arXiv:2303.02047v1 [cs.LG] 3 Mar 2023

希尔伯特空间中 PAC 学习的复杂性*

Sergei Chubanov†

2023年3月6日

摘要

我们从在希尔伯特空间中学习凸多面体的角度研究二元分类问题，任何二元分类问题都可以归约为该问题。有限维空间中的凸多面体学习问题在文献中已有充分研究。我们将该问题推广到希尔伯特空间，并提出一种学习多面体的算法，使其能够以至少 $1-\delta$，的概率正确分类该分布中至少 $1-\varepsilon$ 的样本，其中 ε 和 δ 为给定参数。此外，作为推论，我们改进了有限维空间中多面体分类的一些已有界。

1 引言

一般而言，我们所讨论的分类问题可以表述如下：寻找一个能够一致地对训练数据进行分类的二元分类器，使得描述该分类器所涉及的基本运算数量尽可能少。限制分类器复杂性的直觉依据是广为人知的奥卡姆剃刀原理；已有研究证明（例如 Blumer等人 [1987, 1989]），分类器的复杂性与其预测准确率之间存在联系，这种联系体现在可能近似正确（PAC）分类的意义下。

我们从多面体分类的角度研究这一问题，其中待学习的概念类由给定内积空间中的多面体组成。通过选择合适的核，总能实现多面体可分性；也就是说，我们的结果普遍适用于二元分类的一般情形。

需要注意的是，除非 $P = NP$, 否则我们无法期望找到一种多项式时间方法，用最少数量的半空间来定义多面体分类器，因为即使是在有限维空间中利用两个半空间进行多面体分离的问题也是 NP困难的 Megiddo [1996]。

我们的结果与相关工作。 我们提出了一种算法，用于在内积空间中对多面体概念进行适当 和不适当 学习；这些内积空间可以是有限维欧几里得空间，也可以是无限维希尔伯特空间。在本文中，适当学习是指学习由 t 个半空间定义的多面体（t-多面体），前提是存在一个能够对整个实例空间进行一致分类的 t-多面体。与我们比较结果时所参考的以往论文类似，我们假设定义假设真实多面体的半空间，在其 γ-间隔内不包含实例空间中的任何实例，其中 $\gamma > 0$.

一致多面体分离是一个广为人知的问题，正如我们此前所提到的，即使在 $t = 2$, 这种情形下，也已被证明是难以处理的。尽管在有限维空间中，给定不等式数量的精确分离问题具有 NP 困难性，但关于学习算法运行时间的指数界，已有一些非平凡的结果；参见 Gottlieb 等人 [2018]。我们对这一研究方向的贡献如下：首先，我们提出了一种新算法，可改进有限维情形下 PAC 学习算法的现有运行时间界。其次，我们的算法适用于无限维希尔伯特空间；在这种情况下，算法在一种将内积作为基本运算的计算模型下，保证了类似的复杂度界。

*将发表于 AAAI-2023 会议论文集。

†sergei.chubanov@de.bosch.com 博世人工智能中心，德国雷宁根。

据我们所知，目前适当学习和非恰当学习的最紧界均见于 Gottlieb 等人 [2018]。该文中，作者提出了在有限维空间中对多面体概念进行适当学习和非恰当学习的算法。近期，Gottlieb 等人 [2022] 发表了一篇与 Gottlieb 等人 [2018] 相关的后续研究，并对先前工作做了一些修正。为便于更准确地比较，下面简要总结他们的性能界，并与我们的结果进行对比：

- Gottlieb 等人 [2022]:
  - 恰当分类的运行时间：

$$m^{O(t\gamma^{-2}\log(\gamma^{-1}))},$$

    其中 m 为样本复杂度。
  - 非恰当分类的运行时间：

$$m^{O(\gamma^{-2}\log(\gamma^{-1}))},$$

  - 样本复杂度：

$$m = O(D\varepsilon^{-1}\log^2(\varepsilon^{-1}) + \log(\delta^{-1})),$$

    其中 $D = (\gamma^{-2}t\log\ t)$。

  - 对希尔伯特空间的适用性：就目前而言，该方法不适用于无限维空间，因为其中一些关键组件，例如 Gottlieb 等人在引理 2.2 中分析的所谓 fat-shuttering 维数 [2022] ，取决于相关向量空间的维数。

