arXiv:2611.29737v1 [cs.LG] 22 Nov 2026

确定性适当正例学习独立于 ZFC

Shai Ben-David
滑铁卢大学
Vector Institute
shai@uwaterloo.ca

Farnam Mansouri
滑铁卢大学
Vector Institute
f5mansou@uwaterloo.ca

Anay Mehrotra
斯坦福大学
anaymehrotra1@gmail.com

摘要

基于仅正样本进行二元分类，是 PAC 学习的一种变体：学习器仅从未知目标概念的正区域接收独立同分布样本，但评估时使用的是完整分布。该模型中的非适当可学习性已有教科书级的刻画，而适当可学习性数十年来一直悬而未决。近期研究刻画了可访问随机性的学习器所具备的可学习性，并指出，与 PAC 学习不同，有些类可以通过随机性实现适当可学习，却无法在没有随机性的情况下实现，从而使确定性情形仍未解决。我们证明，这一问题无法在标准数学公理（ZFC）体系内解决：存在一个 VC 维数为二的类，其确定性适当可学习性独立于 ZFC，尽管有证明表明随机化学习器可以学习该类。此外，有限特征维数无法刻画确定性适当可学习性，而有限特征维数是学习理论中大多数维数共有的一项理想性质。我们还证明，定义域的大小起决定性作用。在 ZFC 中，对于基数低于某一阈值的每个定义域（包括所有可数域），确定性适当可学习性都有组合刻画。相比之下，在基数足够大的每个定义域上，这种刻画都不成立。在广义连续统假设下，该阈值是精确的：刻画正是在该基数处开始失效。最后，我们的证明引入了基于 Erd˝os–Rado 分割定理的新工具，这些工具可能具有更广泛的应用价值。

目录

1 引言 3 1.1 我们的贡献 .4 1.2 技术概述 .7

2 预备知识 10
2.1 仅正例学习 .10 2.2 集合论基础 .12

3 我们结果的形式化陈述 13

4 结论 15

A 扩展预备知识 19

B 相关工作 20

C 定理 3.1 的证明 21 C.1 引理 C.1 的证明 .22 C.2 定理 C.2 的证明 .22 C.3 定理 C.3 的证明 .24

D 定理 3.2 的证明 25

E 定理 3.3 的证明 25
引理 E.1 的证明 .28 引理 E.2 的证明 .29 引理 E.3 的证明 .30 定理 3.3 证明中使用的技术结果 .
31

F 定理 3.4 的证明 32

1 引言

在 PAC 学习的基础模型中 [Val84a], 学习器观察从未知分布中抽取、并由未知目标概念标记的样本，然后必须输出一个误分类误差较小的分类器。该模型隐含地假设学习器可以同时获得正例和负例。然而在实践中，负例往往成本高昂、数量稀少，或者根本无法获得。Valiant 已经注意到这一问题，他写道：“讨论大象属性的分布或许是合理的，但我们可能更愿意不讨论非大象属性的分布”[Val84b]。这一问题也出现在许多现代应用中。在医学诊断中，记录会列出已确认的病例，但很少能证明某位患者没有该疾病 [EN08]。在推荐系统中，一次交互提供了正面证据，而缺少交互并不一定表示负面偏好 [LDLL+03]。仅正例学习是 PAC 模型的一个自然变体，它所探讨的是：当这些正例观察构成学习器获取数据的全部来源时，能够学到什么。

形式上，样本从定义域 $\mathcal{X}$ 上的未知分布 $\mathcal{D}$ 中抽取，并由假设类 $\mathcal{H}$ 中的未知目标概念 ℓ 进行标记。正例学习器只能看到来自 $\mathcal{D}_+$ 的独立样本，即给定 $\ell(X) = 1$ 时 $X \sim \mathcal{D}$ 的分布，但其评估基于完整分布$\mathcal{D}$。我们研究以下问题。

问题1（适当的正例-only PAC 学习）。给定一个假设类 $\mathcal{H}$，寻找一个学习器和一个函数 $m_{\mathcal{H}}$，称为该学习器的样本复杂度，使其仅依赖于 $\mathcal{H}$，并满足以下条件。对于任意 $\varepsilon, \delta \in (0,1)$、任意分布 $\mathcal{D}$ 以及满足 $\Pr_{X\sim\mathcal{D}}[\ell(X) = 1] > 0$ 的任意目标概念 $\ell \in \mathcal{H}$，学习器获得来自 $\mathcal{D}_+$ 的 $n \geq m_{\mathcal{H}}(\varepsilon, \delta)$ 个独立样本后，返回一个假设 $h \in \mathcal{H}$，满足 $\Pr_{X\sim\mathcal{D}}[h(X) \neq \ell(X)] \leq \varepsilon$，其概率至少为 $1 - \delta$；该概率取自样本以及学习器内部任意随机性的联合分布。

请注意，学习器从未见过负例，但假阳性也会计入其误差。例如，如果假设类包含全 1 分类器，那么该分类器与每个样本都一致，但会在目标概念之外的每个点上出错。

问题1 要求进行 适当 学习，这要求输出属于与目标概念相同的假设类，从而满足相同的结构性假设。例如，半空间的适当学习器必须返回一个半空间。在普通的二元 PAC 学习中，适当性是自然具备的，因为经验风险最小化是每个有限 VC 维数类的确定性适当学习器，而这些类恰好就是可学习的类。相比之下，在仅正例学习中，适当性会使问题困难得多。不要求适当性时，问题已有很好的理解：其刻画甚至被列为教科书习题 [KV94，习题 3。7], 而且，每个可学习类都可以通过简单地输出所有包含样本的假设的交集来学习。这个输出不会产生假阳性，但通常不属于该类（对于半空间而言，它就是样本的凸包）。事实上，适当性可能会改变哪些类是可学习的：有些类可以由非适定学习器学习，但无法由任何适当学习器学习 [BMMZ26]。

在 Natarajan 的工作之后，适当学习的刻画这一问题悬而未决数十年[Nat87]。最近，Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26] 通过确定适当可学习性的充要组合条件，解决了随机情形。他们还表明，与 PAC 学习不同，随机性会产生影响：有些类可以由随机学习器进行适当学习，却无法由确定性学习器进行适当学习。随着非适当情形和随机情形的问题都已解决，一个基本情形仍未解决：

确定性适当正例学习的刻画是什么？

1.1 我们的贡献

我们证明，答案可能取决于数学公理：即使对于 VC 维数为二的类，确定性适当学习器的存在性也可能独立于标准集合论公理。

这些公理称为 *ZFC*（带选择公理的策梅洛–弗兰克尔集合论），为数学（包括学习理论）提供了通常的基础 [Jec03]。如果某个命题及其否定都无法由这些公理证明，则称该命题独立于 ZFC。最著名的例子是连续统假设，我们将在本节稍后回顾这一假设 [Coh63； Göd40]。对于我们的类，存在一些 ZFC 模型，即满足其全部公理的数学宇宙，其中存在确定性适当学习器；也存在另一些模型，其中不存在这样的学习器。因此，学习理论的标准工具无法确定该类是否可确定性学习。我们要问的是，是否存在任何学习规则，即使是不可计算的学习规则。

由于我们允许不可计算的学习器，这种独立于 ZFC 的概念与停机问题等问题的计算不可判定性有很大不同 [Tur37]。任意学习规则不一定是可测的，因此我们考虑具有可数支撑的分布；对于这类分布，问题1 中的保证对每条规则都有明确定义。将分布限制为具有可数或有限支撑，是关于可学习性独立性研究中的标准做法 [BHMS+17； FR20； HY23]。相应地，如果某个学习器的输出是无序样本的函数，并且对每个这样的分布都满足这一保证，则称一个类是可确定性学习的。

非正式定理 1.1 (可学习性独立于 ZFC；参见 定理 3.1)。 假设 *ZFC* 一致，则存在一个 *VC* 维数为二的类，其确定性可学习性独立于 *ZFC*。

此处，确定性是必要条件，因为在 ZFC 中，同一个类可以由随机化学习器学习。Ben-David、Hrubeš、Moran、Shpilka 等人 [BHMS+17； BHMS+19] 最早证明了可学习性可能独立于 ZFC。他们研究了一种称为最大值估计(EMX) 的问题，并证明其可学习性取决于连续统的基数。此后，人们发现其他若干问题中的可学习性也独立于 ZFC，包括类别比例学习[FR20], 老虎机学习 [HY 23], 和适当多分类学习 [ADDS+25]。这些结果建立在 EMX 及其单调压缩机制之上，而该机制将可学习性与定义域的基数联系起来。相比之下，我们的证明采用了不同的思路，直接分析学习器，并运用 Erd˝os–Rado 分割定理 [ER56], ，这是一种拉姆齐定理的无限类比。我们根据学习器的行为对定义域的有限子集进行着色。在足够大的定义域上，分割定理保证存在一个大子集，使得学习器的行为在其上保持一致。随后，适当性对学习器的输出施加了足够的限制，从而可以构造一个使其失败的分布（详情参见 1.2 节）。这一论证也可能有助于证明这一特定类之外的不可学习性。

组合刻画，例如 PAC 学习中的有限 VC 维数，通过假设类的性质来表达可学习性。理想情况下，这些性质应尽可能“简单”。为捕捉一种简单性概念，Ben-David、Hrubeš、Moran、Shpilka 等人 [BHMS+17； BHMS+19] 引入了有限特征：如果对某个维数的每个下界都可以由有限多个定义域点和假设作为见证，则称该维数具有有限特征。学习理论中现有的大多数维度都具有这一性质。例如，对于 VC 维数，见证是一个可打散集合，以及实现其标记的假设；Natarajan 维数、图维数和 Littlestone 维度也具有类似的表现。

人们可能会认为，非正式定理 1.1已经足以排除通过有限特征维数进行刻画的可能性。然而，转到 ZFC 的另一个模型可能会改变定义域或假设类，而这也可能改变维数。因此，仅凭独立性并不能排除这种刻画。我们的下一个结果比较了两个具有相同定义域和假设的模型，在这两个模型中，每个有限特征维数的取值都相同 [BHMS+17]，并表明尽管如此，我们的类的确定性可学习性仍会发生变化。

非正式定理 1.2 (不存在有限特征刻画；参见定理 3.2)。 假设 *ZFC* 一致，则存在两个 *ZFC* 模型，它们包含相同的定义域，以及非正式定理 *1.1*中该类的假设，但该类在一个模型中可确定性学习，在另一个模型中却不可。因此，不存在任何有限特征维数能够刻画确定性可学习性。

定义域大小与可学习性。 非正式定理 1.1 背后的划分论证需要较大的定义域。所需大小会随样本大小增长，因此定义域必须足够大，才能同时满足所有样本大小的要求。因此，很自然会问，较小定义域能否避免这一障碍：

限制定义域大小能否恢复 *ZFC* 中可学习性的刻画？

我们给出了这样一种刻画，其大小界限用两种标准的无限基数序列表示，即 aleph 数和 beth 数。我们简要回顾它们；第 2.2 节 提供了更多背景。

▶ **基数与阿列夫数**。 基数 $\aleph_0$ 是自然数的大小。下一个基数 $\aleph_1$是大于 $\aleph_0$的最小基数，而 $\aleph_2, \aleph_3, \ldots$ 则通过反复取下一个更大的基数来定义。它们的上确界 $\aleph_\omega$ 是大于每个有限 n 对应的 $\aleph_n$ 的最小基数。实数集是不可数集，因此其实数基数至少为 $\aleph_1$。断言其实数基数恰好为 $\aleph_1$，就是连续统假设 （CH）。

▶ **幂集与 beth 数**。 实数集的基数与自然数集的所有子集构成的集合的基数相同。这个基数记作 $\beth_1 := 2^{\aleph_0}$。反复取幂集便得到 beth 数： $\beth_0 = \aleph_0$ 以及 $\beth_{n+1} = 2^{\beth_n}$。因此 $\beth_1 = |\mathbb{R}|$，而 $\beth_2$ 是实数集的所有子集构成的集合的基数。这列数的上确界为 $\beth_\omega$，且 $\aleph_\omega \leq \beth_\omega$。所谓广义连续统假设 （GCH）断言，对于每个无限基数 κ， 2^κ 都是下一个更大的基数。由此可得，对于所有有限的 $\beth_n = \aleph_n$ n，并进而得到 $\beth_\omega = \aleph_\omega$。假设 ZFC 一致，则 CH 和 GCH 都独立于 ZFC [Jec03]。

因此，仅凭 ZFC 无法确定实数集的基数是否为 $\aleph_1$，也无法确定 $\aleph_\omega = \beth_\omega$。 非正式定理 1.3 因此分别陈述了在 ZFC 下成立的结论，以及在广义连续统假设下可以推出的结论。

非正式 **定理 1.3** （域大小与可学习性；参见 定理 3.3）。 以下结 *ld.*

（*i*）在 *ZFC* 中，对于基数小于 $\aleph_\omega$的每个定义域，确定性可学习性等价于下方 定义 *1* 中的三个组合条件。（*ii*）在 *ZFC* 中，每个基数至少为 $\beth_\omega$ 的定义域都支撑一个 *VC* 维数为二、满足 定义 *1* 中的全部三个条件但不可确定性学习的类。

（*iii*）在广义连续统假设下，$\beth_\omega = \aleph_\omega$，因此这些界相重合：对于定义域为 $\mathcal{X}$ 的每个类，这三个条件刻画确定性可学习性，当且仅当 $|\mathcal{X}| < \aleph_\omega$。

仅在 ZFC 下，我们的界限在 $\aleph_\omega$ 和 $\beth_\omega$之间留下了一个可能的间隙。第 (ii) 部分中的论域远大于学习理论中常见的论域：基数为 $\beth_\omega$ 的集合，比从 $\mathbb{R}$ 经过有限次取幂集得到的任何集合都大。即便如此，在我们的下界构造中，每个确定性学习器都会在一个具有有限支撑集的分布上失败。

接下来，我们将介绍非正式定理 1.3第 (i) 部分中的三个条件。我们使用样本的闭包这一学习理论中的标准概念。我们将每个假设视为其标记为正的点集。若存在某个假设包含集合S，则称该集合为可实现的，其版本空间$\mathcal{H}_S = \{h \in \mathcal{H} : S \subseteq h\}$ 由所有可能的目标概念组成。其闭包是其版本空间中所有假设的交集，即 $\mathrm{CLOS}_{\mathcal{H}}(S) = \bigcap_{h \in \mathcal{H}_S} h$，也就是$S$ 所迫使其标记为正的点。

定义 1（组合条件）。我们考虑关于 $\mathcal{H}$ 的以下条件。

(i) 有限 *VC* 维数：$\mathrm{VCdim}(\mathcal{H}) < \infty$。

(ii) 一致外部分离性：对于每个 $\eta > 0$，存在 $M \in \mathbb{N}$，使得任意有限非空可实现集 S 都存在 $h_1, \ldots, h_M \in \mathcal{H}_S$，满足 $|\{j : y \in h_j\}| \leq \eta M$，对每个 $y \notin \mathrm{CLOS}_{\mathcal{H}}(S)$；见公式 (2.2)。

(iii) ^{P} 适当的单点闭包：对于每个可实现点 x，$\mathrm{CLOS}_{\mathcal{H}}(\{x\})$ 属于 $\mathcal{H}$.

