arXiv:2302.06512v1 [cs.LG] 13 Feb 2023

高斯边缘分布下不可知学习半空间与 ReLU 回归的近最优密码学困难性

Ilias Diakonikolas* DanielM. Kane† Lisheng Ren‡威斯康星大学麦迪逊分校
加州大学圣迭戈分校 威斯康星大学麦迪逊分校
ilias@cs.wisc.edu dakane@ucsd.edu lren29@wisc.edu

2023年2月14日

摘要

我们研究高斯分布下半空间的无关学习任务。具体而言，给定来自定义在 $\mathbb{R}^n \times \{\pm 1\}$ 上的未知分布的带标签样本 (x y)，其在 x 上的边缘分布为标准高斯分布，且标签 y 可以任意取值，目标是输出一个 0-1 损失为 OPT $+\epsilon$ 的假设，其中 OPT 是拟合最佳的半空间的 0-1 损失。我们证明，在广受认可的误差学习（LWE）问题的次指数时间困难性假设下，该任务具有近乎最优的计算困难性。此前的困难性结果要么在定性上并非最优，要么仅适用于受限的算法类别。我们的技术还可推广，为包括 ReLU 回归在内的相关问题给出近乎最优的下界。

*由美国国家科学基金会中型项目资助（CCF-2107079）、美国国家科学基金会项目资助（CCF-1652862，CAREER）、斯隆研究奖学金以及美国国防高级研究计划局少标签学习（LwLL）项目资助。
†由美国国家科学基金会项目资助（CCF-2107547）、美国国家科学基金会项目资助（CCF-1553288，CAREER）、斯隆研究奖学金以及 CasperLabs 的资助。
‡由美国国家科学基金会 CCF-1652862 奖项（CAREER）和美国国防高级研究计划局少标签学习（LwLL）项目资助。

1 引言

半空间或线性阈值函数（LTF）是任意形如 $f(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x}\rangle - t)$ 的布尔值函数 $f : \mathbb{R}^n \to \{\pm 1\}$，其中 $\mathbf{w} \in \mathbb{R}^n$ 称为权重向量，而 $t \in \mathbb{R}$ 称为阈值。这里，一元函数 $\text{sign} : \mathbb{R} \to \{\pm 1\}$ 定义为：若 $u \geq 0$，则 $\text{sign}(u) = 1$；否则 $\text{sign}(u) = -1$。学习未知半空间是机器学习中的经典问题，自20世纪50年代以来便得到广泛研究，最初由感知机算法 [Ros58]，开始，并催生了支持向量机 [Vap98] 和 AdaBoost [FS97] 等具有重要实践意义的技术。在可实现设定中 [Val84]，，已知半空间可以在不作分布假设的情况下高效学习（例如参见 [MT94]）。相较之下，在无分布假设的非特定模型中 [Hau92, KSS94]，，即使是弱学习在计算上也很困难 [GR06, FGKP06, Dan16, Tie22]。鉴于这种计算难解性，大量研究聚焦于特定分布设定下的半空间无关学习。直观而言，数据分布的内在结构有望得到利用，从而设计出能够抵御对抗性标签噪声的非平凡高效算法。

本文聚焦于一个经过充分研究的任务：当样本所服从的底层分布假设为高斯分布时，如何无关学习半空间。具体而言，我们获得来自关于带标签样本 D 的联合分布的独立同分布样本 $(\mathbf{x}\ y)$，其中 $\mathbf{x} \in \mathbb{R}^n$ 是样本，$y \in \mathbb{R}$ 是相应的标签；目标是计算一个假设，使其表现不逊于拟合最佳的半空间。此外，我们假设 $\mathbb{R}^n$ 上的边缘分布 $D_{\mathbf{x}}$ 是标准高斯分布 $\mathcal{N}(0, \mathbf{I})$。正如我们随后将要说明的，与分布无关的设定相比，这一分布假设使学习问题在计算上更容易。有趣的是，即使是该问题的高斯版本，也呈现出信息-计算权衡；本文将对此进行探究，并基本加以解决。

为使表述具体，我们先介绍一些记号，随后给出上述问题的定义。对于布尔值假设 $h : \mathbb{R}^n \to \{\pm 1\}$ 以及一个支撑于$\mathbb{R}^n \times \{\pm 1\}$上的分布 D，我们用 $R_{0-1}(h; D)$ 表示 h 相对于 D的 0-1 误差，即 $R_{0-1}(h;\ D) \overset{\text{定义}}{=} \Pr_{(\mathbf{x},y)\sim D}[h(\mathbf{x}) \neq y]$。对于定义在 $\mathbb{R}^n$上的布尔值函数类 $\mathcal{C}$，我们用 $R_{0-1}(\mathcal{C}; D)$ 表示任意 $h \in \mathcal{C}$的最小 0-1 误差，即 $R_{0-1}(\mathcal{C}; D) \overset{\text{定义}}{=} \min_{h\in\mathcal{C}} R_{0-1}(h; D)$。

问题 1.1（高斯边缘分布下半空间的不可知学习）。设 LTF 为 $\mathbb{R}^n$ 上半空间的类。给定误差参数 $0 < \epsilon < 1$，以及从分布 D 中抽取的独立同分布样本 $(\mathbf{x}\ y)$，其中 D 在 $\mathbb{R}^n \times \{\pm 1\}$ 上的边缘分布 $D_{\mathbf{x}}$ 是标准高斯分布 $\mathcal{N}(0, \mathbf{I})$，且对标签 y 不作任何假设，学习算法 $\mathcal{A}$ 的目标是输出一个假设 h: $\mathbb{R}^n \to \{\pm 1\}$，使得以高概率有 $R_{0-1}(h;\ D) \leq R_{0-1}(\text{LTF}; D) + \epsilon$。我们称算法 $\mathcal{A}$ 以加性误差 ϵ 在高斯边缘分布下不可知地学习半空间（或 LTF）。

问题 1.1 的先前工作 根据标准结果 [Hau92, KSS94]，，可知半空间不可知学习问题的样本复杂度为 $O(n/\epsilon^2)$。L_1-回归算法 [KKMS08] 以样本复杂度和运行时间 $n^{O(1/\epsilon^2)}$ [DGJ+10, DKN10] 求解问题 1.1。虽然 L_1-回归算法并非适当学习器，但近期研究提出了一种在样本复杂度和时间复杂度方面定性相近的适当学习器（即 $n^{\text{poly}(1/\epsilon)}$）[DKK+21]。重要的是，$L_1$-回归算法仍是该问题已知最高效的算法。

鉴于该问题的样本复杂度与已知算法的复杂度之间存在差距，自然会让人想知道，已知高效算法的局限性是否是内在的。建立统计问题中信息-计算权衡通常有两种方法。一种方法着眼于受限的算法族（例如统计查询算法或低次数多项式检验）。需要指出的是，这类结果对所有多项式时间算法构成的算法族并无任何启示

。另一种或许更有说服力的方法，是通过已知（平均情形）困难问题的高效归约来实现。本文采用的正是这种方法。

回到问题 1.1，一系列研究 [GGK20、DKZ20、DKPZ21] 已经在统计查询（SQ）模型中确立了紧确困难性。统计查询算法 [Kea98] 是一类算法，只允许查询分布上有界函数的期望，而不能直接访问样本。 [DKPZ21] 证明，对于该问题，任何统计查询算法要么需要$2^{n^{\Omega(1)}}$ 次查询，要么至少需要一次精度极高的查询（这意味着样本复杂度下界为 $n^{\Omega(1/\epsilon^2)}$）。有趣的是，已知（例如，参见 [DFT+15]）， L_1-回归算法可以在统计查询模型中高效实现。然而，由于统计查询模型受到限制，这一下界对一般的高效算法并无影响。

在我们的工作之前，关于问题 1.1，唯一已知的计算困难性结果来自 Klivans 和 Kothari [KK14]。该工作给出了从带噪声稀疏奇偶函数学习问题到问题 1.1 的归约。在合理的假设下，即在 $\{0,1\}^n$ 上学习带噪声的 k-稀疏奇偶函数需要 $n^{\Omega(k)}$时间， [KK14] 的归约意味着问题 1.1 存在 $n^{\Omega(\log(1/\epsilon))}$ 计算下界。有趣的是，这一下界无法改进，因为相应的困难实例可以在 $n^{O(\log(1/\epsilon))}$时间内求解。

最后，我们指出，对于定性上较弱的误差保证 $C\cdot\mathrm{OPT}+\epsilon$，当普适常数 $C>1$ 足够大时，已有已知算法可在 poly(d/ϵ) 时间内运行 [ABL17, Dan15, DKS18]。

总而言之，问题 1.1 的当前最佳已知算法具有样本复杂度和运行时间$n^{\mathrm{poly}(1/\epsilon)}$，而当前最佳已知计算困难性结果给出了一个 $n^{\Omega(\log(1/\epsilon))}$ 下界。此外，对于受限的统计查询算法类，已有紧确下界。由此引出一个自然的问题：

我们能否为问题 1.1 建立近乎最优的计算困难性结果 ？

本文通过给出从一个经典密码学问题出发的计算困难性归约，对这一问题作出肯定回答，并表明当前算法在本质上已达到最优。具体而言，我们证明了复杂度下界 $n^{\mathrm{poly}(1/\epsilon)}$ （定理 1.3），其前提是广泛被认为成立的误差学习（LWE）问题（定义 2.2）具有次指数困难性。

学习半空间是一个更一般设定的特例，在该设定中，底层函数具有 $\sigma((\langle \mathbf{w},\mathbf{x}\rangle - t))$ 的形式，其中 $\sigma\colon \mathbb{R}\to\mathbb{R}$ 是一个单变量激活函数。如果激活函数的性质优于符号函数，具体而言，如果 σ 是单调且 Lipschitz 连续的（即广义线性模型设定），那么该学习问题在计算上可能更容易。本文将证明，我们的技术也可以扩展到其中一些情形，并给出近乎最优的困难性结果。具体而言，我们聚焦于久经研究的 ReLU 回归问题。

ReLU 是任意具有 $f\colon \mathbb{R}^n\to\mathbb{R}_+$形式 $f(\mathbf{x}) = \mathrm{ReLU}(\langle \mathbf{w},\mathbf{x}\rangle - t)$ 的函数，其中 $\mathbf{w}\in\mathbb{R}^n$ 称为权重向量， $t\in\mathbb{R}$ 称为阈值。激活函数 $\mathrm{ReLU}:\mathbb{R}\to\mathbb{R}_+$ 定义为 $\mathrm{ReLU}(u) = \max\{0,u\}$。ReLU 是现代深度神经网络中最常用的激活函数。此外，寻找相对于平方损失拟合效果最佳的 ReLU，是神经网络理论中的一项基础原语。一系列研究从算法和下界两个角度探讨了这一问题，参见例如 [Sol17, GKKT17, MR18, GKK19, FCG20, DGK+20, DKTZ22, ATV22]。与半空间的情形类似，在可实现设定下，ReLU 回归可以高效求解；而在与分布无关的非特设定下，即使误差保证较弱，计算上也很困难 [MR18, DKMR22a]。本文研究具有高斯边缘分布的非特设定。

由于 ReLU 回归是一项实值任务，我们需要采用相应的术语。对于实值假设 $h\colon \mathbb{R}^n\to\mathbb{R}$ 以及一个支撑于 $\mathbb{R}^n\times\{\pm 1\}$ 上的分布 D，我们用 $R_2(h;D)$

表示 h 相对于 D 的 L_2^2-误差，即 $R_2(h; D)$ 定义为 $= \mathrm{E}_{(\mathbf{x},y)\sim D}[(h(\mathbf{x}) - y)^2]$。对于 $\mathbb{R}^n$ 上的实值函数类 $\mathcal{C}$，我们用 $R_2(\mathcal{C}; D)$ 表示任意 $h \in \mathcal{C}$ 的最小 L_2^2-误差，即 $R_2(\mathcal{C}; D)$ 定义为 $= \min_{h\in\mathcal{C}} R_2(h; D)$。

问题 1.2（高斯边缘分布下的 ReLU 回归）。令 ReLU 表示定义在 $\mathbb{R}^n$ 上的 ReLU 类，其权重向量属于集合 $\{\mathbf{w} \in \mathbb{R}^n : \|\mathbf{w}\|_2 \le 1\}$。给定加性误差参数$0 < \epsilon < 1$ 以及来自分布 D 在 $\mathbb{R}^n \times \mathbb{R}$ 上的独立同分布样本 $(\mathbf{x}\ y)$，其中在 $\mathbb{R}^n$上的边缘分布 $D_{\mathbf{x}}$ 为标准高斯分布 $\mathcal{N}(0, \mathrm{I})$，且标签 y 有界，学习算法 $\mathcal{A}$的目标是输出一个假设 $h\colon \mathbb{R}^n \to \mathbb{R}$，使得 $R_2(h; D) \le R_2(\mathrm{ReLU}; D) + \epsilon$，且以高概率成立。我们称算法 $\mathcal{A}$以加性误差 ϵ在高斯边缘分布下无知学习 ReLU。

问题 1.2 的先前工作 尽管问题 1.2 与问题 1.1 之间不存在黑盒关系，但这两个问题的情形是类似的。 [DGK+20] 为问题 1.2 给出了一种样本复杂度和运行时间为 $n^{\mathrm{poly}(1/\epsilon)}$的算法。尽管已知运行时间为 poly($n/\epsilon$)、但保证较弱的算法 [GKK19、DGK+20、DKTZ22], ，已知的具有 OPT $+ \epsilon$ 误差的最快算法仍是 [DGK+20]提出的算法。在计算困难性方面， [GKK19] 给出了从稀疏带噪奇偶问题出发的归约，从而为问题 1.2 带来了 $n^{\Omega(\log(1/\epsilon))}$ 的计算下界。在受限 SQ 模型中， $n^{\mathrm{poly}(1/\epsilon)}$ 的（近乎最优）SQ 下界已由 [GGK20、DKZ20、DKPZ21]证明。

