs 等人 [27] 证明，在对某些有限图类进行图遍历时，随机游走查询与连续时间量子游走查询（CQWQ）之间存在指数级分离。由此自然会想到以下问题：

开放问题 1 （局部 MQ 与 CQWQ 查询学习之间的分离）。 能否找到一个自然的概念类，使得使用 *RWQ*（或局部 *MQ*）查询进行学习与使用 *CQWQ* 查询进行学习之间存在分离？

开放问题 2 （使用 CQWQ 查询模拟 QMQ）。 *CQWQ* 查询能否相对于 $\mathcal{U}$ （$\{0,1\}^n$） 精确或近似地模拟 *QMQ*？

4 样本复杂度分离

在无噪声设定下，Blumer 等人 [28] 和 Hanneke [29] 证明，学习 VC 维为 d 的概念类 $\mathcal{C}$ 的经典（ε， δ）-PAC 样本复杂度为 $\Theta\left(\frac{d}{\varepsilon}+\frac{\log(1/\delta}{\varepsilon}\right)$。在不可知设定下，Vapnik 和 Chervonenkis [7] 以及 Talagrand[30] 证明，学习 VC 维为 d的概念类$\mathcal{C}$的（ε， δ）-不可知 PAC 样本复杂度为 $\Theta\left(\frac{d}{\varepsilon^2}+\frac{\log(1/\delta}{\varepsilon^2}\right)$.

我们回顾一下，在可实现情形下，能够访问 k 经典样本（或量子样本）分别等价于向 EX 预言机（或 QEX 预言机）进行 k 查询；而在无关情形下，则分别等价于向 AEX 预言机（或 QAEX 预言机）进行 k 查询。关于量子与经典样本复杂度的首批结果由 Servedio 和 Gortler [31] 给出，他们证明，在无噪声设定下，对于任何能够使用 D 次 EX 预言机查询和 Q次 QEX 预言机查询进行 PAC学习的概念类 $\mathcal{C}$ ，量子样本复杂度与经典样本复杂度之间存在多项式关系，即为 $D = \mathrm{O}(nQ)$。

这一结果后来由 Arunachalam 和 de Wolf [20], 加以改进，他们证明，在无噪声设定与不可知设定下，量子样本复杂度均等于相应的经典样本复杂度（至多相差一个常数因子），从而证明

[5]在实例空间 $\mathcal{X}$上固定一个度量，例如，当实例从布尔超立方体中采样时，考虑汉明距离 d 。现在，我们只允许查询训练集中的任意实例周围半径为 d 的开球内的样本。

量子样本并不天然比经典样本更强大。其证明技术 [20]包括使用“近乎最优测量”和傅里叶分析的状态识别论证。Hadiashar 等人 [32] 后来重新证明了[20]的结果，所用论证为信息论论证。

引理 4.1 (量子样本复杂度 [20, 32])。给定基准概念类 $\mathcal{C}$，其 *VC* 维为 d，学习$(\varepsilon,\ \delta)$-量子 *PAC* 样本复杂度为$\mathcal{C}$，$\Theta\left(\frac{d}{\varepsilon}+\frac{\log(1/\delta}{\varepsilon}\right)$；而学习$(\varepsilon,\ \delta)$-量子不可知 *PAC* 样本复杂度为$\mathcal{C}$ $\Theta\left(\frac{d}{\varepsilon^2}+\frac{\log(1/\delta}{\varepsilon^2}\right)$.

然而，假设实例上的边缘分布为引理 4.1可能并不成立。例如，Caro [33] 指出，在有界噪声设定下，当样本从乘积分布中抽取时，学习布尔线性函数可获得量子样本复杂度加速。这引出了以下问题：

开放问题 3 (非均匀边缘分布下的量子样本复杂度分离)。是否存在一个自然的非线性布尔函数概念类，使我们能够证明在非均匀边缘分布下的量子样本复杂度加速？

对于非均匀边缘分布的情形，我们可以将注意力集中在一些特殊的边缘分布上，例如乘积分布、平滑乘积分布或各向同性对数凹分布。值得注意的是，即使对开放问题 3的答案是否定的，仍然存在实现计算分离的可能性，后续章节将对此进行讨论。

4.1 广义量子样本预言机

最近，Caro 等人 [34]提出了一种新型量子不可知样本预言机，称为叠加态混合（MOS）预言机 $\mathcal{O}_{\mathfrak{D}}$。对于任意分布 $\mathfrak{D}$ （定义在 $\mathcal{X}\times\mathbb{F}_2$ 上），查询 $\mathcal{O}_{\mathfrak{D}}$ 可得到混合态

$$\rho_{\mathfrak{D}}=\mathop{\mathbb{E}}_{f\sim\mathcal{F}_{\mathfrak{D}}}\left[\left|\psi_{\mathfrak{D}_{\mathcal{X}},f}\right\rangle\left\langle\psi_{\mathfrak{D}_{\mathcal{X}},f}\right|\right], \tag{22}$$

其中，$\mathfrak{D}_{\mathcal{X}}$ 是定义在 $\mathcal{X}$ 上的边缘分布，而 $\mathcal{F}_{\mathfrak{D}}$ 是由 $\mathcal{D}$诱导的、定义在所有标记函数 f: $\mathcal{X}\mapsto\{0,1\}$ 上的概率分布。测量$\rho_{\mathfrak{D}}$ 的全部 $n+1$ 个量子比特，可得到如下来自 $\mathfrak{D}$ 的样本：

$$\langle x,y|\,\rho_{\mathfrak{D}}\,|x,y\rangle=\mathop{\mathbb{E}}_{f\sim\mathcal{F}_{\mathfrak{D}}}\left[\left|\langle x,y|\psi_{\mathfrak{D}_{\mathcal{X}},f}\rangle\right|\right]=\mathop{\mathbb{E}}_{f\sim\mathcal{F}_{\mathfrak{D}}}\left[\mathfrak{D}_{\mathcal{X}}(x)\cdot\mathbb{I}_{f(x)=y}\right] \tag{23}$$

$$\Longrightarrow\ \langle x,y|\,\rho_{\mathfrak{D}}\,|x,y\rangle=\mathfrak{D}_{\mathcal{X}}(x)\cdot\Pr_{f\sim\mathcal{F}_{\mathfrak{D}}}\mathbb{I}_{f(x)=y}]=\mathfrak{D}(x,y). \tag{24}$$

MOS 预言机直接推广了 QEX、QREX 和 QAEX 预言机。诚然，它比 QAEX 预言机更一般，因为它允许我们对.....进行建模

相关标签噪声。稍后在 第 5.2 节，我们将看到表明 MOS 预言机威力的结果。然而，就样本复杂度而言， [34] 证明，在不可知设定下，量子（MOS）样本复杂度与经典样本复杂度仅相差对数因子。具体而言，MOS 量子不可知样本复杂度为 $\Omega\left(\frac{d}{\varepsilon^2}\text{对数}\,d+\frac{1/\delta}{\varepsilon^2}\right)$。这自然引出了以下问题。

开放问题4 （最优叠加混合样本复杂度下界）。 叠加混合样本复杂度中的 $\frac{d}{\varepsilon^2}\text{对数}\,d$ 项是否是紧的？

如果对 开放问题4的回答是肯定的，这将表明，即使从样本复杂度的角度来看，MOS 模型相较于标准量子示例模型也具有轻微但不可忽略的优势。

4.2 多分类设定中的样本复杂度

Mohan 和 Tewari [35] 研究了多种不同的学习设定，例如带噪声和不可知的离线与在线量子 PAC 学习，其中 $|\mathcal{Y}|<\infty$，并表明在除一种情况之外的所有情况下，量子与经典样本复杂度至多相差常数因子（在线多类别非特定设置中的界并不紧确，与已知的最佳经典界相差一个对数因子[6]）。Mohan 和 Tewari [35] 的研究自然引出了以下问题。

开放问题5 （多分类无先验样本复杂度）。 在在线多类别非特定设置中，经典查询/样本复杂度与量子查询/样本复杂度是否等价（至多相差常数因子）？

开放问题6 (无界标签的样本复杂度分离)。 在无界标签情形下，即使是在可实现离线 *PAC* 设置中，经典与量子查询/样本复杂度是否等价（相差至多一个常数因子）？

开放问题7 (多分类叠加态混合样本复杂度)。 在多分类可实现设定和不可知设定中，叠加态混合样本复杂度 *[34]* 是多少？

最后，我们也可以从近期提出的其他量子在线学习框架的角度来探讨上述问题。

开放问题8 （不同框架中的样本复杂度分离）。 *Mohan* 和 *Tewari* 的在线学习框架 *[35]* 能否扩展到 *Bansal* 等人 *[36]*提出的更一般的量子在线学习框架？

[6]关于这一设定的详细讨论，请参阅 [35] 第5.4节

4.3 通过酉访问实现的样本复杂度分离

在 定义 3.4 和 3.5中，学习器无需假设能够查询 QEX 和 QAEX 预言机，而可以改为访问方程(3) 和(5)右侧对应量子态的副本。的结果在这一等价设定下同样成立。本综述中的学习算法均假设采用此设定。

另一方面，若我们假设处于白盒设定中，即量子学习器可以酉访问直接生成量子样本的量子谕言机（而不只是能够访问量子样本），那么 Salmon等人 [37]证明了，对概念类 ε, δ)-（无噪声）PAC 学习的样本复杂度$\mathcal{C}$ 在 VC 维数为 d时为 $\tilde{\mathrm{O}}\left(\frac{d}{\sqrt{\varepsilon}}+\log\left(\frac{1/\delta}{\sqrt{\varepsilon}}\right)\right)$.

开放问题 9 (白盒不可知样本复杂度分离)。 *Salmon*等人的量子样本复杂度优势 *[37]* 是否延伸至不可知设定？

Salmon等人 [37] 的证明关键地用到了这样一个事实：拥有酉访问意味着可以访问酉算子的逆。我们可以从更偏灰盒的设定出发，提出以下问题。

开放问题 10 (灰盒样本复杂度分离)。 当量子学习器能够访问一个生成量子示例的量子信道时，学习的量子（可实现／不可知）样本复杂度是多少？

在这里，我们假设，能够访问一个量子信道并不一定意味着能够访问其逆信道。鉴于 Tang 和 Wright [38], 最近的研究，这个问题变得非常有趣；他们指出，在涉及放大和估计的量子加速中，访问逆信道是关键组成部分。最后，我们采用这一新方法重新审视 Mohan 和 Tewari [35] 的样本复杂度结果。

开放问题 11 (白盒多分类样本复杂度分离)。 *Salmon*等人 *[37]* 所展示的量子样本复杂度优势，能否扩展到 *Mohan* 和 *Tewari* *[35]*探讨的多分类情形？

5 时间复杂度分离

第 4 节 中的结果是信息论性质的，并未排除量子与经典学习器在计算意义上的分离。换言之，第 4 节 中的结果表明，对于每个概念类 $\mathcal{C}$ ，如果量子学习器能够使用多项式数量的样本/查询来学习它，那么经典学习器也能够使用多项式数量的样本/查询来学习它。然而，引理 4.1并不意味着

如果概念类 $\mathcal{C}$ 能够由量子学习器高效学习，那么它也必定能够由经典学习器高效学习。本节将探讨量子学习器与经典学习器在运行时间上的分离。

5.1 通过隐藏子群问题实现的分离

Servedio 和 Gortler [31] 首先通过使用一个 $\mathcal{C}$ 基于因数分解 Blum整数的概念类[7] [39]，为量子学习与经典学习之间存在分离提供了有力证据。因数分解 Blum整数被认为在经典计算中具有计算难度 [39, 40]，，但对于阿贝尔隐藏子群问题（Abelian Hidden Subgroup Problem(HSP)），量子算法能够高效地解决该问题 [5]。

随后，Liu 等人通过使用一个 [41], 基于离散对数问题（DLP）的概念类 $\mathcal{C}_{\mathrm{DLP}}$，展示了类似的经典学习与量子学习之间的分离。

定义 5.1（DLP）：**给定一个大素数 、一个的生成元 以及一个输入，在时间 $\log_g x$** poly（$\lceil \log_2 p \rceil$）内计算对数。