非正式地说，一致外部分离性要求存在一个长度有界的相容假设列表，使得闭包之外的每个点都很少出现。前两个条件刻画了随机适当学习器的可学习性 [BMMZ26，定理 4]，；而确定性学习器还需要第三个条件 [BMMZ26，命题 18]。因此，在每个论域上，这三个条件都是必要的（注 2.2 解释了各条件的作用）。非正式定理 1.3 表明，在基数小于 $\aleph_\omega$的每个论域上，这三个条件也足以保证可学习性；但在基数至少为 $\beth_\omega$的任何论域上则不然。在基数为 $\aleph_\omega$的论域上，随机学习器与确定性学习器之间的差距甚至可能取决于公理：非正式定理 1.1 中的类在某些 ZFC 模型中需要随机性，在另一些模型中则不需要。

在基数小于 $\aleph_\omega$ 的论域上，非正式定理 1.3 给出了一种不同于 非正式定理 1.2 所排除的那些刻画。对论域的限制至关重要：在 ZFC 的每个模型中，非正式定理 1.1 中的类都满足全部三个条件，但其确定性可学习性取决于所处的模型。每个条件都是一阶陈述，其量词遍历定义域点以及 $\mathcal{H}$ 的成员。从这个意义上说，这些条件具有组合性质，并满足 Ben-David、Hrubeš、Moran、Shpilka 等人提出的组合刻画要求 [BHMS+19]。然而，外部分离性和单点闭包似乎都不存在有限失败证据，因此它们并非 有限性 条件（按 Lechner 和 Ben-David 的定义 [LB24，第 3 节）。3]。

一个一般性的障碍。 非正式定理 1.3 （ii）源自一个关于类的不可学习性的更一般判据。

命题 1.4（见 定理 3.4）。 设 $\mathcal{H}$ 是一个概念类，A 是一个可实现集，且 $|A| \geq \beth_\omega$ 与$\mathrm{CLOS}_{\mathcal{H}}(A) \not\in \mathcal{H}$。如果对于某个基数 $\kappa < \beth_\omega$，至多 κ 个假设包含 A，而其他每个假设与 A 至多交于 κ 个点，那么 $\mathcal{H}$ 不是可确定性学习的。

1.2 技术概述

本节将重点介绍我们证明背后的关键技术思想。在解释为何在没有随机性的情况下选择适当假设十分困难之后，我们使用一个单一的类来说明我们结果的两个方面：在较大的论域上，没有确定性学习器能够成功；而在较小的论域上，我们可以构造出一个。随后，我们将这些论证推广到一般类，最后解释这两个方面如何结合起来，得到与 ZFC 的独立性。

适当纯正例学习的障碍。 我们从贯穿所有结果的核心困难开始。正例学习器根据完整分布进行评估，但其样本 S 从不会揭示某个未观测点是否为负点。它能够安全标记为正的唯一点，是属于闭包 $\mathrm{CLOS}_{\mathcal{H}}(S)$的点；每个与 S 一致的目标都会将这些点标记为正。当该闭包不属于 $\mathcal{H}$时，适当输出就不可能同时包含样本并避开闭包之外的所有点。学习器必须输出一个假设 h 满足 $h \setminus \mathrm{CLOS}_{\mathcal{H}}(S) \neq \varnothing$，从而面临假阳性误差；或者省略某些观测到的正例，从而面临假阴性错误。在 UES 下，随机化学习器可以在与样本一致的同时控制第一种风险：从列表 $h_1, \ldots, h_M \in \mathcal{H}_S$ 中均匀选择一个假设；定义 1 保证闭包之外的每个点被纳入的概率至多为 η。由于闭包包含于目标概念中，每个负点都位于闭包之外，因此对负区域上的这一点态保证取平均，即可界定期望假阳性误差。一致性和有限 VC 维数则控制假阴性错误。Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26，定理 4] 表明，UES 与有限 VC 维数刻画了随机适定可学习性。

确定性学习器无法从这个列表中进行独立随机选择。Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26定理 7]以及下文讨论的正向结果中都采用了一种策略：利用数据中已有的随机性，让假设的选择随可能的样本而变化。学习器看到的是一个无序多重集，因此它的选择只能取决于观测点及其重数。当 $\mathcal{D}_+$ 集中在单个点 x上时，这种变化来源便不复存在：对于每个样本大小 n，学习器总是会观察到n 个 x的副本。在这种情况下，为避免这两种误差，适定单点闭包是必要的 [BMMZ26，命题 18]。

核心构造。 适定单点闭包可以处理集中在单个点的样本，但无法解决一般情况下的困难：以下类满足全部三个条件，但它是否可确定性学习取决于其定义域的基数。该构造扩展了 Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26定理 7]中的例子。设 A 是一个无限集，并设 $B = \{b_1, b_2, \ldots\}$ 是一个与 A不相交的可数无限集。定义

$$\mathcal{X}_A = A \sqcup B, \qquad h_j = A \cup \{b_j\} \quad (j \geq 1), \qquad \mathcal{M}_A = \{\{x\} : x \in \mathcal{X}_A\} \cup \{h_j : j \geq 1\}. \tag{1.1}$$

每个假设h_j 由A 和一个额外的标记 b_j组成，我们将 $\mathcal{M}_A$ 其称为标记类。该类的 VC 维数为二，并满足 UES 和适当单例闭包。然而，任何

支撑在A 上的样本若至少包含两个不同的点，其闭包就是A，而A并不是$\mathcal{M}_A$中的假设。因此，适当的学习器必须要么输出一个单元素集，舍弃除一个观测点之外的所有点；要么输出某个h^j，其中包含一个标记b^j ，而该标记的标签无法从样本中得知。

不可学习性结果。 我们首先证明，当 $\mathcal{M}_A$ 且 $|A| \geq \beth_\omega$时，不能进行确定性学习，非正式定理 1.3 （ii）由此得证；随后将该论证推广至命题 1.4。

考虑 $\mathcal{M}_A$ 以及支撑在A上的样本。如果正分布分散在许多点上，则单元素集会产生较大的假阴性错误，而每个h^j 都可能在b^j处产生假阳性。为了将这种风险分散到各个标记元素上，确定性学习器因此必须在不同样本之间改变j 。由于h^j只有可数多个，当A 足够大时，分割定理排除了这种可能性。

固定样本大小 n，并为 A 的每个 n 元子集赋予颜色：如果学习器在该子集上输出一个单元素集，则赋予 0；如果输出 h^j，则赋予 j。当 $|A| \geq \beth_\omega$ 时，Erd˝os–Rado 定理给出 A 的一个无限子集 F $\subseteq$，使得其中所有 n 元子集都具有相同颜色。这样的集合称为齐次集。该定理将拉姆齐定理扩展到无限多种颜色，但代价是需要大得多的底集。界 $|A| \geq \beth_\omega$ 使我们能够对每个固定的有限 n 应用该定理。集合 F 可能取决于 n 和学习器。我们取正分布为充分大的 F 的有限子集上的均匀分布，从而使样本以高概率不包含重复观测。如果公共颜色为 0，学习器将返回一个单元素集，并遗漏大部分正质量。否则，在这类样本上，它总是返回同一个 h^j 。于是，我们可以选择一个目标概念 h^{j+1} ，并在 b^j 上放置大量负质量，从而迫使假阳性误差很大。因此，对于每个所提出的样本大小，都存在一个有限支撑分布可以击败该学习器（定理 C.3）。

命题 命题 1.4 的证明将这一论证推广到假设不必具有上述简单形式的类。设 A 和 κ 如该命题中所述，因此 $|\mathcal{H}_A| \leq \kappa$ 且 $|h \cap A| \leq \kappa$ 对每个 $h \notin \mathcal{H}_A$都成立。与 $\mathcal{M}_A$中的单元素集输出不同， $\mathcal{H}_A$ 之外的假设现在可能包含 A中的许多点。因此，除了记录输出的类型外，我们还记录它对一个未纳入点的预测。

再次固定样本大小 n，并为 A选定一个良序。我们根据学习器在依次删除每个点后的行为，为每个 $(n+1)$ 元子集着色，并按此顺序列出被删除的点。对每次删除，我们记录返回的是 $\mathcal{H}_A$ 中的哪个假设；若输出不属于 $*$，则记录 $\mathcal{H}_A$，同时记录该输出是否包含被删除的点。至多有 种可能的颜色。Erd˝os–Rado 定理（定理 2.3）的基数形式由此保证存在一个足够大的齐次集 $F \subseteq A$。

齐次性使得在 F 的未观测点处的预测仅取决于该点相对于样本点的位置。选择一个来自 F 的样本，使得每个间隙——即各点之前、点与点之间以及最后一个点之后——都包含多于 κ 个 F 中的点。若学习器在该样本上的输出位于 $\mathcal{H}_A$之外，则其预测在每个间隙上都是常数，因此在任一间隙上的正预测都会把超过 κ个 A 中的点纳入 $\mathcal{H}_A$之外的假设，这与 $|h \cap A| \leq \kappa$矛盾。因此，所有这些预测都必须为负。比较删除操作还会迫使输出类型在来自 n个元素的 F样本之间保持一致。因此，要么学习器在每个这样的样本上都返回 $\mathcal{H}_A$ 中的同一个固定假设，要么其输出不包含样本之外的 F 中的任何点。在第一种情况下，该固定假设包含一个位于 $\mathrm{CLOS}_{\mathcal{H}}(A)$之外的点，而该点可以被赋予较大的负质量。在第二种情况下，将正质量分散到许多点上会迫使假阴性误差很大。因此，该论证归结为与 $\mathcal{M}_A$相同的两类失败，并且再次仅使用有限支撑分布（定理 3.4）。

控制假阴性。 类 $\mathcal{M}_A$ 在满足 $|A| \le |\mathbb{R}|$(定理 C.2）时是可确定性学习的。这给出了独立性结果中可学习的一面。当 $|\mathcal{X}| < \aleph_\omega$时，满足三个组合条件的每个类都是可确定性学习的，正如非正式定理 1.3（i）所述。两个证明采用相同的策略，类似于 Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26、定理 7]。对于足够大的 n，令 $t = n^{2/3}$。如果某个点 x 出现至少 $n - t$ 次，则返回 $\mathrm{CLOS}_{\mathcal{H}}(\{x\})$；否则，返回一个与正样本一致的假设。第一个输出根据单点闭包是适当的，并且包含于目标概念中，因此不会产生假阳性。它至多遗漏 t 个观测，而一致输出一个也不会遗漏。因此，经验假阴性误差至多为 $t/n = n^{-1/3}$，而 VC 一致收敛可以界定总体假阴性误差。因此，主要任务是选择一致输出，使其假阳性误差也足够小。

学习标记类。 处理高度重复的点后，观测到的标记 b^j 会将目标概念标识为 h^j。其余样本的支撑集包含于 A。我们将每个这样的样本分配给某个 h^j，使得选中任意特定 j 的概率都很小。

固定一个单射 $\iota : A \to \mathbb{R}$。有理数切分 是一个有理数阈值 r，用于将满足 $\iota(x) < r$ 的观测与满足 $\iota(x) \ge r$的观测分开。这正是使用界 $|A| \le |\mathbb{R}|$ 的地方：通过 ι，可数多个有理数切分能够分离每个样本中的各个点。按照固定的枚举，选择第一个使两侧各至少有 $t/2$ 个观测的有理数切分，并将重复观测计入其中。没有任何点出现 $n - t$ 次或更多，因此这样的切分存在。我们将该切分的指标以及其下方的观测数量编码到输出索引 j中。因此，固定 j 就同时固定了一个切分和一个计数。我们不必以所选切分为条件：该分支输出 h^j 的事件包含于这个固定切分具有这个固定计数的事件中。计数是平衡的，因此二项式界给出的概率为$O(t^{-1/2})$。由于h^j 只是在A中添加了点b^j，这就控制了假阳性误差。

在较小定义域上学习一般类。对于一般类，我们使用 UES 在有限组样本中指定相容假设，而定义域大小界使我们能够构造这些有限组。有限选择引理（引理 E.1）提供了所需的指定方式：对于任意有限列表的可实现样本 $\mathrm{S}_1, \dots, \mathrm{S}_K$，我们可以同时选择 $\mathrm{h}_i \in \mathcal{H}_{S_i}$，使得对于每个点 y，至多有 $\mathrm{r} = \mathrm{d}^\star + 1$ 个索引满足 y 属于 $\mathrm{h}_i \setminus \mathrm{CLOS}_{\mathcal{H}}(\mathrm{S}_i)$，其中 $\mathrm{d}^\star$ 是对偶类（A 节）的 VC 维数。我们称这类样本是 y 的坏样本。该界对所有 y 同时成立，且与列表长度无关。对于负点 y，每个正样本的闭包都不包含 y，因此，这些恰好是所选输出在 y 处产生假阳性的样本。证明采用概率方法：从 UES 列表中独立选择每个 h_i，则任何固定点处出现过度重叠的可能性都很小。列表大小有界且对偶类的 VC 维数有限，这限制了不同重叠事件的数量，因此，在足够小的 UES 尺度下应用并集界，即可得到合适的元组。在观察样本之前，我们先固定这样一个元组。

假设 $|\mathcal{X}| < \aleph_\omega$ 提供了一个有限重构映射 [BHMS+17]。每个样本都可以用一条消息来表示，其中至多包含 k 个观测点，这里 k 是固定的有限值；根据这条消息，可以重构出一个包含该样本支撑集的有限集合（引理 E.3）。例如，在可数域上固定一个枚举：指标最大的观测点可以确定一个包含整个样本的有限初始段。这类映射恰好存在于基数小于$\aleph_\omega$ [BHMS+17]的论域上，因此我们的刻画针对这些论域给出。一般而言，我们按所选消息对大小为 n 的样本分组。每组都是有限的，因为其中的样本具有

相同的大小，且都属于同一个有限重构集合。在每组中应用 引理 E.1，即可为每个样本确定一个一致的输出；对每组中的每个点 r，至多有 y 个坏样本。

为了将其转化为概率界，考虑任意一个包含 U 中 m 个点的有限集合。支撑集位于 U 中的样本最多可以使用 $\Pi_k(m) = \sum^{\min\{k,m\}}_{j=0} \binom{m}{j}$ 条消息，因此对于固定的 $r\Pi_k(m)$ 条消息是坏的。这个数量随 y 的变化呈多项式增长，且与样本大小 m 无关。汇集引理（n引理 E.2）将这个数量转化为一个趋于零的概率界，前提是排除包含高度重复点的样本。其证明汇集 $2n$ 个观测值，并界定随机抽取其中一半构成坏样本的概率。这种汇集仅是分析手段：学习器不使用额外的样本或随机比特。对负区域上的逐点界取平均，即可界定预期假阳性误差。结合此前的假阴性界和马尔可夫不等式，即可证明可学习性（定理 3.3 (i)。

独立于 ZFC 与有限特征。 我们将正向和负向结果应用于两个 ZFC 模型中的同一个类。固定 $A_\star = \aleph_\omega \times \{0\}$ 以及一个与之不相交的可数标记元素集合。假设 ZFC 一致，则从一个满足广义连续统假设的模型出发。在这个模型中，$|A_\star| = \aleph_\omega = \beth_\omega$，因此$\mathcal{M}_{A_\star}$ 不可确定性学习。

标准的强迫构造通过添加新集合（包括新的实数）来扩展该模型，同时保留基数以及旧定义域和假设中的元素。在扩展模型中，$|\mathbb{R}| > \aleph_\omega$，因此我们的有理数切割构造为同一个类给出了一个学习器。该类保持不变，只有可用学习规则的集合扩展了：扩展模型添加了新的函数，例如从$A_\star$到 $\mathbb{R}$的单射，但没有添加新的点或假设。这就证明了非正式定理 1.1；有关集合论细节，请参见C节。

最后，一个量词仅涉及未改变的定义域点和假设的性质，在两个模型中具有相同的真值。这样的性质无法刻画确定性可学习性，因为后者在两个模型中有所不同。这两个模型还具有相同的自然数，并且关于有限特征维数的每个下界在两个模型中都具有相同的真值。因此，这样的维数要么在两个模型中都是有限的，要么在两个模型中都不是有限的，从而证明了非正式定理 1.2。

2 预备知识

在本节中，我们介绍记号和重要的预备知识（另见 A 节）。

记号。我们将 $\mathbb{N} = \{1, 2, \ldots\}$ 和 $\omega = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$。记号 $[A]^n$ 表示 A 的所有大小为 n 的子集构成的族。指示函数 $\mathbb{1}\{E\}$ 在 E 成立时取 1，否则取 0。对于集合 A、B，它们的对称差为 $A \triangle B = (A \setminus B) \cup (B \setminus A)$。用 $\mathcal{X}^{[*]}$ 表示 $\mathcal{X}$ 上有限无序多重集的集合。对于多重集 S，令 $|S|$ 表示其大小（包括重复项），并令 $\mathrm{supp}(S) = \{x : x \in S\}$ 表示其不同元素构成的集合。