总而言之，问题 1.2 的已知最佳算法具有样本复杂度和运行时间$n^{\mathrm{poly}(1/\epsilon)}$，而已知最佳计算困难性结果给出了一个 $n^{\Omega(\log(1/\epsilon))}$ 下界。因此，自然会问能否建立 $n^{\mathrm{poly}(1/\epsilon)}$ 的计算困难性。与线性阈值函数的情形类似，我们在 LWE 的次指数困难性假设下证明了这一结论（定理 1.4）。

1.1 我们的结果与技术

我们先对 LWE 问题给出一个非正式定义。在 LWE 问题中，我们获得样本 $(\mathbf{x}_1, y_1) \ldots, (\mathbf{x}_m, y_m)$，目标是在以下两种情形之间进行区分：

- 每个 $\mathbf{x}_i$ 均从 $\mathbb{Z}_q^n$中均匀随机（u.a.r.）抽取，并且存在一个隐藏秘密向量 $\mathbf{s} \in \mathbb{Z}_q^n$，使得 $y_i = \langle \mathbf{x}_i, \mathbf{s}\rangle + z_i$，其中 $z_i \in \mathbb{Z}_q$ 是离散高斯噪声（与 $\mathbf{x}_i$独立）。
- 每个 $\mathbf{x}_i$ 和每个 y_i 相互独立，分别从$\mathbb{Z}_q^n$ 和 $\mathbb{Z}_q$ 中均匀随机采样。

LWE 的形式化定义（定义 2.2）以及精确的计算困难性假设（假设 2.3）均在第 2 节中给出。

对于问题 1.1，我们证明：

定理 1.3（高斯半空间的无关学习困难性）。假设 LWE 无法在 $2^{n^{1-\Omega(1)}}$ 时间内求解。那么对于任意常数 $c > 0$ 和 $\alpha < 2$，以下结论成立：若$\epsilon \le 1/\log^{1/2+c}(n)$，则任何在具有高斯边缘分布的 $\mathbb{R}^n$ 上以至多 ϵ的加性误差无知学习线性阈值函数的算法，其运行时间至少为 $\min\{n^{\Omega(1/(\epsilon\sqrt{\log n})^{\alpha})}, 2^{n^{0.99}}\}$。

有必要对这一表述作一些说明。两个项中的较小值~~˜~~是处理 ϵ 非常小这一情形所必需的，具体而言，即 $\epsilon = O(1/\sqrt{n})$。（由于该问题总能通过蛮力搜索在 $2^{O(n)}$ 时间内解决，因此在这一范围内，第一个项不能作为时间下界 ϵ。）另一方面，对于 $\tilde{\Omega}(1/\sqrt{n}) = \epsilon \le 1/\log^{1/2+c}(n)$，定理 1.3 对任意常数 $\alpha < 2$给出的时间下界为 $n^{\Omega(1/(\epsilon\sqrt{\log n})^{\alpha})}$。这一界几乎与上界 $n^{O(1/\epsilon^2)}$ [KKMS08], 相匹配，指数中相差一个$\sqrt{\log n}$ 因子。注意，~~—~~多余的

$\sqrt{\log}\ n$ 因子在 ϵ 足够小时可以忽略。例如，如果 $\epsilon \le 1/\log\ n$，那么对任意常数 $\alpha < 1$，所隐含的下界为 $n^{\Omega(1/\epsilon^{\alpha})}$。对于 $\epsilon = O(n^{-c})$，取一个较小的常数 $c > 0$，我们得到一个下界 $n^{\Omega(1/\epsilon^{\alpha})}$，其中 $\alpha < 2$ 为任意常数。

对于问题 1.2，我们证明：

定理 1.4（高斯 ReLU 回归的困难性）。假设学习带误差无法在$2^{n^{1-\Omega(1)}}$ 时间内求解。那么，对于任意常数 $c > 0$ 和 $\alpha < 1/2$，以下结论成立：若 $\epsilon \le 1/\log^{2+c}(n)$，则任何在高斯边缘分布下对 $\mathbb{R}^n$ 进行 ReLU 回归且加性误差为 ϵ 的算法，其运行时间至少为 $\min\{n^{\Omega(1/(\epsilon\ 对数^2\ n)^{\alpha})}, 2^{n^{0.99}}\}$。

直观而言，上述结论表明，如果 ϵ 足够小（例如 $\epsilon = O(1/\log^3 n)$），但又不至于太小（否则后一项会超过显然可行的蛮力算法），那么任何求解问题 1.2 的算法都至少需要 $n^{(1/\epsilon)^{\Omega(1)}}$时间。这个运行时间下界在定性上与 $n^{\mathrm{poly}(1/\epsilon)}$ [DGK+20] 的上界相匹配，并且将已知最佳计算下界 $n^{\Omega(\log(1/\epsilon))}$ [GKK19]提高了指数级。

1.2 技术

我们的计算困难性归约建立在两个主要思想之上。第一个思想受 [DKMR22b] 所采用方法的启发。我们注意到，[DKMR22b] 建立了从 LWE 到带 Massart噪声的无分布 PAC 学习半空间的困难性归约。尽管 Massart噪声模型在技术上比对抗性标签噪声模型更容易处理，但我们在此关注的是边缘分布为高斯的（简单得多的）情形。事实上，[DKMR22b] 的结果对高斯情形没有任何影响。然而，他们的一个思想对我们的场景很有用。

[DKMR22b] 的关键思想是：对支撑于 $\mathbb{R}^n$ 上的 LWE 连续变体应用拒绝采样（在 [GVV22] 中已证明，该变体与支撑于 $\mathbb{Z}_q^n$ 上的标准 LWE 问题同样困难），可以得到以下两种结果之一：(i) 在零假设情形下得到一个标准高斯分布；或 (ii) 在备择假设情形下得到一个近似的离散高斯分布，并在隐藏方向上加入少量噪声、在正交方向上加入一个标准高斯分布。通过对这类拒绝采样分布进行混合，[DKMR22b] 成功构造了定义在 $\mathbb{R}^n \times \{\pm 1\}$ 上的 (x y) 联合分布，使得：

(i) 在零假设情形下，y 与 x 相互独立；

(ii) 在备择假设情形下，[1] y 由作用于 x 的多项式阈值函数（PTF）在 Massart噪声下给出。

鉴于上述结果，[DKMR22b] 得出结论：任何用于 Massart 半空间线性阈值函数的学习器都可用于区分备择假设情形与零假设情形，从而解决 LWE 问题。

本文将类似的技术应用于高斯边缘分布下非投机地学习半空间和 ReLU 的任务。我们设定中的一个关键区别是，要求 x 的分布为标准高斯分布——而上述构造本身并不满足这一性质。粗略地说，[DKMR22b] 证明了，区分标准高斯分布与另一种分布是学习带误差困难的：后一种分布在除隐藏方向之外的所有方向上都是标准高斯分布，而在隐藏方向上则近似为指定的离散高斯分布混合，再加上少量噪声。[DKMR22b] 中的学习应用，是通过构造一个带有 Massart噪声的 PTF 得到的；在该构造中，$y = 1$ 和 $y = -1$ 上的条件分布都是这种（含噪的）离散高斯分布混合。在我们的设定中，需要构造不同的条件分布对。

[1] 这利用了从 [DK22]构造此类分布的方法。

具体做法如下。设 x 服从标准高斯分布，并考虑函数$f_{\mathbf{s}}(\mathbf{x}) = (-1)^{\lfloor \langle \mathbf{x},\ \mathbf{s} \rangle \rfloor}$，其中 s 是某个范数相对较大的未知向量。考虑给定 $f_{\mathbf{s}}(\mathbf{x}) = 1$ 时 x 的分布，可得到一种分布，它满足：(i) 在与 s 正交的方向上为高斯分布；(ii) 在 s 方向上为给定 $\lfloor \langle \mathbf{x},\mathbf{s} \rangle \rfloor$ 为偶数条件下的高斯分布。可以看出，这是一种离散高斯分布的混合。对于给定 $f_{\mathbf{s}}(\mathbf{x}) = -1$时 s 的分布，也可以作出同样的论证。因此，利用上述技术，我们可以证明：给定带标签样本 $(\mathbf{x}\ y)$，其中 x 服从标准高斯分布，要区分以下两种情况是学习带误差困难的：(i) y 与 x 相互独立；(ii) 对于某个未知向量 s，$y = f_{\mathbf{s}}(\mathbf{x})$。

这一结果构成了我们两种学习应用的基础。具体而言，对于高斯线性阈值函数的无先验学习问题，不难证明存在一个线性阈值函数 g，使得 $\mathrm{E}_{\mathbf{x}\sim\mathcal{N}(\mathbf{0},\mathbf{I})}[f_{\mathbf{s}}(\mathbf{x})g(\mathbf{x})] = \epsilon = \Omega(1/\|\mathbf{s}\|_2)$。这意味着，任何能够以误差 $\mathrm{OPT} + \epsilon/3$ 无关学习线性阈值函数的算法，其中 $\mathrm{OPT} = R_{0-1}(\mathrm{LTF}; D)$，都可用于区分以下两种情况：$y$ 与 x 相互独立（此时 $\mathrm{OPT} = 1/2$）；以及上述情况（即，$y = f_{\mathbf{s}}(\mathbf{x}) = (-1)^{\lfloor \langle \mathbf{x},\ \mathbf{s} \rangle \rfloor}$），此时 $\mathrm{OPT} = 1/2 - \epsilon$。这意味着，高斯线性阈值函数的不可知学习是 LWE 困难的。

对于 ReLU 回归，我们证明存在一个 ReLU g，使得 $\mathrm{E}_{\mathbf{x}\sim\mathcal{N}(\mathbf{0},\mathbf{I})}[f_{\mathbf{s}}\ g\ ] =$ $\epsilon = \mathrm{E}(\mathbf{x})\ (\mathbf{x})\ \Omega(1/\|\mathbf{s}\|_2^2)$。具体而言，这一相关性意味着，$L_2^2$-距离在 f 与其适当缩放后的版本 g 之间，在负方向上与 1 保持一定距离。因此，要区分以下两种情况是 LWE 困难的：$y = f_{\mathbf{s}}(\mathbf{x})$（因此，ReLU 的最小 L_2^2-误差至多为 $1 - \epsilon^2$）；以及 y 与 x 相互独立（此时，任何 ReLU 的最小 L_2^2-误差至少为 1）。

上述概述忽略了以下重要的技术问题。通过将前面所述的归约直接应用于标准版本的（连续）LWE 问题 [BRST21]（其秘密向量 s 满足 $\|\mathbf{s}\|_2 = \sqrt{n}$），我们只有在加性误差 ϵ 非常小，即 $\epsilon = O(1/\sqrt{n})$ 时，才能为我们的不可知~学习问题得到时间下界。为了证明适用于更大范围 ϵ 的下界，我们需要改为从连续 LWE 问题的小范数版本出发，其中秘密向量 s 大致满足 $\|\mathbf{s}\|_2 \approx 1/\epsilon$。我们通过对 [GVV22]，中归约的非平凡修改来实现这一点，并将其视为本文的另一项技术贡献。具体而言，[GVV22] 将标准离散 LWE 问题归约为秘密稀疏的离散 LWE 问题（即，秘密向量 s $\in \{0, \pm 1\}^n$ 满足$\|\mathbf{s}\|_1 = k$）。（这本身利用了 [Mic18] 中的一个思想。）随后，[GVV22] 进一步将秘密稀疏的离散 LWE 问题归约为连续 LWE 问题，其秘密向量具有较小的 ℓ_2-范数。这里的局限在于，他们给出的 ℓ_2-范数界中包含一个 $\sqrt{\log m}$ 因子，其中 m 是样本数量。不幸的是，这种定量依赖性阻碍了我们为学习线性阈值函数任务获得近乎最优的下界。为了解决这一问题，我们提出了一种（略有）改进的归约（见引理 B.5），去除了秘密向量范数中的 $\sqrt{\log m}$ 因子。这样，我们便可以将归约技术应用于小范数连续 LWE 问题，从而为我们的学习问题给出几乎紧的下界。

2 预备知识

符号约定 我们用 $\langle \mathbf{x}, \mathbf{y} \rangle$ 表示向量 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$的内积。对于 $p \geq 1$ 和 $\mathbf{x} \in \mathbb{R}^n$，我们用 $\|\mathbf{x}\|_p$ 定义为$= (\sum_{i=1}^n |\mathbf{x}_i|^p)^{1/p}$ 表示 x 的 ℓ_p-范数。我们用 $\mathbb{S}^{n-1}$ 表示 $\mathbb{R}^n$中的单位球面，即集合 $\mathbb{S}^{n-1}$ 定义为$= \{\mathbf{x} \in \mathbb{R}^n \colon \|\mathbf{x}\|_2 = 1\}$。对于 $q \in \mathbb{N}$，我们记 $\mathbb{Z}_q$ 定义为$= \{0, 1, \cdots, q-1\}$ 和$\mathbb{R}_q$ 定义为$= [0, q)$。我们用 $\mathrm{mod}_q \colon \mathbb{R}^n \to \mathbb{R}_q^n$ 表示对向量 x 的每个坐标分别应用 mod_q 运算的函数。对于集合 $S \subset \mathbb{R}^n$，我们用 $U(S)$表示 S上的均匀分布。我们用 $\mathbf{x} \sim D$ 表示服从分布 D的随机变量 x。对于随机变量

x（或分布 D），我们用 $P_{\mathbf{x}}$ （或 P_D）表示随机变量 x（或分布 D）的概率密度函数或概率质量函数。我们将需要以下部分支撑高斯分布的概念。