- 我们的算法（针对有限维空间）：
  - 恰当分类的运行时间： $m^{O(t\gamma^{-2}}$).– 非恰当分类
    的运行时间：

$$O\left(d \cdot m \cdot (t\gamma^{-2})^{4\gamma^{-2}}\right).$$

  - 样本复杂度 m:
    * 适当学习： $O(dt\log\ t)$, 其中 d 是空间的维数。* 非恰当学习： $O(\varepsilon^{-1}D + \varepsilon^{-1}$ $\log(\delta^{-1}))$，其中 $D = O((\gamma^{-2} + \log t)(t\gamma^{-2})^{4\gamma^{-2}})$.

对于恰当分类，与 Gottlieb 等人 [2022]的方法相比，我们将指数的幂次降低了 $\log(\gamma^{-1})$。对于非恰当分类，我们算法的运行时间显著更短，但样本复杂度关于 γ^{-2}. 呈指数级增长

Vempala [2010]和 Klivans 等人 [2008]提出了一种适用于有限维空间的非恰当学习算法；该算法的运行时间随空间维度呈指数增长，不需要间隔假设，但对概率分布提出了额外假设。

Bennett 和 Bredensteiner [2000], 、Mavroforakis 和 Theodoridis [2006] 将支持向量机作为大间隔分类器，这当然也与我们的工作相关。特别是，Bennett 和 Bredensteiner [2000] 提出了 SVM 的一种相关几何解释。除了 SVM 等核方法之外，在函数型数据分析领域，人们还开始研究直接应用于 Hilbert Rossi 和 Villa [2006] 以及 Banach Der 和 Lee [2007] 空间的大间隔分类。

欧几里得空间中的多面体分类问题有多种称法，例如，学习凸多胞体 Gottlieb 等人 [2018], 线性决策列表 Jeroslow [1975],Anthony [2004], Mangasarian [1968], 神经列表Long 和 Servedio [2006], 或阈值函数 Anthony 和 Ratsaby [2012]。许多学习

此类结构的方法（综述见 Anthony [2004] ）可视为切平面方法：在每次迭代中，通过依次添加并重新组织新的半空间或超平面来解决不一致问题。这些方法有多种变体，包括 Mangasarian [1968], Mangasarian 等人 [1990], 提出的多曲面方法。在每个阶段，该方法会确定两个彼此接近的平行超平面，使未被夹在两平面之间的点具有相同的分类。

在 Manwani 和 Sastry [2010], 中，作者提出了一种基于逻辑函数学习多面体分类器的方法。Kantchelian 等人 [2014] 提出了一种使用随机梯度下降学习多面体概念的算法。Anthony [2004] 的研究分析了基于线性决策列表的技术的泛化性质，并在标准概率学习框架下给出了泛化误差的界。Chattopadhyay 等人 [2019]最近提出了一种针对线性决策列表的下界技术，并用它证明了一个显式下界。

2 预备知识

在希尔伯特空间 $\mathcal{H}$，半空间可表示为 $\{z \in \mathcal{H} : \langle h, z\rangle + d \geq 0\}$ 其中 $h \in \mathcal{H}$ 且 $d \in \mathbb{R}$. 。这里，$\langle \cdot, \cdot \rangle$ 表示内积。由于根据上下文总能确定相应元素所属的空间，我们不会使用额外的记号来指定相应内积所关联的空间。例如，对于笛卡尔积 $\mathcal{H} \times \mathbb{R}$，中的元素 $h_i \in \mathcal{H}$ 和 $d_i \in \mathbb{R}$，，若 $\langle (h_1,\ d_1), (h_2,\ d_2) \rangle = \langle h_1, h_2\rangle + d_1 d_2$.，我们将元素 $(h_i,\ d_i)\ i = 1,\ 2$，的内积记作 $\langle (h_1,\ d_1), (h_2,\ d_2)\rangle$