2.1 仅正例学习

学习模型。现在我们将问题1的模型精确地表述如下。令$(\mathcal{X}, \Sigma)$ 为一个可测空间，其 σ-代数包含所有单元素集，并令 $\varnothing \neq \mathcal{H} \subseteq \Sigma$ 为一个概念类；我们将每个假设与其正区域对应。令 $\mathcal{P}$ 为定义在该空间上的概率分布族。问题1对应于所有分布构成的族，而我们的结果涉及可数

有支撑分布（见下文的假设1）。对于目标概念 $\ell \in \mathcal{H}$ 和满足 $\mathcal{D}(\ell) > 0$的分布 $\mathcal{D} \in \mathcal{P}$，学习器从 $\mathcal{D}_+ = \mathcal{D}(\cdot \mid \ell)$中接收样本，而假设 $h \in \mathcal{H}$ 的误差为 $\mathrm{err}_{\mathcal{D}}(h, \ell) = \mathcal{D}(h \triangle \ell) = \Pr_{X\sim\mathcal{D}}[h(X) \neq \ell(X)]$。

假设1。 $\mathcal{P}$ 是 $\mathcal{X}$ 上可数支撑分布的族，即存在某个可数集 $C \subseteq \mathcal{X}$，使得对于其中的分布 $\mathcal{D}$，有 $\mathcal{D}(C) = 1$。

注2.1（分布的选择）。我们的结果对所有学习规则进行量化，包括不可测的学习规则；对于可数支撑分布，问题1中的保证对每个规则都有明确定义。类似的限制很常见 [MY16], ，尤其是在研究可学习性独立于 ZFC 的工作中 [BHMS+17], [FR20，推论 8], [HY23，附录 C.1]。这一限制不会影响我们的下界，因为下界只使用有限支撑分布，因此对任何包含这些分布的分布族都成立。定理 3.3(i)中的刻画也可推广到所有分布，前提是 $\mathcal{H}$ 满足 VC 一致收敛成立所需的可测性条件 [KW24] ，且其证明中构造的学习器具有联合可测的预测；也就是说，映射 $(x_1, \ldots, x_n, y) \mapsto \mathbb{1}\{y \in \mathcal{A}(\{x_1, \ldots, x_n\})\}$ 是可测的。

学习器。 学习器观察到一个无序样本：从$\mathcal{D}_+$中独立抽取n 次形成的多重集，记作$S \sim \mathcal{D}_+^n$。确定性适当学习器是一个函数 $\mathcal{A} : \mathcal{X}^{[*]} \to \mathcal{H}$。如果存在样本复杂度界$m_{\mathcal{H}} : (0,1)^2 \to \mathbb{N}$ ，使得对任意 $\ell \in \mathcal{H}, \mathcal{D} \in \mathcal{P}$ （满足 $\mathcal{D}(\ell) > 0$）、 $\varepsilon, \delta \in (0,1)$，以及 $n \geq m_{\mathcal{H}}(\varepsilon, \delta)$，

$$\Pr_{S\sim\mathcal{D}_+^n}[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S), \ell) > \varepsilon] \leq \delta. \tag{2.1}$$

适当性只要求 $\mathcal{A}(S) \in \mathcal{H}$；输出不必包含样本。 随机化适当学习器还会接收一个均匀随机种子 $r \in \{0,1\}^{\mathbb{N}}$，且该种子独立于样本，因此$\mathcal{A} : \mathcal{X}^{[*]} \times \{0,1\}^{\mathbb{N}} \to \mathcal{H}$；此时，公式（2.1）中的概率是针对 S 和 r共同取值的。若某类可由确定性适当 PAC 学习器学习，我们称其 $\mathcal{H}$可确定性学习；若某类可由随机化学习器学习，则称其可由随机化学习器学习。多重集这一限制只对确定性学习器有影响：给定多重集后，无重复点的独立同分布样本的顺序是一个均匀随机排列，否则确定性学习器可以利用该顺序作为随机比特的来源。

版本空间与闭包。 对于集合 $A \subseteq \mathcal{X}$，其版本空间 为 $\mathcal{H}_A = \{h \in \mathcal{H} : A \subseteq h\}$。对于多重集$S$，令 $\mathcal{H}_S = \mathcal{H}_{\mathrm{supp}(S)}$。如果一个集合或样本的版本空间非空，则称其为可实现。对于非空版本空间 V，令$L(V) = \bigcap_{h\in V} h$，并定义

$$\mathrm{CLOS}_{\mathcal{H}}(A) = L(\mathcal{H}_A) \qquad \text{and} \qquad \mathrm{CLOS}_{\mathcal{H}}(S) = L(\mathcal{H}_S)$$

只要相应的版本空间非空即可。由于来自 $\mathcal{D}_+$ 的样本几乎必然属于目标概念， $\ell \in \mathcal{H}_S$ 因此 $\mathrm{CLOS}_{\mathcal{H}}(S) \subseteq \ell$。反过来，每个 $y \notin \mathrm{CLOS}_{\mathcal{H}}(S)$ 都会被某个 $h \in \mathcal{H}_S$排除，因此样本无法排除 y 为负类的可能性。因此，闭包恰好是样本能够确定为正类的点的集合。非适定学习器可以输出闭包[KV94，练习 3。7], 但闭包不一定属于 $\mathcal{H}$。例如，若 $\mathcal{H}$ 由 $\mathbb{R}^d$中的闭半空间组成，那么有限非空集合的闭包就是它的凸包，而凸包并非半空间。

组合条件。 类 $\mathcal{H}$ 具有 一致外部分离性 (UES) [BMMZ26]当且仅当，对于每个 $\eta > 0$，都存在 $M(\eta) \in \mathbb{N}$ ，使得每个有限非空的可实现 S 都存在 $h_{S,1},\ldots,h_{S,M(\eta)} \in \mathcal{H}_S$ 满足

$$\sup_{y \notin \mathrm{CLOS}_{\mathcal{H}}(S)} \frac{1}{M(\eta)} \sum_{j=1}^{M(\eta)} \mathbb{1}\{y \in h_{S,j}\} \le \eta. \tag{2.2}$$

此列表中的假设可能重复。我们称 $\mathcal{H}$ 具有 真单点闭包，若 $\mathrm{CLOS}_{\mathcal{H}}(\{x\}) \in \mathcal{H}$ 对每个可实现的 x都成立。这两个条件都放宽了要求每个闭包都属于 $\mathcal{H}$的限制。在该要求下，UES 对 $M(\eta) = 1$ 和 $h_{S\text{成立},\,1} = \mathrm{CLOS}_{\mathcal{H}}(S)$。一般而言，它只要求存在一个有界的假设列表，使得闭包外的每个点在列表中出现的次数都很少。真单点闭包条件保留了这一要求，但仅针对单个点。加上有限 VC 维数，这便是定义 1 中的三个条件（另见注 2.2）。

注 2.2 （关于定义 1的讨论）。 有限 VC 维数刻画了普通的二元 PAC 学习 [KV94], 而仅正例学习提供的信息更少，因此它仍然是必要的。1.2 节解释了另外两个条件的作用：一致外部分离性使随机化学习器能够控制假阳性，而适定单点闭包则处理集中于单个点的正分布，在这种情况下，样本不提供随机性。若没有适定单点闭包，每个确定性适当学习器都会在支撑集至多含两个点的分布上失效[BMMZ26，命题 18]。随机化学习器则可以在不同假设之间“权衡”，这正是 Ben-David、Mansouri、Mehrotra 和 Zampetakis 的随机化刻画中不包含该条件的原因 [BMMZ26]。

2.2 集合论基础

我们的结果取决于无限集的大小，其中一些结果还依赖于 ZFC 之外的公理。我们将回顾相关基础知识，并假设读者对集合论并不熟悉。关于这一主题的全面论述，请参阅 Jech [Jec03]。有关 ZFC 和独立性的补充背景见
A 节。

基数。 集合 的基数 $|A|$ 衡量其大小：若A 与 B之间存在双射，则 $|A| = |B|$ ；若存在从 A 到 B的单射，则 $|A| \le |B|$ 。根据选择公理，任意两个基数都可比较，且每个非空基数集合都有最小元素。对于基数 κ，其后继 κ^+ 是严格大于 κ的最小基数。我们使用标准记号：

$$\aleph_0 = |\mathbb{N}|, \qquad \aleph_{j+1} = \aleph_j^+, \qquad \aleph_\omega = \sup_{j<\omega} \aleph_j.$$

因此，一个无限集的基数小于 $\aleph_\omega$ ，当且仅当其基数是某个有限j的 $\aleph_j$ 。与每个满足 $j \ge 1$ 的 $\aleph_j$ 不同，基数 $\aleph_\omega$ 是较小基数集合的可数并。它是具有这一性质的最小不可数基数。

幂集与阿列夫数。对于基数为 κ 的集合，其幂集的基数为 2^κ 。康托尔定理表明，对于任意基数 κ，都有 $2^\kappa > \kappa$ 。实数的基数等于 $\mathbb{N}$ 的幂集的基数。定义

$$\beth_0 = \aleph_0, \qquad \beth_{j+1} = 2^{\beth_j}, \qquad \beth_\omega = \sup_{j<\omega} \beth_j, \qquad \mathfrak{c} = 2^{\aleph_0} = |\mathbb{R}| = \beth_1. \tag{2.3}$$

因此， $\beth_2$ 是 $\mathbb{R}$ 的幂集的基数，而基数为 $\beth_\omega$ 的集合大于通过有限次取幂集从 $\mathbb{R}$ 获得的任何集合。由于 $\beth_{j+1} > \beth_j$，归纳可得，对于每个 j，都有 $\aleph_j \leq \beth_j$，从而 $\aleph_\omega \leq \beth_\omega$。

广义连续统假设（GCH）。连续统假设（CH）断言 $\mathfrak{c} = \aleph_1$，也就是说，每个不可数实数集的基数都等于 $\mathbb{R}$。广义连续统假设（GCH）断言，对于每个无限基数 κ，都有 $2^\kappa = \kappa^+$。哥德尔证明了 GCH（因而也包括 CH）与 ZFC 相容，而科恩证明了 CH 的否定与 ZFC 相容 [Coh63；Göd40]。因此，CH 和 GCH 都独立于 ZFC。仅有限 j 的恒等式 $2^{\aleph_j} = \aleph_{j+1}$ 就意味着 $\beth_j = \aleph_j$ 和 $\beth_\omega = \aleph_\omega$。所以在 GCH 下，非正式定理 1.3 中的两个阈值相同。不满足 GCH 时，二者可能不同：若 $\mathfrak{c} > \aleph_\omega$，则 $\aleph_\omega < \mathfrak{c} = \beth_1 < \beth_\omega$。与 ZFC 相容的是，对于任意有限 j ≥ 1，都有 $\mathfrak{c} = \aleph_j$；同样也与 ZFC 相容的是 $\mathfrak{c} > \aleph_\omega$ [Jec03]。因此，ZFC 无法判定 $\mathbb{R}$，也无法判定任何基数为 $\mathfrak{c}$ 的定义域，其基数是否小于 $\aleph_\omega$。有一个值可以排除：根据柯尼希定理， $\mathfrak{c}$ 不是可数多个基数更小的集合的并集 [Jec 03]，，所以$\mathfrak{c} \neq \aleph_\omega$。

良序与序数。 A 良序 是一种线性序，其中每个非空子集都有最小元素。选择公理蕴含每个集合都具有良序。我们将每个无限基数与其初始序数相对应，即与其具有相同基数的最小序数。 A 节回顾了这些概念。

埃尔˝德什–拉多分割定理。 对 $[A]^n$ 的一个着色是一个从 $[A]^n$ 到一个“颜色”集合的映射。集合$F \subseteq A$ 称为着色下的齐次集 χ，如果 χ 在 $[F]^n$上为常值。以下是拉姆齐定理的推广，详见1.2 节，其中使用了迭代指数：对于无限基数 λ，定义 $\exp_0(\lambda) = \lambda$ 和 $\exp^{j+1}(\lambda) = 2^{\exp_j(\lambda)}$，使得 $\exp^j(\aleph_0) = \beth_j$。

定理 2.3（埃尔˝德什–拉多 [ER56；杰克03]）。 设 λ 为一个无限基数且 $n \geq 0$。如果 $|A| \geq \exp_n(\lambda)^+$，那么 $[A]^{n+1}$ 的任意至多使用 λ 种颜色的着色都有一个基数为 λ^+的齐次集。

取 $\lambda = \aleph_0$ 可得以下特例，它足以证明非正式定理 1.3的第 (ii) 部分。

推论 2.4 (无限齐次集)。 对于每个有限的 $n \geq 1$，每个满足 $|A| \geq \beth^+{}_{n-1}$ 的集合 A，以及每个 $\chi : [A]^n \to \mathbb{N}_0$，都存在一个无限集 $F \subseteq A$，使得 χ 在 $[F]^n$ 上为常数。特别地，对于每个有限的 n，$|A| \geq \beth_\omega$ 已足够。

最后一个断言使用了 $\beth^+{}_{n-1} \leq \beth_n < \beth_\omega$。推论 2.4 中的规模要求会随着 n 增长，而且确实需要一定的增长。对于 $n = 2$，将每一对不同的无限二进制序列按它们首次出现差异的位置着色。这种着色使用可数多个颜色，但不存在三个序列，使得它们构成的三对序列都具有相同的颜色。因此，基数为 $\mathfrak{c} = \beth_1$ 的集合未必包含大小为三的齐次集，而推论 2.4 适用于基数为 $\beth_1^+$ 的集合。

3 我们结果的正式表述

本节将正式陈述我们的结果。

在全文中， $\mathcal{P}$ 表示可数支撑分布的族（假设1）。如第 2.1 节中所述，如果某个确定性适当学习器对问题1 在 $\mathcal{P}$中的每个分布上都能达到相应保证，则称一个类可确定性学习；如果某个随机化适当学习器能够做到这一点，则称其可由随机化学习器学习。

我们先给出非正式定理 1.1的形式化版本，其证明见C 节。该定理涉及 $\mathcal{M}_A$ 在方程 (1.1)中的一个特定实例，其定义域的基数为 $\aleph_\omega$。

定理 3.1（可学习性独立于 ZFC）。 令 $\mathcal{H}_\star = \mathcal{M}_{A_\star}$ 为 方程 （1.1）中满足 $A_\star = \aleph_\omega \times \{0\}$ 且 $B = \mathbb{N} \times \{1\}$的类。那么 $\mathcal{H}_\star$ 的 *VC* 维数为二，并且以下结论成立。

（*i*）在 *ZFC* 中， $\mathcal{H}_\star$ 可由随机化学习器进行学习。

（*ii*）假设 *ZFC* 一致，则 $\mathcal{H}_\star$ 的确定性可学习性独立于 *ZFC*。

因此，ZFC 无法判定这个类是否具有确定性可学习性，尽管它的 VC 维数为二，即使 ZFC 证明了随机化学习器的存在。特别地，在某些 ZFC 模型中，该类需要随机性，而在另一些模型中则不需要，因此随机性是否必要本身可能取决于公理。

接下来，我们将非正式定理 1.2 形式化。我们遵循 Ben-David、Hrubeš、Moran、Shpilka 等人的工作来定义有限特征。 [BHMS+17；BHMS+19]。组合刻画（例如 PAC 学习中的有限 VC 维数）之所以有用，是因为它们很简单，而有限特征刻画了简单性的一种概念。非正式地说，如果某个维数的每个下界都能由有限多个定义域点和假设见证，则称该维数具有有限特征。例如，$\mathrm{VCdim}(\mathcal{H}) \geq \mathrm{s}$ 当且仅当存在 s 个点和 2^s 个假设，使得 $\mathcal{H}$ 中的这些点实现了所有标记。形式地说，性质是集合论的一阶语言中的公式 $\mathrm{P}(\mathcal{X}, \mathcal{H})$；该语言中唯一的关系是相等和属于，并带有自由变量 $\mathcal{X}$和 $\mathcal{H}$。我们将 $\mathcal{X}$视为定义域，将 $\mathcal{H}$ 视为其上的类。如果 ZFC 证明某个性质等价于一个有界公式，则该性质具有有限特征；也就是说，这是一个带有自由变量 $\mathcal{X}$和 $\mathcal{H}$ 的公式，其中每个量词都具有 $\exists \mathrm{x} \in \mathcal{X}$、 $\forall \mathrm{x} \in \mathcal{X}$、 $\exists \mathrm{h} \in \mathcal{H}$ 或 $\forall \mathrm{h} \in \mathcal{H}$ 的形式。