定义 2.1（部分支撑高斯分布）。对于 $\sigma \in \mathbb{R}_+$ 和 $\mathbf{x} \in \mathbb{R}^n$，令 $\rho_\sigma(\mathbf{x}) \stackrel{\text{def}}{=} \sigma^{-n} \exp\left(-\pi(\|\mathbf{x}\|_2/\sigma)^2\right)$。对于任意可数集[2] $S \subseteq \mathbb{R}^n$，令 $\rho_\sigma(S) \stackrel{\text{def}}{=} \sum_{\mathbf{x}\in S} \rho_\sigma(\mathbf{x})$，并令 $D^{\mathcal{N}}_{S,\sigma}$为支撑在 S 上的分布，其概率质量函数为 $P_{D^{\mathcal{N}}_{S,\sigma}}(\mathbf{x}) = \rho_\sigma(\mathbf{x})/\rho_\sigma(S)$。

为保持一致，我们使用 $D^{\mathcal{N}}_{\mathbb{R}^n,\sqrt{2\pi}\sigma}$ 表示 n维高斯分布 $\mathcal{N}(0, \sigma^2\mathrm{I})$。

带误差学习 带误差学习（LWE）问题由 [Reg05]提出。为便于后续在不同 LWE问题的变体之间进行归约，这里我们采用一个稍微更一般的定义。

定义 2.2（广义带误差学习）。设 m、$n \in \mathbb{N}$、$q \in \mathbb{R}_+$，且令 D_{sample}、D_{secret}、D_{noise} 分别为定义在 $\mathbb{R}^n$、$\mathbb{R}^n$、$\mathbb{R}$ 上的分布。在 LWE$(m, D_{\text{sample}}, D_{\text{secret}}, D_{\text{noise}}, \mathrm{mod}q)$ 问题中，给定 m 个独立样本 $(\mathbf{x}\ y)$，目标是在以下两种情形之间进行区分：

(i) 备择假设：从 D_{secret} 中抽取一个向量 s（称为“秘密向量”）。随后，通过取 $\mathbf{x} \sim D_{\text{sample}}$，$z \sim D_{\text{noise}}$，并令 $y = \mathrm{mod}q(\langle \mathbf{x}, \mathbf{s}\rangle + z)$，生成每个样本 $(\mathbf{x}\ y)$。

(ii) 零假设：随机变量 x 与 y 相互独立。此外，x 的边缘分布与备择假设中的相同，而 y 的边缘分布为$U(S)$，其中 S 是备择假设中 y 的边缘分布的支撑集。

若 A 满足 $p_{\text{alternative}} - p_{\text{null}} \geq \alpha$，则称算法 A 以优势 $\alpha > 0$解决 LWE 问题，其中$p_{\text{alternative}}$（分别为 p_{null}）表示：当输入分布来自备择假设（分别为零假设）时，A 输出“备择假设”的概率。当 LWE 中的分布在某个集合 S 上均匀分布时，我们可以将 $U(S)$ 简写为 S。

我们的困难性假设如下：

假设 2.3（次指数带误差学习假设）。令 $c > 0$ 为一个足够大的常数，并令 $q \in \mathbb{N}$。对于任意常数 $\beta \in (0, 1)$，$\kappa \in \mathbb{N}$，问题 LWE$(2^{O(n^\beta)}\mathbb{Z}_q^n, \mathbb{Z}_q^n, D^{\mathcal{N}}_{\mathbb{Z},\ \sigma}, \mathrm{mod}q)$ 在满足$q \leq n^\kappa$ 和 $\sigma = c\sqrt{n}$ 的条件下，无法在 $2^{O(n^\beta)}$ 时间内以 $2^{-O(n^\beta)}$ 优势求解。

这是一个广受认可的猜想，也得到了我们当前对该领域理解的支持。[Reg05、Pei09] 给出了一个多项式时间量子归约，将近似最短向量问题（判定版本）（GapSVP）归约到 LWE（具有类似的 n, q, σ 参数）。我们注意到，目前已知最快的 GapSVP 算法需要 $2^{O(n)}$ 时间 [ALNS20]。因此，推翻这一猜想将是一个重大突破。 [GVV22] 和 [DKMR22b]也采用了类似的假设，以确立高斯混合的计算困难性以及 Massart 半空间的与分布无关的学习难度。

除了上述标准 LWE 问题外，我们还将考虑 LWE 问题的一个连续变体（由 [BRST21] 引入），其中分布的支撑集是连续的。具体而言，我们证明的第一部分是以下命题，它对 [GVV22] 中的证明稍作修改，并给出了从标准 LWE 到连续 LWE 的归约。证明见附录 B。

[2]我们将把 集合 S 取为格的平移，从而保证 $\rho_\sigma(S)$ 有限且分布定义良好 d.

命题 2.4（具有小范数秘密的连续 LWE（cLWE）的困难性）。在假设 2.3 下，对于任意 $n \in \mathbb{N}$、任意常数 $\beta \in (0, 1)$、$\kappa \in \mathbb{N}$ $\gamma \in \mathbb{R}_+$，以及任意满足 $c > 0$ 为充分小的普适常数的 $\log^\gamma n \le k \le cn$，问题 $\mathrm{LWE}(n^{O(k^\beta)}, D^{\mathcal{N}}_{\mathbb{R}^n,\,1}, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}, \mathrm{mod}_T)$，其中 $\sigma \ge k^{-\kappa}$ 且 $T = 1/(c'\sqrt{k}\log n)$，$c' > 0$ 为充分大的普适常数，无法在 $n^{O(k^\beta)}$ 时间内以 $n^{-O(k^\beta)}$ 优势求解。

3 高斯线性阈值函数无关学习的困难性

本节将从命题 2.4（其证明推迟至附录 B）继续展开。该命题是我们归约的第一步；接下来我们给出归约的第二部分，也是主要部分。由此，我们确立了高斯分布下无关学习线性阈值函数的密码学困难性。

总体思路如下。给定来自分布 D 在 $\mathbb{R}^n \times \mathbb{R}_T$ 上的样本 $(\mathbf{x}\ y)$，该分布是命题 2.4 中连续 LWE 问题 $\mathrm{LWE}(n^{O(k^\beta)}, D^{\mathcal{N}}_{\mathbb{R}^n,\,1}, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}, \mathrm{mod}_T)$ 的一个实例（注意，T 是隐藏方向上周期信号的“周期”），我们高效地生成来自分布 D' 在 $\mathbb{R}^n \times \{\pm 1\}$ 上的样本 $(\mathbf{x}\ y')$（我们保持 $\mathbf{x}$ 不变），使得：

(i) 若 D 属于备择假设情形，则存在一个线性阈值函数 $h : \mathbb{R}^n \to \{\pm 1\}$，使得 $R_{0-1}(h;\ D') \le 1/2 - \Omega(T)$。

(ii) 若 D 属于零假设情形，则对于 $(\mathbf{x}\ y') \sim D'$，我们以概率 $1/2$ 得到 $y' = +1$，以概率 $1/2$ 得到 $y' = -1$，且结果与 $\mathbf{x}$ 无关；因此，任何假设都无法达到比 $1/2$显著更低的误差。

鉴于上述性质，如果某个算法能够将具有高斯边缘分布的 LTF 的无关学习误差降至 $R_{0-1}(\mathrm{LTF};\ D') + o(T)$，那么它就能区分上述两种情形并解决 LWE 问题。

本节正文将介绍我们的归约，并对上述内容作形式化表述。本节的主要定理如下所述并在下文证明，它确立了非知情学习 LTF 的一个自然判定版本的困难性。

定理 3.1（非知情学习高斯线性阈值函数的密码学困难性）。在假设 2.3 下，对于任意 $n \in \mathbb{N}$、任意常数 $\beta \in (0, 1)$、$\gamma \in \mathbb{R}_+$ 以及任意 $\log^\gamma n \le k \le cn$（其中 c 是一个充分小的常数），不存在一个运行时间为 $n^{O(k^\beta)}$ 的算法，能够区分联合分布 D 的以下两种情形；该分布定义在 $(\mathbf{x}\ y)$ 上，支撑集为 $\mathbb{R}^n \times \{\pm 1\}$，边缘分布为 $D_{\mathbf{x}} = D^{\mathcal{N}}_{\mathbb{R}^n,\,1}$，且具有 $n^{-O(k^\beta)}$ 优势：

(i) 备择假设：存在一个线性阈值函数，其 0-1 误差非平凡地小于 $1/2$，即 $R_{0-1}(\text{线性阈值函数};\ D) \le 1/2 - \Omega\left(1/\sqrt{k} \text{ 对数 } n\right)$。

(ii) 零假设：一个样本 $(\mathbf{x}\ y) \sim D$ 满足以下条件：$y = +1$ 以概率 $1/2$，且 $y = -1$ 以概率 $1/2$，后者独立于 $\mathbf{x}$。

证明。我们给出一种高效方法，其输入为来自分布 D' 的样本——该分布要么来自命题 2.4 中 $\mathrm{LWE}(n^{O(k^\beta)}, D^{\mathcal{N}}_{\mathbb{R}^n,1}, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}, \mathrm{mod}_T)$ 的备择假设，要么来自其零假设——并生成来自另一分布 D 的样本，该分布具有以下性质：若 D' 来自 LWE 问题的备择（分别为零）假设，那么所得分布 D 将满足该定理对无关先验的线性阈值函数学习判定问题所要求的备择（分别为零）假设条件。

归约过程如下：对于来自分布 D' 的样本 $(\mathbf{x}\ y')$，该分布是命题 2.4 中问题 $\mathrm{LWE}(n^{O(k^\beta)} D^{\mathcal{N}}_{\mathbb{R}^n,}$ $_1, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}, \mathrm{mod}_T)$ 的一个实例，我们只需输出 $(\mathbf{x}\ y) \sim D$，其中若 $y' \le T/2$ 则 $y = +1$，否则为 $y = -1$。我们证明 D 满足上述所述的要求。首先注意， D 的边缘分布 $D_{\mathbf{x}}$ 满足 $D_{\mathbf{x}} = D^{\mathcal{N}}_{\mathbb{R}^n,}$ $_1$，因此只需验证，对于各自的情形， R_{0-1}(线性阈值函数; D) ~~$= 1/2 -$~~ $\Omega\left(1/\sqrt{k}\right.$ 对数 $\left.n\right)$，并且 $y = +1$ 以概率 $1/2$ 独立于 $\mathbf{x}$。

对于备择假设情形，令 D' 来自 LWE 的备择假设情形。令 $\mathbf{s}$ 为 LWE 问题中的秘密向量。我们考虑以下两个线性阈值函数： $h_1(\mathbf{x}) = \mathrm{sign}(\langle \mathbf{s}, \mathbf{x}\rangle - T/6)$ 和 $h_2(\mathbf{x}) = \mathrm{sign}(-\langle \mathbf{s}, \mathbf{x}\rangle + T/3)$。如果我们能证明 $R_{0-1}(h_1; D) + R_{0-1}(h_2; D) \le 1 - \Omega(T)$，那么 $h = h_1$ 或 $h = h_2$ 满足 $R_{0-1}(h; D) \le 1/2 - \Omega(T)$，这意味着根据 T 的定义，R_{0-1}(线性阈值函数; ~~$D) \le$~~ $R_{0-1}(h; D) \le 1/2 - \Omega(1/\sqrt{k} \log n)$。

为了证明 $R_{0-1}(h_1; D) + R_{0-1}(h_2; D) \le 1 - \Omega(T)$，我们考察定义域中 h_1 与 h_2 一致的子集，即区域

$$\begin{aligned} B &\stackrel{\mathrm{def}}{=} \{\mathbf{t} \in \mathbb{R}^n \mid h_1(\mathbf{t}) = h_2(\mathbf{t})\} \\ &= \{\mathbf{t} \in \mathbb{R}^n \mid \langle \mathbf{s}, \mathbf{t}\rangle \in [T/6, T/3]\} \ . \end{aligned}$$

由于对于任意 $\mathbf{t} \in B$，总有 $h_1(\mathbf{t}) = h_2(\mathbf{t}) = +1$，因此我们可以写成

$$\begin{aligned} & R_{0-1}(h_1; D) + R_{0-1}(h_2; D) \\ =& \mathbf{Pr}_{(\mathbf{x},y)\sim D}[y \ne h_1(\mathbf{x})] + \mathbf{Pr}_{(\mathbf{x},y)\sim D}[y \ne h_2(\mathbf{x})] \\ =& \mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \notin B \wedge y \ne h_1(\mathbf{x})] \\ & + \mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \notin B \wedge y \ne h_2(\mathbf{x})] \\ & + 2\mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \in B \wedge y = -1] \ . \end{aligned}$$

由于对于任意 $\mathbf{x} \notin B$，我们有 $h_1(\mathbf{x}) \ne h_2(\mathbf{x})$，前两项之和为 $\mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \notin B]$。因此，我们有

$$\begin{aligned} & R_{0-1}(h_1; D)) + R_{0-1}(h_2; D) \\ =& \mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \notin B] + 2\mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \in B \wedge y = -1] \\ =& 1 + \mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \in B \wedge y = -1] \\ & - \mathbf{Pr}_{(\mathbf{x},y)\sim D}[\mathbf{x} \in B \wedge y = +1] \\ =& 1 - \mathbf{Pr}[\mathbf{x} \in B](1 - 2\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y = -1 \mid \mathbf{x} \in B]) \ . \end{aligned}$$

根据 $\mathbf{t}$ he B 和 $\mathbf{x} \sim D^{\mathcal{N}}_{\mathbb{R}^n,}$ $_1$的定义，我们有 $\mathbf{Pr}[\mathbf{x} \in B] = \Omega(T)$。因此，我们得到

$$R_{0-1}(h_1; D) + R_{0-1}(h_2; D) = 1 - \Omega(T)\left(1 - 2\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y = -1 \mid \mathbf{x} \in B]\right) \ . \tag{1}$$