大间隔与多面体。关于大间隔多面体分类，我们将部分采用先前一些关于欧几里得空间中多面体分类的研究所引入的术语。半空间 H 的外 γ-间隔是指不属于 H 且到其边界 ∂H 的距离不大于 γ. 的点集。H 的内γ-间隔是指属于 H 且到 ∂H 的距离不大于 γ. 的点集。H 的 γ-间隔是其内间隔与外 γ-间隔的并集。我们用 H^γ 表示 H 与其 γ-间隔的并集。

如果存在一个半空间 H，使得 $H \cap X^- = \emptyset$ 和 $X^+ \subset H$，并且 H 的 γ-间隔不包含 $X^- \cup X^+$. 中的任何点，我们就称两个集合 X^- 和 X^+ 线性 γ-可分。简而言之，我们也可以称 X 线性 γ-可分。

多面体 F 是有限个半空间 H_α. 的交集 $\cap_\alpha H_\alpha$。设 F^γ 为多面体$\cap_\alpha H_\alpha^\gamma$. 。$F$ 的 γ-间隔为 $F^\gamma \setminus F$. 。如果对于某个集合 X，多面体 F 包含 X^+ 且不与 X^- 相交，同时 F 的 γ-间隔也不与整个集合 X 相交，我们就称 $F\ \gamma$ 将 X^- 与X^+. 分离。t-多面体是由 t 个半空间相交而成的多面体。

实例空间。假设实例空间 $\mathcal{X} \subset \mathcal{H}$ 包含在以原点为中心的单位球内**0** 中，其中$\mathcal{H}$. 也就是说，$\|x\| \leq 1$ 对所有 $x \in \mathcal{X}$. 均成立。我们假设有两个类别标签 -1 和 **1**，并且实例空间 $\mathcal{X}$配备了一个函数 $y:\ \mathcal{X} \longrightarrow \{-1, 1\}$，该函数具有如下性质：存在 $\rho > 0$ 使得

$$\forall x, x' \in \mathcal{X} : y(x) \neq y(x') \Longrightarrow \|x - x'\| \geq \rho. \tag{1}$$

该条件表明，不同类别的两个实例不可能任意接近。也就是说，$\mathcal{X}^-$与 $\mathcal{X}^+$之间的距离非零，其中 $\mathcal{X}^-$ 和 $\mathcal{X}^+$ 分别是它们所属的类别。

此外，当考虑 $\mathcal{X}$的一个子集 X，我们用 X^- 表示所有负例的集合，即那些 $x \in X$ 的 ith $y(x) = -1$属于 并用 X^+ 表示 X中所有正例的集合，即满足 $y(x) = 1$.的那些元素

机器学习框架。谈到概率分布时，我们假设我们的空间是某个再生核希尔伯特空间，该空间基于欧几里得空间上的一个性质足够良好的核构成，并且希尔伯特空间中的实例空间是该欧几里得空间的一个紧子集 Ω 的像。在 $\mathcal{X}$ 中抽取样本（实例的有限子集），意味着从 Ω 上定义的固定概率分布中抽取样本，然后考察其在希尔伯特空间中的像。

如果 $y(x) = 1$ 蕴含 $x \in H$，且 $y(x) = -1$ 蕴含 $x \notin H$.，我们称 H of $\mathcal{H}$ 正确分类 $x \in \mathcal{X}$。在这一作用下，我们通常将 H 称为分类器。

为了估计我们的算法构造的多面体分类器的预测准确率，我们采用基于 VC 维数的可能近似正确（PAC）学习经典框架；$1 - \varepsilon$ 和 $1 - \delta$ 分别表示