维数 D 为定义域 $\mathcal{X}$ 上的每个类 $\mathcal{H}$ 赋予一个值 $\mathrm{D}(\mathcal{X}, \mathcal{H}) \in \mathbb{N}_0 \cup \{\infty\}$；如果存在一个有界公式的可计算序列 $(\varphi_s)_{s \in \mathbb{N}}$，使得 ZFC 能证明以下陈述，那么该维数具有有限特征：对于每个 $\mathrm{s} \in \mathbb{N}$ 以及所有 $\mathcal{X}, \mathcal{H}$，$\mathrm{D}(\mathcal{X}, \mathcal{H}) \geq \mathrm{s}$ 当且仅当 φ_s 对 $(\mathcal{X}, \mathcal{H})$ 成立。这类性质很自然，因为其真值仅取决于定义域和假设本身：在任何两个包含相同定义域和相同假设 [BHMS+17] 的 ZFC 模型中，具有有限特征的性质真值相同。若 ZFC 能证明，对于每个定义域 $\mathcal{X}$ 及其上的每个类 $\mathcal{H}$（定义域为 $\mathcal{X}$），$\mathrm{P}(\mathcal{X}, \mathcal{H})$ 成立当且仅当 $\mathcal{H}$ 可确定性学习，则性质 P 刻画确定性可学习性。若维数 D 满足性质 $\mathrm{D}(\mathcal{X}, \mathcal{H}) < \infty$ 也成立，则该维数刻画确定性可学习性。现在可以陈述非正式定理 1.2 的正式版本，其证明见 D 节。

定理 3.2（不存在有限特征刻画）。 假设 *ZFC* 是一致的。那么，不存在任何有限特征性质或有限特征维数能够刻画确定性可学习性。

定理 3.2表明，与 PAC 学习不同，确定性适当正例学习不存在类似 VC 维数所提供的刻画。其证明使用了定理 定理 3.1证明中的两个模型；这两个模型具有相同的定义域和类，但对于该类是否可确定性学习意见相左。任何有限特征性质都无法区分这两个模型。有限特征维数同样无法做到这一点：每个性质 $\mathrm{D}(\mathcal{X}, \mathcal{H}) \geq s$ 在两个模型中都具有相同的真值，而这两个模型具有相同的自然数，因此在一个模型中维数有限，当且仅当它在另一个模型中也有限。

我们的第三个结果，即非正式定理 1.3的形式化版本，涉及定义域的大小。第 (i) 和 (ii) 部分的证明分别见于E 节和C 节。

定理 3.3 （域大小与可学习性）。 以下结论成立。

（*i*）在 *ZFC* 中，对于每个满足 $|\mathcal{X}| < \aleph_\omega$ 的定义域 $\mathcal{X}$，类 $\mathcal{H} \subseteq 2^{\mathcal{X}}$ 可确定性学习，当且仅当它满足定义 *1*中的三个条件。若这些条件成立，且 $|\mathcal{X}| \leq \aleph_{k-1}$ 对某个 $k \geq 1$ 成立，则 $O(2^{3d}(k+1)^6/(\varepsilon\delta)^3)$ 个样本就足够，其中 $d = \text{VC 维数}(\mathcal{H})$。

（*ii*）在 *ZFC* 中，对于每个满足 $|A| \geq \beth_\omega$ 的集合 A，$\mathcal{M}_A$ 中的 方程 (1.1) 定义的类具有 *VC* 维数为二，并满足定义 *1*中的三个条件，但不可确定性学习。特别地，每个基数至少为 $\beth_\omega$ 的定义域都支持这样一个类。

（*iii*）在广义连续统假设下，$\beth_\omega = \aleph_\omega$。因此，对于每个定义域 $\mathcal{X}$，定义 *1* 中的三个条件刻画了其上所有类的确定性可学习性，当且仅当 $\mathcal{X}$。 $|\mathcal{X}| < \aleph_\omega$。

定理 3.3 的补定理 3.1，其定义域的基数恰好为 $\aleph_\omega$。在任何更小的论域上，包括所有可数论域，ZFC 都能证明确定性可学习性的组合刻画；除 Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26] 的随机化刻画之外，唯一额外的条件是适定单点闭包。在基数至少为 $\beth_\omega$ 的论域上，这些条件不再充分；在广义连续统假设下，两个阈值相等。

第 (ii) 部分中的不可学习性是以下一般判据的一个实例，也是 命题 1.4 的形式化版本。其证明见F 节。

定理 3.4 (基数条件导致的不可学习性)。 设 $\mathcal{H}$ 是定义域为 $\mathcal{X}$ 的一个类。假设存在某个可实现集 $A \subseteq \mathcal{X}$ 满足

$$|A| \geq \beth_\omega, \quad \mathrm{CLOS}_{\mathcal{H}}(A) \notin \mathcal{H}, \quad \textit{and} \quad \max\{|\mathcal{H}_A|, \sup_{h \in \mathcal{H} \setminus \mathcal{H}_A} |h \cap A|\} < \beth_\omega . \tag{3.1}$$

那么 $\mathcal{H}$ 不可确定性学习。对于任意确定性适当学习器 $\mathcal{A}$ 和任意样本大小 $n \geq 1$，存在 $\ell \in \mathcal{H}$ 以及一个有限支撑分布 $\mathcal{D}$，使得 $\mathcal{D}(\ell) > 0$，并且

$$\Pr_{S \sim \mathcal{D}^n_+}[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S), \ell) \geq 1/2] \geq 3/4 . \tag{3.2}$$

定理 3.4 对 VC 维数或外部分离性均不作任何假设。其条件仅涉及基数。它所给出的失败结论也很强：对于任意样本大小，在某个有限支撑分布下，学习器以至少 3/4 的概率产生至少为 1/2 的误差。我们预计这一判据有助于证明其他类在大定义域上的不可学习性。

4 结论

在这项工作中，我们研究随机化的适当正例学习的刻画留下的一个未解问题：哪些类可以在不使用随机性的情况下学习？我们证明，答案可能取决于数学公理：假设 ZFC 一致，则存在一个 VC 维数为二的类，其确定性可学习性独立于 ZFC，尽管它可以由一个

随机化学习器学习。我们的结果还排除了用有限特征维数进行刻画的可能性。在 ZFC 中，对于基数小于 $\aleph_\omega$ 的定义域，我们给出了由三个组合条件构成的刻画；而在基数至少为 $\beth_\omega$ 的任何定义域上，这些条件都不充分。在广义连续统假设下，$\aleph_\omega = \beth_\omega$，因此定义域大小的阈值 $\aleph_\omega$ 是精确的。

我们的工作留下了几个尚未解决的问题。在仅有 ZFC（不含 GCH）的情况下，我们的界限在 $\aleph_\omega$ 和 $\beth_\omega$之间留下了一个可能的间隙。更广泛地说，是否存在对所有论域上的确定性可学习性的刻画？根据非正式定理 1.2，它不可能是有限特征维数。基数小于 $\aleph_\omega$的论域上，确定性适当正例学习的最优样本复杂度是多少？埃尔德˝什–拉多分割论证能否在其他学习模型中证明其独立于 ZFC？

人工智能披露

本文的大部分想法由作者提出。GPT-6-Pro 建议使用埃尔德˝什–拉多分割定理来证明非正式定理 1.1中不可学习的一侧。

参考文献

[ADDS+25] Julian Asilis、Siddartha Devic、Shaddin Dughmi、Vatsal Sharan 和 Shang-Hua Teng。《适当可学习性与未标记数据的作用》。载于：第36届算法学习理论国际会议论文集。第272卷。机器学习研究论文集。PMLR，2025年，第112–133页。网址：https://proceedings.mlr.press/v272/asilis25b.html（引用见第4、21页）。[Ass83] Patrick Assouad。《密度与维数》。载于：傅里叶研究所年刊 33.3（1983），第233–282页。网址：https://www.numdam.org/item/AIF_1983__33_3_233_0/（引用见第19页）。[BD20] Jessa Bekker 和 Jesse Davis。《从正例与未标记数据中学习：综述》。载于：《机器学习》109.4（2020），第719–760页。网址：https://doi.org/10.1007/s10994-020-05877-5（引用见第20页）。[BHMS+17] Shai Ben-David、Pavel Hrubeš、Shay Moran、Amir Shpilka 和 Amir Yehudayoff。《一个独立于集合论 ZFC 公理的学习问题》。arXiv:1711.05195。版本1。2017年。网址：https://arxiv.org/abs/1711.05195v1（引用见第4、5、9、11、14、20、21、25、26页）。[BHMS+19] Shai Ben-David、Pavel Hrubeš、Shay Moran、Amir Shpilka 和 Amir Yehudayoff。《可学习性可能不可判定》。载于：自然·机器智能 1.1（2019），第44–48页。网址：https://doi.org/10.1038/s42256-018-0002-3（引用见第4、6、14、20、21页）。[BLS10] Gilles Blanchard、Gyemin Lee 和 Clayton Scott。《半监督新颖性检测》。载于：机器学习研究期刊 11（2010），第2973–3009页。网址：https://jmlr.org/papers/v11/blanchard10a.html（引用见第20页）。[BMMZ26] Shai Ben-David、Farnam Mansouri、Anay Mehrotra 和 Manolis Zampetakis。《适当纯正例学习中的意外发现》。arXiv:2606.28309。2026年。网址：https://arxiv.org/abs/2606.28309（引用见第3、6、7、9、12、15、20、21、26、28页）。[BPDS25] Nataly Brukhim、Aldo Pacchiano、Miroslav Dudík 和 Robert E. Schapire。《老虎机学习的难度》。载于：第三十八届学习理论会议论文集。第291卷。机器学习研究论文集。PMLR，2025年，第4452–4485页。网址：https://proceedings.mlr.press/v291/brukhim25a.html（引用见第21页）。

[CDS20] Clément L. Canonne、Anindya De 和 Rocco A. Servedio。《连续分布下从满足赋值中学习》。载于：2020年 ACM-SIAM 离散算法研讨会论文集。工业与应用数学学会，2020年，第82–101页。网址：https://epubs.siam.org/doi/abs/10.1137/1.9781611975994.6（引用见第20页）。[Coh63] Paul J. Cohen。《连续统假设的独立性》。载于：美国国家科学院院刊 50.6（1963），第1143–1148页。网址：https://doi.org/10.1073/pnas.50.6.1143（引用见第4、13、19页）。[DDS15] Anindya De、Ilias Diakonikolas 和 Rocco A. Servedio。《从满足赋值中学习》。载于：第二十六届 ACM-SIAM 离散算法年度研讨会论文集。工业与应用数学学会，2015年，第478–497页。网址：https://www.cs.columbia.edu/~rocco/papers/soda15.html（引用见第20页）。[Den98] François Denis。《从正统计查询进行 PAC 学习》。载于：算法学习理论。第1501卷。计算机科学讲义。Springer，1998年，第112–126页。网址：https://link.springer.com/chapter/10.1007/3-540-49730-7_9（引用见第20页）。[EN08] Charles Elkan 和 Keith Noto。《仅凭正例与未标记数据学习分类器》。载于：第14届 ACM SIGKDD知识发现与数据挖掘国际会议论文集。美国计算机协会，2008年，第213–220页。网址：https://doi.org/10.1145/1401890.1401920（引用见第3、20页）。[ER56] Paul Erd˝os 和 Richard Rado。《集合论中的分划演算》。载于：美国数学会通报 62.5（1956），第427–489页。网址：https://doi.org/10.1090/S0002-9904-1956-10036-0（引用见第4、13页）。[FR20] Benjamin Fish 和 Lev Reyzin。《从未标记数据中学习类别比例的复杂度》。载于：人工智能研究杂志 69（2020），第1333–1349页。网址：https://doi.org/10.1613/jair.1.12013（引用见第4、11、21页）。[Ger89] Mihály Geréb-Graus。《并行、分布式与自动机计算的下界》。博士论文。哈佛大学，1989年（引用见第20页）。[Göd40] 库尔特·哥德尔。《选择公理和广义连续统假设与集合论公理的一致性》。第3卷。数学研究年刊。普林斯顿大学出版社，1940年（引用见第4、13页）。[Han24] Steve Hanneke。《星数与 Eluder 维度：关于分歧维度的初步观察》。载于：第三十七届学习理论会议论文集。第247卷。机器学习研究论文集。PMLR，2024年，第2308–2359页。网址：https://proceedings.mlr.press/v247/hanneke24a.html（引用见第20页）。[HY23] Steve Hanneke 和 Liu Yang。《老虎机可学习性可能不可判定》。载于：第三十六届学习理论会议论文集。第195卷。机器学习研究论文集。PMLR，2023年，第5813–5849页。网址：https://proceedings.mlr.press/v195/hanneke23d.html（引用见第4、11、21页）。[Jec03] Thomas Jech。《集合论：第三千年版，修订扩充版》。数学专著丛书。Springer，2003年。网址：https://link.springer.com/book/10.1007/3-540-44761-X（引用见第4、5、12、13、19、22页）。[Kiv95] Jyrki Kivinen。《可靠学习与单侧误差》。载于：数学系统理论 28.2（1995），第141–172页。网址：https://doi.org/10.1007/BF01191474（引用见第20页）。[KTZ19] Vasilis Kontonis、Christos Tzamos 和 Manolis Zampetakis。《未知截断条件下的高效截断统计量》。载于：2019 IEEE 第 60 届计算机科学基础年度研讨会（FOCS）。IEEE，2019年，第1578–1595页。网址：https://arxiv.org/abs/1908.01034（引用见第20页）。[KV94] Michael J. Kearns 和 Umesh Vazirani。《计算学习理论导论》。麻省理工学院出版社，1994年。网址：https://mitpress.mit.edu/9780262111935/an-introduction-to-computational-learning-theory/（引用见第3、11、12、20页）。