人们普遍认为，离散对数问题在经典计算中难以求解 [42]，，但我们可以使用 Shor算法高效地解决离散对数问题 [5]。利用这一技术，[41] 获得了一种量子核估计流程，用于学习 $\mathcal{C}$离散对数问题。

5.1.1 有限阿贝尔群上的隐藏子群问题

[31, 41] 得到的结果并不令人意外，尤其是考虑到 Simon算法和 Shor算法——它们是量子加速的典型范例。Simon算法和 Shor算法都解决了一个更一般问题的特定实例，这个问题称为隐藏子群问题（HSP），定义如下。

定义 5.2（隐藏子群问题）。设 $\mathcal{G}$ 是一个已知有限群，$\mathcal{S}$是一个有限集，且有一个黑盒函数 f：$\mathcal{G} \mapsto \mathcal{S}$，满足 $f(x) = f(y) \iff x\mathcal{H} = y\mathcal{H}$，其中 $\mathcal{H} \leq \mathcal{G}$ 是某个未知子群，且 $h \in \mathcal{H}$。求出$\mathcal{H}$的一个生成集。

在定义 5.2中，黑盒函数 f 被称为隐藏子群 $\mathcal{H}$。或者，f 也可以理解为区分 $\mathcal{H}$的陪集。量子计算与经典计算之间的大多数指数级分离，都源于量子算法在阿贝尔群上的隐藏子群问题方面实现了量子加速。

例 5.1。可实现量子加速的阿贝尔群隐藏子群问题的受限版本包括

- 多伊奇问题：$\mathcal{G} = \mathbb{Z}_2$ 和 $\mathcal{H}$ 要么为0 （平衡），要么为 $\mathbb{Z}_2$ （常数），
- ***Simon***问题：$\mathcal{G} = \mathbb{Z}_2^n$ 与 $\mathcal{H} = \{0, s\}$，其中 $s \in \mathbb{Z}_2^n$ 是某个秘密，并且

[7] 布鲁姆整数是一个整数 $N = pq$，其中 p，q 是与 4 模 3 同余的 m位素数。

• 离散对数问题： $\mathcal{G}=\mathbb{Z}_p\times\mathbb{Z}_p$，并且$\mathcal{H}=\{(0,\ 0),\ (1,\ \log_g x)\ .....,\ (p-1,\ (p-1)\log_g x)\}$。

事实上，我们知道如何高效求解任意阿贝尔隐藏子群问题。

引理 5.1（高效求解阿贝尔隐藏子群问题，[43]）。设 $\mathcal{G}$ 是一个已知的有限群，$\mathcal{S}$ 是一个有限集，$f:\ \mathcal{G}\mapsto\mathcal{S}$ 是一个隐藏子群$\mathcal{H}\leq\mathcal{G}$的黑盒函数。存在一种量子算法，其规模为 $\mathrm{O}(\mathrm{polylog}\,(|\mathcal{G}|))$，运行时间为 $\log|\mathcal{G}|$，并以至少 $1-1/|\mathcal{G}|$ 的概率输出 $\mathcal{H}$ 的一个生成集。

高效求解有限阿贝尔群上的隐藏子群问题的一个关键组成部分，是存在对 $\mathcal{G}$ 高效执行的量子傅里叶变换（QFT）。Cleve 和 Watrous [44] 证明了存在一个深度为 O (log $n+\mathrm{polylog}\,(1/\varepsilon)$)、规模为 O ($n\log n/\varepsilon$) 的电路，可用于近似计算群 $\mathbb{Z}_{2^n}$上的量子傅里叶变换。Cheung 和 Mosca[45] 提出了一种高效的量子分解算法：给定有限阿贝尔群 $\{g_1,\ldots,g_s\}$ 的一个生成集 $\mathcal{G}$，我们可以在 polylog ($|G|$) 时间内输出一组元素d_1，.....，d_ℓ使得 $\mathcal{G}=\mathbb{Z}_{d_1}\times\ldots\times\mathbb{Z}_{d_\ell}$。结合 Cleve 和 Watrous [44],以及 Cheung 和 Mosca 的结果 [45]，即可得到 Shor 整数分解算法（借助引理 D.1）。

5.1.2 作为概念类的隐藏子群问题

现在我们定义一个由隐藏子群问题诱导的新概念类。

定义 5.3 (隐藏子群问题诱导的概念类)。任何关于隐藏子群问题实例 $\mathsf{P}=(\mathcal{G},\mathcal{H})$ 的判定问题$\mathcal{X}$，只要可以表述为关于实例集合的成员资格问题，就称为由隐藏子群问题诱导的概念类 $\mathcal{C}:\ \mathcal{X}\mapsto\{0,1\}$ $_{\mathsf{P}}$。

例 5.2。设 $\mathsf{P}=(\mathcal{G},\mathcal{H})$ 是一个隐藏子群问题实例，且满足 $|\mathcal{G}|=2^n$。以下判定问题 $\mathcal{C}:\ \mathbb{F}_2^n\mapsto\{0,1\}$ 都是由隐藏子群问题诱导的概念类$_{\mathsf{P}}$:

• 给定一个字符串 $x\in\mathbb{F}_2^n$，$|\mathcal{H}|\geq|x|$是否成立?

• 给定一个字符串 $x\in\mathbb{F}_2^n$，$\mathcal{H}$ 是否具有规模至少为 $|x|$ 的生成集?

注 5.1。定义 5.3并不局限于阿贝尔隐含子群问题的情形，也可以针对任意有限群上的隐含子群问题来定义。

经典地，为任意有限阿贝尔群上的隐含子群问题所诱导的概念类获得高效学习器，需要依赖各种困难性假设；而有限阿贝尔群上的隐含子群问题所诱导的概念类似乎允许高效的量子学习算法。这引出了以下开放问题。

开放问题12（阿贝尔 HSP 诱导的概念类的分离）。如果 $\mathcal{C}:\ \mathbb{F}_2^n\mapsto\{0,1\}$ 是由有限阿贝尔群 $\mathcal{G}$上的某个 *HSP* 诱导的任意概念类，我们能否证明存在一个高效的量子学习算法来学习 $\mathcal{C}$? 在受限和非受限噪声模型下，结论是否也成立?

备注 5.2。 我们猜想 开放问题*12* 至少在无噪声情形下成立。

观察 5.1。 假设 *DLP* 的困难性 *[39, 40]*,开放问题*12*意味着，对于若干 *HSP* 诱导的概念类，即使在无噪声情形下，量子学习与经典学习之间也存在分离（如 *[31, 41]*中所示）。

5.1.3 任意群上的隐藏子群问题

尽管已知的阿贝尔群上 HSP 的高效算法需要高效的 QFT 子程序，但目前尚不清楚，能否在群 G 上高效执行 QFT 是在非阿贝尔情形下高效求解 HSP 的必要条件或充分条件。Ettinger 等人 [46] 表明，存在基于 QFT 的量子算法，可以求解任意有限群 $\mathcal{G}$ 上的 HSP；这些算法只需对 O_f 进行 $\log|G|$ 次查询，但运行时间是指数级的。

对于某些性质良好的非阿贝尔群类，例如元循环群[8]，甚至元阿贝尔群[9] [47]，存在能够精确计算或近似计算 QFT 的高效算法。建议读者参阅 Lomont [48] 和 Wang [49]撰写的综述，以全面了解任意有限群上的 HSP；即使脱离学习理论的背景，这本身也是一个重要问题。在某些有限非阿贝尔群类上高效求解 HSP 的量子算法，会在理论计算机科学中带来非常有趣的结果，例如高效求解图同构（GI）和最短向量问题（最短向量问题（SVP））。特别是，SVP 是后量子密码学领域的核心问题，其与学习的联系将在下一节讨论。

5.1.4 其他开放问题

考虑对称群 S_n。由于 S_n 是一个具有 $n!$ 个元素的群，定义任何能体现量子与经典可学习性之间分离的、由隐含子群问题诱导的概念类，都会非常有意义。研究由 S_n 的子群诱导、具有更多结构的概念类，以寻求高效的量子学习算法，也可能很有价值。

开放问题 13（非阿贝尔隐含子群问题诱导的概念类的分离）。 是否存在由隐含子群问题在 S_n 或其任意子群（例如置换群或交错群）上诱导的概念类，使得量子学习算法与经典学习算法之间存在超多项式分离？

我们在此指出，开放问题 13 的答案很可能是否定的，前提是我们对量子算法额外要求具备高效可学习性。我们特别强调二面体群这一重要情形（见定义 D.1），因为它们与基于格的密码学密切相关。

---

[8]元循环群 $\mathcal{G}$ 是指具有循环正规子群 $\mathcal{N}$ 的群，且 $\mathcal{G}/\mathcal{N}$ 也是循环群。[9]元阿贝尔群 $\mathcal{G}$ 是指具有正规子群 $\mathcal{N}$ 的群，且 $\mathcal{N}$ 和 $\mathcal{G}/\mathcal{N}$ 均为阿贝尔群。

非正式地说，若存在一个高效的量子算法来求解二面体隐含子群问题，则意味着 SVP $\in$ 有界错误量子多项式时间 [50]。这基本上意味着NP $\subseteq$ 有界错误量子多项式时间，因为已有研究表明，在某些受限情形下，SVP 是NP-困难的 [51]。目前已知求解二面体隐含子群问题的最佳算法，是 Kuperberg 提出的一种亚指数时间算法 [52]。以下开放问题进一步凸显了上述讨论。

开放问题 14 （学习二面体 HSP 诱导的概念类的困难性）。我们能否为在任意二面体群 $\mathcal{G}$ 上高效量子学习任何由 *HSP* 诱导的概念类提供困难性保证 $N \geq 3$?

5.2 基于格问题的分离

在讨论计算困难性时，我们通常关注已知具有最坏情况困难性保证的问题，因为我们希望设计出即使面对最坏输入也能高效运行的算法。然而，要确保密码方案安全，困难性保证必须对随机密钥也成立，也就是说，我们需要平均情况困难性保证。这是一项具有挑战性的任务，因为即使问题具有最坏情况困难性保证，我们也无法直接看出如何构造困难实例。事实上，对于许多NP困难问题（例如图着色，见 [53]），并不存在这样的保证。

5.2.1 带误差学习问题简介

Ajtai [54] 证明，对于某些基于格的问题，存在从最坏情况困难性到平均情况困难性的此类归约。此后，Regev [50]引入了带误差学习（LWE）问题，并证明，求解LWE问题的高效算法可以为我们提供一个求解最短向量问题（SVP）的高效算法[10] 。现在，我们来定义 LWE问题的判定版本。

定义 5.4 (LWE判定版本)。给定 m 个独立样本$(x, y) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$，其中 $q \geq 2$为某个素数，并保证这些样本是从以下分布之一中抽取的：

1. 均匀分布：$(x, y) \sim \mathfrak{U}\left(\mathbb{Z}_q^n \times \mathbb{Z}_q\right)$。*2. LWE*分布：$x \in \mathbb{Z}_q^n$ 的边缘分布是均匀分布，且 $y = \langle s,x\rangle + \varepsilon$ 模 q，其中 $s \in \mathbb{Z}_q^n$ 是秘密字符串，$\varepsilon \sim \mathfrak{D}$ 是某个噪声参数。

判断这些样本是否来自 *LWE*分布。

LWE问题也是另一种重要学习问题——带噪声奇偶学习（LPN）的一种推广，其中 $q = 2$ 与 $\mathfrak{D}$ 是 $\mathbb{Z}_2$上的伯努利分布。LPN是译码线性码 [55]这一NP难问题的平均情形版本。

[10]高效求解 LWE 实际上意味着可以高效求解间隙最短向量问题。

观察 5.2. 如果没有随机噪声项 ε，就可以使用高斯消元法对样本高效地求解*LWE*和*LPN*。

备注 5.3. 除非另有明确说明，*LWE* 和 *LPN* 问题均假设处于 *RCN* 设定，即 ε 是随机分类噪声。

在上一节中，我们看到 SVP 可归约到二面体隐藏子群问题，这与 LWE 类似。下面我们将明确说明这两类问题之间的联系。

备注 5.4（LWE 与二面体隐藏子群问题之间的联系）。*LWE* 可通过平均情形到最坏情形归约，归约到二面体隐藏子群问题的一个更强变体，即鲁棒二面体隐藏子群问题。由于已知用于求解二面体隐藏子群问题的亚指数时间（量子）算法 *[52]* 并不能完全适用于鲁棒二面体隐藏子群问题，因此目前尚无已知的用于 *LWE* 的亚指数时间（量子）学习算法。