如果我们能证明 $\mathbf{Pr}[y = -1 \mid \mathbf{x} \in B] \le 1/3$，那么结论成立，因为这意味着$R_{0-1}(h_1; D) + R_{0-1}(h_2; D) \le 1 - \Omega(T)$。

注意，根据 LWE 问题的备择情形分布的定义，我们有

$$y' = \mathrm{mod}_T(\langle \mathbf{s}, \mathbf{x}\rangle + z) \ ,$$

并且仅当 $y = -1$ $y' > T/2$ 时才成立，而这又仅当

$$\langle \mathbf{s}, \mathbf{x}\rangle + z > T/2 \text{ or } \langle \mathbf{s}, \mathbf{x}\rangle + z < 0 \ .$$

对于 $\mathbf{x} \in B$，我们有 $\langle \mathbf{s}, \mathbf{x} \rangle \in [T/6, T/3]$，因此仅当 $y = -1$ $|z| \geq T/6$ 时才成立。注意，$z \sim D^{\mathcal{N}}_{\mathbb{R},\sigma}$ 命题 2.4 指出，对于任意固定常数 $\kappa \in \mathbb{N}$ 和 $\sigma \geq k^{-\kappa}$，LWE 问题都是困难的。给定本定理中的常数 $\gamma \in \mathbb{R}^+$，我们取 $\kappa = \lceil 1/(2\gamma) + 1/2 + 1 \rceil$，它是一个固定常数。于是，根据命题 2.4，~~对于~~ $\sigma = k^{-\kappa} = 1/(k^{3/2}\sqrt{\text{对数 } n}) = o(T)$，LWE 问题是困难的。因此，我们有

$$\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y = -1 \mid \mathbf{x} \in B] \leq \mathbf{Pr}_{z \sim D^{\mathcal{N}}_{\mathbb{R},\sigma}}[|z| \geq T/6] = o(1) .$$

因此，将上述结果代入 (1)，我们可以得出

$$R_{0-1}(h_1; D) + R_{0-1}(h_2; D) = 1 - \Omega(T)\left(1 - 2\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y = -1 \mid \mathbf{x} \in B]\right) \leq 1 - \Omega(T) .$$

然后，如上所述，如果 $h = h_1$ 和 $h = h_2$ 都不满足 $R_{0-1}(h; D) \leq 1/2 - \Omega(T)$，那么 $R_{0-1}(h_1; D) + R_{0-1}(h_2; D) > 1 - \Omega(T)$，这就产生了矛盾。因此，$h = h_1$ 或 $h = h_2$ 满足 $R_{0-1}(h; D) \leq 1/2 - \Omega(T) \leq 1/2 - \Omega\left(1/\sqrt{k \text{ 对数 } n}\right)$。这就完成了备择假设情形的证明。

对于零假设情形，显然 $y = +1$ 以概率 $1/2$ 成立，且与 $\mathbf{x}$ 无关，因为在 LWE 问题的零假设情形下，$y' \sim U([0, T))$ 与 $\mathbf{x}$ 无关。至此，正确性证明完毕。

还需要验证无关学习线性阈值函数的时间下界和区分优势。根据命题 2.4，我们知道，在假设 2.3 下，对于问题 $\mathrm{LWE}(n^{O(k^\beta)}, D^{\mathcal{N}}_{\mathbb{R}^n,1}, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}, \mathrm{mod}_T)$，对于任意 $\sigma \geq k^{-\kappa}$（其中 $\kappa \in \mathbb{N}$ 是一个常数）以及 $T = 1/(c'\sqrt{k \text{ 对数 } n})$，其中 $c' > 0$ 是一个足够大的普适常数，该问题无法在 $n^{O(k^\beta)}$ 时间内以 $n^{-O(k^\beta)}$ 优势求解。因此，在相同假设下，不存在能以 $n^{-O(k^\beta)}$ 优势在 $n^{O(k^\beta)}$ 时间内求解不可知学习线性阈值函数问题判定版本（定理陈述中定义）的算法。

□

以下推论可由定理 3.1 直接得出。

推论 3.2. 在假设 2.3 下，对于任意常数 $\alpha \in (0, 2)$、$\gamma > 1/2$ 以及任意 $c/(\sqrt{n \text{ 对数 } n}) \leq \epsilon \leq 1/\text{对数}^{\gamma} n$（其中 c 是一个足够大的常数），不存在能够在具有高斯边缘分布的 $\mathbb{R}^n$ 上以加性误差 ϵ 无关学习线性阈值函数，并且运行时间为 $n^{O(1/(\epsilon\sqrt{\log n})^\alpha)}$ 的算法。

证明。我们在定理 3.1 中将参数 k 设为满足 $\epsilon = c/\sqrt{k \log n}$ 的值，其中 c 是一个充分小的常数。于是，任何以加性误差 ϵ 无知学习线性阈值函数的算法，都能以概率 $2/3$ 解决定理 3.1 中的检验问题。因此，任何此类算法的运行时间都不应为 $n^{O(k^\beta)}$，其中 $\beta \in (0, 1)$。由于 $\epsilon = c/\sqrt{k \log n}$，并且若我们取 $\beta = \alpha/2$，则时间下界可改写为 $n^{O(k^\beta)} = n^{O(1/(\epsilon\sqrt{\log n})^{2\beta})} = n^{O(1/(\epsilon\sqrt{\log n})^\alpha)}$。证明完毕。

□

4 高斯边缘分布下 ReLU 回归的计算困难性 s

本节将确立高斯边缘分布下 ReLU 回归的近乎最优计算困难性。值得指出的是，该困难性结果同样适用于任意 L-Lipschitz 激活函数 $f: \mathbb{R} \to \mathbb{R}$，对于 $L = O(1)$，存在 $t \in \mathbb{R}$，使得对于任意 $x \leq t$，$f(x)$ 为常数。粗略地说，我们的结果表明，在高斯边缘分布下，任何以误差 $\mathrm{OPT} + \epsilon$ 解决该问题的算法，都需要 $n^{\mathrm{poly}(1/(\epsilon \log^2 n))}$ 时间。

我们的思路是证明，第 3 节中的同一个困难实例可以被 ReLU 回归算法区分。本节的主要定理在下文陈述并证明，它确立了无关学习 ReLU 的一种自然判定版本的困难性。

定理 4.1。在假设 2.3 下，对于任意常数 $\beta \in (0,1)$、$\gamma \in \mathbb{R}_+$ 以及任意 $\log^\gamma n \le k \le cn$，其中 c 是一个足够小的常数，不存在一个运行时间为 $n^{O(k^\beta)}$的算法，能够区分以下两种情形：定义在 $(\mathbf{x}, y)$ 上、支撑集为 $\mathbb{R}^n \times \{\pm 1\}$ 且边缘分布为 $D_{\mathbf{x}} = D^{\mathcal{N}}_{\mathbb{R}^n,1}$ 的联合分布 D，并具有 $n^{-O(k^\beta)}$ 优势：

(i) 备择假设：存在一个 ReLU，其 L_2^2-误差非平凡地小于 1，即 $R_2(\mathrm{ReLU}; D) \le 1 - \Omega\left(1/(k \log n)^2\right)$。

(ii) 零假设：样本 $(\mathbf{x}, y) \sim D$ 满足以下条件：以 $1/2$的概率 $y = +1$，以 $1/2$ 的概率 $y = -1$，且与 $\mathbf{x}$ 无关。

证明。我们先给出以下中间引理。该引理大致说明，若存在一个与某个分布非平凡相关的 ReLU，那么必定还存在另一个具有非平凡 L_2^2-误差的 ReLU。

引理 4.2。设 $\epsilon \in (0,1)$，且 D 是 (u, y) 的一个联合分布，其支撑集包含于 $\mathbb{R} \times \{\pm 1\}$，满足边缘分布 $D_u = \mathcal{D}^{\mathcal{N}}_{\mathbb{R},1}$ 且 $\mathbf{E}_{(u,y)\sim D}[y] = 0$。假设存在一个形如 $f(u) = \mathrm{ReLU}(u - t)$ 的 ReLU，使得 $t \ge 0$ 且 $\left|\mathbf{E}_{(u,y)\sim D}[y f(u)]\right| \ge \epsilon$。那么存在 $k \in (-1, 1)$，使得 ReLU $g(u) = \mathrm{ReLU}(ku - kt)$ 满足 $\mathbf{E}_{(u,y)\sim D}[(y - g(u))^2] \le 1 - \epsilon^2$。

证明。我们首先注意到 $g(u) = k f(u)$，因此

$$\begin{aligned}
&\mathbf{E}_{(u,y)\sim D}[(y - g(u))^2] \\
=&\mathbf{E}_{(u,y)\sim D}[y^2] + \mathbf{E}_{(u,y)\sim D}[g(u)^2] - 2\mathbf{E}_{(u,y)\sim D}[y g(u)] \\
=&\mathbf{E}_{(u,y)\sim D}[y^2] + k^2 \mathbf{E}_{(u,y)\sim D}[f(u)^2] - 2k\mathbf{E}_{(u,y)\sim D}[y f(u)] \,.
\end{aligned}$$

由于 y 的支撑集为 $\{\pm 1\}$，我们有第一项满足 $\mathbf{E}_{(u,y)\sim D}[y^2] = 1$。

为了界定第二项，我们证明对于任意 u，$f(u)^2 \le u^2$。注意，对于 $u \ge t$，由于根据假设有$t \ge 0$，我们有 $f(u)^2 = (u - t)^2 \le u^2$。对于 $u < t$，我们有 $f(u)^2 = 0 \le u^2$。因此，结合 $u \sim \mathcal{D}^{\mathcal{N}}_{\mathbb{R},1}$ 这一事实，我们可以得出

$$\mathbf{E}_{(u,y)\sim D}[g(u)^2] = k^2 \mathbf{E}_{(u,y)\sim D}[f(u)^2] \le k^2 \mathbf{E}_{(u,y)\sim D}[u^2] = k^2 \,.$$

总而言之，我们得到

$$\mathbf{E}_{(u,y)\sim D}[(y - g(u))^2] \le 1 + k^2 - 2k\mathbf{E}_{(u,y)\sim D}[y f(u)] \,.$$

现在我们选择 k 的取值。如果 $\mathbf{E}_{(u,y)\sim D}[y f(u)] > 0$，那么取 $k = \epsilon$；否则取$k = -\epsilon$，此时我们始终有 $k \in (-1,1)$（因为 $\epsilon \in (0,1)$），并且

$$\begin{aligned}
\mathbf{E}_{(u,y)\sim D}[(y - h(u))^2] \le& 1 + \epsilon^2 - 2\epsilon|\mathbf{E}_{(u,y)\sim D}[y f(u)]| \\
\le& 1 - \epsilon^2 \,.
\end{aligned}$$

□

现在，我们使用命题 2.4 给出一个类似于定理 3.1 证明的归约。我们知道，在假设 2.3 下，以下结论成立：对于任意 $\sigma \ge k^{-\kappa}$ ($\kappa \in \mathbb{N}$ 为常数) 和 $T = 1/(c'\sqrt{k} \log n)$，问题 $\mathrm{LWE}(n^{O(k^\beta)} D^{\mathcal{N}}_{\mathbb{R}^n,1}, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}, \text{模 } T)$（其中 $c' > 0$ 是一个足够大的普适常数）无法在 $n^{O(k^\beta)}$ 时间内以 $n^{-O(k^\beta)}$ 优势求解。我们将给出一个从 LWE 问题到此处问题的高效归约。

对于来自分布 D' 的样本 $(\mathbf{x}, y')$，其中 D' 是 $\mathrm{LWE}(n^{O(k^\beta)} D^{\mathcal{N}}_{\mathbb{R}^n,1}, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}, \mathrm{mod}_T)$ 问题的一个实例，我们将直接输出 $(\mathbf{x}, y)$，使得：(i) $y = +1$ if $y' \le$

$T/2$，以及 (ii) $y=-1$ 否则，将其作为来自另一分布 D 的样本。我们论证，D 将满足以下性质：如果 D' 来自 LWE 问题的备择（分别为原假设）假设，那么所得分布 D 将满足定理 4.1 中 ReLU 回归判定问题的备择（分别为原假设）要求。

由于 $D_{\mathbf{x}}$ 的边缘分布 D 满足 $D_{\mathbf{x}}=D_{\mathbb{R}^n,1}^{\mathcal{N}}$，只需证明在备择假设情形下，有 $R_2(\mathrm{ReLU};D)=1-\Omega(1/(k\log n)^2)$；而在零假设情形下，有 $y=+1$，其概率为 $1/2$，且独立于 $\mathbf{x}$。

对于备择假设情形，我们首先引入以下引理。

引理 4.3. 对于任意 $\mathbf{s}\in\mathbb{S}^{n-1}$, $\sigma, T\in\mathbb{R}_+$，令 D 为 $(\mathbf{x},y)$ 的联合分布，其支撑集为 $\mathbb{R}^n\times\{\pm1\}$，且每个样本 $(\mathbf{x},y)$ 按以下方式生成。我们取 $\mathbf{x}\sim D_{\mathbb{R}^n,\sigma}^{\mathcal{N}}$，若 $|\langle\mathbf{s},\mathbf{x}\rangle+z|\le T/2$，令 $y=+1$，否则令 $y=-1$。给定 $\mathbf{x}$，则存在一个形式为 $h(\mathbf{x})=\mathrm{ReLU}(\langle\mathbf{s},\mathbf{x}\rangle-t)$ 的 ReLU，使得 $t\ge0$ 并且

$$\left|\mathbf{E}_{(\mathbf{x},y)\sim D}[yh(\mathbf{x})]\right|=\Omega(T^2)\,.$$