[KW24] Lothar Sebastian Krapp 和 Laura Wirth。《统计学习基本定理》中的可测性。arXiv:2410.10243。2024。URL：https://arxiv.org/abs/2410.10243（引自第11页）。[LB24] Tosca Lechner 和 Shai Ben-David。“用于刻画分布类可学习性的维度所固有的局限性”。载于：《第三十七届学习理论会议论文集》。第247卷。《机器学习研究论文集》。PMLR，2024年，第3353–3374页。URL：https://proceedings.mlr.press/v247/lechner24a.html（引自第6、21页）。[LDLL+03] Bing Liu、Yang Dai、Xiaoli Li、Wee Sun Lee 和 Philip S. Yu。“利用正例与未标记样本构建文本分类器”。载于：《第三届 IEEE 国际数据挖掘会议论文集》。IEEE计算机学会，2003年，第179–186页。URL：https://doi.org/10.1109/ ICDM.2003.1250918（引自第3页）。[LLYL02] Bing Liu、Wee Sun Lee、Philip S. Yu 和 Xiaoli Li。“文本文件的部分监督分类”。载于：《第十九届国际机器学习会议论文集》。2002年，第387–394页。URL：https://www.cs.uic.edu/~liub/NSF/New-PSC.html（引自第20页）。[LMZ24] Jane H. Lee、Anay Mehrotra 和 Manolis Zampetakis。“未知截断下的高效统计：多项式时间算法及超越高斯分布”。载于：2024年IEEE第65届计算机科学基础年会（FOCS）。IEEE，2024年，第988–1006页。URL：https: //arxiv.org/abs/2410.01656（引自第20页）。[LMZ26] Jane H. Lee、Anay Mehrotra 和 Manolis Zampetakis。“从正样本中学习的平滑分析”。载于：第58届ACM年度计算理论研讨会论文集。美国计算机协会，2026年，第42–53页。URL：https://arxiv.org/abs/2504. 10428（引自第20页）。[MB25] Farnam Mansouri 和 Shai Ben-David。“从正例与未标记样本中学习——有限样本量界”。载于：神经信息处理系统进展。第38卷。2025年，第57481–57516页。URL：https://proceedings.nips.cc/paper_files/paper/2025/hash/4a30b165d71ca6cbb8f849252b1ef483-Abstract-Conference.html（引自第20页）。[MY16] Shay Moran 和 Amir Yehudayoff。“VC类的样本压缩方案”。载于：ACM会刊 63.3（2016），21:1–21:10。URL：https://doi.org/10.1145/2890490（引自第11页）。[Nat87] Balas K. Natarajan。“关于布尔函数的学习”。载于：第十九届ACM年度计算理论研讨会论文集。美国计算机协会，1987年，第296–304页。URL：https://doi.org/10.1145/28395.28427（引自第3、20页）。[PNS15] Marthinus C. du Plessis、Gang Niu 和 Masashi Sugiyama。“从正例与未标记数据中学习的凸形式化”。载于：第32届国际机器学习会议论文集。第37卷。《机器学习研究论文集》。PMLR，2015年，第1386–1394页。URL：https://proceedings.mlr.press/v37/plessis15.html（引自第20页）。[Rob55] Herbert Robbins。“关于斯特林公式的一条注记”。载于：《美国数学月刊》62.1（1955），第26–29页。URL：https://www.jstor.org/stable/2308012（引自第23页）。[Sau72] Norbert Sauer。“关于集合族的稠密度”。载于：《组合理论杂志》A辑 13.1（1972），第145–147页。URL：https://doi.org/10.1016/0097-3165(72)90019-2（引自第19页）。[Shv90] Haim Shvaytser。“从正例学习的一个必要条件”。载于：《机器学习》5.1（1990），第101–113页。URL：https://doi.org/10.1007/BF00115896（引自第20页）。[Tur37] Alan M. Turing。“论可计算数及其在判定问题中的应用”。载于：《伦敦数学会会刊》s2-42.1（1937），第230–265页。URL：https: //doi.org/10.1112/plms/s2-42.1.230（引自第4页）。[Val84a] Leslie G. Valiant。“可学习对象理论”。载于：《ACM通讯》27.11（1984），第1134–1142页。URL：https://doi.org/10.1145/1968.1972（引自第3页）。[Val84b] Leslie G. Valiant。“演绎学习”。载于：伦敦皇家学会哲学汇刊。A辑，数学与物理科学 312.1522（1984），第441–446页。URL：https: //doi.org/10.1098/rsta.1984.0069（引自第3页）。

A 扩展预备知识

本节将介绍证明中用到的其他预备知识。

附加记号。我们将 $[K] = \{1,\ldots,K\}$ 记为$K \in \mathbb{N}$。我们将 $[A]^{\leq k}$和 $[A]^{<\omega}$ 分别记为A 中大小至多为k 的子集族和有限大小子集族。对于多重集S，令$N_x(S)$表示x 在S中的重数。对于$x \in \mathcal{X}$， δ_x 是x处的点质量；$\mathrm{Unif}(F)$ 表示有限集合F上的均匀分布。

VC 维数与对偶类。我们将$d =$ 记为 $\mathrm{VCdim}(\mathcal{H})$。 $\mathcal{H}$ 在集合U 上的迹为$\{h \cap U : h \in \mathcal{H}\}$。对偶类 $\mathcal{H}^\star = \{e_x : x \in \mathcal{X}\}$ 的定义域为 $\mathcal{H}$，且$e_x(h) = \mathbb{1}\{x \in h\}$。我们将$d^\star =$ 记为 $\mathrm{VCdim}(\mathcal{H}^\star)$。我们还使用

$$\Pi_k(m) = \sum_{j=0}^{\min\{k,m\}} \binom{m}{j}.$$

Sauer 引理给出了 $\mathcal{H}$ 在有限集合U 上的迹的大小上界 $\Pi_d(|U|)$ [Sau72], , 而 Assouad 界给出$d^\star < 2^{d+1}$ 当$d < \infty$ [Ass83]。

ZFC 与独立性。ZFC 表示带选择公理的策梅洛–弗兰克尔集合论，这是数学公理的标准体系。我们的可学习性与不可学习性结果都是 ZFC 的定理，除非我们明确假设了额外的假设，例如 GCH。若 ZFC 既不能证明某个命题，也不能证明其否定，则称该命题独立于 ZFC。根据哥德尔完备性定理，这恰好意味着存在两个 模型 满足 ZFC，也就是说，存在两个满足其全部公理的数学宇宙，使得该命题在其中一个宇宙中为真，而在另一个宇宙中为假。独立性结果假设 ZFC 是一致的，因为不一致的理论可以证明所有命题。我们的独立性结果使用了科恩的 强迫法，这是一种构造模型的标准方法 [Coh63；Jec03]。它通过添加新集合（例如新的实数）来扩展给定的模型，同时每个旧集合仍保留其元素。因此，在模型及其扩展中，一个类及其定义域是相同的，正如 非正式定理 1.2 所要求的那样。我们使用的是 基数保持的扩展：旧模型中的每个基数仍然是基数，因此 $\aleph_\omega$（定义见 第 2.2 节）在两个模型中相同。于是，我们的类的定义域仍保持基数 $\aleph_\omega$，而连续统可以增长到超过该基数（C 节）。

良序与序数。良序是这样一种线性序：其中每个非空子集都有最小元。例如， $\mathbb{N}$ 上的通常序是良序，但 $\mathbb{Z}$ 和 $[0,1]$ 上的通常序则不是。选择公理蕴含每个集合都可以赋予良序。然而，对于 $\mathbb{R}$这类不可数集合，目前尚不知道任何显式的良序，而且根据这种序定义的子集不一定是勒贝格可测的。序数是典范的良序集合：每个良序集合都与唯一的序数序同构，这个序数称为其序型。每个序数都是比它小的序数所组成的集合。有限序数为 0、1、2、……，而 ω 是最小的无限序数。我们将每个无限基数与其初始序数等同，即与其基数相同的最小序数等同，并用ω_j 和 ω_ω 表示 $\aleph_j$ 和 $\aleph_\omega$ 的初始序数。因此，基数为 κ 的集合可以枚举为 $\{a_\alpha : \alpha < \kappa\}$，正如可数集合可以列为$a_0, a_1, \ldots$ 一样。在这样的枚举中，每个初始段 $\{a_\beta : \beta < \alpha\}$ （其中 $\alpha < \kappa$ ）的基数都小于 κ。这一性质推动了引理 E.3中的有限重构。

B 相关工作

本节介绍从正例学习的相关模型，以及关于可学习性的独立性和刻画的进一步研究。

单侧误差下的学习。 Natarajan [Nat87] 开创了在单侧误差模型下从正例学习的研究。在这种模型中，学习器绝不能将负点标记为正例。他刻画了该模型中的适当可学习性，而 Shvaytser [Shv90] 和 Kivinen[Kiv95] 则刻画了非适当可学习性，并给出了其样本复杂度的上界。禁止假阳性使该模型比我们的模型限制更严格；在我们的模型中，学习器可能在两个方向上都出错，并按总误差进行评估。Natarajan 猜想，他的刻画可以推广到我们双侧模型中的适当学习。Ben-David、Mansouri、Mehrotra 和 Zampetakis[BMMZ26] 推翻了这一猜想。

从正例学习。 在本文研究的双侧模型中，非适当可学习性由假设的有限交所构成的类具有有限 VC 维数这一性质刻画 [Ger89]。这一刻画作为一道习题出现在 Kearns 和 Vazirani 的教科书中 [KV94，习题 3。7]。Hanneke [Han24] 根据 1-中心星数确定了非适当学习的最优样本复杂度。这两个结果都涉及非适当学习器；即使样本闭包不属于该类，它们也可以将其作为输出。对于适当学习，Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26] 通过随机学习器刻画了可学习性，区分了随机学习器与确定性学习器，并研究了一致学习器和稳定学习器。他们未能给出确定性学习器的刻画，而这正是我们研究的问题。

分布假设。 另一类工作研究在分布假设下的仅正例学习。对于高斯分布，Kontonis、Tzamos 和 Zampetakis [KTZ19]以及 Lee、Mehrotra 和 Zampetakis [LMZ24] 为具有有界高斯表面积的类给出了高效算法。Lee、Mehrotra 和 Zampetakis [LMZ26] 证明，当未知分布满足光滑性条件时，可以学习每个有限 VC 维数的类。其他工作假设底层分布已知，例如均匀分布 [CDS20； DDS15]。这些工作获得了高效算法，但依赖于关于分布的假设。相比之下，我们的结果是分布无关的，并且我们关注哪些类至少允许存在学习器，而不考虑计算效率。

从正例与未标记样本中学习。 在正例-未标记（PU）学习中，学习器还会接收来自完整分布的未标记样本 [BD20； BLS10； Den98； EN08； LLYL02；LMZ24； LMZ26； PNS15]。未标记样本揭示了分布在哪些位置分配了质量，因此也揭示了假阳性会造成高代价的区域。Mansouri 和 Ben-David [MB25] 给出了 PU 学习所需正例与未标记样本数量的上界和下界。他们的下界表明，对于某些类而言，无论使用多少正例，都无法替代未标记样本。我们的学习器不会接收未标记样本，因此只能通过类的结构来控制假阳性。

可学习性的独立性。在 Ben-David、Hrubeš、Moran、Shpilka 等人提出的 EMX 问题中[BHMS+17；BHMS+19]在引言中讨论了这个问题：学习器必须输出类中的一个成员，其概率质量接近该类中的最大值。对于由 $[0,1]$,的有限子集组成的类，他们证明 EMX可学习性等价于单调压缩方案的存在；而这两者成立当且仅当 $2^{\aleph_0} < \aleph_\omega$，这一点无法由 ZFC 判定。独立性结果

此后，类别比例学习 [FR20], 老虎机学习 [HY23], 以及适当多分类学习 [ADDS+25]的独立性结果也相继得到确立；这些结果都是通过将相应问题与 EMX 联系起来而得出的。我们的独立性证明则直接分析学习器，采用 Erd˝ős–Rado 分割定理（1.2 节）。我们预计，这种方法对其他学习问题也有用，因为它不需要将这些问题与 EMX 联系起来。我们的可学习性结果确实用到了 Ben-David、Hrubeš、Moran、Shpilka 等人提出的有限重构引理 [BHMS+17]；该引理是他们的压缩方案的基础，并且适用于基数小于 $\aleph_\omega$的每个定义域；这个基数也决定了 EMX可学习性。

基于维度的刻画的局限性。 Ben-David、Hrubeš、Moran、Shpilka 等人[BHMS+17；BHMS +19] 还通过比较两个 ZFC 模型证明，没有任何有限特征维数能够刻画 EMX可学习性。非正式定理 1.2 则是关于确定性适当正例学习的类似结果。其他研究也排除了针对其他学习问题的基于维度的刻画。Lechner 和 Ben-David [LB24] 证明，没有任何维数概念能够刻画学习分布类的样本复杂度，也没有任何可学习性刻画能满足一组自然要求。他们还针对受限边缘分布下的分类以及实值函数类证明了类似结果。我们在 定义 1之后的讨论中使用了他们提出的有限性条件概念。Brukhim、Pacchiano、Dudík 和 Schapire [BPDS25]证明，按照 Ben-David、Hrubeš、Moran、Shpilka 等人 [BHMS+19]的定义，没有任何维数能够刻画老虎机可学习性，即使对于有限类也是如此。相比之下，我们的障碍取决于定义域的大小：对于基数小于 $\aleph_\omega$的任何定义域，定义 1 中的条件都能刻画确定性可学习性。

C 定理 3.1的证明

本节将证明定理 3.1，现重述如下。

定理 3.1（可学习性独立于 ZFC）。 设 $\mathcal{H}_\star = \mathcal{M}_{A_\star}$ 是 方程 （1.1）中满足 $A_\star = \aleph_\omega \times \{0\}$ 且 $B = \mathbb{N} \times \{1\}$的类。则 $\mathcal{H}_\star$ 的 *VC* 维数为二，且以下结论成立。

(*i*) 在 *ZFC* 中， $\mathcal{H}_\star$ 可由随机化学习器学习。

(*ii*) 假设 *ZFC* 一致，则 $\mathcal{H}_\star$ 的确定性可学习性独立于 *ZFC* .

以下引理验证了 $\mathcal{M}_A$，特别是 $\mathcal{H}_\star$满足定义 1中的三个条件。其证明见 C.1 节。

引理 C.1。 方程 $\mathcal{M}_A$ 中的类 公式 (1.1) 具有 *VC* 维数为二、一致外部分离性和真单点闭包。

根据 引理 C.1 和 Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26中的定理 4], $\mathcal{H}_\star$ 在 ZFC 中可由随机化学习器学习。这证明了 定理 3.1的第 (i) 部分。对于第 (ii) 部分，我们使用关于 $\mathcal{M}_A$的确定性可学习性的以下两个结果。它们的证明分别见C.2 节和C.3 节。

定理 C.2。 在公式 (1.1) 中定义的类 $\mathcal{M}_A$ 在 $|A| \le \mathfrak{c}$ 时可确定性学习，样本复杂度为 $m^{\mathcal{M}_A}(\varepsilon, \delta) = O((\varepsilon\delta)^{-3})$。

定理 C.3。 类 $\mathcal{M}_A$ 定义于公式(1.1)在以下情况下不可确定性学习： $|A| \geq \beth_\omega$。

现在我们结合 定理 C.2 和C.3 来证明 定理 3.1 的第 (ii) 部分。

定理 *定理 3.1* 的证明。设

$$A_\star = \aleph_\omega \times \{0\}, \qquad B = \mathbb{N} \times \{1\}, \qquad \mathcal{H}_\star = \mathcal{M}_{A_\star}, \qquad \mathcal{X}_\star = \mathcal{X}_{A_\star}. \tag{C.1}$$

假设 ZFC 一致。集合论中的标准结果给出一个 M_0 模型，它满足 ZFC 和广义连续统假设，以及一个基数保持扩张 M_1 ，其中 $2^{\aleph_0} > \aleph_\omega$ [Jec03第 13 章和 15]。在 M_0, 中，广义连续统假设给出 $|A_\star| = \aleph_\omega = \beth_\omega$，根据 定理 C.3， $\mathcal{H}_\star$ 不可确定性学习。在 M_1, 中，同一个集合满足 $|A_\star| = \aleph_\omega < 2^{\aleph_0}$。因此，根据定理 C.2， $\mathcal{H}_\star$ 可确定性学习。因此，断言 $\mathcal{H}_\star$ 可确定性学习在这两个模型中具有相反的真值，所以它独立于 ZFC。

□

C.1 引理 引理 C.1 的证明

证明*引理 C.1*。 由于每个单元素集都属于该类，因此每个单元素集都是自身的闭包，且单点闭包都是真子集。为验证 UES，固定一个非空有限可实现集 S ，并令$M = \lceil 1/\eta \rceil$。若 $|S| = 1$，则重复单点集假设M 次。若 $|S| \geq 2$ ，且S 包含b_j，则h_j 是唯一的一致假设，因此将其重复M 次。在剩下的情形中， $|S| \geq 2$ 且$S \subseteq A$。它的版本空间由所有h_j 组成，其交集为A。列表$h_1, \ldots, h_M$ 中每个不属于A 的点至多出现一次，因此每个这样的点在列表中所占比例至多为 $1/M \leq \eta$ 。由此证明了 UES。最后，VC 维数$(\mathcal{M}_A) = 2$。

□

C.2 定理 C.2 的证明

证明*定理 C.2*。 由于 $|A| \leq \mathfrak{c}$，固定一个单射 $\iota : A \to \mathbb{R}$。将有理数枚举为 $r_1, r_2, \ldots$，并定义

$$C_m = \{a \in A : \iota(a) < r_m\}, \qquad v_m(S) = \sum_{x \in C_m} N_x(S).$$

同时固定单射编码 $\pi(m, v) = 2^{m-1}(2v - 1)$ ，其中m、$v \in \mathbb{N}$。因此，输出索引 $\pi(m, v)$同时确定m和v。

步骤 1（定义学习器）。对于大小为$n \geq 27$的样本S ，令$t = n^{2/3}$ ，并按以下顺序应用规则：

（i）如果$N_x(S) \geq n - t$ 对某个x成立，则返回 $\{x\}$。

（ii）否则，如果 $\mathrm{supp}(S) \cap B \neq \varnothing$，令$j = \min\{i : b_i \in \mathrm{supp}(S)\}$ ，并返回h_j.