5.2.2 LWE 和 LPN 的学习算法

Blum等人 [56] 给出了首个用于 LWE（因此也适用于 LPN）的经典亚指数时间与样本学习算法，该算法需要 $2^{O(n/\log n)}$ 时间和样本。有趣的是，Feldman等人 [57]证明了在不可知设定 下，针对均匀边缘分布的 LPN 可归约到 RCN 设定下针对均匀边缘分布的 LPN。因此， [56] 的算法也意味着，在不可知设定下（针对均匀边缘分布）存在一个用于 LPN 的亚指数时间/样本学习算法。

然而，在量子设定下，存在针对 LWE [58]和 LPN [59], 的高效量子学习算法，它们可以访问一种特殊预言机，我们将其记为 $\mathrm{QREX}^{s,\ \varepsilon}_{\mathrm{LWE}}$ 预言机，并定义如下：

$$\mathrm{QREX}^{s,\varepsilon}_{\mathrm{LWE}}\left|0\right\rangle\left|0\right\rangle \mapsto \frac{1}{\sqrt{q^n}}\sum_{x\in\mathbb{F}_q^n}\left|x\right\rangle\left|\langle s,x\rangle+\varepsilon \mod q\right\rangle ,\varepsilon\sim\mathfrak{D}.$$

需要注意的是，Grilo 等人 [58] 的结果仅在向量子学习器提供 $\mathrm{QREX}^{s,\ \varepsilon}_{\mathrm{LWE}}$ 预言机访问权限时才成立；这与经典情形不同，在经典情形下，我们只能访问 m 经典带标签样本。

注 5.5. *LWE* 和 *LPN* 的高效学习结果 *[58, 59]*也可以从 *MOS* 模型 *[34]*推出，具体如第 *4.1* 节所述。

尽管强大的预言机模型能够提供分离结果，但理解量子学习器实现加速的确切途径仍然很重要。因此，自然会想到以下问题：

开放问题15（随机样本条件下LWE的量子学习）。给定对形式为：
$\frac{1}{\sqrt{|\mathcal{S}|}}\sum_{x\in\mathcal{S}\subseteq[m]}\left|x\right\rangle\left|\langle s,x\rangle+\varepsilon \text{ 模 } q\right\rangle$的量子样本的访问权限，其中 $\mathcal{S}$未知且 $x\sim\mathfrak{U}\left(\mathbb{Z}_q^n\right)$，
是否存在针对*LWE*的高效量子学习算法？

我们注意到，对开放问题15 的回答很可能是否定的，因为出于多种原因，人们认为LWE是困难的。如果在随机样本条件下存在针对LWE的高效量子算法，那么就会存在破解学习带错误密码系统的高效量子算法。这极不可能，因为这样的量子算法也意味着存在求解最短向量问题的高效算法。如果最短向量问题被认为是NP完全的，那么对开放问题15 的肯定回答将意味着NP ⊆有界错误量子多项式时间。

与 LWE 不同，LPN 的密码学应用范围一直较为有限 [60]。最近，一系列工作 [61, 62] 表明，在极低噪声率下，最近码字问题（NCP）的最坏情况困难性可由极高噪声率下 LPN 的困难性推出。然而，与 SVP 的困难性保证不同，人们并不认为 NCP 是NP困难的 [61]。因此，有必要将是否存在高效的 LPN 量子算法作为一个独立于 LWE 的问题来探讨。

开放问题16 （随机样本下针对 LPN 的量子学习）。 给定以下形式的量子样本：

$$\frac{1}{\sqrt{|\mathcal{S}|}} \sum_{x \in \mathcal{S} \subseteq [m]} |x\rangle \, |\langle s, x\rangle + \varepsilon \mod 2\rangle \,,$$

其中 $\mathcal{S}$ 未知， $x \sim \mathfrak{U}(\mathbb{Z}_2^n)$ ，且 $\varepsilon \sim \mathrm{Ber}(\mathbb{Z}_2)$ ，是否存在用于 *LPN* 的高效量子学习算法？

5.3 半空间学习分离

半空间是一类布尔函数，对于学习算法领域极为重要。提升算法、支持向量机和神经网络在底层都采用了半空间学习问题的某种变体。半空间是一种布尔函数 $c \in \mathcal{C}$: $\mathbb{R}^d \mapsto \mathbb{F}_2$ 定义如下：

$$c(x) = \operatorname{sign}\left(\langle a, x\rangle - \theta\right) = \operatorname{sign}\left(\sum_{i \in [m]} a_i x_i - \theta\right), \quad a, x, \theta \in \mathbb{R}^d.$$

任何学习器只要能接收从分布 $\mathfrak{D}$ 上$\mathbb{R}^d \times \mathbb{F}_2$ 抽取的带标签样本，并输出一个假设 h 使得误差$_{\mathfrak{D},}$ $c\,(h) \leq$ 小于等于误差$_{\mathfrak{D}}(\mathcal{C})+\beta+\varepsilon$ ，就是一个 $beta$-最优学习器。如果其运行时间为多项式时间 poly $\left(d, \frac{1}{\varepsilon}\right)$，则认为该学习器是高效的。

容易看出，在可实现情形／无噪声设定下，可以利用线性规划技术高效地学习半空间。对于有界噪声模型，例如随机分类噪声 [63]，、Massart噪声 [64]和 Tsybakov噪声 [65] ，在对数凹边缘分布下，也存在经典的高效半空间学习器。

半空间学习中最有趣的情形似乎超出了有界噪声模型。在不可知设定下，假设边际分布均匀，学习半空间便可直接推出 LPN 的学习算法。因此，半空间学习似乎比 LPN 难得多。对于不可知设定下半空间的非适当 [66, 67] 经典学习和适当 [57, 68, 69] 经典学习，已有多种困难性结果。

需要注意的是，这些困难性结果针对的是随机经典样本设定（不含查询）。此前，我们已经看到，在 LWE [58]或 LPN [59] 的情形下，当量子学习器能够查询访问 QREX 预言机时，存在量子与经典学习分离的证据。值得探讨的是，即使对于半空间学习，这种分离是否也有可能实现。

开放问题17 （半空间的高效量子不可知学习）。 假设实例上的边缘分布均匀，是否存在一种能够访问QAEX查询、且在样本、查询和时间方面都高效的量子不可知学习器来学习半空间?

5.4 决策树的学习分离

布尔函数 $f:\mathbb{F}_2^n \mapsto \mathbb{F}_2$ 的决策树表示是一棵二叉树，其中从根结点到任一叶结点的每条路径都对应 f 的一行真值表。叶结点标记为 $\{0,1\}$ 或 $\{\pm 1\}$，内部结点标记为文字：变量 x_i （或其取反），且每个内部结点都有两条分别标记为 true 和 false 的出边。决策树的叶结点数（或内部结点数）称为其**规模**，而从决策树根结点到任一叶结点的最长路径称为该决策树的**深度**。由于布尔函数的决策树表示并不唯一，在计算上，我们更希望用“最小”的决策树来表示 f。然而，Hyafil与Rivest [70] 证明，构造任意布尔函数对应的最优决策树是一个NP-完全问题。

针对深度有多项式界的决策树，高效学习算法的研究由Ehrenfeucht与Haussler [71], 开创。他们在无噪声设定下给出了经典的恰当学习算法，其运行时间和样本复杂度均为拟多项式。Koch等人 [72] 最近证明，Ehrenfeucht与Haussler [71] 的结果基本上是最优的（即使可以访问成员查询）。Kushilevitz与Mansour [73] 在其开创性工作中，通过考虑非真学习设定，绕过了决策树学习的超多项式下界。他们的结果后来被推广到不可知设定 [74–76]。然而，所有这些算法都需要访问成员查询。Chatterjee等人 [23] 给出了适用于所有噪声模型的决策树量子学习算法，其复杂度关于所有参数均为严格多项式级别，且只需访问采样查询预言机。

5.4.1 技术与开放问题

Kushilevitz与Mansour [73] 表明，若布尔函数 f 由规模为 t 的决策树计算，则 $\|\mathrm{FS}(f)\|_1 \leq t$。或者，若布尔函数 f 由至多具有 $\mathrm{O}(t)$ 个节点的决策树计算，则 $\|\mathrm{FS}(f)\|_1 = \mathrm{O}(t)$。因此，我们可以直接应用推论 C.2 得到一个使用成员查询、面向多项式规模决策树的多项式时间学习算法。难点在于消除高效可学习性对成员查询或MQ查询的依赖。下面，我们讨论如何在无噪声情形下实现这一点，并将更具限制性噪声设定下的学习讨论留待Chatterjee等人 [23]。

Bshouty 和 Jackson [10] 展示了如何通过 1 次调用 QEX $(\mathcal{U},\ f)$ 预言机获得近似傅里叶采样态[11]，如下所示：

$$\frac{1}{\sqrt{2}}\sum_{\mathcal{S}\subseteq[n]}\hat{f}(\mathcal{S})\,|\mathcal{S}\rangle\,|1\rangle+\frac{1}{\sqrt{2}}\,|0\ldots0\rangle\,|0\rangle\,. \tag{25}$$

对式 (25)中得到的态应用振幅放大，我们现在可以再对 QEX $(\mathcal{U}, f)$ 预言机进行额外的 $\Theta(1)$ 次查询，从而得到 ˆ 傅里叶采样态 $|\psi\rangle=\sum_{\mathcal{S}\subseteq[n]}f(\mathcal{S})\,|\mathcal{S}\rangle$。本着推论 C.2的精神，现在可以直接通过对傅里叶分布进行采样，得到决策树的弱学习器，并利用（量子或经典）提升算法得到相应的强学习器。

正如细心的读者必定已经推断出的那样，无论在经典设定还是量子设定中，用于非恰当决策树学习的学习算法（尤其是在不可知设定中）都假设实例上的边缘分布为均匀边缘分布。这就引出了如下问题：

开放问题 18 （非均匀边缘分布下决策树的高效量子非知情学习）**.** 我们能否在无法访问 *QMQ* 的情况下，针对实例上具有非均匀边缘分布的多项式有界决策树，得到高效的量子（不可知）学习器？

5.5 析取范式的学习分离

析取范式函数 f 是析取范式子句的析取，而析取范式子句又是布尔文字 $\tilde{x}$的合取。从数学上说，

$$f(x)=C_1(x)\vee C_2(x)\vee\ldots\vee C_m(x),$$

其中，每个子句 $C(x)=\wedge_{i\in\mathcal{S}}\tilde{x}_i$ 都是相对于某个 $\mathcal{S}\in[n]$定义的。析取范式 f 的宽度是其任一子句中文字的最大数量，而析取范式f 的规模则是 f中子句的总数。

当学习器只能访问随机样本时，Daniely 和 Shalev-Shwartz[77] 证明，即使进行非真学习析取范式，也意味着存在一种能够有效判定随机 k-SAT 公式是否不可满足的算法。[12] 由于在可实现/RCN设定下学习析取范式可归约为学习 LPN [57], ，因此 [77] 的结果也给出了使用随机样本经典学习 LPN 的困难性结果。在经典设定下，我们知道，只有当学习器能够进行成员查询 [79]时，才存在适用于无噪声/RCN设定、且假设边缘分布为均匀分布（或“近似”均匀分布）的析取范式（深度为 2 的 AC^0 电路）高效非真学习算法。然而，在量子学习设定中，Bshouty 和 Jackson [10] 给出了高效的量子学习算法：使用可访问 QEX 或 QREX 预言机的量子学习器，在可实现和 RCN 设定下非真学习析取范式。

在非均匀边缘分布设定下，Feldman [80] 证明，在成员查询模型中，当

[11]参见引理 B.1[12]如果 P $\neq$ NP，那么很难证明 k-SAT 公式 [78]。

实例来自任意乘积分布时，存在一种高效的经典学习算法。他们还证明，只有当实例来自平滑乘积分布时，才存在一种仅使用随机样本的 DNF 学习算法 [81]。Kanade 等人 [82] 证明，当样本来自乘积分布时，能够访问 QEX 的量子学习器可以高效地 PAC 学习析取范式。