证明。令 $h_t(\mathbf{x})$ 定义为 $=\mathrm{ReLU}(\langle\mathbf{s},\mathbf{x}\rangle-t)$，并令 $r(t)=\mathrm{E}_{(\mathbf{x},y)\sim D}[yh_t(\mathbf{x})]$。那么我们只需证明存在一个 $t>0$，使得 $|r(t)|=\Omega(T)$。注意到 $r(t)$ 的导数为

$$\begin{aligned}r'(t)=&\frac{d\mathbf{E}_{(\mathbf{x},y)\sim D}[yh_t(\mathbf{x})]}{dt}\\=&\frac{d\mathbf{E}_{(\mathbf{x},y)\sim D}[y(\langle\mathbf{s},\mathbf{x}\rangle-t)\mathbf{1}(\langle\mathbf{s},\mathbf{x}\rangle>t)]}{dt}\\=&-\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y=+1\wedge\langle\mathbf{s},\mathbf{x}\rangle>t]+\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y=-1\wedge\langle\mathbf{s},\mathbf{x}\rangle>t]\,,\end{aligned}$$

$r(t)$ 的二阶导数为

$$\begin{aligned}r''(t)=&\frac{d(-\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y=+1\wedge\langle\mathbf{s},\mathbf{x}\rangle>t]+\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y=-1\wedge\langle\mathbf{s},\mathbf{x}\rangle>t])}{dt}\\=&P_{\langle\mathbf{s},\mathbf{x}\rangle}(t)(\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y=-1\mid\langle\mathbf{s},\mathbf{x}\rangle=t]-\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y=1\mid\langle\mathbf{s},\mathbf{x}\rangle=t])\\=&P_{\langle\mathbf{s},\mathbf{x}\rangle}(t)\left(2\mathbf{Pr}_{(\mathbf{x},y)\sim D}[y=-1\mid\langle\mathbf{s},\mathbf{x}\rangle=t]-1\right)\,.\end{aligned}$$

考虑区间 $t\in[T/6,T/3]$。注意，只有当 $\langle\mathbf{s},\mathbf{x}\rangle+z=t+z\notin[0,T/2]$ 时，$y=-1$ 才成立。因此，只有当 $|z|\ge T/6$ 时，$y=-1$ 才成立。注意 $z\sim D_{\mathbb{R},\sigma}^{\mathcal{N}}$ 以及 $\sigma=o(T)$。因此，对于 $t\in[T/6,T/3]$，我们有

$$\begin{aligned}r''(t)=&P_{\langle\mathbf{s},\mathbf{x}\rangle}(t)\left(2\mathbf{Pr}_{(\mathbf{x},y)\sim D'}[y=-1\mid\langle\mathbf{s},\mathbf{x}\rangle=t]-1\right)\\\le&P_{\langle\mathbf{s},\mathbf{x}\rangle}(t)\left(2\mathbf{Pr}_{z\sim D_{\mathbb{R},\sigma}^{\mathcal{N}}}[|z|\ge T/6]-1\right)\\=&-\Omega(1)\,,\end{aligned}$$

其中最后一个等式由 $\sigma=o(T)$ 和 $P_{\langle\mathbf{s},\mathbf{x}\rangle}(t)=\Omega(1)$ 得出，因为 $\langle\mathbf{s},\mathbf{x}\rangle\sim D_{\mathbb{R},1}^{\mathcal{N}}$ 且 $t\in[T/6,T/3]$ 对于 $T<1$ 成立。

接下来我们证明，要么 $r(T/3)-r(T/4)=\Omega(T^2)$，要么 $r(T/6)-r(T/4)=\Omega(T^2)$。首先注意，要么 $r'(T/4)\le0$，要么 $r'(T/4)>0$。如果 $r'(T/4)\le0$，那么

$$\begin{aligned}r(T/3)-r(T/4)=&r'(T/4)(T/12)+\int_{T/4}^{T/3}r''(t)(T/3-t)dt\\\le&\int_{T/4}^{T/3}r''(t)(T/3-t)dt=-\Omega(T^2)\,.\end{aligned}$$

如果 $r'(T/4) > 0$，那么

$$r(T/6) - r(T/4) = r'(T/4)(-T/12) + \int_{T/4}^{T/6} r''(t)(T/6 - t)dt$$
$$\leq \int_{T/4}^{T/6} r''(t)(T/6 - t)dt = -\Omega(T^2) .$$

由于 $r(T/4) - r(T/3) = \Omega(T^2)$ 或 $r(T/4) - r(T/6) = \Omega(T^2)$，因此 $|r(T/6)|$、$|r(T/4)|$、$|r(T/3)|$ 中至少有一个为 $\Omega(T^2)$。证明至此完成。 □

这里我们将在 $(\mathbf{x}, y)$ 的联合分布上应用引理 4.3。回顾一下，命题 2.4 指出，对于任意固定常数 $\kappa \in \mathbb{N}$ 和 $\sigma \geq k^{-\kappa}$，LWE 问题都是困难的。给定本定理中的常数 $\gamma \in \mathbb{R}^+$，我们取 $\kappa = \lceil 1/(2\gamma) + 1/2 + 1 \rceil$，它是一个固定常数。于是，由命题 2.4 可知，对于 $\sigma = k^{-\kappa} = 1/(k^{3/2}\sqrt{\log n}) = o(T)$，LWE 问题是困难的。因此，根据引理 4.3，存在一个形式为 $h(\mathbf{x}) = f(\langle \mathbf{s}, \mathbf{x} \rangle) = \mathrm{ReLU}(\langle \mathbf{s}, \mathbf{x} \rangle - t)$ 的 ReLU，使得 $t \geq 0$ 和 $\left|\mathbf{E}_{(\mathbf{x},y)\sim D}[yh(\mathbf{x})]\right| = \left|\mathbf{E}_{(\mathbf{x},y)\sim D}[yf(\langle \mathbf{x}, \mathbf{s} \rangle)]\right| = \Omega(T^2) = \Omega(1/(k \log n))$。若我们将引理 4.2 应用于 $(\langle \mathbf{x}, \mathbf{s} \rangle, y)$ 的联合分布和 ReLU 函数 f，就可得存在一个形式为 $h'(\mathbf{x}) = kf(\langle \mathbf{x}, \mathbf{s} \rangle) = \mathrm{ReLU}(\langle k\mathbf{s}, \mathbf{x} \rangle - kt)$ 的 ReLU，使得 $k < 1$ 并且

$$\mathbf{E}_{(\mathbf{x},y)\sim D}[(y - h'(\mathbf{x}))^2] \leq 1 - \Omega(1/(k \log n)^2) .$$

由于 $k < 1$，我们有 $\|k\mathbf{s}\|_2 \leq \|\mathbf{s}\|_2 = 1$，因此 $h' \in \mathrm{ReLU}$。这意味着 t

$$R_2(\mathrm{ReLU}; D) \leq 1 - \Omega\left(1/(k \log n)^2\right) .$$

对于零假设情形，显然，$y = +1$ 以概率 $1/2$，且 $y = -1$ 以概率 $1/2$，并且与 $\mathbf{x}$ 无关，因为在 LWE 问题的零假设情形下，$y' \sim U([0, T))$ 与 $\mathbf{x}$ 无关。证明完毕。

□

以下推论可直接由定理 4.1 得出。

推论 4.4。在假设 2.3 下，对于任意常数 $\alpha \in (0, 1/2)$、$\gamma > 2$，以及任意 $c/(n \text{对数 } n)^2 \leq \epsilon \leq 1/\text{对数}^{\gamma} n$，其中 c 是一个足够大的常数，不存在一种算法能够在$\mathbb{R}^n$ 上针对高斯边缘分布进行 ReLU 回归，使误差达到 $R_2(\mathrm{ReLU}; D) + \epsilon$，且运行时间为 $n^{O(1/(\epsilon \log^2 n)^{\alpha})}$。

证明。我们在定理 4.1 中选择参数 k，使其取值满足 $\epsilon = c/(k \log n)^2,$，其中 c 是一个充分小的常数。于是，任何能够将 ReLU 无知学习到加性误差ϵ的算法，都能以概率 $2/3$解决定理 4.1 中的检验问题。因此，对于任意 $\beta \in (0, 1)$，此类算法的运行时间都不应为 $n^{O(k^{\beta})}$。由于 $\epsilon = c/(k \log n)^2,$，并且如果我们选择 $\beta = 2\alpha$，那么时间下界可以改写为 $n^{O(k^{\beta})} = n^{O(1/(\epsilon \text{对数}^2 n)^{\beta/2})} = n^{O(1/(\epsilon \log^2 n)^{\alpha})}$。证明完毕。 □

参考文献

[ABL17] P. Awasthi、M. F. Balcan 和 P. M. Long。利用局部化高效学习带噪声线性分隔器的能力。《ACM期刊》，63(6):50:1–50:27，2017年。[ALNS20] D. Aggarwal、J. Li、P. Q. Nguyen 和 N. Stephens-Davidowitz。滑动约化再探——填补 SVP近似中的空白。载于《密码学进展》——CRYPTO 2020——第40届国际密码学年会，CRYPTO 2020，《计算机科学讲义》第12171卷，第274–295页。Springer，2020年。

[ATV22] P. Awasthi、A. Tang 和 A. Vijayaraghavan。使用梯度下降对一般 ReLU 激活进行不可知学习。CoRR，abs/2208.02711，2022。[BRST21] J. Bruna、O. Regev、M. J. Song 和 Y. Tang。连续 LWE。载于 STOC '21：第53届ACM SIGACT计算理论研讨会，第694–707页。ACM，2021年。[Dan15] A. Daniely。不可知学习半空间的 PTAS。载于第28届学习理论会议论文集，COLT 2015，第484–502页，2015年。[Dan16] A. Daniely。学习半空间的复杂性理论限制。载于第48届计算理论研讨会年会论文集，STOC 2016，第105–117页，2016年。[DFT+15] D. Dachman-Soled、V. Feldman、L.Y. Tan、A. Wan 和 K. Wimmer。近似弹性、单调性与不可知学习的复杂性。载于第二十六届ACM-SIAM离散算法研讨会论文集，SODA 2015，第498–511页。SIAM，2015年。[DGJ+10] I. Diakonikolas、P. Gopalan、R. Jaiswal、R. Servedio 和 E. Viola。有界独立性能够欺骗半空间。《SIAM计算杂志》，39(8):3441–3462，2010年。[DGK+20] I. Diakonikolas、S. Goel、S. Karmalkar、A. R. Klivans 和 M. Soltanolkotabi。ReLU 回归的近似方案。载于学习理论会议，COLT，《机器学习研究论文集》第125卷，第1452–1485页。PMLR，2020年。[DK22] I. Diakonikolas 和 D. Kane。带 Massart噪声学习半空间的近最优统计查询困难性。载于学习理论会议，《机器学习研究论文集》第178卷，第4258–4282页。PMLR，2022。[DKK+21] I. Diakonikolas、D. M. Kane、V. Kontonis、C. Tzamos 和 N. Zarifis。在高斯边缘分布下对半空间进行不可知适当学习。载于学习理论会议，COLT 2021，《机器学习研究论文集》第134卷，第1522–1551页。PMLR，2021年。[DKMR22a] I. Diakonikolas、D. Kane、P. Manurangsi 和 L. Ren。带对抗性标签噪声学习单个神经元的难度。载于人工智能与统计国际会议，AISTATS 2022，《机器学习研究论文集》第151卷，第8199–8213页。PMLR，2022。[DKMR22b] I. Diakonikolas、D. M. Kane、P. Manurangsi 和 L. Ren。带 Massart噪声学习半空间的密码学困难性。CoRR，abs/2207.14266，2022。会议版本见 NeurIPS'22。[DKN10] I. Diakonikolas、D. M. Kane 和 J. Nelson。有界独立性能够欺骗二次阈值函数。载于第51届IEEE计算机科学基础研讨会年会，FOCS 2010，第11–20页。IEEE 计算机学会，2010年。[DKPZ21] I. Diakonikolas、D. M. Kane、T. Pittas 和 N. Zarifis。在统计查询模型下，高斯边缘分布中的不可知学习采用多项式回归的最优性。载于学习理论会议，COLT 2021，《机器学习研究论文集》第134卷，第1552–1584页。PMLR，2021年。