（iii）否则，令

$$m(S) = \min\{m : t/2 \leq v_m(S) \leq n - t/2\},$$

并返回

$$\mathcal{A}(S) = h_{\pi(m(S), v_{m(S)}(S))}.$$

对于 $n < 27$，包括空样本，返回 h_1。
第一条规则至多选取一个点，因为 $n - t > n/2$。为说明第三条规则有定义，按 ι-值对不同的样本点 $u_1, \ldots, u_q$ 进行排序，并令 i 为满足 $\sum_{j=1}^{i} N^{u_j}(S) \geq t/2$ 的最小指标。由于第一条规则不适用，$N^{u_i}(S) < n - t$，而 $\sum_{j=1}^{i-1} N^{u_j}(S) < t/2$。因此 $t/2 \leq \sum_{j=1}^{i} N^{u_j}(S) < n - t/2 < n$，所以 $i < q$。任何满足 $\iota(u_i) < r_m < \iota(u_{i+1})$ 的有理数切分都能实现这一计数，因此第三条规则有定义。所有选择都是多重集的固定函数，并且每个输出都属于 $\mathcal{M}_A$。

步骤 2（界定固定计数）。对于 $1 \leq v \leq n - 1$,

$$\sup_{p \in [0,1]} \Pr[\mathrm{Bin}(n,p) = v] = \binom{n}{v} (v/n)^v (1 - v/n)^{n-v} \leq \sqrt{\frac{n}{v(n-v)}}\,. \tag{C.2}$$

最大值在 $p = v/n$ 处取得。该不等式由 Robbins 的斯特林界推出[Rob55]，，其给出了更紧的界 $e^{1/(12n)}\sqrt{n/[2\pi v(n-v)]}$。
对于固定的 $C \subseteq \mathcal{X}_A$，令 E 表示事件：$\sum_{x \in C} N_x(S) = v$。根据假设1，存在某个可数的 $K \subseteq \mathcal{X}_A$，使得 $\mathcal{D}_+(K) = 1$。集合 C 不一定可测，但 $C \setminus K$ 包含于零测集 $\mathcal{X}_A \setminus K$ 中，因此 C 对 $\mathcal{D}_+$ 的完备化可测，且 $\mathcal{D}_+(C) = \mathcal{D}_+(C \cap K)$。因此，此计数服从参数为 n 和 $\mathcal{D}_+(C)$ 的二项分布，所以

$$\Pr_{S \sim \mathcal{D}_+^n}[E] = \Pr[\mathrm{Bin}(n, \mathcal{D}_+(C)) = v] \leq \sup_{p \in [0,1]} \Pr[\mathrm{Bin}(n,p) = v]\,. \tag{C.3}$$

特别地，如果 $t/2 \leq v \leq n - t/2$，那么 $v(n-v) \geq nt/4$，并且 方程（C.2）和 （C.3）给出

$$\Pr[E] \leq \frac{2}{\sqrt{t}}\,. \tag{C.4}$$

现在固定 $\ell \in \mathcal{M}_A$ 和 $\mathcal{D}$，其中 $\mathcal{D}(\ell) > 0$。如果 $\ell = \{x\}$，则所有观测值都等于 x，因此第一条规则会原样返回ℓ。剩下需要考虑的是 $\ell = h_s$。

步骤3（界定假阳性）。对于一个正样本，第一条规则返回 ℓ 的一个子集。第二条规则返回 ℓ，因为只有标记 b_s 可能被观测到。因此，假阳性只能由第三条规则导致。

固定$j \neq s$，其中 $\mathcal{D}(\{b_j\}) > 0$，并将事件E_j 定义为

$$E_j = \{S : \mathrm{supp}(S) \subseteq \ell,\ b_j \in \mathcal{A}(S)\}\,.$$

对于正样本，仅当第三条规则返回h_j时，事件E_j 才会发生。
将 $j = \pi(m, v)$ 记为满足 $\in \mathbb{N}^2$的唯一数对 (m, v)。如果 $v < t/2$ 或 $v > n - t/2$，则第三条规则永远不会返回 h_j，因此 $E_j = \varnothing$。否则，$E_j \subseteq \{S : v_m(S) = v\}$。将方程 (C.4) 应用于固定的切割 C_m，得到 $\Pr[E_j] \leq 2/\sqrt{t}$。此界不以依赖于样本的选择 m(S) 为条件。由于对于 $j \neq s$，$h_j \setminus h_s = \{b_j\}$，

$$\mathbb{E}\,\mathcal{D}(\mathcal{A}(S) \setminus \ell) = \sum_{\substack{j \neq s \\ \mathcal{D}(\{b_j\}) > 0}} \mathcal{D}(\{b_j\}) \Pr[E_j] \leq \frac{2}{\sqrt{t}} \sum_{j \neq s} \mathcal{D}(\{b_j\}) \leq \frac{2}{\sqrt{t}}\,. \tag{C.5}$$

步骤 4（界定假阴性）。学习器的每个输出至多错分t 个正样本点。标准 PAC 界给出一个普适常数c_1，使得

$$\mathbb{E}\left[\mathcal{D}(\ell \setminus \mathcal{A}(S))\right] \le \mathbb{E}\left[\mathcal{D}_+(\ell \setminus \mathcal{A}(S))\right] \le t/n + c_1\sqrt{\frac{2\log(en)}{n}}. \tag{C.6}$$

结合 E 式 (C.5) 与式 (C.6)，并利用$t = n^{2/3}$ 和 $\log(en) \le 3n^{1/3}$，得到 s

$$\mathbb{E}_{S\sim\mathcal{D}_+^n} \operatorname{err}_{\mathcal{D}}(\mathcal{A}(S),\ell) \le (3+3c_1)\, n^{-1/3}.$$

因此，对于$n \ge \left(\frac{3+3c_1}{\varepsilon\delta}\right)^3$，由马尔可夫不等式可得所需的界

$$\Pr_{S\sim\mathcal{D}_+^n}\left[\operatorname{err}_{\mathcal{D}}(\mathcal{A}(S),\ell) > \varepsilon\right] \le \delta.$$

□

C.3 定理定理 C.3的证明

定理定理 *C.3*的证明。固定一个确定性适当学习器 $\mathcal{A}$ 和一个样本大小 $n \ge 1$。对于 $S \in [A]^n$，将 S 视为每个点恰好出现一次的多重集，并根据学习器的输出对其进行着色：

$$\chi(S) = \begin{cases} 0, & \mathcal{A}(S) \text{ is a singleton,} \\ j, & \mathcal{A}(S) = h_j \quad (j \ge 1) \end{cases}. \tag{C.7}$$

这种着色至多使用可数多个颜色。根据 推论 2.4，存在一个无限的 $F \subseteq A$，其中各点具有相同的颜色 $j_\star \in \mathbb{N}_0$，因此

$$\chi(S) = j_\star \qquad \text{for every } S \in [F]^n. \tag{C.8}$$

取一个大小为 $F_0 \subseteq F$，且满足 $N \ge \max\{2, 4\binom{n}{2}\}$ 的有限集合。我们将选择 ℓ 和 $\mathcal{D}$，使得 $\mathcal{D}_+ =$ 均匀分布(F_0)。设 E 为 n 个采样点互不相同这一事件。那么

$$\Pr[E] \ge 1 - \binom{n}{2}/N \ge 3/4. \tag{C.9}$$

在 E 上，样本是 n 元子集 F，根据 公式 (C.8)，其输出的颜色为 $j_\star$。我们分两种情况讨论。

情况 *1*：$j_\star = 0$。取 $\ell = h_1$，并令 $\mathcal{D} =$ 服从均匀分布(F_0)。由于 $F_0 \subseteq \ell$，我们有 $\mathcal{D} = \mathcal{D}_+ =$ 服从均匀分布(F_0)。在 E 上，学习器返回一个单元素集 $\{z\}$，其中 $\mathcal{D}(\{z\}) \le {}^1 N$，因此

$$\operatorname{err}_{\mathcal{D}}(\{z\},\ell) \ge 1 - \frac{1}{N} \ge \frac{1}{2}.$$

情况 *2*：$j_\star \ge 1$。取 $\ell = h_{j_\star+1}$，并令 $\mathcal{D} = \frac{1}{2}$服从均匀分布$(F_0) + \frac{1}{2}\delta_{b_{j_\star}}$。由于 $F_0 \subseteq A \subseteq \ell$ 且 $b_{j_\star} \notin \ell$，我们有$\mathcal{D}(\ell) = 1/2$，并且 $\mathcal{D}_+ =$ 服从均匀分布(F_0)。在 E 上，输出恰好是 $h_{j_\star}$。点 $b_{j_\star}$ 属于 $h_{j_\star} \setminus \ell$，且其 $\mathcal{D}$-质量为 1/2，因此

$$\operatorname{err}_{\mathcal{D}}(h_{j_\star},\ell) \ge \mathcal{D}(\{b_{j_\star}\}) = \frac{1}{2}.$$

在这两种情况下，公式 (C.9) 给出了所需的界

$$\Pr_{S\sim\mathcal{D}_+^n}\left[\operatorname{err}_{\mathcal{D}}(\mathcal{A}(S),\ell) \ge \frac{1}{2}\right] \ge \frac{3}{4}. \tag{C.10}$$

□

D 定理 3.2 的证明

本节中，我们将证明定理 3.2，现将其重述如下。

定理 3.2(不存在有限特征刻画)。 假设 *ZFC* 是一致的。那么，不存在任何有限特征性质或有限特征维数能够刻画确定性可学习性。

我们使用以下有界公式的不变性 [BHMS+17]。

断言 1 (有界公式的不变性)。 设 M_1 是 *ZFC* 模型 M_0 的一个强迫扩张，并且$\mathcal{X}$， $\mathcal{H} \in M_0$。

(i) 对每个有界公式 φ，都有 $M_0 \models \varphi(\mathcal{X}, \mathcal{H})$ 当且仅当 $M_1 \models \varphi(\mathcal{X}, \mathcal{H})$。

(ii) 对每个有界公式的可计算序列 $(\varphi_s)_{s\in\mathbb{N}}$，$M_0$ 满足“φ_s 对每个 $s \in \mathbb{N}$都成立”$(\mathcal{X}, \mathcal{H})$ 当且仅当 M_1 满足。

证明。 集合 $\mathcal{X}$ 和 $\mathcal{H}$，以及每个 $h \in \mathcal{H}$，在 M_0 和 M_1, 中具有相同的元素，并且这两个模型在这些元素上的隶属关系相同。因此，两个模型对有界公式的求值都基于同一个结构 $(\mathcal{X} \cup \mathcal{H} \cup \{\mathcal{X}, \mathcal{H}\}; \in)$，该结构以 $\mathcal{X}$ 和 $\mathcal{H}$ 为元素，故由对 φ 的归纳可得 (i)。对于 (ii)，两个模型的自然数也相同，序列 $(\varphi_s)_{s\in\mathbb{N}}$也相同，并且固定集合大小结构的满足关系在 M_0和 M_1之间是绝对的。

□

定理 3.2 的证明。设 $\mathrm{M}_0 \subseteq \mathrm{M}_1$ 是定理 3.1 证明中的模型：M_0 满足广义连续统假设，而 M_1 是 M_0 的保基数强迫扩张，且 $\mathrm{M}_1 \models 2^{\aleph_0} > \aleph_\omega$。定义域 $\mathcal{X}_\star$ 和方程 (C.1) 中定义的类 $\mathcal{H}_\star$ 在两个模型中是相同的集合。根据定理 C.2 和定理 C.3，$\mathcal{H}_\star$ 在 M_0 中不可确定性学习，而在 M_1中可确定性学习。

首先，假设一个P 具有有限特征的性质刻画了确定性可学习性，并令 φ 为一个有界公式，ZFC 证明该公式等价于P。两个模型都满足 ZFC，因此$M_0 \models \neg\varphi(\mathcal{X}_\star, \mathcal{H}_\star)$ 以及$M_1 \models \varphi(\mathcal{X}_\star, \mathcal{H}_\star)$。这与断言 1 (i) 矛盾。

现在假设一个具有有限特征的维数 D 刻画了确定性可学习性，并令$(\varphi_s)_{s\in\mathbb{N}}$ 为相应的有界公式序列。由于 $\mathcal{H}_\star$ 在M_0, 中不可确定性学习，我们有$M_0 \models \mathrm{D}(\mathcal{X}_\star, \mathcal{H}_\star) = \infty$。等价地，$M_0$ 满足“φ_s 对$(\mathcal{X}_\star, \mathcal{H}_\star)$对每个$s \in \mathbb{N}$均成立”。根据断言 1 (ii)，$M_1$ 满足同一陈述，因此$M_1 \models \mathrm{D}(\mathcal{X}_\star, \mathcal{H}_\star) = \infty$。根据该刻画， $\mathcal{H}_\star$ 在M_1,中不可确定性学习，矛盾。

□

E 定理 3.3 的证明

本节证明定理 3.3，现将其重述如下。

定理 3.3 （域大小与可学习性）。 以下结论成立。

(i) 在 *ZFC* 中，对于每个具有 $|\mathcal{X}| < \aleph_\omega$ 的定义域 $\mathcal{X}$，每个类 $\mathcal{H} \subseteq 2^{\mathcal{X}}$ 可确定性学习，当且仅当它满足定义 *1*中的三个条件。如果这些条件成立，且 $|\mathcal{X}| \leq \aleph_{k-1}$ 对某个 $k \geq 1$ 成立，那么只需 $O(2^{3d}(k+1)^6/(\varepsilon\delta)^3)$ 个样本，其中 $d = \mathrm{VC}$ 维数$(\mathcal{H})$。

(ii) 在 *ZFC* 中，对于每个满足 $|A| \geq \beth_\omega$ 的集合 A，$\mathcal{M}_A$ 在方程(1.1) 中的类具有 *VC* 维数为二，且满足定义 *1*中的三个条件，但不可确定性学习。特别地，任何基数至少为 $\beth_\omega$ 的定义域都支撑这样一个类。

(iii) 在广义连续统假设下，$\beth_\omega = \aleph_\omega$。因此，对于每个定义域 $\mathcal{X}$，定义 *1* 中的三个条件刻画了 $\mathcal{X}$ 上所有类的确定性可学习性，当且仅当 $|\mathcal{X}| < \aleph_\omega$。

第 (ii) 部分由 定理 C.3 得出；在广义连续统假设下，第 (iii) 部分由第 (i) 和第 (ii) 部分得出。我们通过以下三个引理证明第 (i) 部分：一个引理选择假设，一个引理将样本计数转换为概率界，还有一个引理从较短的消息中重构有限样本。

Ben-David、Mansouri、Mehrotra 和 Zampetakis [BMMZ26，附录 B，3] 在公式 (2.2) 中将 UES 列表大小界定为 $M(\eta) = O\left(\frac{(d^\star+1)\log(e/\eta)}{\eta}\right)$。利用这个界，下面的引理给出了同时选择性质。对于每个点，它界定了在对应闭包之外包含该点的所选假设的总数，其上界为 $d^\star + 1$；而 UES 仅界定单个列表中此类假设所占的比例。由于有限 VC 维数蕴含$d^\star < \infty$，因此可得到一个固定的有限界，且与版本空间的数量无关。证明见E.1 节。