5.5.1 技术与开放问题

Jackson [83] 证明，任意一个 s项析取范式在其定义域至多 $1/2 + 1/\mathrm{poly}(s)$ 的比例上具有谱集中性。因此，我们可以直接应用 引理 C.2，在无噪声设定下经典地学习析取范式（在有界噪声设定下，s和噪声参数会带来多项式级的开销）。借鉴决策树学习部分所述的类似技术，Bshouty 和 Jackson [10] 通过使用量子采样预言机，消除了对成员查询的依赖。这引出了以下关于在无限制噪声设定下学习析取范式的开放问题：

开放问题 19 (析取范式的高效量子不可知学习)。 是否存在一种高效的不可知量子学习算法（可访问QAEX 预言机），能够在均匀边缘分布下学习析取范式公式？

我们还想提出以下开放问题，作为对 Feldman [80] 和 Kanade 等人 [82]研究结果的后续探讨。

开放问题 20 （非均匀边缘分布下高效量子不可知学习析取范式）。 当实例上的边缘分布为平滑乘积分布时，是否存在一种高效的不可知量子学习算法（可访问 QAEX 预言机），能够学习析取范式公式？

5.6 Junta 的学习分离

Junta 函数是一种仅依赖于部分输入变量的布尔函数。更正式地说，k-Junta 函数是定义在 n 个变量上的布尔函数，它仅依赖于某个未知的 k 个变量子集，并且对其余 $n-k$ 个变量保持不变。

事实 5.1。 每个 k-*Junta* 函数都可以表示为规模为 2^k 的决策树，或规模为 k 的析取范式公式。每棵规模为 k 的决策树都可以表示为规模为 k 的析取范式公式。

尽管学习 k-Junta 函数的问题可以直接归约为学习宽度为 k 的析取范式函数，但我们在此单独介绍与学习 k-Junta 函数相关的结果。显然，利用 $2^k \log n$ 个随机经典样例并花费 $\mathrm{O}\left(n^k\right)$ 时间（通过求解线性方程），即可学习一个 k-Junta 函数。

Mossel 等人 [84] 给出了首个用于学习 k-Junta 函数的非平凡算法，其运行时间为 $\mathrm{O}\left(n^{\omega k/\omega+1}\,\mathrm{poly}\,(n)\right)$，其中矩阵乘法可以在以下时间内求解

$O(n^{6k})$。[13] 后来，Valiant [85] 将其改进还给出了一个 $O\left(n^{0.8k}\,\mathrm{poly}\,(n,\ 1/\varepsilon)\right)$ 的算法，用于在随机分类噪声设定下，以噪声率 $\varepsilon > 0$学习 k-Junta 函数。

在量子设定下， [10] 的析取范式学习算法意味着，可以使用 QEX查询，在 $O\left(2^{6k}\,\mathrm{poly}\,(n,\ 1/\varepsilon)\right)$ 时间和样本复杂度下学习 k-Junta 函数。Atıcı 和 Servedio [86] 通过给出一个使用 QEX查询的 $O\left(\frac{k}{\varepsilon}\log\ k\right)$ 时间和 $2^k\log\ 1/\varepsilon$ 样本复杂度的学习算法，改进了这一结果。这两个结果在随机分类噪声设定下、噪声率为 η 时同样成立，但运行时间还需额外乘以 poly $(1/\eta)$因子。Chatterjee等人 [23] 设计了第一个能够在不可知设定下学习 k-Junta 函数的量子算法。然而，他们的算法的运行时间取决于 n，并且对 k 和 ε的依赖性要差得多。这引出了以下问题：

开放问题21 （不可知学习Junta 函数的分离）。 在不使用成员查询的不可知设定下，（量子）学习 k-*Junta* 函数的最佳时间复杂度和样本复杂度是多少？

5.7 浅层电路的学习分离

在本节中，我们将浅层经典电路作为概念类进行学习。我们首先考虑最简单的经典电路——NC_d^0 类。NC^0 电路是在 n 个比特上运行的、具有常数深度和多项式规模的布尔电路，由非门以及扇入有界的与门和或门组成。当电路深度为 $d = O(1)$ 时，我们使用记号 NC_d^0。由于 NC_d^0电路具有有界扇入（例如为 $b = O(1)$）和常数深度（例如为 $d = O(1)$），因此输出比特[14] 至多受 $b^d = O(1)$ 个输入变量的影响。因此，NC_d^0电路就是 $O\left(b^d\right)$ Junta 函数！关于学习由 NC^0 电路生成的概念类的更详细介绍，请参阅前文关于学习 Junta 函数的讨论，即第5.6节。

现在，我们将关于量子与经典学习算法之间分离的讨论，转向将AC^0 和TC^0 电路类作为概念来学习。这些电路类在复杂性理论、密码学和学习理论中无处不在。下面我们将正式定义这些电路类。

定义5.5 （AC_d^0 电路和TC_d^0 电路）。AC^0 电路是一个作用于 n比特、深度为常数且规模为多项式的布尔电路，由非门以及无界扇入的与门和或门组成。TC^0 电路是一个作用于 n比特、深度为常数且规模为多项式的布尔电路，由非门以及无界扇入的与门、或门和多数门组成。当电路深度为 $d = O(1)$时，我们分别使用AC_d^0 和TC_d^0 的记法。

事实 5.2。对于任意 $d > 0$，$\mathrm{AC}_d^0 \subset \mathrm{TC}_d^0$。

在布尔函数和学习理论的背景下，AC^0 与 TC^0 电路类相当重要。我们知道，析取范式 $\in\mathrm{AC}_2^0$ 以及半空间的交集/多数函数都可以由 TC^0 电路计算。因此，针对 AC^0 和 TC^0 电路的高效学习算法，将意味着针对析取范式和交集的高效学习算法

[13]这里， $\omega < 2.376$ 是矩阵乘法指数。因此， [84] 需要 $O\left(n^{2k/3}\right)$ 时间。[14]请回忆一下，在本综述中，我们只考虑输出为 1 比特的布尔函数，即作为概念类的判定问题。

的半空间。然而，Arunachalam 等人 [87] 证明了以下关于量子学习 AC_0 和 TC_0 电路的不可能性结果。

引理 5.2（量子学习浅层经典电路的不可能性 [87]）。 以下不可能性结果适用于学习 TC_0 电路。
1. 如果不存在能够利用随机样本求解 LWE问题的有界错误量子多项式时间 算法，那么即使可以访问成员查询，也不存在能够高效学习TC^0_2 的量子学习器。*2.* 如果不存在能够利用随机样本求解环LWE问题[15]的有界错误量子多项式时间 算法，那么即使可以访问成员查询，也不存在能够高效学习TC^0 的量子学习器。

此外，如果不存在能够利用随机样本求解环LWE问题的强次指数时间量子算法，那么即使可以访问成员查询，也不存在用于学习AC^0 的拟多项式时间量子学习器。

我们注意到，引理 5.2意味着不存在用于学习半空间交集的高效量子学习算法（这类布尔函数属于TC^0_2），即使可以访问成员查询，前提是假设LWE的困难性。

与学习TC^0 电路的强困难性不同，在AC^0 电路的情形下，希望并未完全破灭。我们此前已经看到，仅使用 QEX 和 QREX 查询 [10], ，量子学习AC^0_2 类要容易得多；而经典学习器只有在能够访问成员查询时，才有已知的高效算法。对于在分布无关设定下学习具有有界顶扇入的AC^0_3 电路（DNF的与），目前已知最好的经典学习算法源自 Ding等人 [89] 。这引出了以下问题：

开放问题 22（学习AC^0_d 电路的分离）。 对于具有有界顶扇入的AC^0_3 电路，在均匀边缘分布下，是否存在一种可高效运行的量子学习器（可访问查询）？或者说，对于具有有界顶扇入的AC^0_d电路，在均匀边缘分布下，不存在可高效运行的量子学习器（可访问查询）的最小 $d > 1$ 是什么？

注记 5.6. 尽管本节的自然延伸是聚焦于学习浅层电路的量子对应物，但我们指出，此类讨论不在本综述的范围之内。

6 结论

仔细审视上述综述的大量结果后，我们开始发现一个规律。对于许多本质上属于经典的问题而言，与相应的经典对应物相比，量子预言机/量子样本似乎能更自然地捕捉学习任务的本质。我们通过以下开放问题来阐释上述观点。

[15]参见 [88] ，以了解 RLWE 问题的形式定义及其在密码学中的应用。

开放问题 23（经典设定中的采样预言机与随机样本）。是否存在这样一种学习任务：经典学习算法能够证明，访问分布的简洁表示（例如，以访问采样预言机的形式）相较于使用随机样本存在分离？

事实上，乍一看，在经典 PAC 学习的语境下，要在形式上区分这两种模型似乎极为困难。另一方面，对于量子学习算法而言，这是一个完全自然的问题，因为在测量／采样步骤之前，量子算法会自然地操作简洁表示。无论答案是肯定还是否定，对于经典学习问题回答这个问题，都将成为未来研究学习预言机及其分离的相对能力的一项有趣进展。若对上述问题的回答是否定的（这更有可能），则会进一步确立量子算法在学习量子编码的经典函数方面的理论优势。

开放问题列表

1 局部 MQ 与 CQWQ 查询学习之间的分离 . . . . . 15 2 使用 CQWQ 查询模拟 QMQ.15 3 非均匀边缘分布下的量子样本复杂度分离 16 4 最优叠加混合样本复杂度下界 . 17 5 多类别非特定样本复杂度分离.17 6 无界标签样本复杂度分离.17 7 多类别叠加混合样本复杂度 . . . . . 17 8 不同框架中的样本复杂度分离 . . . . . 17 9 白盒不可知样本复杂度分离 . . . . . 18 10 灰盒样本复杂度分离 .18 11 白盒多分类样本复杂度分离 . . . . . 18 12 阿贝尔 HSP 诱导的概念类的分离 . . . . . 20 13 非阿贝尔 HSP 诱导的概念类的分离 . . . . . 21 14 学习二面体 HSP 诱导的概念类的困难性 . . . . . 22 15 基于随机样本的 LWE 量子学习.23 16 基于随机样本的 LPN 量子学习.24 17 半空间的高效量子不可知学习.25 18 非均匀边缘分布下决策树的高效量子不可知学习.26 19 DNF 的高效量子不可知学习.27 20 非均匀边缘分布下 DNF 的高效量子不可知学习.27 21 析取项的非特定学习的分离.28 22 学习 AC^0_d 电路的分离.29 23 经典设定下采样预言机与随机样本的比较 30

致谢

作者感谢 Debajyoti Bera、Ambuj Tewari、Piyush Srivastava 和 MinHsiu-Hsieh 为本文早期草稿提供了详尽的意见。

参考资料

[1]Feynman, R.P.：用计算机模拟物理。《国际理论物理学杂志》21(6), 467–488 (1982) https://doi.org/10.1007/BF02650179[2]Feynman, R.P.：量子力学计算机。《光学新闻》11(2), 11–20 (1985) https://doi.org/10.1364/ON.11.2.000011[3] 多伊奇, D.：量子理论、丘奇–图灵原理与通用量子计算机。《伦敦皇家学会会刊 A 辑》400, 97–117 (1985) https://doi.org/10.1098/rspa.1985.0070[4] Grover, L.K.：用于数据库搜索的快速量子力学算法。载于：第二十八届计算理论年度 ACM 研讨会论文集，第 212–219 页 (1996)[5] Shor, P.W.：量子计算算法：离散对数与因数分解。载于：第 35 届计算机科学基础年会论文集，第 124–134 页 (1994)。https://doi.org/10.1109/SFCS.1994.365700[6] Vapnik, V.N., Chervonenkis, A.Y.：关于事件相对频率一致收敛至其概率。《概率论及其应用》16(2), 264–280 (1971) https://doi.org/10.1137/1116025[7] Vapnik, V., Chervonenkis, A.：《模式识别理论》 [（俄文）]。科学出版社，莫斯科 (1974)。（德文译本：W. 瓦普尼克与 A. 切尔沃年基斯，《字符识别理论》，学术出版社，柏林，1979）[8] Vapnik, V.：《基于经验数据的依赖关系估计》。施普林格，??? (1982)[9]Valiant, L.G.：可学习性理论。Commun. ACM 27(11), 1134–1142 (1984) https://doi.org/10.1145/1968.1972[10] Bshouty, N.H., Jackson, J.C.：利用量子样本预言机在均匀分布下学习析取范式。载于：第八届计算学习理论年会论文集。COLT '95，第 118–127 页。计算机协会，美国纽约州纽约市 (1995)。https://doi.org/10.1145/225298. 225312[11] Arunachalam, S., Wolf, R.：特邀专栏：量子学习理论综述。SIGACT新闻 48(2), 41–67 (2017) https://doi.org/10.1145/3106700.3106710