[DKS18] I. Diakonikolas、D.M. Kane 和 A. Stewart。《在恶意噪声下学习几何概念》。载于《第50届 ACM SIGACT 计算理论年会论文集》，STOC 2018，第1061–1073页。ACM，2018。[DKTZ22] I. Diakonikolas、V. Kontonis、C. Tzamos 和 N. Zarifis。《通过梯度下降在对抗性标签噪声下学习单个神经元》。载于《学习理论会议》，《机器学习研究论文集》第178卷，第4313–4361页。PMLR，2022。[DKZ20] I. Diakonikolas、D. Kane 和 N. Zarifis。《高斯边缘分布下无关学习半空间和 ReLU 的近最优 SQ 下界》。载于《神经信息处理系统进展 33：2020 年神经信息处理系统年会》，NeurIPS 2020，2020。[FCG20] S. Frei、Y. Cao 和 Q. Gu。《通过梯度下降不可知地学习单个神经元》。载于《神经信息处理系统进展 33：2020 年神经信息处理系统年会》，NeurIPS 2020，2020。[FGKP06] V. Feldman、P. Gopalan、S. Khot 和 A. Ponnuswami。《学习含噪奇偶函数和半空间的新结果》。载于 FOCS 论文集，第563–576页，2006。[FS97] Y. Freund 和 R. Schapire。《在线学习的决策理论推广及其在提升中的应用》。《计算机与系统科学杂志》，55(1)：119–139，1997。[GGK20] S. Goel、A. Gollakota 和 A. R. Klivans。《通过函数梯度得到的统计查询下界》。载于《神经信息处理系统进展 33：2020 年神经信息处理系统年会》，NeurIPS 2020，2020。[GKK19] S. Goel、S. Karmalkar 和 A. R. Klivans。《高斯边缘分布下学习 ReLU 的时间与准确率权衡》。载于《神经信息处理系统进展 32：2019 年神经信息处理系统年会》，NeurIPS 2019，2019。[GKKT17] S. Goel、V. Kanade、A. R. Klivans 和 J. Thaler。《在多项式时间内可靠地学习 ReLU》。载于《第30届学习理论会议论文集》，COLT 2017，第1004–1042页，2017。[GR06] V. Guruswami 和 P. Raghavendra。《带噪声学习半空间的困难性》。载于《第47届 IEEE 计算机科学基础研讨会（FOCS）论文集》，第543–552页。IEEE计算机学会，2006。[GVV22] A. Gupte、N. Vafa 和 V. Vaikuntanathan。《连续 LWE 与 LWE 一样困难，以及在学习高斯混合中的应用》。arXiv 预印本 arXiv:2204.02550，2022。会议版本发表于 FOCS'22。[Hau92] D. Haussler。《神经网络及其他学习应用的 PAC 模型的决策理论推广》。《信息与计算》，100：78–150，1992。[Kea98] M. J. Kearns。《通过统计查询实现高效抗噪学习》。《ACM 杂志》，45(6)：983–1006，1998。

[KK14] A. R. Klivans 和 P. Kothari。《将困难学习问题嵌入高斯空间》。载于《近似、随机化与组合优化：算法与技术》，APPROX/RANDOM 2014，第793–809页，2014。[KKMS08] A. Kalai、A. Klivans、Y. Mansour 和 R. Servedio。《不可知学习半空间》。《SIAM计算杂志》，37(6)：1777–1805，2008。[KSS94] M. Kearns、R. Schapire 和 L. Sellie。《迈向高效无关学习》。《机器学习》，17(2/3)：115–141，1994。[Mic18] D. Micciancio。《二进制秘密下带误差学习的困难性》。《理论计算》，14(1)：1–17，2018。[MR18] P. Manurangsi 和 D. Reichman。《训练 ReLU 的计算复杂性》。arXiv 预印本 arXiv:1810.04207，2018。[MT94] W. Maass 和 G. Turan。《阈值门学习速度有多快？》载于 S. Hanson、G. Drastal 和 R. Rivest 编，《计算学习理论与自然学习系统》，第381–414页。麻省理工学院出版社，1994。[Pei09] C. Peikert。《基于最坏情况最短向量问题的公钥密码系统：扩展摘要》。载于《第41届 ACM 年度计算理论研讨会论文集》，STOC 2009，2009，第333–342页。ACM，2009。[Reg05] O. Regev。《关于格、带误差学习、随机线性码与密码学》。载于《第37届 ACM 年度计算理论研讨会论文集》（STOC）。ACM出版社，2005。[Ros58] F. Rosenblatt。《感知器：一种用于信息存储及其在大脑中组织的概率模型》。《心理学评论》，65：386–407，1958。[Sol17] M. Soltanolkotabi。《通过梯度下降学习 ReLU》。载于《神经信息处理系统进展》，第2007–2017页，2017。[Tie22] S. Tiegel。《基于最坏情况格问题的无关学习半空间的困难性》。CoRR，abs/2207.14030，2022。[Val84] L. Valiant。《可学习性理论》。《ACM通讯》，27(11)：1134–1142，1984。[Vap98] V. Vapnik。《统计学习理论》。Wiley-Interscience，纽约，1998。

附录

A 其他技术背景

对于 $n, k \in \mathbb{N}$ 满足 $k \le n$ 的情况，我们用 $S_{n,k}$ 表示 k-稀疏集合 $S_{n,k}$ 定义= $\{\mathbf{x} \in \{0, \pm 1\}^n : \|\mathbf{x}\|_1 = k\}$。我们用 negl($\lambda$) 表示 $\lambda^{-\omega(1)}$。

离散高斯分布的定义在此处也很有用。从本质上说，离散高斯是一种单变量离散分布，其支撑集是 $\mathbb{R}$ 上等间距的点，并且其支撑集内任意一点的概率质量都与高斯分布在该点的概率密度成正比。根据定义 2.1，离散高斯分布可写为如下形式。

定义 A.1（离散高斯分布）。对于 $T \in \mathbb{R}_+, y \in \mathbb{R}$ 和 $\sigma \in \mathbb{R}_+$，我们将“尺度为 σ 的 T-间隔、y-偏移离散高斯分布”定义为 $D^{\mathcal{N}}_{T\mathbb{Z}+y,\sigma}$ 的分布。

在整个证明过程中，我们需要对模 1 的高斯分布以及添加了噪声的高斯分布进行处理。因此，以下定义将很有用。

定义 A.2（由 $\mathbb{R}^n_1$ 得到的扩展高斯分布）。对于 $\sigma \in \mathbb{R}_+$，令 $D^{\text{expand}}_{\mathbb{R}^n_1\,\sigma}$，表示按以下方式抽取的 $\mathbf{x}'$ 的分布：首先抽取 $\mathbf{x} \sim U(\mathbb{R}^n_1)$，然后抽取 $\mathbf{x}' \sim D^{\mathcal{N}}_{\mathbb{Z}^n+\mathbf{x}\,\sigma}$。

定义 A.3（ $\mathbb{R}^n_1$ 上的折叠高斯分布）。对于 $\sigma \in \mathbb{R}_+$，我们用 $D^{\text{collapse}}_{\mathbb{R}^n_1\,\sigma}$，表示 $\text{mod}_1(\mathbf{x})$ 在 $\mathbb{R}^n_1$ 上的分布，其中 $\mathbf{x} \sim D^{\mathcal{N}}_{\mathbb{R}^n,\ \sigma}$。

B 小范数秘密的连续 LWE 困难性

这里给出命题 2.4 的证明，这是我们困难性归约的第一步。具体而言，我们将假设 2.3 中的标准离散 LWE 问题——其中 D_{sample} 的支撑集是离散集合 $\mathbb{Z}^n_q$ ——归约为连续 LWE（cLWE）问题——其中 D_{sample} 的支撑集为 $\mathbb{R}^n$。这种 cLWE 问题最早由 [BRST21]，引入，该论文给出了从逼近最短向量问题（GapSVP）的判定版本到 cLWE 的量子归约。随后， [GVV22] 给出了从经典 LWE 问题到 cLWE 问题的经典归约，表明 cLWE 问题至少与 LWE 问题一样难。

值得注意的是，我们在此不会直接使用连续 LWE 的困难性陈述，而是将标准离散 LWE 归约到连续 LWE。这样归约的优势在于，我们可以从一个稀疏离散 LWE 实例出发，其中秘密向量 $\mathbf{s}$ 从 $S_{n,k}$ 中均匀采样；归约后得到一个维数为 n 、且秘密向量的 ℓ_2 范数约为 $\sqrt{k}$ ($\sqrt{k} \approx$ 对数$^{0.01}\, n$ 的连续 LWE 实例。相比之下， [BRST21] 中的秘密向量具有 $\sqrt{n}$ ℓ_2 范数。

为此，我们对 [GVV22] 中的一个思路稍作修改，以消除秘密向量 ℓ_2 范数中由对数 m （其中 m 是样本数量）导致的膨胀。

为证明该命题，我们先给出以下引理，它将标准 LWE 归约为秘密向量具有 k-稀疏性的 LWE（即，秘密向量 $\mathbf{s} \in S_{n,k}$）。

引理 B.1（ [GVV22]中的推论4）。对于任意 $n, m, q, l, \lambda, k \in \mathbb{N}, \sigma \in \mathbb{R}_+$，假设 $\log(q)/2^l = \text{negl}(\lambda)$，$\sigma \ge 4\sqrt{\omega(\log\ \lambda) + \ln\ n + \ln\ m}$ 且 $k \log_2(n/k) \ge (l+1)\log_2(q) + \omega(\log\ \lambda)$。那么，如果测试问题 $\text{LWE}(n, \mathbb{Z}^l_q, \mathbb{Z}^l_q, D^{\mathcal{N}}_{\mathbb{Z}}, \sigma, \text{mod}q)$不存在具有优势 ϵ的 $T + \text{poly}(n, m, q, \lambda)$ 时间区分器，则问题 $\text{LWE}(m, \mathbb{Z}^n_q, S_{n,k}, D^{\mathcal{N}}_{\mathbb{Z}}, \sigma', \text{mod}q)$不存在具有优势$2\epsilon m + \text{negl}(\lambda)$的 T 时间区分器，其中 $\sigma' = 2\sigma\sqrt{k+1}$。

上述引理将 LWE($n, \mathbb{Z}_q^l, \mathbb{Z}_q^l, D_{\mathbb{Z}}^{\mathcal{N}}, \sigma$, mod$q$) 归约为 LWE($m, \mathbb{Z}_q^n, S_{n,k}, D_{\mathbb{Z}}^{\mathcal{N}}, \sigma'$, mod$q$)。这里的 λ 充当安全参数。注意，原问题 LWE($n, \mathbb{Z}_q^l, \mathbb{Z}_q^l, D_{\mathbb{Z}}^{\mathcal{N}}, \sigma$, mod$q$) 有 $2^l \log q$ 种可能的秘密向量选择，而新问题 LWE($m, \mathbb{Z}_q^n, S_{n,k}, D_{\mathbb{Z}}^{\mathcal{N}}, \sigma'$, mod$q$) 大约至少有 $2^{k \log_2(n/k)}$ 种可能的秘密向量选择。从秘密向量熵的角度来看，这直观地解释了为何引理中要求 $k \log_2(n/k) \geq (l+1)\log_2(q) + \omega(\log \lambda)$。

接着，我们加入少量额外的高斯噪声，将噪声分布从离散高斯分布 $D_{\mathbb{Z}, \sigma}^{\mathcal{N}}$ 调整为连续高斯分布 $D_{\mathbb{R}, \sigma'}^{\mathcal{N}}$，其中 σ' 将略大于 σ。由此得到以下引理：

引理 B.2（[GVV22] 中的引理 15）。令 $n, m, q, \lambda \in \mathbb{N}$, $\sigma \in \mathbb{R}_+$, $\epsilon \in (0, 1]$ 并假设 $\sigma > \sqrt{4 \ln m + \omega(\log \lambda)}$。对于任意 $S \subseteq \mathbb{R}^n$，假设不存在针对 LWE($m, \mathbb{Z}_q^n, S, D_{\mathbb{Z}, \sigma}^{\mathcal{N}}$, mod$q$) 问题的 $T + \text{poly}(m, n, \log(q), \log(\sigma))$ 时间区分器，其优势为 ϵ。那么，不存在针对 LWE($m, \mathbb{Z}_q^n, S, D_{\mathbb{R}, \sigma'}^{\mathcal{N}}$, mod$q$) 问题的 T 时间区分器，其优势为 $\epsilon + \text{negl}(\lambda)$，其中我们设

$$\sigma' = \sqrt{\sigma^2 + 4 \ln(m) + \omega(\log \lambda)} = O(\sigma) .$$

我们首先注意到，引理 B.2 中对参数的两个要求 $\sigma > \sqrt{4 \ln m + \omega(\log \lambda)}$ 和 $\sigma' = \sqrt{\sigma^2 + 4 \ln(m) + \omega(\log \lambda)}$ 蕴含 $\sigma' = \sqrt{\sigma^2 + 4 \ln(m) + \omega(\log \lambda)} = O(\sigma)$。这表明，我们对噪声尺度的放大至多为一个普适常数倍。应用此引理后，我们再次加入少量额外的高斯噪声，将样本分布 D_{sample} 从 $U(\mathbb{Z}_q^n)$ 调整为 $U(\mathbb{R}_q^n)$。由此得到以下结果：

引理 B.3（[GVV22] 中的引理 16）。设 $n, m, q, \lambda \in \mathbb{N}$, $\sigma, r \in \mathbb{R}_+$ 和 $\epsilon \in (0, 1]$。令 $S \subseteq \mathbb{R}^n$，其中支撑集中的所有元素都具有固定的 ℓ_2-范数 r，并假设 $\sigma \geq 3r\sqrt{\ln n + \ln m + \omega(\log \lambda)}$。假设不存在运行时间为 $T + \text{poly}(m, n, \log(q), \log(\sigma))$ 的区分器，能够以优势 ϵ 区分 LWE($m, \mathbb{Z}_q^n, S, D_{\mathbb{R}, \sigma}^{\mathcal{N}}$, mod$q$)；那么也不存在运行时间为 T 的区分器，能够以优势 $\epsilon + \text{negl}(\lambda)$ 区分 LWE 问题 LWE($m, \mathbb{R}_q^n, S, D_{\mathbb{R}, \sigma'}^{\mathcal{N}}$, mod$q$)，其中我们设

$$\sigma' = \sqrt{\sigma^2 + 9r^2(\ln n + \ln m + \omega(\log \lambda))} = O(\sigma) .$$

类似地，陈述 $\sigma \geq 3r\sqrt{\ln n + \ln m + \omega(\text{对数 } \lambda)}$ 和 $\sigma' = \sqrt{\sigma^2 + 9r^2(\ln n + \ln m + \omega(\text{对数 } \lambda))}$ 表明 $\sigma' = O(\sigma)$。因此，为了使样本连续，我们再次将噪声尺度最多放大一个常数倍。接下来，我们给出引理 18 的一个修改版本，见[GVV22]。我们首先介绍以下来自 [DKMR22b] 的事实。