引理 E.1（同时有限选择）。假设 $d < \infty$ 且 *UES*，并令 $r = d^\star + 1$。对非空有限可实现样本的版本空间的每一个有限列表 $V_1, \ldots, V_K$，存在 $h_i \in V_i$，使得

$$|\{i \in [K] : y \in h_i \setminus L(V_i)\}| \leq r \qquad (y \in \mathcal{X}). \tag{E.1}$$

版本空间可以重复。

下述引理在思想上类似于 Ben-David、Mansouri、Mehrotra 和 Zampetakis[BMMZ26，引理 23], 他们用该引理证明了其定理 7。该引理在关于分布的最大点质量的假设下，对固定多重集的概率给出上界；而在这里，我们施加条件 $\max_x N_x(S) \leq n - t$ 于该族中的多重集上 $\mathcal{B}$。这一重数界使我们能够控制任一特定 $S \in \mathcal{B}$的条件概率，在给定扩大样本 $T = S \uplus S'$ 时，其中 S' 是一个相同规模的独立样本，且 T 仅用于分析。该引理的证明见E.2 节。

引理 E.2（从有限集计数到概率界）。设 $n \geq 2, 1 \leq t \leq n/2$，并设 $\mathcal{B}$ 是一个由大小为 n 的多重集构成的族，且对所有 $S \in \mathcal{B}$,$\max_x N_x(S) \leq n - t$。假设 $b : \mathbb{N} \to [0, \infty)$ 满足

$$|\{S \in \mathcal{B} : \mathrm{supp}(S) \subseteq U\}| \leq b(|U|) \qquad \textit{for every finite nonempty } U \subseteq \mathcal{X}. \tag{E.2}$$

Then

$$\Pr_{S \sim \mathcal{D}_+^n}[S \in \mathcal{B}] \leq \frac{12}{\sqrt{t}} \sup_{m \geq 2} \frac{m b(m)}{2^m}. \tag{E.3}$$

大小为 k 的有限重构映射是一个函数 $\eta : [\mathcal{X}]^{\leq k} \to [\mathcal{X}]^{<\omega}$，满足对每个有限 $U \subseteq \mathcal{X}$，都存在一个子集 J，使得 $|J| \leq k$ 且$U \subseteq \eta(J)$。输出是一个有限集合，其大小可能取决于输入。

下一个引理是 Ben-David、Hrubeš、Moran、Shpilka 等人提出的定理 5 的集合输入形式。 [BHMS+17，定理 5。1]。为完整起见，我们在 E.3 节 中给出其证明。

引理 E.3（有限重构下界 $\aleph_\omega$）。 如果 $|\mathcal{X}| \le \aleph_{k-1}$ 对某个整数 $k \ge 1$，则存在一个大小为 k 的有限重构映射。有限定义域允许大小为零。

现在我们证明定理 3.3 (i)。

定理 *3.3* （*i*）的证明。假设满足定义 1中的三个条件。由于 $|\mathcal{X}| < \aleph_\omega$，存在一个整数$k \ge 1$ 使得 $|\mathcal{X}| \le \aleph_{k-1}$。固定一个大小为 η 的重构映射k，根据引理 E.3该映射存在。令$r = d^\star + 1$，并固定 $\preceq$的一个良序 $[\mathcal{X}]^{\le k}$。固定这些选择后，我们定义一个确定性适当学习器，并分别界定其假阳性错误和假阴性错误。

步骤 1（选择有限表）。对于每个有限的U，定义唯一的消息

$$\rho(U) = \min_{\preceq}\{J \subseteq U : |J| \le k,\ U \subseteq \eta(J)\}. \tag{E.4}$$

由于 η 是一个重构映射，且 $\preceq$ 是一个良序，因此最小值存在。根据定义，$\rho(U) \subseteq U$， $|\rho(U)| \le k$，且$U \subseteq \eta(\rho(U))$。对于$n \ge 1$ 以及每条消息J，令

$$\mathcal{F}_{n,J} = \{S : |S| = n,\ \mathcal{H}_S \neq \varnothing, \rho(\mathrm{supp}(S)) = J\}. \tag{E.5}$$

此族中的每个多重集都以有限集合 $\eta(J)$为支撑，因此该族是有限的。对于固定的n，所有非空族构成了大小为n的全部可实现样本的划分。
应用引理 E.1 于$(\mathcal{H}_S : S \in \mathcal{F}_{n,J})$，得到$h_{n,J,S} \in \mathcal{H}_S$，满足

$$|\{S \in \mathcal{F}_{n,J} : y \in h_{n,J,S} \setminus \mathrm{CLOS}_{\mathcal{H}}(S)\}| \le r \qquad (y \in \mathcal{X}). \tag{E.6}$$

对于每个大小为n的可实现样本S，令

$$g_n(S) = h_{n,\rho(\mathrm{supp}(S)),S} \in \mathcal{H}_S. \tag{E.7}$$

步骤 2（定义学习器）。对于$n \ge 9$，令$t = n^{2/3}$；对于可实现样本S，定义

$$\mathcal{A}(S) = \begin{cases} \mathrm{CLOS}_{\mathcal{H}}(\{x\}), & N_x(S) \ge n - t \text{ for some } x, \\ g_n(S), & \text{otherwise} \end{cases}. \tag{E.8}$$

对于$n < 9$，令 $\mathcal{A}(S)$ 为 $\mathcal{H}$中的任意一个概念。由于$t < n/2$，至多有一个点满足$N_x(S) \ge n - t$，因此第一个分支定义良好。由于 $\mathcal{H}$ 具有适定单点闭包，其输出是适当的。第二个分支返回一个与S相容的适当假设。

步骤 3（计算含有假阳性的样本）。对于$y \in \mathcal{X}$，令$\mathcal{B}_{n,y}$ 为大小为n 且满足以下条件的可实现多重集族

$$\max_x N_x(S) < n - t, \qquad y \in g_n(S) \setminus \mathrm{CLOS}_{\mathcal{H}}(S).$$

固定一个有限非空的$U \subseteq \mathcal{X}$并令 $m = |U|$。根据 公式 (E.6)，

$$\begin{aligned} \big|\{S \in \mathcal{B}_{n,y} : \mathrm{supp}(S) \subseteq U\}\big| &= \sum_{\substack{J \subseteq U \\ |J| \le k}} \big|\{S \in \mathcal{B}_{n,y} \cap \mathcal{F}_{n,J} : \mathrm{supp}(S) \subseteq U\}\big| \\ &\le r\Pi_k(m) \le 2^{d+1}\Pi_k(m). \end{aligned} \tag{E.9}$$

步骤 4（界定假阳性）。应用 引理 E.2 with $b(m) = 2^{d+1}\Pi_k(m)$ 可得

$$\Pr_{S\sim\mathcal{D}_+^n}[S \in \mathcal{B}_{n,y}] \le \frac{12\,2^{d+1}\sup_{m\ge 2} m\Pi_k(m)2^{-m}}{\sqrt{t}}. \tag{E.10}$$

固定 k 时， $\Pi_k(m)$ 是关于 m 的多项式，因此 $\sup_{m\ge 2} m\Pi_k(m)2^{-m}$ 有限。 第 E.4 节给出显式界

$$12\,2^{d+1}\sup_{m\ge 2} m\Pi_k(m)2^{-m} \le 48\,2^d(k+1)^2. \tag{E.11}$$

固定 $\ell \in \mathcal{H}$ 和 $\mathcal{D}$ ，使得 $\mathcal{D}(\ell) > 0$。对于 $y \notin \ell$，每个正样本都满足 $y \notin \mathrm{CLOS}_{\mathcal{H}}(S)$。公式 (E.8)的第一个分支返回 ℓ 的一个子集，因为它选出的点 x 属于 ℓ ，且 $\mathrm{CLOS}_{\mathcal{H}}(\{x\}) \subseteq \ell$。因此，$y$ 处的一个假阳性必定来自第二个分支，此时$S \in \mathcal{B}_{n,\ y}$。因此

$$\Pr[y \in \mathcal{A}(S)] \le 48\,2^d(k+1)^2 n^{-1/3}.$$

由于 $\mathcal{D}$ 具有可数支撑，将此界对 $\mathcal{D}$ 的负点质量求和可得

$$\mathbb{E}\,[\mathcal{D}(\mathcal{A}(S)\setminus\ell)] \le 48\,2^d(k+1)^2 n^{-1/3}. \tag{E.12}$$

步骤 5（界定假阴性）。学习器的每个输出至多会错分 t 个正样本点。标准 PAC 界给出一个通用常数 c_1 ，使得

$$\mathbb{E}\,[\mathcal{D}(\ell\setminus\mathcal{A}(S))] \le \mathbb{E}\,[\mathcal{D}_+(\ell\setminus\mathcal{A}(S))] \le t/n + c_1\sqrt{\frac{d\log(en)}{n}}. \tag{E.13}$$

结合公式 (E.12)和(E.13)，代入$t = n^{2/3}$，并利用 $\log(en) \le 3n^{1/3}$ ，对于某个通用常数c,

$$\mathbb{E}[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S),\ell)] \le \left(48\,2^d(k+1)^2 + 1 + c_1\sqrt{3(d+1)}\right) n^{-1/3} \le c\,2^d(k+1)^2 n^{-1/3}.$$

对于 $n \ge \left(\frac{c\,2^d(k+1\ ^2}{\varepsilon\delta}\right)^3)$，由马尔可夫不等式可得所需的界

$$\Pr[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S),\ell) > \varepsilon] \le \frac{c\,2^d(k+1)^2}{\varepsilon n^{1/3}} \le \delta.$$

□

E.1 引理的证明：引理 E.1

证明：*引理 E.1*。 若 $K \le r$，则从每个 V_i 中任选一个假设即可。假设 $K > r$。对于每个 i，取一个 UES 列表 $f_{i,1},\ldots,f_{i,M} \in V_i$，其尺度为 ηBen-David、Mansouri、Mehrotra 和 Zampetakis[BMMZ26，附录 B。3] 给出了列表大小界

$$M = O\left(\frac{(d^\star+1)\log(e/\eta)}{\eta}\right). \tag{E.14}$$

从 $[M]$ 中独立且均匀地选取 $J_1,\ldots,J_K$。

对于固定的 y，每个事件 $y \in f_{i,J_i} \setminus L(V_i)$ 的概率至多为 η。这些事件在i之间相互独立。因此

$$\Pr\left[\left|\{i : y \in f_{i,J_i} \setminus L(V_i)\}\right| \geq r+1\right] \leq \binom{K}{r+1}\eta^{r+1}. \tag{E.15}$$

令E_y 表示y 至少属于$r+1$ 个集合$f_{i,J_i} \setminus L(V_i)$的事件。为了界定不同事件的数量，考虑在每个候选假设$f_{i,j}$和每个闭包$L(V_i)$中具有相同隶属关系的点y和z 。对于任意选择的J_1,.....、J_K，计入y 和z处重叠计数的索引相同，因此$E_y = E_z$。因此，即使 $\mathcal{X}$ 是无限的，这些事件也由这些有限的隶属模式决定。

候选假设最多有KM 个。由于 VC 维数$(\mathcal{H}^\star) = d^\star$，根据 Sauer 引理，关于 $\mathcal{H}^\star$的隶属模式数量至多为$(KM+1)^{d^\star}$。属于$L(V_1)$、.....、$L(V_K)$ 的情况构成一个长度为K的二进制向量，其可能取值最多有 2^K 种。

因此，不同坏事件的数量至多为 $2^K(KM+1)^{d^\star}$。对于每一组实际出现的模式对，选择一个代表点。利用方程 (E.15) 应用并集界，可得

$$\Pr\left[\bigcup_{y\in\mathcal{X}} E_y\right] \leq 2^K(KM+1)^{d^\star}\binom{K}{r+1}\eta^{r+1} = O_{K,d^\star}\left(\eta^2[\log(e/\eta)]^{d^\star}\right).$$

最后一个等式用到了 方程 (E.14) 以及 $r+1 = d^\star+2$。固定 K,$d^\star$，该界随 η 趋于零。因此，存在一个不包含坏事件的元组。对于这样的元组，假设 $h_i = f_{i,J_i}$ 满足方程 (E.1)。

□

E.2 引理的证明引理 E.2

为证明 引理 E.2，我们先建立以下二项式界。我们将利用它们来界定会产生各个候选样本的指标子集所占比例。

引理 E.4。 对于 n,$N \in \mathbb{N}$ 且 $s \leq N$，我们有

$$2^{-N}\binom{N}{s} \leq \min\left\{\frac{1}{2}, \sqrt{\frac{1}{N}}\right\} \quad (N \geq 1) \qquad \textit{and} \qquad 2^{-2n}\binom{2n}{n} \geq \frac{1}{2\sqrt{n}}.$$

引理的证明 引理 *E.4*。 对于$j \geq 0$，定义 $b_j := 4^{-j}\binom{2j}{j} = (2j)!\,4^j(j!)^2$。则 $b_0 = 1$ 且 $b_1 = 1/2$，约去阶乘后得到 $b_{j+1} b_j = \frac{2j+1}{j+2}$。对于每个 $j \geq 0$，

$$\sqrt{\frac{j}{j+1}} \leq \frac{2j+1}{2j+2} \leq \sqrt{\frac{j+1}{j+2}}.$$

迭代这些不等式可得 $\frac{1}{2\sqrt{n}} \leq b_n \leq \frac{1}{\sqrt{2(n+1)}}$，从而证明了第二个界。对于第一个界，若 N 为偶数，则 $2^{-N}\binom{N}{s} \leq b_{N/2} \leq \sqrt{\frac{1}{N}}$。若 $N = 2j+1$ 为奇数，则

$$2^{-N}\binom{N}{s} \leq 2^{-(2j+1)}\binom{2j+1}{j} = \frac{2j+1}{2j+2}b_j = b_{j+1} \leq \frac{1}{\sqrt{2(j+2)}} \leq \frac{1}{\sqrt{N}}.$$

□

引理 *E.2* 的证明。独立地抽取 x_1，……，x_{2n}，并令 S_I 为包含 x_i 的多重集，其中每个 $i \in I$ 对应一个 $I \subseteq [2n]$，且 $|I| = n$。
每个 S_I 都服从从 $\mathcal{D}_+$ 中独立抽取 n 个观测值的分布，因此

$$\Pr_{S\sim\mathcal{D}_+^n}[S \in \mathcal{B}] = \mathbb{E}\left[\frac{1}{\binom{2n}{n}} \sum_{\substack{I\subseteq[2n]\\|I|=n}} \mathbb{1}\{S_I \in \mathcal{B}\}\right]. \tag{E.16}$$

我们针对每个固定的、包含 $2n$ 个观测值的总体来界定这个平均值。令 T 为这些观测值构成的多重集。如果 T 在 $\mathcal{B}$ 中没有大小为 n 的子多重集，则平均值为零。否则，写作 $\operatorname{supp}(T) = \{u_1, \ldots, u_m\}$，其中 u_i 在 T 中的重数为 N_i，且 $N_1 \geq \cdots \geq N_m \geq 1$。$\mathcal{B}$ 中的一个样本最多包含 $n - t$ 个 u_1 副本，其在 T 中的补集大小为 n，因此 $N_1 \leq 2n - t$。存在这样的样本也意味着 $m \geq 2$。由此可得

$$N_1 \geq \frac{2n}{m}, \qquad N_2 \geq \frac{2n - N_1}{m-1} \geq \frac{t}{m-1}. \tag{E.17}$$

对于一个大小为 n 的子多重集 $R \subseteq T$，若其在 u_i 处的重数为 s_i，则恰有 $\prod_i \binom{N_i}{s_i}$ 个指标子集会生成 $S_I = R$。因此，所有指标子集中生成 R 的比例为

$$w_T(R) = \frac{\prod_{i=1}^m \binom{N_i}{s_i}}{\binom{2n}{n}}. \tag{E.18}$$

引理 E.4 中证明的初等中心二项式界给出

$$2^{-N}\binom{N}{s} \leq \min\left\{\frac{1}{2}, \sqrt{\frac{1}{N}}\right\} \quad (N \geq 1) \qquad \text{and} \qquad 2^{-2n}\binom{2n}{n} \geq \frac{1}{2\sqrt{n}}. \tag{E.19}$$