预测准确率和置信度估计，其中 ε 和 δ 是在 $(0, 1/2)$. 中的给定值。对于有限维情形，线性概念的 VC 维数由空间的维数 d 界定。由 t 个半空间定义的多面体族的 VC 维数由 $d \cdot t\log\ t,$ 界定，这是一个众所周知的事实。无限维情形则更为困难，不过我们也可以给出类似的估计：在假设存在一个 t-多面体能够对实例空间进行 γ-分离的情况下，用 $O(\gamma^{-2})$替代 d。然而，我们的算法方法虽然适用于有限维情形，却不能保证算法学习到的一致多面体属于该概念类。为克服这一困难，我们采用离散化，并额外假设所讨论的希尔伯特空间是一个再生核希尔伯特空间（RKHS），且其核具有足够良好的性质。离散化使我们能够从有限概念类中进行学习。

可实现性。 任何带有二元标签的有限维数据集的像都能在某个合适的再生核希尔伯特空间中实现多面体可分性。例如，我们可以选择 RBF核 K，或选取给定核 K' 的某个合适函数 K，以保证对于所有 $x,\ x' \in \mathcal{X},\ x \neq x'.,\ K(x,\ x) = 1,\ K(x,\ x') < 1,$ 。然后，根据（1），我们可以证明，原始$\mathcal{X}$在再生核希尔伯特空间中的像可由某个 t 和 γ，所对应的 t-多面体进行 γ-可分；详见附录。

计算模型。 基本运算包括计算所涉及向量空间中的内积和范数，以及算术运算。

3 希尔伯特空间中的线性可分性

本节考虑与 $\mathcal{X}$ 的子集 X 相关联的以下严格线性不等式组 $P(X)$:

$$P(X) : y(x)(\langle x, h\rangle + d) > 0, \forall x \in X.$$

其中我们要寻找 $(h, d)\ \in \mathcal{H} \times \mathbb{R}.$。一对 (h, d) 定义了一个半空间

$$H = \{z \in \mathcal{H} : \langle z, h\rangle + d \geq 0\},$$

可用作线性分类器；如果 $x \in H,$，则它为 $x,$ 赋予标签 1；否则，赋予标签 -1。如果 (h, d) 对于 $P(X),$ 是可行的，那么 H 会为所有 $x \in X$.赋予正确标签

为求解 $P(X),$，我们使用算法 1，它是对 Dantzig [1992]中介绍的冯·诺伊曼线性规划算法的修改。它以 X 和一个按如下方式定义的状态 S 作为输入:

定义 1 状态 S 是空状态（记作 $\emptyset$），或一对(h_S, d_S)，其中 $h_S \in \mathcal{H}$ 且 $d_S \in \mathbb{R}$，它们定义了一个半空间，我们将其记作 $H_S.$。对于 $S = \emptyset$，令 $H_\emptyset = \mathcal{H}.$。也就是说， $H_\emptyset$ 是整个空间。

如果 (h_S, d_S) 是 $\{y(x)(x,1) : x \in X\}.$ 中有限子集的凸组合，则状态 $S \neq \emptyset$ 与 X 兼容。给定一个与 X 兼容的状态 S，线性规划算法会返回另一个与 $X.$ 兼容的状态 S'。线性规划算法以 (S, X, γ) 为输入所得到的最终状态 S' 记为 $\mathrm{LP}(S, X)$。 这里省略 γ，因为它始终相同。

该算法在乘积空间 $\mathcal{H} \times \mathbb{R}$ 中运行，其内积定义为 $\langle(h_1, d_1),(h_2, d_2)\rangle = \langle h_1, h_2\rangle_{\mathcal{H}} + d_1 d_2.$ 其中， $\langle\cdot, \cdot\rangle_{\mathcal{H}}$ 指原空间 $\mathcal{H}.$ 中的内积。乘积空间中的范数由其内积诱导。