[12]Anshu, A., Arunachalam, S.：量子态学习复杂性综述。《自然综述：物理学》6(1), 59–69 (2024) https://doi.org/10.1038/ s42254-023-00662-4[13] Dunjko, V., Briegel, H.J.：量子领域中的机器学习与人工智能：近期进展综述。《物理学进展报告》81(7), 074001 (2018) https://doi.org/10.1088/1361-6633/aab406[14] Nietner, A.：统一（量子）统计算法与参数化（量子）算法 (2023)。https://arxiv.org/abs/2310.17716[15] Arunachalam, S., Chakraborty, S., Lee, T., Paraashar, M., Wolf, R.：关于量子精确学习的两个新结果。Quantum 5, 587 (2021) https://doi.org/ 10.22331/q-2021-11-24-587[16] Shalev-Shwartz, S., Ben-David, S.：《理解机器学习：从理论到算法》。剑桥大学出版社，??? (2014)。https://doi.org/10. 1017/CBO9781107298019[17] Kalai, A.T., Mansour, Y., Verbin, E.：关于不可知提升与奇偶函数学习。载于：第四十届 ACM 理论计算研讨会论文集。STOC ’08，第 629–638 页。计算机协会，美国纽约州纽约市 (2008)。https://doi.org/10.1145/1374376.1374466[18] Arunachalam, S., Dutt, A., Gheorghiu, A., Oliveira, M.：a (2025)。https://arxiv. org/abs/2509.14461[19] Bshouty, N.H.：通过单调理论精确学习布尔函数。Information and Computation 123(1), 146–153 (1995) https://doi.org/10.1006/inco. 1995.1164[20] Arunachalam, S., Wolf, R.：学习算法的最优量子样本复杂度。载于：第 32 届计算复杂性会议论文集。CCC ’17。达格斯图尔莱布尼茨信息学中心，德国 Dagstuhl (2017)[21] Apolloni, B., Gentile, C.：基于算法复杂度理论的 PAC 学习样本量下界。Theoretical ComputerScience 209(1-2),141–162 (1998) https: // doi.org/10.1016/S0304-3975(97)00102-3[22] Iwama, K., Raymond, R., Yamashita, S.：量子有偏谕示的一般界。IPSJ数字速递 1, 415–425 (2005)[23] Chatterjee, S., SAPV, T., Bera, D.：决策树的高效量子不可知不当学习。载于：Dasgupta, S., Mandt, S., Li, Y. (编)《第 27 届人工智能与统计国际会议论文集》。机器学习研究论文集，第 238 卷，第 514–522 页。PMLR, ??? (2024)。https://proceedings.mlr.press/v238/chatterjee24a.html

[24]Awasthi, P., Feldman, V., Kanade, V.：利用局部成员查询进行学习。载于：Shalev-Shwartz, S., Steinwart, I.（编）《第26届年度学习理论会议论文集》。机器学习研究论文集，第30卷，第398–431页。PMLR，美国新泽西州普林斯顿（2013年）。https://proceedings.mlr.press/v30/Awasthi13.html

[25]Bary-Weisberg, G., Daniely, A., Shalev-Shwartz, S.：利用局部查询进行分布无关学习。载于：Kontorovich, A., Neu, G.（编）《第31届国际算法学习理论会议论文集》。机器学习研究论文集，第117卷，第133–147页。PMLR，???（2020年）。https://proceedings.mlr.press/v117/bary-weisberg20a.html

[26] Bshouty, N., Mossel, E., O'Donnell, R., Servedio, R.A.：从随机游走中学习析取范式。在：计算机科学基础研讨会第44届IEEE年会，2003年。会议论文集，第189–198页（2003年）。https://doi.org/10.1109/SFCS.2003.1238193[27] Childs, A.M., Cleve, R., Deotto, E., Farhi, E., Gutmann, S., Spielman, D.A.：量子游走带来的指数级算法加速。在：第35届ACM计算理论年度研讨会论文集。STOC '03，第59–68页。计算机协会，美国纽约州纽约市（2003年）。https://doi.org/10.1145/780542.780552 。https://doi.org/10.1145/780542.780552[28] Blumer, A., Ehrenfeucht, A., Haussler, D., Warmuth, M.K.：可学习性与 Vapnik–Chervonenkis 维数。J. ACM **36**(4), 929–965 (1989)https://doi.org/10.1145/76359.76371[29] Hanneke, S.：PAC学习的最优样本复杂度。《机器学习研究杂志》**17**(38),1–15(2016)[30] Talagrand, M.：高斯过程与经验过程的更紧下界。《概率年刊》**22(1), 28–76 (1994).** 访问日期：2024年7月15日[31] Servedio, R.A., Gortler, S.J.：量子与经典可学习性。在：第16届IEEE计算复杂性年度会议论文集，第138–148页（2001年）。https://doi.org/10.1109/CCC.2001.933881[32] Hadiashar, S.B.,Nayak, A., Sinha, P.：基于信息论的量子学习最优下界。《IEEE信息论汇刊》**70**(3),1876–1896(2024) https://doi.org/10.1109/TIT.2023.3324527[33] Caro, M.C.：关于乘积分布的布尔线性函数量子学习。《量子信息处理》**19**(6), 172 (2020)https://doi.org/10.1007/s11128-020-02661-1[34]Caro, M.C., Hinsche, M., Ioannou, M., Nietner, A., Sweke, R.：量子学习的经典验证。在：Guruswami, V.（编）第15届

理论计算机科学创新会议（ITCS 2024）。莱布尼茨国际信息学论文集（LIPIcs），第287卷，第24–12423页。达格斯图尔城堡 – 莱布尼茨信息学中心¨，Dagstuhl，德国（2024年）。https://doi.org/10.4230/LIPIcs.ITCS.2024.24 . https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2024.24[35] Mohan, P., Tewari, A.：超越批量二元分类的量子学习理论。《量子》**9**，1813（2025年）https://doi.org/10.22331/q-2025-07-29-1813[36] Bansal, A., George, I., Ghosh, S., Sikora, J., Zheng, A.：多种量子对象的在线学习（2024年）。https://arxiv.org/abs/2406.04245[37] Salmon, W., Strelchuk, S., Gur, T.：量子 PAC 学习中的可证明优势。在：Agrawal, S., Roth, A.（编）《第三十七届学习理论会议论文集》。机器学习研究论文集，第247卷，第4487–4510页。PMLR，???（2024年）。https://proceedings.mlr.press/v247/salmon24a.html[38] Tang, E., Wright, J.：振幅放大与估计需要逆（2025年）。https://arxiv.org/abs/2507.23787[39] Kearns, M., Valiant, L.：密码学对布尔公式和有限自动机学习的限制。J. ACM **41**(1), 67–95 (1994) https://doi.org/10.1145/174644.174647[40] Angluin, D., Kharitonov, M.：成员查询何时不起作用？《计算机与系统科学杂志》**50**(2), 336–355 (1995)https://doi.org/10.1006/jcss.1995.1026[41] Liu, Y., Arunachalam, S., Temme, K.：监督式机器学习中严谨且稳健的量子加速。《自然物理》**17**(9), 1013–1017 (2021) https://doi.org/10.1038/s41567-021-01287-z[42]Blum, M., Micali, S.：如何生成密码学安全的伪随机比特序列。SIAM计算杂志**13**(4), 850–864 (1984)https://doi.org/10.1137/0213053[43] Ettinger, M., Høyer, P.：关于非交换隐子群的量子算法。《应用数学进展》**25**(3), 239–251 (2000) https://doi.org/10.1006/aama.2000.0699[44] Cleve, R., Watrous, J.：用于量子傅里叶变换的快速并行电路。在：第41届计算机科学基础年度研讨会论文集，第526–536页（2000年）。https://doi.org/10.1109/SFCS.2000.892140[45] Cheung, K.K.H., Mosca, M.：有限阿贝尔群的分解。《量子信息与计算》**1**(3), 26–32 (2001)

[46] Ettinger, M., Høyer, P., Knill, E.：隐藏子群问题的量子查询复杂度为多项式。《信息处理快报》91(1)，43–48（2004年）
https://doi.org/10.1016/j.ipl.2004.01.024[47]Moore, C., Rockmore, D., Russell, A.：通用量子傅里叶变换。 ACM算法汇刊 2(4)，707–723（2006年）
https://doi.org/10.1145/1198513. 1198525[48] Lomont, C.：隐藏子群问题——综述与开放问题（2004年）。 https://arxiv.org/abs/quant-ph/0411037[49] Wang, F.：隐藏子群问题（2010年）。 https://arxiv.org/abs/1008. 0010[50] Regev, O.：关于格、带误差学习、随机线性码与密码学。 J. ACM 56(6)（2009年）
https://doi.org/10.1145/1568318.1568324[51] Bennett, H.：最短向量问题的复杂度。SIGACT 新闻 54(1)，37–61 （ 2023 年）
https://doi.org/10.1145/3586165.3586172[52] Kuperberg, G.：求解二面体隐藏子群问题的亚指数时间量子算法。 SIAM计算杂志 35(1)，170–188（2005年） https: // doi.org/10.1137/S0097539703436345[53] Dyer, M.E., Frieze, A.M.：若干随机NP难问题的多项式期望时间解法。算法期刊 10(4)，451–489（1989） https: // doi.org/10.1016/0196-6774(89)90001-1[54] Ajtai, M.：生成格问题的困难实例（扩展摘要）。载于：第二十八届ACM计算理论年度研讨会论文集。STOC ’96，第99–108页。计算机协会，纽约州纽约市，美国（1996年）。
https://doi.org/10.1145/237814.237838 . https://doi.org/10.1145/237814.237838
[55] Lyubashevsky, V.：噪声环境下的奇偶性问题、随机线性码译码与子集和问题。载于：组合优化近似算法国际研讨会，第378–389页（2005年）。施普林格[56] Blum, A., Kalai, A., Wasserman, H.：抗噪声学习、奇偶性问题与统计查询模型。 J. ACM 50(4)，506–519（2003年） https://doi.org/ 10.1145/792538.792543[57] Feldman, V., Gopalan, P., Khot, S., Ponnuswami, A.K.：含噪奇偶性与半空间学习的新结果。载于：2006年第47届IEEE计算机科学基础年会（FOCS’06），第563–574页（2006年）。 https://doi.org/10. 1109/FOCS.2006.51