事实 B.4（[DKMR22b] 中的事实 A.4）。设 $n \in \mathbb{N}, \sigma \in \mathbb{R}_+, \epsilon \in (0, 1/3)$ 满足 $\sigma \geq \sqrt{\ln(2n(1 + 1/\epsilon))/\pi}$。那么，我们有

$$\frac{P_{D_{\mathbb{R}_1^n, \sigma}^{\text{expand}}/\sigma}(\mathbf{t})}{P_{D_{\mathbb{R}^n, 1}^{\mathcal{N}}}(\mathbf{t})} = \frac{P_{U(\mathbb{R}_1^n)}(\text{mod}_1(\sigma \mathbf{t}))}{P_{D_{\mathbb{R}_1^n, \sigma}^{\text{collapse}}}(\text{mod}_1(\sigma \mathbf{t}))} = 1 \pm O(\epsilon) ,$$

对所有 $\mathbf{t} \in \mathbb{R}^n$，并且

$$d_{\text{TV}}\left(\frac{D_{\mathbb{R}_1^n, \sigma}^{\text{expand}}}{\sigma}, D_{\mathbb{R}^n, 1}^{\mathcal{N}}\right), d_{\text{TV}}\left(D_{\mathbb{R}_1^n, \sigma}^{\text{collapse}}, U(\mathbb{R}_1^n)\right) = \exp\left(-\Omega(\sigma^2)\right) .$$

本质上，事实 B.4 表明，给定 $\mathbf{x} \sim D_{\mathbb{R}^n, \sigma}^{\mathcal{N}}$，当 σ 足够大时，$\text{mod}_1(\mathbf{x})$ 的分布在逐点意义下（就其概率密度函数而言）接近于 $U(\mathbb{R}_1^n)$。因此，如果我们考虑这一过程的逆过程

给定 $\mathbf{v} \sim U(\mathbb{R}_1^n)$，我们对 $\mathbf{u} \sim D^{\mathcal{N}}_{\mathbb{R}^n + \mathbf{v}\,\sigma}$ 进行采样，则 $\mathbf{u}$ 的分布将充分接近于 $D^{\mathcal{N}}_{\mathbb{R}^n,\ \sigma}$。我们可以利用这一事实，将 LWE 问题中的样本分布从 $U(\mathbb{R}_q^n)$ 改为 $D^{\mathcal{N}}_{\mathbb{R}^n,\ 1}$，因为经过重新缩放后，$U(\mathbb{R}_q^n)$ 基本上就是 $U(\mathbb{R}_1^n)$。这里的区别在于，原始的引理 18 取了一个较大的 σ，使得 $d_{\mathrm{TV}}\left(\frac{D^{\mathrm{expand}}_{\mathbb{R}^n_1\,\sigma}}{\sigma},\ D^{\mathcal{N}}_{\mathbb{R}^n,\ 1}\right) \approx 1/m$，因此 m 个样本无法察觉其中的差异。然而，由于这两个分布实际上在逐点意义下非常接近，我们可以改为取一个较小的 σ，并对 $\mathbf{u}$ 额外执行一步拒绝采样，使其分布完全成为高斯分布。这样，我们便能为具有高斯边缘分布的不可知学习线性阈值函数给出近乎最优的下界。现在，我们给出 [GVV22] 中引理 18 的修改版本。

引理 B.5（ [GVV22] 中引理 18 的修改版）。设 $n, m, q \in \mathbb{N}, \sigma, r, \alpha \in \mathbb{R}_+$。设 $S \subseteq \mathbb{Z}^n$，其中支撑集中的所有元素都具有固定的 ℓ_2-范数 r。假设不存在运行时间为 $T +$ 多项式$(n, m,$对数$(q))$ 的区分器，能够以 ϵ 优势区分 LWE$(m, \mathbb{R}_q^n, S, D^{\mathcal{N}}_{\mathbb{R},\sigma},$ 模 $q)$ 问题。那么，不存在运行时间为 T- 的区分器，能够以 $\epsilon + 2^{-\Omega(m)}$ 优势区分 LWE$(m', D^{\mathcal{N}}_{\mathbb{R}^n,1}, S/r, D^{\mathcal{N}}_{\mathbb{R},\alpha\sigma/q},$ 模 $\alpha)$ 问题，其中

$$\alpha = c/\left(r\sqrt{\log n}\right)\ ,$$

$$m' = cm\ ,$$

且 $c > 0$ 是一个足够小的普适常数。

证明。我们将给出一个归约论证。给定来自 LWE$(m, \mathbb{R}_q^n, S, D^{\mathcal{N}}_{\mathbb{R},\sigma},$ 模 $q)$ 的样本 $(\mathbf{x}\ y)$，我们可以按以下方式生成来自 LWE$(m', D^{\mathcal{N}}_{\mathbb{R}^n,1}, S/r, D^{\mathcal{N}}_{\mathbb{R},\alpha\sigma/q},$ 模 $\alpha)$ 问题的样本 $(\mathbf{x}', y')$，且成功概率至少为一个常数。

我们取一个 $\tilde{\sigma} = 1/r\alpha$ 并取样本 $\tilde{\mathbf{x}} \sim D^{\mathcal{N}}_{\mathbb{Z}^n + \mathbf{x}/q,\ \tilde{\sigma}}/\tilde{\sigma}$。我们定义函数 $f\colon\ \mathbb{R}^n \to \mathbb{R}$ as

$$f(\mathbf{t}) \stackrel{\mathrm{def}}{=} \frac{P_{D^{\mathcal{N}}_{\mathbb{R}^n,1}}(\mathbf{t})}{P_{D^{\mathrm{expand}}_{\mathbb{R}^n_1,\tilde{\sigma}}/\tilde{\sigma}}(\mathbf{t})}\ .$$

以概率 $f(\tilde{\mathbf{x}})/\max_{\mathbf{t}\in\mathbb{R}^n} f(\mathbf{t})$，我们取 $\mathbf{x}' = \tilde{\mathbf{x}}$ 和 $y' = y/(qr\tilde{\sigma})$，并将 $(\mathbf{x}', y')$ 作为学习带误差$(m', D^{\mathcal{N}}_{\mathbb{R}^n,\ 1}, S/r, D^{\mathcal{N}}_{\mathbb{R},\ \alpha\sigma/q},$ mod$\alpha)$ 的样本输出。否则，我们输出失败。

我们将证明，若 $(\mathbf{x}\ y)$ 来自备择假设情形，则必有 $\mathbf{x}' \sim D^{\mathcal{N}}_{\mathbb{R}^n,\ 1}$ 且 $y' =$ mod$_\alpha(\langle \mathbf{s}', \mathbf{x}'\rangle + z')$，其中 $\mathbf{s}' \sim U(S/q)$ 且 $z' \sim D^{\mathcal{N}}_{\mathbb{R},\ \alpha\sigma/q}$。由于 $(\mathbf{x}\ y)$ 来自备择假设情形，它必满足 $\mathbf{x} \sim U(\mathbb{R}_q^n)$ 且 $y =$ mod$_q(\langle \mathbf{s}, \mathbf{x}\rangle + z)$，其中 $\mathbf{s} \sim U(S)$ 且 $z \sim D^{\mathcal{N}}_{\mathbb{R},\ \sigma}$。接着，事实 $\tilde{\mathbf{x}} \sim D^{\mathcal{N}}_{\mathbb{Z}^n + \mathbf{x}/q,\ \tilde{\sigma}}/\tilde{\sigma}$ 意味着 $\tilde{\sigma}(\tilde{\mathbf{x}} - \mathbf{x}/q) \in \mathbb{Z}^n$ 且 $q\tilde{\sigma}(\tilde{\mathbf{x}} - \mathbf{x}) \in q\mathbb{Z}^n$；结合 $\mathbf{s} \in \mathbb{Z}^n$，我们得到

$$\mathrm{mod}_q(\langle \mathbf{s}, \mathbf{x}\rangle) = \mathrm{mod}_q(\langle \mathbf{s}, q\tilde{\sigma}\tilde{\mathbf{x}}\rangle + \langle \mathbf{s}, \mathbf{x} - q\tilde{\sigma}\tilde{\mathbf{x}}\rangle) = \mathrm{mod}_q(\langle \mathbf{s}, q\tilde{\sigma}\tilde{\mathbf{x}}\rangle)\ .$$

于是我们可以写成

$$\begin{aligned}
y' =& y/(qr\tilde{\sigma}) \\
=& \mathrm{mod}_q(\langle \mathbf{s}, \mathbf{x}\rangle + z)/(qr\tilde{\sigma}) \\
=& \mathrm{mod}_q(\langle \mathbf{s}, q\tilde{\sigma}\tilde{\mathbf{x}}\rangle + z)/(qr\tilde{\sigma}) \\
=& \mathrm{mod}_1(\langle \mathbf{s}, \tilde{\sigma}\tilde{\mathbf{x}}\rangle + z/q)/(r\tilde{\sigma}) \\
=& \mathrm{mod}_{1/(r\tilde{\sigma})}\left(\langle \mathbf{s}/r, \tilde{\mathbf{x}}\rangle + z/(qr\tilde{\sigma})\right) \\
=& \mathrm{mod}_\alpha\left(\langle \mathbf{s}/r, \mathbf{x}'\rangle + \alpha z/q\right)\ ,
\end{aligned}$$

其中，最后一个等式由事实 $\tilde{\sigma} = 1/(r\alpha)$ 推出。注意，上述表达式中的三个项 $\mathbf{s}/r$、$\mathbf{x}'$ 和 $\alpha z/q$ 彼此独立（因为 $\mathbf{x}'$、$\mathbf{s}$ z 相互独立）。现在只需验证它们各自的分布。

显然，$\mathbf{s}/r \sim U(S/r)$。对于另外两项，我们首先定义如下记号。对于函数 $f, g\colon U \to \mathbb{R}$，若存在常数 $c \in \mathbb{R} \setminus \{0\}$，使得对所有$u \in U$，均有 $f(u) = cg(u)$，则记作 $f(u) \propto g(u)$。对于 $\mathbf{x}'$，首先注意到 $\mathbf{x}/q \sim \mathbb{R}_1^n$，因此 $\tilde{}\mathbf{x} \sim D_{\mathbb{R}_1^n}^{\mathrm{expand}}, \tilde{\sigma}$ $/\tilde{\sigma}$。结合我们执行的拒绝采样过程，可得

$$P_{\mathbf{x}'}(\mathbf{u}) \propto \frac{f(\mathbf{u})}{\max_{\mathbf{t}\in\mathbb{R}^n} f(\mathbf{t})} P_{\tilde{\mathbf{x}}}(\mathbf{u}) = \frac{f(\mathbf{u})}{\max_{\mathbf{t}\in\mathbb{R}^n} f(\mathbf{t})} P_{D_{\mathbb{R}_1^n,\tilde{\sigma}}^{\mathrm{expand}}/\tilde{\sigma}}(\mathbf{u}) = \frac{P_{D_{\mathbb{R}^n,1}^{\mathcal{N}}}(\mathbf{t})}{\max_{\mathbf{t}\in\mathbb{R}^n} f(\mathbf{t})} \propto P_{D_{\mathbb{R}^n,1}^{\mathcal{N}}}(\mathbf{u}) \,.$$

因此，我们得出 $\mathbf{x}' \sim D_{\mathbb{R}^n,\ 1}^{\mathcal{N}}$。对于 $\alpha z/q$，注意到 $z \sim D_{\mathbb{R}}^{\mathcal{N}},\ \sigma$，因此 $\alpha z/q \sim D_{\mathbb{R}}^{\mathcal{N}}$ $_{,\alpha\sigma/q}$.

对于零假设情形，容易看出边缘分布满足 $D_{\mathbf{x}'} = D_{\mathbb{R}^n,\ 1}^{\mathcal{N}}$ 且$D_{y'} = U(\mathbb{R}_\alpha)$，并且 $\mathbf{x}'$ 与 y' 相互独立——因为 $\mathbf{x}$ 与 y 相互独立，且 $\mathbf{x}'$ （分别为 y'）仅依赖于 $\mathbf{x}$（分别依赖于 y）。

还需验证该采样过程以 $1 - 2^{-\Omega(m)}$ 的概率至少产生 m' 个样本。我们首先证明，每次单独的拒绝采样都至少以一个正常数概率成功。根据事实 B.4，当 $\tilde{\sigma} = 1/r\alpha = \sqrt{\log n}/c$ 对充分小的正常数 $c > 0$ 成立时，我们有

$$f(\mathbf{t}) = \frac{P_{D_{\mathbb{R}^n,1}^{\mathcal{N}}}(\mathbf{t})}{P_{D_{\mathbb{R}_1^n,\tilde{\sigma}}^{\mathrm{expand}}/\tilde{\sigma}}(\mathbf{t})} \in (1/2, 3/2) \,.$$

注意，对于任意 $\mathbf{x}$，我们以 $f(\tilde{}\mathbf{x})/\max_{\mathbf{t}\in\mathbb{R}^n} f(\mathbf{t})$ 的概率接受样本；根据上述界，该概率至少为 $1/3$。然后，通过应用切尔诺夫界可知，拒绝采样以至少 $1 - 2^{-\Omega(m)}$ 的概率成功至少 $m' = cm$ 次，其中 $c > 0$ 是充分小的正常数。证明至此完成。

□

我们注意到，引理 B.5 比 [GVV22] 中的引理 18 更强，具体而言，原引理 18 的界为 $\alpha = c/\left(r\sqrt{\log\, n + \log\, m + \omega(\log\, \lambda)}\right)$，而此处的界为 $\alpha = c/\left(r\sqrt{\log\, n}\right)$。对于学习线性阈值函数的任务，如果使用引理 18 而非引理 B.5，并对证明其余部分采用相同论证，仍可得到一个 $n^{\Omega(1/(\epsilon\sqrt{\log n})^{0.99})}$ 下界——相比之下，我们在此建立的是 $n^{\Omega(1/(\epsilon\sqrt{\log n})^{1.99})}$ 近乎最优的下界。