在线性规划算法的 while 循环中，有一个步骤，我们称之为贡献进展的步骤。之所以这样命名，是因为每当线性规划算法执行到这一步时，它都会更新 $(h',\ d')$，使得 $\|(h',\ d')\|^{-2}$至少增加一个确定的值。由于 $\|(h',\ d')\|^{-2}$ 有一个固定的上界，该上界取决于 γ，因此我们可以估算线性规划算法的迭代次数（即其 while 循环的迭代次数）；在此，我们针对有限维情形，调整了 Chubanov [2015] 提出的算法进展分析。相关论述表述如下引理:

引理 1 关于线性规划算法，以下论述成立:

(a) 在一个贡献进展的步骤中，(h, d) 的长度平方的倒数至少增加 $y(x)(x,1)$:的长度平方的倒数。

$$\|(h', d')\|^{-2} \geq \|(h, d)\|^{-2} + \|y(x)(x, 1)\|^{-2}. \tag{2}$$

算法 1 线性规划（LP）算法 **hm**

输入：系统 $P(X)$、与 X兼容的状态 S 和 γ.输出：状态 $S'.S' : = S.$当 $S' = \emptyset$ 或 $(h_{S'}, d_{S'})$ 对于 $P(X)$ 和 $\|(h_{S'}, d_{S'})$ 不可行时$\| \geq \gamma/2$ 执行 若 $S' \neq \emptyset$则（贡献进展的步骤）$(h, d) : = (h_{S'}, d_{S'})$ 找到满足 $y(x)(\langle x, h_{S'}\rangle + d_{S'})$ 的 $x \in X \ \leq 0.$令 (h', d') 为 $(0, 0)$ 在 $[(h, d)\ y(x)(x, 1)$ 上的正交投影] 并令 S' 由 (h', d') 定义。否则 任意选取 $x \in X.(h', d') : = y(x)(x, 1)$ 令 S' 由 (h', d') 定义。结束若 若 $\|($ $h_{S'}, d_{S'})\| < \gamma/2$ 则 报告 X 不是线性 γ-可分的。结束若 结束当 返回 $S'.$

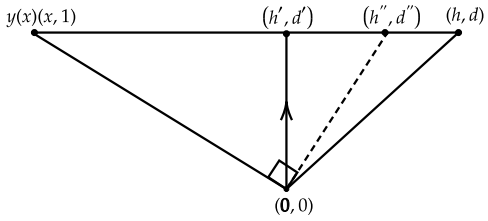

图 1：线性规划算法的迭代。

（*b*）以下 性质在算法 *1* 的执行过程中保持不变，前提是 X 是线性 γ-可分的 *:*

$$\|(h', d')\| \geq \gamma/2 \tag{3}$$

（*c*）如果算法 *1* 报告 X 不是线性 γ-可分的，则该判断是正确的。

（*d*）线性规划算法执行的迭代次数上界为 O（γ^{-2}）。

(e) 如果 X 是线性 γ-可分的，那么该算法会返回一个状态 S'，使得相应的半空间 $H_{S'}$ 能正确地分离 X^- 和 $X^+.$

证明。 (a) 这由逆勾股定理得出，该定理在希尔伯特空间中也成立。由于 x 对应于 P（X）的一个不满足的不等式，因此有

$$\langle (h, d), y(x)(x, 1)\rangle = y(x)(\langle x, h\rangle + d) \leq 0.$$

因此，该 顶点为（h， d），（$\mathbf{0}$，0），以及 y（x）（x，1）的三角形包含一个直角三角形，直角位于（$\mathbf{0}$， $0)$ 且其 另外两个顶点分别是 $y(x)(x$，$1)$（前一个三角形的一个顶点）和某个点 $(h''$， $d'')$ $\in$