[58] Grilo, A.B., Kerenidis, I., Zijlstra, T.：带误差学习问题借助量子样本变得容易。《物理评论A》**99**，032314（2019年）https://doi.org/10.1103/PhysRevA.99.032314[59] Cross, A.W., Smith, G., Smolin, J.A.：抗噪声的量子学习。《物理评论A》**92**，012327（2015年）https://doi.org/10.1103/PhysRevA.92.012327[60]Pietrzak, K.：基于带噪声奇偶学习的密码学。载于：Bielikov´a,M., Friedrich, G., Gottlob, G., Katzenbeisser,S., Tur´an, G.（编）：SOFSEM 2012：计算机科学理论与实践，第99–114页。施普林格，柏林、海德堡（2012年）[61]Brakerski, Z., Lyubashevsky, V., Vaikuntanathan, V., Wichs, D.：通过码平滑实现LPN的最坏情况困难性与密码学哈希。载于：密码学进展——EUROCRYPT 2019：第38届密码学技术理论与应用国际年会，德国达姆施塔特，2019年5月19日至23日，论文集，第三部分，第619–635页。施普林格，柏林、海德堡（2019年）。https://doi.org/10.1007/978-3-030-17659-4 21[62]Yu, Y., Zhang, J.：平滑二元线性码及LPN的最坏情况亚指数困难性。载于：密码学进展——CRYPTO 2021：第41届国际密码学年会，CRYPTO 2021，线上活动，2021年8月16日至20日，论文集，第三部分，第473–501页。施普林格，柏林、海德堡（2021年）。https://doi.org/10.1007/978-3-030-84252-9 16[63] Blum, A., Frieze,A.：学习含噪线性阈值函数的多项式时间算法。载于：第37届计算机科学基础年会论文集。FOCS '96，第330页。IEEE计算机学会，美国（1996年）[64] Diakonikolas, I., Kontonis, V., Tzamos, C., Zarifis, N.：在结构化分布下利用Massart噪声学习半空间。载于：Abernethy, J., Agarwal, S.（编），第三十三届学习理论会议论文集。《机器学习研究论文集》第125卷，第1486–1513页。PMLR，???（2020年）。https://proceedings.mlr.press/v125/diakonikolas20c.html[65] Diakonikolas, I., Kane, D.M., Kontonis, V., Tzamos, C., Zarifis, N.：利用Tsybakov噪声高效学习半空间。载于：第53届ACM SIGACT计算理论研讨会论文集。STOC 2021，第88–101页。计算机协会，美国纽约州纽约市（2021年）。https://doi.org/10.1145/3406325.3450998[66] Daniely, A.：不可知学习半空间的多项式时间近似方案。载于：Gr¨unwald, P., Hazan, E., Kale, S.（编）：第二十八届学习理论会议论文集。《机器学习研究论文集》第40卷，第484–502页。PMLR，法国巴黎（2015年）。https://proceedings.mlr.press/v40/Daniely15.html

[67] Tiegel, S.：从最坏情形格中不可知学习半空间的困难性

问题。载于：Neu, G.、Rosasco, L.（编）：第三十六届学习理论会议论文集。《机器学习研究论文集》，第195卷，第3029–3064页。PMLR，???（2023年）。https://proceedings.mlr.press/v195/tiegel23a.html[68] Guruswami, V.、Raghavendra, P.：带噪半空间学习的困难性。《SIAM计算杂志》 **39**（2），742–765（2009年）[69] Diakonikolas, I.、Kane, D.M.、Pittas, T.、Zarifis, N.：高斯边缘分布下统计查询模型中不可知学习的多项式回归最优性。载于：Belkin, M.、Kpotufe, S.（编）：第三十四届学 习理论会议论文集。《机器学习研究论文集》，第134卷，第1552–1584页。PMLR，???（2021年）。https://proceedings.mlr.press/v134/diakonikolas21c.html[70]Hyafil, L.、Rivest, R.L.：构造最优二叉决策树是NP完全问题。《信息处理快报》 **5**（1），15–17（1976年）https://doi.org/10.1016/0020-0190(76)90095-8[71] Ehrenfeucht, A.、Haussler, D.：从随机样本中学习决策树。《信息与计算》**82**（3），231–246（1989）[72] Koch, C.、Strassle, C.、Tan, L.-Y.：决策树学习与测试的超多项式下界。载于：2023年ACM-SIAM离散算法年会（SODA）论文集，第1962–1994页（2023年）。SIAM[73] Kushilevitz, E.、Mansour, Y.：使用傅里叶谱学习决策树。载于：第二十三届ACM计算理论年会论文集。STOC '91，第455–464页。计算机协会，美国纽约州纽约市（1991年）。 https://doi.org/10.1145/103418.103466 。https://doi.org/10.1145/103418.103466[74] Gopalan, P.、Kalai, A.T.、Klivans, A.R.：不可知学习决策树。载于：第四十届ACM计算理论年会论文集。STOC '08，第527–536页。计算机协会，美国纽约州纽约市（2008年）。https://doi.org/10.1145/1374376.1374451[75] Feldman, V.：特定分布的无关提升。载于：Yao, A.C.（编）：计算机科学创新——ICS 2010，清华大学，北京，中国，2010年1月5–7日。论文集，第241–250页。清华大学出版社，???（2010年）[76]Kanade, V.、Kalai, A.：基于势函数的无关提升。载于：Bengio, Y.、Schuurmans, D.、Lafferty, J.、Williams, C.、Culotta, A.（编）：神经信息处理系统进展，第22卷。Curran Associates, Inc.，???（2009年）。 https://proceedings.neurips.cc/paper-files/paper/2009/file/13f9896df61279c928f19721878fac41-Paper.pdf[77]Daniely, A.、Shalev-Shwartz, S.：关于学习的计算复杂性理论限制

析取范式。在：Feldman, V., Rakhlin, A., Shamir, O.（编）第29届学习理论年会。机器学习研究论文集，第49卷，第815–830页。PMLR，???（2016）。https://proceedings.mlr.press/v49/daniely16.html[78] H˚astad, J.：一些最优不可近似性结果。ACM 会刊（JACM）48(4)，798–859（2001）[79] Jackson, J.C.：一种高效的成员查询算法，用于在均匀分布下学习析取范式。计算机与系统科学杂志 55(3)，414–440（1997）https://doi.org/10.1006/jcss.1997.1533[80]Feldman, V.：从傅里叶谱学习析取范式表达式。在：Mannor, S., Srebro, N., Williamson, R.C.（编）COLT 2012——第25届学习理论年会，2012年6月25–27日，爱丁堡，苏格兰。JMLR论文集，第23卷，第17–11719页。JMLR.org，???（2012）。http://proceedings.mlr.press/v23/feldman12b/feldman12b.pdf[81] Spielman, D.A., Teng, S.-H.：算法的平滑分析：为什么单纯形算法通常能在多项式时间内运行。J. ACM 51(3)，385–463（2004）https: //doi.org/10.1145/990308.990310[82] Kanade, V., Rocchetto, A., Severini, S.：µ-偏置量子傅里叶采样。《量子信息处理》19(15&16)，1261–1278（2019）https://doi.org/10.26421/ QIC19.15-16-1[83] Jackson, J.C.：一种高效的成员查询算法，用于在均匀分布下学习析取范式。计算机与系统科学杂志 55(3)，414–440（1997）https://doi.org/10.1006/jcss.1997.1533[84] Mossel, E., O'Donnell, R., Servedio, R.A.：学习含有 k 个相关变量的函数。计算机与系统科学杂志 69(3)，421–434（2004）https: //doi.org/10.1016/j.jcss.2004.04.002。STOC 2003 特刊[85] Valiant, G.：以次二次时间寻找相关性，并应用于学习奇偶函数和最近点对问题。J. ACM 62(2)（2015）https://doi.org/ 10.1145/2728167[86] Atıcı, A., Servedio, R.A.：用于学习和测试Junta函数的量子算法。《量子信息处理》6(5)，323–348（2007）https://doi.org/10.1007/ s11128-007-0061-6[87] Arunachalam, S., Grilo, A.B., Sundaram, A.：学习浅层经典电路的量子困难性。SIAM计算杂志 50(3)，972–1013（2021）https://doi.org/10.1137/20M1344202[88] Lyubashevsky, V., Peikert, C., Regev, O.：关于理想格和环上的带误差学习。J. ACM 60(6)（2013）https://doi.org/10.1145/2535925

[89]Ding, N., Ren, Y., Gu, D.：PAC学习深度为3的 AC^0 有界顶扇入电路。在：Hanneke, S., Reyzin, L.（编）第28届国际算法学习理论会议论文集。机器学习研究论文集，第76卷，第667–680页。PMLR，???（2017）。https://proceedings.mlr.press/v76/ding17a.html

[90] Beals, R., Buhrman,H., Cleve, R., Mosca, M., 沃尔夫, R.：通过多项式得到量子下界。J. ACM 48(4), 778–797 (2001年) https://doi.org/10.1145/502090.502097

[91]Brassard, G., Hø yer, P., Mosca, M., Tapp, A.：量子振幅放大与估计。载于：量子计算与信息（华盛顿特区，2000年）。《当代数学》，第305卷，第53–74页。美国数学学会，罗德岛州普罗维登斯，???（2002年）。https://doi.org/10.1090/conm/305/05215 。https://doi.org/10.1090/conm/305/05215[92]Ambainis, A.：具有可变时间的量子搜索。载于：Albers, S., Weil, P.（编）《第25届理论计算机科学专题国际研讨会》。莱布尼茨国际信息学论文集（LIPIcs），第1卷，第49–60页。达格斯图尔城堡 – 莱布尼茨信息学中心¨，德国达格斯图尔（2008年）。https://doi.org/10.4230/LIPIcs.STACS.2008.1333[93] Linial, N., Mansour, Y., Nisan, N.：常数深度电路、傅里叶变换与可学习性。《J. ACM》40(3)，607–620（1993年）https://doi.org/10.1145/174130.174138[94] Bernstein, E., Vazirani,U.：量子复杂性理论。《SIAM计算杂志》**26**(5)，1411–1473（1997年）https://doi.org/10.1137/S0097539796300921[95] Baum, E.B., Lang, K.：使用人类谕示时，查询学习的效果可能很差。载于：国际神经网络联合会议，第8卷，第8页（1992年）。中国北京[96] Angluin, D., Laird, P.：从噪声样本中学习。《机器学习》**2**(4)，343–370（1988年）https://doi.org/10.1023/A:1022873112823[97] Haussler, D.：面向神经网络及其他学习应用的PAC模型的决策论推广。《信息与计算》**100**(1)，78–150（1992年）https://doi.org/10.1016/0890-5401(92)90010-D[98] Kearns, M.J., Schapire, R.E., Sellie, L.M.：迈向高效的不可知学习。载于：第五届计算学习理论年度研讨会论文集。COLT '92，第341–352页。计算机协会，美国纽约州纽约市（1992年）。https://doi.org/10.1145/130385.130424[99] Goldreich, O., Levin, L.A.：适用于所有单向函数的硬核谓词。载于：

第二十一届ACM计算理论年度研讨会论文集。STOC '89，第25–32页。计算机协会，美国纽约州纽约市（1989）。https://doi.org/10.1145/73007.73010

附录 A 量子计算

经典布尔电路是带有与门、或门和非门的有限有向无环图，它们接受 n个输入位，并输出 m个输出位，以计算某个函数$f \in \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$。经典电路以 n个比特作为输入，并输出 m个比特。量子电路是经典电路的推广，其中输入是一个 n量子比特态，输出是一个 m量子比特态，并且与门、或门和非门被替换为量子门。在本综述中，我们将以下增广的（非最小）通用门集作为典范量子基门： $\{X, Y, Z, H, CNOT, S, T\}$。有时，在量子电路中，我们假设可以获得对酉算子（也称为**预言机**）的黑盒访问，该酉算子编码了某些仅能通过查询获取的信息。参见图 A1 中的预言机示例。

Phase Oracles. $O_x^{\pm} : |i, a\rangle \mapsto (-1)^{a.x_i} |i, a\rangle$

Function Oracles. $O_f : |x, a\rangle \mapsto |x, a \oplus f(x)\rangle$

Probability Oracles. $O_p : |x, 0, 0\rangle \mapsto \sqrt{p_x} \left|x, 1, \psi_x^1\right\rangle + \sqrt{1 - p_x} \left|x, 0, \psi_x^0\right\rangle$

图 A1 量子算法中使用的一些常见预言机

在 Beals 等人形式化引入的标准量子查询模型中，量子查询算法有 3 组寄存器：辅助量子比特 $\mathcal{A}$、工作区 $\mathcal{W}$和查询寄存器 $\mathcal{Q}$，它们都初始化为均匀分布 $|0\rangle$。一个 k 次查询的量子查询算法通过交替执行酉操作并调用某个量子查询预言机，进行 k 轮。算法最后测量辅助量子比特寄存器。下面列出两个重要的量子子程序，它们历来为量子学习理论带来了大部分加速。

假设我们可以使用一个酉算子 U 来制备一个量子态

$$|\psi\rangle = \sqrt{\alpha_{\text{good}}} |\psi_{\text{good}}\rangle + \sqrt{1 - \alpha_{\text{good}}} |\alpha_{\text{bad}}\rangle ,$$