Usi使用 N_1,N_2 的逆平方根界，并对其他因子使用 1/2，得到 es

$$w_T(R) \leq \frac{2\sqrt{n}}{\sqrt{N_1 N_2}} 2^{-(m-2)} \leq 2\sqrt{\frac{m(m-1)}{2t}} 2^{-(m-2)} < \frac{12m\,2^{-m}}{\sqrt{t}}. \tag{E.20}$$

至多 $b(m)$ 个成员 $\mathcal{B}$ 的支撑集包含于 $\operatorname{supp}(T)$，因此在 T 的子多重集中，候选项至多有这么多个。对它们求和公式 (E.20)，即可将公式 (E.16) 中的平均值上界定为 $12mb(m)2^{-m}/\sqrt{t}$。对 $m \geq 2$ 取上确界，并对样本池求平均，即可证明公式 (E.3)。

□

E.3 引理 E.3 的证明

引理 *E.3* 的证明。如果定义域是有限的，令 $\eta(\varnothing) = \mathcal{X}$。此时，空消息可以重构每个有限子集，因此所需大小为零。若要允许更大的消息大小，则在其余输入上任意定义 η。

对于可数无限定义域以及 $k = 1$，枚举 $\mathcal{X} = \{x_1, x_2, \ldots\}$。令 $\eta(\varnothing) = \varnothing$，并定义

$$\eta(\{x_j\}) = \{x_1, \ldots, x_j\}.$$

对于非空的 $U = \{x_{i_1}, \ldots, x_{i_m}\} \subseteq \mathcal{X}$，具有最大指标的点给出消息，因为 $U \subseteq \eta(\{x_{\max\{i_1,\ldots,i_m\}}\})$。

现在假设 $k \geq 2$，并假设关于 $k-1$的结果成立。用 $\prec$ 将 $\mathcal{X}$ 良序化，使其序型为基数 $|\mathcal{X}|$的初始序数。对于每个 $x \in \mathcal{X}$，真初始段 $\mathcal{X}_{<x} := \{z \in \mathcal{X} : z \prec x\}$的基数至多为 $\aleph_{k-2}$。因此，归纳假设为每个x给出一个重构映射$\eta_x : [\mathcal{X}_{<x}]^{\leq k-1} \longrightarrow [\mathcal{X}_{<x}]^{<\omega}$。令 $\eta(\varnothing) = \varnothing$。对于非空的 $J \in [\mathcal{X}]^{\leq k}$，令 $x =$ 为$\max_{\prec} J$，并定义

$$\eta(J) = \{x\} \cup \eta_x(J \setminus \{x\}) . \tag{E.21}$$

输出是有限的。为验证重构，取一个非空有限集 U，并令 $x = \max_{\prec} U$。归纳假设给出 $J' \subseteq U \setminus \{x\}$，其大小至多为 $k-1$，且 $U \setminus \{x\} \subseteq \eta_x(J')$。于是，消息 $J = J' \cup \{x\} \subseteq U$ 满足 $|J| \leq k$，而公式（E.21）给出 $U \subseteq \eta(J)$，归纳得证。

□

E.4 定理 3.3 证明中用到的技术结果定理 3.3

引理 E.5。 对于每个 $k \in \mathbb{N}$，我们有

$$\sup_{m \geq 2} m\Pi_k(m)2^{-m} \leq 2(k+1)^2 .$$

引理 *E.5* 的证明。 对于 $0 < z < 1$和 $j \geq 0$,

$$\sum_{m=0}^{\infty} \binom{m}{j} z^m = \frac{z^j}{(1-z)^{j+1}} .$$

对z 求导，得到

$$\sum_{m=0}^{\infty} m\binom{m}{j} z^{m-1} = \left(\frac{j}{z} + \frac{j+1}{1-z}\right) \frac{z^j}{(1-z)^{j+1}} .$$

当$z = 1/2$ 时，我们得到

$$\sum_{m=0}^{\infty} m\binom{m}{j} 2^{-m} = 4j + 2 .$$

所有求和项均非负，因此

$$\begin{aligned}
\sup_{m \geq 2} m\Pi_k(m)2^{-m} &\leq \sum_{m=0}^{\infty} m\Pi_k(m)2^{-m} \\
&= \sum_{j=0}^{k} \sum_{m=0}^{\infty} m\binom{m}{j} 2^{-m} \\
&= \sum_{j=0}^{k} (4j+2) \\
&= 2(k+1)^2 .
\end{aligned}$$

□

F 证明 定理 3.4

在本节中，我们证明定理 3.4，其内容重述如下。

定理 3.4 (基数条件导致的不可学习性)。 设 $\mathcal{H}$ 是定义域为 $\mathcal{X}$上的一个类。假设某个可实现集 $A \subseteq \mathcal{X}$ 满足

$$|A| \geq \beth_\omega, \quad \mathrm{CLOS}_{\mathcal{H}}(A) \notin \mathcal{H}, \quad \textit{and} \quad \max\{|\mathcal{H}_A|, \sup_{h \in \mathcal{H} \setminus \mathcal{H}_A} |h \cap A|\} < \beth_\omega . \tag{3.1}$$

则 $\mathcal{H}$ 不是可确定性学习的。对于每个确定性适当学习器 $\mathcal{A}$ 和每个样本大小 $n \geq 1$，存在 $\ell \in \mathcal{H}$ 以及一个有限支撑集分布 $\mathcal{D}$，满足 $\mathcal{D}(\ell) > 0$，使得

$$\Pr_{S \sim \mathcal{D}^n_+}[\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S), \ell) \geq 1/2] \geq 3/4 . \tag{3.2}$$

证明 *定理 3.4* 。我们可以用一个满足以下条件的方程 (3.1) 的单个无限基数 $\kappa < \beth_\omega$ 来表示

$$0 < |\mathcal{H}_A| \leq \kappa, \qquad \text{for all } h \in (\mathcal{H} \setminus \mathcal{H}_A),\ |h \cap A| \leq \kappa . \tag{F.1}$$

固定一个确定性适当学习器 $\mathcal{A}$ 和一个样本大小 $n \geq 1$。选择 κ 使其满足 方程 (F.1)，并枚举

$$\mathcal{H}_A = \{h_\xi : \xi < \theta\}, \qquad 0 < \theta \leq \kappa ,$$

不重复；作为指标时，将基数视为其初始序数。

步骤 1（按学习器输出对样本着色）。对每个 $S \in [A]^n$，定义

$$c(S) = \begin{cases} \xi, & \mathcal{A}(S) = h_\xi \in \mathcal{H}_A, \\ *, & \mathcal{A}(S) \notin \mathcal{H}_A \end{cases} . \tag{F.2}$$

这里，$*$ 与每个指标 $\xi < \theta$均不同。指标 ξ 标识确切的输出 h_ξ。符号 $*$ 将不包含 A 的输出归为一组，这些输出可能随样本而变化。为了控制这些输出，我们还记录它们在一个额外点上的预测。

固定一个良序 $\prec$，其定义在 A 上。对于 $T = \{x_1 \prec \cdots \prec x_{n+1}\} \in [A]^{n+1}$，令 $S_i = T \setminus \{x_i\}$，并设

$$\chi(T) = (c(S_i),\ \mathbb{1}\{x_i \in \mathcal{A}(S_i)\})_{i=1}^{n+1} . \tag{F.3}$$

坐标i 记录删除x_i 后的输出类型，以及所得假设是否包含x_i。良序指定坐标位置，而学习器只接收无序多重集S_i。

颜色 $\chi(T)$ 属于$((\theta \cup \{*\}) \times \{0,1\})^{n+1}$。由于 κ 是无限的，而n 是有限的，该集合的基数至多为

$$(\kappa \cdot 2)^{n+1} = \kappa .$$

因此，颜色数至多为 $\kappa < \beth_\omega$。

步骤 2（寻找一个齐次集）。我们使用分割定理（定理 2.3），其中 $\lambda = \kappa^+$。大小为 λ 的集合过大，无法包含在任何交集$h \cap A$ 中$h \not\in \mathcal{H}_A$ 内（根据 κ，每个集合的基数至多为方程 (F.1)）。要对 $[A]^{n+1}$应用分割定理，我们检查其基数要求。由于 $\kappa < \beth_\omega$，存在一个有限的s 使得 $\kappa \le \beth_s$，因此

$$\lambda = \kappa^+ \le \beth_s^+ \le \beth_{s+1} < \beth_\omega .$$

取 一个有限的s_0，满足 $\lambda \le \beth_{s_0}$。由于取幂集保持基数不等式 ，

$$\exp_n(\lambda)^+ \le \beth_{s_0+n}^+ \le \beth_{s_0+n+1} < \beth_\omega \le |A| . \tag{F.4}$$

着色至多使用 $\kappa \le \lambda$ 种颜色，因此可应用分割定理，得到 A 的一个基数为 λ^+ 的齐次子集。其继承的良序的序型至少为初始序数 λ^+。取其前 λ^+ 个元素，这会保持齐次性，并记

$$F = \{f_\alpha : \alpha < \lambda^+\}, \qquad \alpha < \beta \implies f_\alpha \prec f_\beta .$$

设其公共颜色为

$$\chi(T) = ((c_1, e_1), \dots, (c_{n+1}, e_{n+1})) \qquad (\text{for all } T \in [F]^{n+1}) . \tag{F.5}$$

该向量与T无关，尽管其坐标可能彼此不同。接下来我们证明所有坐标 c_i 都相等。

步骤 3（固定输出类型）。对于 $i = 1$，……，n，令

$$p_i = f_{\lambda \cdot i}, \qquad P = \{p_1, \dots, p_n\} .$$

这里 $\lambda \cdot i$ 表示序数乘法： λ的连续i 个副本的序型。对于每个正有限i，序数 $\lambda \cdot i$ 的基数为 λ，因此 $\lambda \cdot i < \lambda^+$。所以，点$p_i$ 均有定义，且按递增顺序排列。

这些点将F 中其余的点分成$n+1$ 个间隙：

$$\begin{aligned} I_1 &= \{z \in F : z \prec p_1\}, \\ I_i &= \{z \in F : p_{i-1} \prec z \prec p_i\} \quad (2 \le i \le n), \\ I_{n+1} &= \{z \in F : p_n \prec z\} . \end{aligned}$$

每个间隙的基数至少为 λ。第一个间隙包含所有由 $\alpha < \lambda$ 标记的点。每个中间间隙包含 $\lambda \cdot (i-1)$ 与 $\lambda \cdot i$ 之间的区块，但不包含其端点。最后一个间隙包含 $\lambda \cdot n$ 与 $\lambda \cdot (n+1)$ 之间的类似区块，这两个点都位于 λ^+ 以下。每个区块的基数均为 λ，去掉有限个端点后基数仍不变。

固定 $i \in [n+1]$ 以及 $z \in I_i$。点 z 在 $P \cup \{z\}$的递增枚举中位于第 i 位；删除该点后得到 P。因此，比较 i个坐标在 公式 (F.3)和 公式 (F.5)可得

$$c_i = c(P), \qquad e_i = \mathbb{1}\{z \in \mathcal{A}(P)\} \qquad (z \in I_i) . \tag{F.6}$$

由于每个间隔都非空，每个c_i 都等于$c(P)$。将c 记作公共值$c_1 = \cdots = c_{n+1}$。

现在任取$S \in [F]^n$ 以及任意$z \in F \setminus S$。如果z 在S 的递增枚举中位置为$i \cup \{z\}$，则删除z 的坐标给出

$$c(S) = c_i = c, \qquad \mathbb{1}\{z \in \mathcal{A}(S)\} = e_i \,. \tag{F.7}$$

该着色在 $(n+1)$ 元子集上是齐次的，因此，从 F 中取出的每个 n 元样本都具有相同的输出类型。

步骤 4（学习器在 F 上的行为）。若 $c = \xi < \theta$，则根据 $c(S)$ 的定义以及等式（F.7）可得

$$\mathcal{A}(S) = h_\xi \qquad \text{for every } S \in [F]^n \,. \tag{F.8}$$

反之，假设 $c = *$。由于 $c(P) = *$，适当性给出 $\mathcal{A}(P) \in \mathcal{H} \setminus \mathcal{H}_A$。一致交集界意味着

$$|\mathcal{A}(P) \cap A| \le \kappa < \lambda \,.$$

If $e^i = 1$ 对某个 i 成立，则等式（F.6）表明整个间隙 I_i 都包含在 $\mathcal{A}(P) \cap A$ 中，从而给出 g

$$\lambda \le |I_i| \le |\mathcal{A}(P) \cap A| \le \kappa \,,$$

这就产生矛盾。因此，对每个 i，都有 $e_i = 0$。
对任意 $S \in [F]^n$ 以及 $z \in F \setminus S$，现在由等式（F.7）中的第二个等式可得 $z \notin \mathcal{A}(S)$。等价地，

$$\mathcal{A}(S) \cap F \subseteq S \qquad (S \in [F]^n) \,. \tag{F.9}$$

输出不包含样本之外的F 中的任何点。它是否包含S 中的全部点是不受限制的，因为该论证未对学习器作出一致性假设。

步骤 5（强制常数误差）。选择一个有限的$F_0 \subseteq F$，其大小为$N \ge \max\{2n, 4\binom{n}{2}\}$。我们将选择$\ell$ 和 $\mathcal{D}$，使得 $\mathcal{D}_+ =$ 均匀分布(F_0)。令E 表示所抽取的n 个样本点互不相同这一事件。那么

$$\Pr[E] \ge 1 - \binom{n}{2} / N \ge 3/4 \,. \tag{F.10}$$

在E上，样本属于 $[F_0]^n \subseteq [F]^n$，因此步骤 4 的结论适用。
如果$c = *$，则任取 $\ell \in \mathcal{H}_A$，这由可实现性保证存在，并令 $\mathcal{D} =$ 均匀分布(F_0)。由于$F_0 \subseteq A \subseteq \ell$，我们有 $\mathcal{D}(\ell) = 1$，从而 $\mathcal{D}_+ = \mathcal{D}$。在$E$上，方程（F.9）给出

$$\mathcal{A}(S) \cap F_0 \subseteq S, \qquad |\mathcal{A}(S) \cap F_0| \le n \,.$$

由于 $\mathcal{D}$ 在F_0 上均匀分布，且$N \ge$ 2n</style},

$$\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S), \ell) = 1 - \frac{|\mathcal{A}(S) \cap F_0|}{N} \ge 1 - \frac{n}{N} \ge \frac{1}{2} \,.$$

若$c = \xi < \theta$，令$L = \mathrm{CLOS}_{\mathcal{H}}(A)$。$\mathcal{H}_A$ 中的每个成员都包含L，因此$L \subseteq h_\xi$。包含关系是严格的，因为$L \notin \mathcal{H}$。选择$z \in h_\xi \setminus L$。根据$L = \bigcap_{h \in \mathcal{H}_A} h$的定义，存在某个 $\ell \in \mathcal{H}_A$ 不包含z。由于$A \subseteq L$，点z 也位于A 和F_0之外。

设 $\mathcal{D} = \frac{1}{2}$均匀分布$(F_0) + \frac{1}{2}\delta_z$。由于 $F_0 \subseteq A \subseteq \ell$ 且 $z \notin \ell$，我们有 $\mathcal{D}_+ =$均匀分布(F_0)。在 E 上，方程（F.8）使输出为 h^ξ，其中包含负点 z。因此

$$\mathrm{err}_{\mathcal{D}}(\mathcal{A}(S), \ell) \geq \mathcal{D}(\{z\}) = \frac{1}{2}.$$

在这两种情况下，E上的误差至少为 1/2。结合方程（F.10）可得方程(3.2)。由于 $\mathcal{A}$ 和n 均为任意取值，因此对于每个确定性适当规则，$\varepsilon = \delta = 1/4$ 处的 PAC 保证都不成立。

□