结合上述引理和假设 2.3，我们给出命题 2.4 的证明。

命题 2.4 的证明。我们通过引理 B.1、引理 B.2、引理 B.3 和引理 B.5，给出从假设 2.3 出发的高效归约。具体而言，归约将按以下步骤进行：

1. 令假设 2.3 中的问题为求解学习带误差($2^{O(l^\beta}$ $'), \mathbb{Z}_q^l, \mathbb{Z}_q^l, D_{\mathbb{Z}}^{\mathcal{N}},\ \sigma'$，模$q$)，其中 $2^{-O(l^{\beta'}}$ $)$优势，其中 l 为维数。2. 接下来，我们使用引理 B.1，将问题归约为求解 LWE($n^{O(k^\beta)}\mathbb{Z}_q^n$, $S_{n,k}, D_{\mathbb{Z}\ ,c\sqrt{k}\sigma'}^{\mathcal{N}}, \mathrm{mod}q$)，其优势为 $n^{-O(k^\beta)}$，其中 c 是一个足够大的正普适常数，n 为维数，且秘密向量取自稀疏集合 $S_{n,k}$。3. 然后应用引理 B.2 和引理 B.3。这两个引理使样本分布和噪声分布变为连续分布。正如之前所述，这两个引理只会使噪声尺度扩大一个普适常数因子，因此我们将问题归约为求解 LWE($n^{O(k^\beta)}\mathbb{R}_q^n, S_{n,k}, D_{\mathbb{R}\ ,c\sqrt{k}\sigma'}^{\mathcal{N}}, \mathrm{mod}q$)，其优势为 $n^{-O(k^\beta)}$，其中 c 是一个足够大的正普适常数。

4. 为完成归约，我们应用引理 B.5，它主要将样本分布从 $U(\mathbb{R}_q^n)$ 改为 $D^{\mathcal{N}}_{\mathbb{R}^n,1}$，并将问题归约为求解 LWE($n^{O(k^\beta)}$ $D^{\mathcal{N}}_{\mathbb{R}^n,1}$, $\mathbb{S}^{n-1}$, $D^{\mathcal{N}}_{\mathbb{R},\sigma}$, mod$\alpha$)，其优势为 $n^{-O(k^\beta)}$。

首先，为了开始归约，我们需要在第一步中选定参数 l, β', q, σ' 的取值。令 $n, k, \beta, \gamma, \kappa$ 为命题 2.4 正文中的参数，也就是我们希望通过归约得到的目标参数。为方便起见，令 $\delta > 0$ 为满足 $1-3\delta = \beta$ 的常数。令 ψ 为满足如下条件的值： $k = \log^\psi n$ (ψ 对 n 和 k 有依赖关系。我们取以下值:

- $l = \log^t n$，其中 $t = 1 + \psi(1-\delta)$;
- $\beta' = \frac{1+\gamma(1-2\delta)}{1+\gamma(1-\delta)}$，这是一个常数，且 $\beta' \in (0, 1)$;
- $q = k^{\kappa+1}$;
- $\sigma' = c\sqrt{l}$，其中 c 是一个充分大的常数。

现在我们来检查归约中每一步的参数是否有效:

1. 我们首先检查参数是否满足假设 2.3 中的要求。注意，

$$q = k^{\kappa+1} = \log^{\psi(\kappa+1)} n = l^{\psi(\kappa+1)/t} \le l^{\psi(\kappa+1)/(\psi(1-\delta))} = l^{(\kappa+1)/(1-\delta)} = l^{O(1)}.$$

2. 接着检查引理 B.1 中的要求。我们将额外参数选为 $\lambda = 2^{l^\beta}$ 和 $m = n^{O(k^\beta)}$。为方便起见，我们先证明 $2^{l^{\beta'}} = n^{\omega(k^\beta)}$。注意，

$$2^{l^{\beta'}} = 2^{\log^{t\beta'} n} = n^{\log^{t\beta'-1} n} = n^{k^{\frac{t\beta'-1}{\psi}}}.$$

由于 $k = \log^\psi n$ 且 $k \ge \log^\gamma n$，因此可得 $\psi \ge \gamma$；所以，$\beta' = \frac{1+\gamma}{(1-2\delta)\ 1+\gamma(1-\delta)} \ge \frac{1+\psi}{1-2\delta)\ 1+\psi(1-\delta)}$。将其代入上式，可得

$$2^{l^{\beta'}} \ge n^{k^{\frac{t\frac{1+\psi(1-2\delta)}{1+\psi(1-\delta)}-1}{\psi}}} = n^{k^{1-2\delta}} = n^{\omega(k^\beta)},$$

其中 th 最后一个等式由事实 $\beta = 1 - 3\delta$ 得出。对于这些要求，我们有:

(a) 显然，$\log(q)/2^l = O(\log l)/2^l = \text{negl}(\lambda)$（因为上一步 $q = l^{O(1)}$）。(b) 对于要求 $\sigma' \ge 4\sqrt{\omega(\log\ \lambda)\ + \ln\ n + \ln m}$，由于 $\sigma' = c\sqrt{l}$，将两边平方后，可改写为

$$l = \omega(\log \lambda + \ln n + \ln m).$$

注意，$\log\ \lambda = O(l^{\beta'})$，其中 $\beta' < 1$；因此，$l = \omega(\log\ \lambda)$。由于 $l = \log^t n$，其中$t = 1 + \psi(1-\delta) \ge 1 + \gamma(1-\delta)$，且 $\gamma(1-\delta)$ 是一个正常数，我们有 $l = \omega(\ln\ n)$。接着，由于 $2^{l^{\beta'}} = n^{\omega(k^\beta)}$ 如上所示，且 $m = n^{O(k^\beta)}$，我们得到 $2^{l^{\beta'}} = \omega(m)$；因此，我们得到 $l = \omega(l^{\beta'}) = \omega(\log\ m)$。结合以上结果可得 $l = \omega(\log\ \lambda + \ln\ n + \ln\ m)$。

(c) 对于要求 $k\log_2(n/k)\ \ge (l+1)\log_2(q)\ + \omega(\log\ \lambda)$，由于 $l = \omega(\log\ \lambda)$ 如上所示，我们可以将其改写为 $k\log_2(n) - k\log_2(k)\ \ge 2l\log_2(q)$。由于在步骤 1 中有 $q = \text{poly}(l)$，因此只需证明 $k\log\ n - k\log\ k = \omega(l\log\ l)$，而对于任意常数 c，这就是 $k\log\ n \ge cl\log\ l + k\log\ k$。我们通过分析两种情况来证明这一点，即 $cl\log l \le k\log k$ 和 $cl\log\ l > k\log k$。

如果 $cl\log\ l \le k\log\ k$，那么由于 $k < c'n$，其中 c' 是一个足够小的普适常数，我们得到 $k\log n \ge 2k\log\ k \ge cl\log\ l + k\log\ k$。
如果 cl 对数 $l > k$ 对数 k，那么只需证明 k 对数 $n = \omega(l$ 对数 $l)$。注意，k 对数 $n =$对数 $^{1+\psi} n$，且 l 对数 $l = t$对数$^t n$ 对数 对数 n。因此，

$$\frac{k\log n}{l\log l} = \frac{\log^{1+\psi-t} n}{t\log\log n}\ .$$

注意，$1+\psi-t = \delta\psi \ge \delta\gamma$（因为 $k \ge$对数$^\gamma n$ 且 $k =$对数$^\psi n$ 蕴含 $\psi \ge \gamma$）至少是一个常数；因此，

$$\frac{k\log n}{l\log l} = \frac{\log^{1+\psi-t} n}{t\log\log n} = \frac{\log^{\delta\psi} n}{t\log\log n} = \omega\left(\frac{\log^{\delta\psi/2} n}{t}\right) = \omega\left(\frac{\log^{\delta\psi/2} n}{1+\psi}\right)\ ,$$

其中，最后一个等式来自这样一个事实：$t = 1+\psi(1-\delta) \le 1+\psi$。因此，我们只需证明 对数$^{\delta\psi/2}_{1+\psi} n$ 至少是一个常~~数。注意~~，对于任意充分大的n，只要对数$^{\delta/2} n \ge e$，我们就有

$$\log^{\delta\psi/2} n = (\log^{\delta/2} n)^\psi \ge e^\psi \ge 1+\psi\ .$$

因此，我们有

$$\frac{k\log n}{l\log l} = \omega(1)\ ,$$

这就是 k 对数 $n = \omega(l$ 对数 $l)$。
因此，要求 k 对数$_2(n/k) \ge (l+1)$ 对数$_2(q) + \omega($对数 $\lambda)$ 在两种情况下都成立。

(d) 现在只需验证引理 B.1 中 $2^{-O(l^{\beta'})}$ 的时间下界以及优势 $2\epsilon m + \mathrm{negl}(\lambda)$，其中 ϵ 是归约前的优势。注意，由于 $2^{l^{\beta'}} = n^{\omega(k^\beta)}$，时间下界至少为任意 $n^{O(k^\beta)}$。对于优势，取 $\epsilon = 2^{-3(l^{\beta'})}$，可得

$$2\epsilon m + \mathrm{negl}(\lambda) = 2^{-3(l^{\beta'})} n^{O(k^\beta)} + \mathrm{negl}(2^{l^{\beta'}}) \le 2^{-2(l^{\beta'})} + \mathrm{negl}(2^{l^{\beta'}}) = n^{-\omega(k^\beta)}\ ,$$

其中，最后一个不等式和等式由上文给出的陈述 $2^{l^{\beta'}} = n^{\omega(k^\beta)}$ 得出。因此，不存在运行时间为 $n^{O(k^\beta)}$、能够以 $n^{-O(k^\beta)}$ 优势求解学习带误差$(n^{O(k^\beta)}\mathbb{Z}_q^n, S_{n,k}, D^N_{\mathbb{Z},c\sqrt{k}\sigma'}, \mathrm{mod} q)$ 的区分器。

3. 接下来，我们检查引理 B.2 和引理 B.3 中的参数要求。注意，只需检查对于足够大的常数 c，$c\sqrt{k}\sigma' \ge 3r\sqrt{\ln\ n + \ln m + \omega(\log\ \lambda)}$。根据其定义，$r = \sqrt{k}$，并且我们已在步骤 2b 中证明 $\sigma' \ge 4\sqrt{\omega(\log\ \lambda)\ + \ln n + \ln\ m}$，因此该不等式成立。

接下来只需验证时间下界和优势。时间下界为 $n^{ck^\beta} - \mathrm{poly}(m, n,$ 对数$(q),$ 对数$(c\sqrt{k}\sigma'))$。由于 $m = n^{O(k^\beta)}$、对数$(q) =$对数$(k^{\kappa+1}) = O($对数 $k)$，以及对数$(c\sqrt{k}\sigma') = O($对数 $k +$对数 $l) = O($对数$^{1+\psi} n) = O(k$ 对数 $n)$，通过选取 c 为充分大的常数，上述下界为任意 $n^{O(k^\beta)}$。类似地，优势为任意 $n^{-O(k^\beta)}$。因此，不存在运行时间为 $n^{O(k^\beta)}$ 的区分器，能够以 $n^{-O(k^\beta)}$ 优势求解 LWE$(n^{O(k^\beta)}\mathbb{R}_q^n, S_{n,k}, D^N_{\mathbb{R},c\sqrt{k}\sigma'},$ 模$q)$。

4. 应用引理 B.5 后，我们得到：不存在运行时间为 $n^{O(k^\beta)}$ 的区分器，能够以 $n^{-O(k^\beta)}$ 优势求解 LWE($m', D^{\mathcal{N}}_{\mathbb{R}^n,1}, S_{n,k}/\sqrt{k}, D^{\mathcal{N}}_{\mathbb{R},c\alpha\sqrt{k}\sigma'/q}$，模 α)，其中 $\alpha = c/\left(\sqrt{k}\text{对数 } n\right)$，$m' = cn^{O(k^\beta)}$，且 $c > 0$ 是一个足够小的普适常数。我们只需检查它是否与命题 2.4 正文中的 σ、m、T 取值相符。对于噪声尺度 σ，我们有

$$c\alpha\sqrt{k}\sigma'/q = c'\sqrt{l}/(\sqrt{\log nq}) = c'\log n^{\psi(1-\delta)/2}/q \le c'k^{1/2}/k^{\kappa+1} = o(k^{-\kappa}) = o(\sigma)\ ,$$

其中，最后一个不等式由 $k = \text{对数}^{\psi} n$ 得出。对于样本数量，我们有 $m' = cn^{c'(k^\beta)}$，通过将 c' 选得足够大，它可以是任意 $n^{O(k^\beta)}$。对于参数 T，我们有 $\alpha = c/\left(\sqrt{k}\text{对数 } n\right) = T$。因此，唯一剩下的差异是，秘密向量分布为 $U(S_{n,k}/\sqrt{k})$，而非 $U(\mathbb{S}^{n-1})$。关键在于，我们可以对所有样本进行随机旋转，这会使秘密向量也随机旋转，并得到我们所需的 $U(\mathbb{S}^{n-1})$ 分布。因此，不存在运行时间为 $n^{O(k^\beta)}$ 的区分器，能够以 $n^{-O(k^\beta)}$ 优势求解 LWE 问题($n^{O(k^\beta)} D^{\mathcal{N}}_{\mathbb{R}^n,1}, \mathbb{S}^{n-1}, D^{\mathcal{N}}_{\mathbb{R},\sigma}$，模 $_T$)。

这证明了命题 2.4。

□