并且我们希望以高概率获得“好”的状态 $|\psi_{\text{good}}\rangle$。仅测量 $|\psi\rangle$ 只能以 $|\psi_{\text{good}}\rangle$ 的概率得到 $|\alpha_{\text{good}}|^2$。因此，从期望来看，我们需要进行 $1/|\alpha_{\text{good}}|^2$ 次独立测量，才能以高概率得到 $|\psi_{\text{good}}\rangle$。然而，Brassard 等人 [91] 指出，可以利用 振幅放大来改进这一点，如下所述。

引理 A.1（振幅放大）。设 $p > 0$ 为常数，且 U 为酉算子，使得 $U|0\rangle = \sqrt{p_0} |\phi_0\rangle |0\rangle + \sqrt{1 - p_0} |\phi_1\rangle |1\rangle$ 对某个未知的 $p_0 \geq p > 0$成立。存在

一个量子算法，该算法对 U 与U^{-1} 的期望调用次数为 $\Theta(\sqrt{p'/p})$ ，并输出状态 $|\phi_0\rangle$，概率为 $p' > 0$。

设 $|\phi_0\rangle = |\psi_{\text{good}}\rangle$，且 $p_0 = |\alpha_{\text{good}}|^2$ 在引理 A.1中。然后，在 $p' = 1$中应用引理 A.1，通过进行 $O(1/|\alpha_{\text{good}}|)$ 次查询 U和 U^{-1}，即可在期望意义下得到。这相较于经典情形实现了二次加速。

振幅放大的一个问题是我们可能不知道 p_0 ，因此也不知道需要调用 U 和 U^{-1} 多少次才能得到 $|\psi_{\text{good}}\rangle$。我们使用 Ambainis 提出的以下结果 [92] 来获得 p_0 的相对误差估计。

引理 A.2（相对振幅估计）。给定常数 $p > 0$、误差参数 $\varepsilon > 0$、常数 $k \geq 1$ 以及酉算子 U，使得 $U|0\rangle = \sqrt{p_0}|\phi_0\rangle|0\rangle + \sqrt{1-p_0}|\phi_1\rangle|1\rangle$，其中 $p_0 \geq p$ 或 $p_0 = 0$。那么，存在一种量子算法，以至少 $1 - 1/2^k$ 的概率生成成功概率 p_0 的估计值 $\tilde{p}_0$ ，使得在 $p_0 \geq p$ 时 $|p_0 - \tilde{p}_0| \leq \varepsilon \cdot p_0$，并且在期望意义下调用 U 和 U^{-1} 共 $O\left(\frac{k}{\varepsilon\sqrt{p}}\left(1 + \log\log\frac{1}{p}\right)\right)$ 次。

经典地，我们可以通过采样 $\tilde{O}(\frac{1}{\varepsilon^2})$ 个 $U|0\rangle$ 的独立副本，并令 $p = O(1/m)$，从而使用乘法型 Chernoff 界以相对误差进行均值估计，其中 $|\phi_0\rangle$ 是 m 个基态上的叠加态。因此，利用 相对振幅估计，我们可以以相对误差 ε 进行均值估计；与经典均值估计技术相比，就对 ε和 m的依赖而言，可获得二次加速。

附录B 布尔函数（续）

布尔函数有许多语义等价的表示方式。在本综述中，我们关注布尔函数的三种主要表示方式，如下所列。

- **真值表**：在这种表示方式中，我们明确列出布尔函数对其全部 2^n 个可能输入的取值。

- **布尔电路**：布尔电路是一个有限有向无环图（DAG），其中明确指定了输入节点、内部结点和输出节点。在本综述中，每个节点的扇入至多为 2。输入节点是输入字符串中的比特，其扇入为 0；内部结点由属于某个布尔函数基集合的基本布尔函数组成（例如与运算、或运算和非门），输出节点（扇出为 0）存储一个二进制值，该值对应于对输入位进行求值后的输出。

- **布尔公式 / 命题公式**：我们可以将任意布尔函数 f 表示为一个代数公式，该公式由属于某个布尔函数基集合的基本布尔函数作用于其输入位构成。形式上，任一每个节点的扇出至多为1的布尔电路都是布尔公式。一些著名的布尔函数规范表示是析取范式（DNF）和合取范式（CNF）表示，其定义相对于

输入文字上的或运算与与运算。与真值表或电路不同，布尔公式对应于对 $f^{-1}(0)$或 $f^{-1}(1)$的覆盖。

注 B.1. 给定一个用于布尔函数的通用基集合 $\mathcal{B}$，我们可以定义布尔电路，使其内部结点只包含来自 $\mathcal{B}$的门。例如，考虑基集合 $\mathcal{B}_1 = \{$与非$\}$ 或 $\mathcal{B}_1 = \{$或非$\}$。与布尔电路类似，我们也可以用布尔函数上的不同通用基集合来表示公式。不过，在本综述中，经典布尔电路和公式的典范基集合是 {与、或、非}。

布尔函数的表示规模可以根据所采用的表示方式来定义。

例 B.1. 布尔函数 f 的一种标准表示方式，是其最小二叉决策树表示的深度 f。另一种标准方式，是其最小宽度析取范式表示的规模 f。

B.1 布尔函数的傅里叶分析

傅里叶分析技术构成了一套极为强大的工具，可用于设计一般的学习算法，尤其适用于在布尔超立方体上进行学习 [93]。本节将讨论范围限定于 $\mathbb{F}_2$值（± 1值）或 $\mathbb{Z}_2$值（$\{0,1\}$值）函数的傅里叶分析，但大多数结果也适用于布尔超立方体上的实值函数。每个布尔函数f（x）都可以唯一地表示为一个多重线性多项式，称为 f的**傅里叶表示**，如下所示。

$$f(x) = \sum_{\mathcal{S}\subseteq[n]} \hat{f}(\mathcal{S})\,\chi_{\mathcal{S}}(x), \quad f : \mathbb{F}_2^n \mapsto \mathbb{F}_2. \tag{B1}$$

在式 (B1)中，$\chi_{\mathcal{S}}(x) = \prod_{i\in\mathcal{S}} x_i$（我们明确地定义了 $\chi_{\mathcal{S}}(\varnothing) = -1$）称为一个^**奇偶单项式**，而 $f(\mathcal{S}) = \mathbb{E}\,[f\cdot\chi_{\mathcal{S}}]$称为$f$ 关于文字子集 $\mathcal{S}$的**傅里叶系数**。注意，如果 $f: \{0,1\}^n \mapsto \mathbb{F}_2$，那么$\chi_{\mathcal{S}}(x) = (-1)^{\sum_{i\in\mathcal{S}} x_i}$。所有奇偶单项式 $\mathbb{P} = \{\chi_{\mathcal{S}} \mid \mathcal{S} \subseteq [n]\}$ 构成了布尔超立方体 $\mathbb{F}_2^n$上所有实值函数的一组基。由 $\mathbb{P}$ 张成的向量空间构成一个内积空间，即，对于任意 f，$g: \mathbb{F}_2^n \mapsto \mathbb{F}_2$，

$$\langle f, g\rangle = \frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n} f(x)\cdot g(x) = \mathop{\mathbb{E}}_{\mathcal{U}(\mathbb{F}_2^n)}[f(x)\cdot g(x)]\,1 - 2\mathop{\Pr}_{\mathcal{U}(\mathbb{F}_2^n)}[f\neq g]. \tag{B2}$$

上述内积定义不仅体现了两个布尔函数之间距离的概念，也表明所有奇偶单项式构成一个标准正交基集合，因为对于所有 $\mathcal{S}\subseteq[n]$，$\langle\chi_{\mathcal{S}},\chi_{\mathcal{S}}\rangle = 1$；而如果 $\mathcal{S}\neq\mathcal{T}$，则 $\langle\chi_{\mathcal{S}},\chi_{\mathcal{T}}\rangle = 0$。由此我们可以得到以下观察。

观察 B.1（布尔函数的傅里叶谱）。 对于任意布尔函数 f，其傅里叶系数的平方构成一个概率分布，称为 f 的傅里叶谱 或傅里叶分布，记作$\mathrm{FS}(f)$。

$$\sum_{\mathcal{S}\subseteq[n]} \hat{f}(\mathcal{S})^2 = 1 \; f, g : \mathbb{F}_2^n \mapsto \mathbb{F}_2.$$

现在我们陈述 傅里叶采样 引理，该引理由 Bernstein 和 Vazirani 提出 [94]。

引理 B.1（傅里叶采样）。 给定对函数预言机 O_f 的访问权限（见图 *A1*），存在一种量子算法（见算法*1*），仅使用1 次对 O_f 的查询和$\mathrm{O}(n)$ 个门，即可生成状态$\hat{}\sum_{\mathcal{S}\in[n]} f(\mathcal{S})\,|\mathcal{S}\rangle$。

与量子情形不同，显式构造分布 FS (f) 在经典情形下更困难，因为任意 $f(\mathcal{S}) = \mathbb{E}\,[f \cdot \chi_{\mathcal{S}}]$都依赖于所有 2^n取值 f。傅里叶采样生成的状态 傅里叶采样 使我们能够从 FS (f) 中采样，即以概率 $\mathcal{S}\,\hat{}\,f(\mathcal{S})^2$进行采样。我们可以使用 傅里叶采样，仅通过一次对 f 的查询来学习线性函数 O_f [94]；而在经典情形下，这需要 O(n) 次查询。

算法1： 傅里叶采样算法

输入： 预言机 O_f：$|x\rangle \mapsto f(x)\,|x\rangle$，提供对 f 的黑盒访问。**1** 从 $|0^n\rangle$ 开始并应用 $H^{\otimes n}$，得到 $|+^{\otimes n}\rangle = \frac{1}{\sqrt{2^n}}\sum_{x\in\mathbb{F}_2^n}|x\rangle$ **.2** 查询 O_f，得到状态 $|\psi\rangle = \frac{1}{\sqrt{2^n}}\sum_{x\in\mathbb{F}_2^n} f(x)\,|x\rangle$ **.3** 将$H^{\otimes n}$ 应用于 $|\psi\rangle$。**ˆ输出：** 状态为 $\sum_{\mathcal{S}\in[n]} f(\mathcal{S})\,|\mathcal{S}\rangle$。

附录 C 计算学习理论

C.1 与学习器相关的假设类

计算学习理论的计算方面研究中一个隐含的核心观点是，每个学习算法 A 都有其对应的假设类 $\mathcal{H}_A$。学习器 A 只能输出来自 $\mathcal{H}_A$的假设。[16] 我们在此还假设，所有假设类 $\mathcal{H}_A$ 都可由 A**在多项式时间内求值**，也就是说，给定任意 $x \in \mathcal{X}$和 $h \in \mathcal{H}_A$，如果 A能在 poly $(n,\mathcal{H}_A$ 的规模$)$ 时间内计算 $h(x)$。

观察 C.1. 我们注意到，由于各自关联的假设类不同，某些学习算法比其他算法的表达能力更强。然而，假设类表达能力最强的学习算法不一定总是最佳选择，因为假设类较大的学习器需要更多计算资源，才能在其关联的假设类中找到最优假设。

[16]当学习器的选择明确时，我们省略假设类的下标。

观察 C.2. 请注意，给定一个在$\mathcal{X}$ 上的任意分布 $\mathfrak{D}$ 以及某个未知的标记概念 $c \in \mathcal{C}$，假设学习算法 A能够输出理想假设并不总是现实的。例如，可以考虑 $\mathcal{H}_A$ 极大，或者概念 c 不属于 $\mathcal{H}_A$的情况。

在无法通过计算或数学方法获得理想假设的设定中，学习算法的主要目标是高效地输出一个假设 h，使得 $h \approx c$，其中衡量标准可以是误差等准确率度量。

C.2 根据输出假设对 PAC 学习器进行分类

C.2.1 适当与不适当 PAC 学习

回顾备注 2.5以及定义可学习性和高效可学习性时的前述讨论。在定义 2.3中，学习算法 A 生成的输出假设 $h \in \mathcal{H}_A$ 的表示方式，不一定与它试图学习的未知目标概念类 $\mathcal{C}$ 的表示方式相同。如果 $\mathcal{H}_A$ 的表示方式与 $\mathcal{C}$ 的表示方式相同，即 $h \in \mathcal{C}$，那么学习器 A 就是 $\mathcal{C}$ 的**适当 PAC 学习器**。反之，当 $h \notin \mathcal{C}$ 时，A 被视为 $\mathcal{C}$ 的**不适当 PAC 学习器**。[17]

与可实现设定类似，在不可知设定中，如果 A 从基准类 $h \in \mathcal{C}$ 中输出一个假设，那么 A 就是 $\mathcal{C}$ 的适当不可知学习器。否则，A 就是 $\mathcal{C}$ 的不适当不可知学习器。

C.2.2 弱学习与强学习

针对未知分布下的未知目标概念，直接设计 PAC 学习器往往颇具挑战。在实践中，设计一个弱 PAC 学习器更容易——它只需以 $\sim 51\%$ 的准确率成功，然后再将此类学习器的成功概率提升至任意接近 100%。换言之：给定一组在弱 意义上属于 PAC 学习器的学习算法，我们可以使用**集成学习算法**将上述弱 学习器集成起来，从而获得强 PAC 学习器[18]。稍后，我们会在前文探讨的各种噪声模型下，对弱 和强PAC 学习器的概念进行形式化。

C.3 对成员查询预言机及其变体的批评

尽管学习理论文献中相当一部分内容都致力于构造能够访问 MQ 预言机的 PAC 学习算法，以处理各种布尔概念类，但这方面的实际研究并不多。这可能是因为 MQ 预言机本身具有悖论般的性质：如果实际上能够为某个目标概念类 $\mathcal{C}$ 构造 MQ 预言机，那么就没有必要为 $\mathcal{C}$ 构造 PAC 学习器。另一方面，EX、REX、AEX 等采样查询预言机（以及它们的量子对应物），只是通过用预言机查询的方式重新表述定义，从而扩展了通常的监督学习概念。

---

[17]（不）恰当学习也称为表示（不）依赖学习。[18]在可实现设定中，强 PAC 学习器满足 (ε, δ)-PAC 可学习性。

备注 C.1。 *Baum* 和 *Lang* 曾进行过一项著名的实验， *[95]* 他们尝试使用人工标注员作为成员查询预言机，学习用于识别手写字符和数字的线性分类器。实验发现，除了进行分布外查询之外，学习器有时也会进行领域外查询。这可以视为支持这样一种假设的轶事证据：成员查询预言机在实践中可能并无用处。

C.4 各种标签噪声模型下的 PAC 学习

C.4.1 可实现／无噪声设定

第 2.3 节 介绍的学习设定称为**可实现设定**，因为我们假设总是存在一个目标概念 $c \in \mathcal{C}$，它**能够实现** 对实例的标注。我们也将此设定称为**无噪声设定**，因为训练集中的标签未经更改即可提供给学习器。

C.4.2 有界噪声设定

我们可以将可实现设定推广，以涵盖更贴近现实的学习场景，如下所示。设 $\mathcal{S} = \{(x_1, c(x_1)) \ldots, (x_m, c(x_m))\}$ 是用于训练学习算法的训练集 A。假设某个对手篡改了训练集的标签$\mathcal{S}$，之后学习器获得该训练集 A 并用它来构造假设。显然，(ε, δ)-PAC 可学习性的样本复杂度和时间复杂度保证必须取决于噪声率。

在 [96]，引入的 RCN设定中，我们假设训练集的标签以 $0 \leq \eta < 1/2$ 的概率被翻转，即给定训练集 $\mathcal{S} = (x_i, y_i)^{i \in [m]}$，以 η 的概率独立地设置每个$y_i = y_i$。在此设定中，PAC 学习器的样本和时间复杂度取决于 poly $(1/\varepsilon, 1/\delta, n,$ 规模 $(\mathcal{C})$ $1/1-2\eta)$。之后，我们将使用带噪声预言机重新定义 RCN设定。更具挑战性的一类有界噪声是 **Massart噪声**，其中每个标签 y_i 以 $\eta_i \leq \eta < 1/2$ 的概率被翻转。

C.4.3 无界噪声设定

一种更具挑战性的噪声模型是无关PAC设定 [97, 98]。在这种噪声模型中，每个标签 y_i 被篡改的概率为 $\eta_i < 1/2$。这里需要注意的是，不存在界限 η 使得 $\eta_i < \eta$, $\forall i \in [|\mathcal{S}|]$。如果标签的篡改率没有界限，那么尚不清楚在训练集中加入噪声后，原本能够高效学习 $\mathcal{C}$的学习算法 A是否仍是 $\mathcal{C}$ 的 PAC 学习器。在无界噪声模型中，我们不再假设存在一个固定的潜在概念$c \in \mathcal{C}$ 来为实例 $x \in \mathcal{X}$赋予标签，因为训练集本身可能不一致。[19][20] 相反，我们假设带标签实例（x, y）是从定义在 $\mathcal{X} \times \mathcal{Y}$上的分布 $\mathfrak{D}$ 中联合采样得到的。因此，对于目标概念的选择，该学习算法是不作预设的。

[19]如果对于某个 $i \in [|\mathcal{S}|]$，噪声率 η_i 非常接近 $1/2$，则训练集可能包含样本 $\{x_i, +1\}$ 和 $\{x_i, -1\}$。因此，实例的标注不再被视为一个函数。[20]有界噪声模型中也会出现不一致性，但可以通过多数投票轻松解决。

C.5 学习的傅里叶分析技术

学习布尔函数 f 的技术，其核心如下：

- 给定有关 f 的傅里叶谱的一些信息，显式构造一个用于 f 的弱学习器。

- 使用 Boosting 等集成学习算法，以高概率为 f 获得一个近似的强学习器，即为 f 获得一个 PAC 学习器。

布尔函数 f 在**次数** k 上的**傅里叶权重** 定义为汉明重量恰好等于 k 的子集的傅里叶系数平方和，即$\mathcal{W}^{=k}f = \quad \sum \hat{f}(S)^2$.我们也可以类似地定义量 $\mathcal{W}^{\leq k}f$ 和 $\mathcal{W}^{\geq k}f$。A$S\subseteq[n],|S|=k$

如果$\sum \hat{f}(S)^2 \leq S\subseteq[n],S\notin\mathcal{A}$，则布尔函数 f 在集合 $\mathcal{A}\subseteq 2^{[n]}$ 上是 ε-谱集中的

ε. 因此，如果$\mathcal{W}^{>k}f \leq \varepsilon.$，则布尔函数 f 在**次数** $\leq k$上是 ε-谱集中的。现在我们概述一些重要结果。

引理 C.1（低阶学习 [93]）。如果布尔函数 f 在集合 $\mathcal{A}$上是 ε-集中的，并且我们明确知道 $\mathcal{A}$，那么存在一个学习算法 A，它可以（ε，δ）-*PAC* 学习 f，所需时间为多项式（$|\mathcal{A}|$，n，$1/\varepsilon$，$1/\delta$），并可查询访问 EX（$\mathcal{U}$，f）预言机。如果 f在次数至多为 k的范围内是 ε-集中的，那么 A 可以在时间为多项式 $(n^k,\ 1/\varepsilon,\ 1/\delta)$ 内学习 f，并可查询访问 EX（$\mathcal{U}$，f）预言机。

注释 C.2。如果 k 不是常数，那么引理 *C.1* 中的学习器不会为 f提供高效 *PAC* 学习器。

引理 C.1 的一个缺点是，它要求我们明确知道布尔函数在哪些文字子集上谱集中。我们可以通过将预言机从 EX ($\mathcal{U}$，f) 升级为 MEMf 预言机，来绕过这一要求。

引理 C.2（从谱集中学习 [73, 99]）。设 $\varepsilon,\delta > 0$ 和 $f:\ \mathbb{F}_2^n \mapsto \mathbb{F}_2$。如果 f 在次数 k 以内是 ε-集中的，那么 $\exists$ 存在一个学习算法 A，该算法能够通过查询访问 MEMf 预言机，在时间 poly $(n^k, 1/\varepsilon, 1/\delta)$ 内对 f 进行 ε,δ)-*PAC* 学习。

引理 C.2 有两个重要推论。第一个结果使我们能够改进引理 C.1：无需明确知道 $\mathcal{S}$，即可学习在某个子集 $\mathcal{S}$上谱集中的函数。

推论 C.1（从 ε-集中的奇偶单项式子集学习）。设 $\varepsilon,\delta >0$ 和 $f:\ \mathbb{F}_2^n \mapsto \mathbb{F}_2$。如果 f 在子集 $\mathcal{S}$ 上是 ε- 集中的，那么 $\exists$ 存在一个学习算法 A，该算法能够通过查询访问 MEMf 预言机，在时间 poly $(|\mathcal{S}|, 1/\varepsilon, 1/\delta)$ 内对 f 进行 ε,δ)-*PAC* 学习。

下一个推论有助于我们定义由具有较大傅里叶系数的奇偶单项式构成的 PAC 学习器。设 f 为一个布尔函数。那么，FS（f）的 ℓ_1 范数定义为

定义为

$$\|\mathrm{FS}(f)\|_1 = \sum_{\mathcal{S}\subseteq[n]} \left|\hat{f}(\mathcal{S})\right|. \tag{C3}$$

推论 C.2 (从具有较大傅里叶系数的奇偶单项式中学习)。 令ε、 $\delta > 0$和 $f:\ \mathbb{F}_2^n \mapsto \mathbb{F}_2$。 $\exists$ 一个学习算法 A ， 它可通过查询访问一个MEMf预言机，以(ε,δ)-*PAC*学习 f，所需时间为poly $(\|\mathrm{FS}(f)\|_1, 1/\varepsilon, 1/\delta)$。

观察 C.3. 推论 *C.2* 告诉我们，如果 ℓ_1 任意布尔函数的范数都有多项式上界，那么我们就可以构造一个 *PAC* 学习器 f，通过向 *MQ* 预言机查询来进行学习。

附录 D 群论

群是一个非空集合 $\mathcal{G}$，即一个带有结合二元运算（$\cdot$）： $\mathcal{G}\times\mathcal{G}\mapsto\mathcal{G}$的集合，使得相对于（$\cdot$）， $\mathcal{G}$ 中存在一个单位元，并且 $\mathcal{G}$ 中的每个元素在 $\mathcal{G}$ 中都有一个逆元（相对于$\cdot$）。如果对于所有 x， $y\in\mathcal{G}$、 $x\ (\cdot)\ y = y\ (\cdot)\ x$，那么 $\mathcal{G}$ 是一个阿贝尔群。 $\mathcal{G}$ 的生成子集 $\mathcal{H}$ 是一个子集，其中 $\mathcal{G}$ 的每个元素都可以表示为 $\mathcal{H}$ 中有限多个元素及其逆的组合。如果 $\mathcal{G}$ 是一个循环群，那么它有一个元素 x 作为该群的生成元。任意有限阿贝尔群都可以如下分解为循环群：

引理 D.1 （有限阿贝尔群基本结构定理）。 每个阿贝尔群 $\mathcal{G}$都同构于形如 $\mathbb{Z}_{p_1^{a_1}}\times\mathbb{Z}_{p_2^{a_2}}\times\ldots\times\mathbb{Z}_{p_k^{a_k}}$ 的群，其中p_1、.....、 p_k 为素数（不一定唯一），且除因子 $\{p_1^{a_1},\ldots,p_k^{a_k}\}$ 的排列顺序外，该分解是唯一的。

注释 D.1。 $\mathbb{Z}_k$ 是一个关于 $+_{\bmod k}$运算、阶为 k 的有限阿贝尔群。若 $k = 1$、2、4 或某个奇素数的幂，则 $\mathbb{Z}_k$ 也是一个循环群。

定义 D.1 （二面体群）。 二面体群 D_{2N} 是一个由两个元素 x 和 y 生成的、阶为 $2N$ 的非阿贝尔群，满足

$$x^N = 1,\quad y^2 = 1,\quad yxy = -x.$$

D_{2N} 是具有 $2N$ 条边的正多边形的对称群。元素 x 和 y分别是关于垂直轴的旋转和反射运算。