$[(h,\ d)\ y(x)(x,\ 1)]$, 即前一个三角形中与 (0，0)，相对的边上的一个点。两个三角形的高度相同，即连接 (0，0) 及其在 $[(h,\ d)\ y(x)(x,\ 1)]$. 上的正交投影 $(h',\ d')$ 的线段。图 1 给出了示意图。

(b) 由于投影到线段上的结果是该线段两个端点的凸组合，因此在每次迭代中，(h', d') 都属于 $\{y(x)(x,$ 1) 的凸包：$x \in X\}$, 这意味着存在 $c\colon X \longrightarrow [0,1]$，使得

$$(h', d') = \left(\sum_{x\in X} c(x)y(x)x, \sum_{x\in X} c(x)y(x)\right), \sum_{x\in X} c(x) = 1.$$

由于 X 是线性 γ-可分的（根据 (b) 的假设），且包含在以 0 为中心的单位球中，因此存在 () 的一个解 (, d^*), 其中 $\|h^*\| = 1$ 且 $|d^*| \leq 1$，使得对所有 $x \in X.$，$y(x)(\langle x, h^*\rangle + d^*) \geq \gamma$。由此可得

$$\gamma = \sum_{x\in X} c(x)\gamma \leq \sum_{x\in X} c(x)y(x)(\langle x, h^*\rangle + d^*)$$

$$= \langle (h', d'), (h^*, d^*)\rangle \leq \|(h', d')\|\|(h^*, d^*)\|.$$

由于 $\|(h^*, d^*)\| \leq 2$，且 (h^*, d^*) 是可行的，并且对所有 $x \in \mathcal{X},\ \|x\| \leq 1$，这意味着 (3)。

(c) 由 (b) 得出，而 (d) 由 (a) 和 (b) 得出。

(e) 在这种情况下，只有找到 $P(X)$ 的一个可行解后，算法才会停止。 ∎

4 希尔伯特空间中的真多面体分离

根据支持向量机理论，总能找到一个合适的核，使得底层数据在相应的再生核希尔伯特空间中表示时是线性可分的。为此，我们可以采用例如径向基函数核。这类核称为通用核。因此，线性可分的类别已经足够一般。然而，在许多情况下，我们更倾向于使用给定的核；或者，在给定的核类中，找到一个能够实现线性分离的核在计算上可能不可行。这促使我们研究 $\mathcal{X}^-$ 和 $\mathcal{X}^+$ 可由一个多面体分离的情形。多面体分离的另一个重要方面是，正如前文所述，通过显式选择一个作为给定核函数的核，可以在合适的再生核希尔伯特空间中实现可实现性。

给定一个集合 $X \subset \mathcal{X}$，多面体分类问题的主要难点在于如何正确地将 X^- 表示为 t 个集合的并 X_i^-，$i \in [t]$，也就是说，存在一个由某些半空间 $H_i,\ i \in [t]$所定义的 γ-分离多面体，使得 $X_i^- \cap H_i = \emptyset.$。如果我们知道这样的表示，就可以通过简单地应用算法 1 t 次来解决该问题。遗憾的是，即使在 $t = 2$的情况下，适当学习的 NP 困难性也表明，除非 $P = NP.$，否则不存在能找到此类正确表示的多项式时间算法。

此时，我们需要固定一个假想的“真实”多面体，以供理论分析之用。也就是说，固定一个 γ-分离的 t-多面体 F^*，它能正确分类整个实例空间 $\mathcal{X}.$。多面体F^* 包含 $\mathcal{X}^+$，且不与 $\mathcal{X}^-.$ 相交。令 F^* 由半空间 $H_i,$ $i \in [t].$ 定义。那么，对于所有$i \in [t].$，$\mathcal{X}^+ \subset H_i$。另一方面，$\mathcal{X}^-$ 可以表示为其若干子集的并 $\mathcal{X}_i^-$，$i \in [t],$ 使得对每个 $i \in [t].$，都有 $H_i \cap \mathcal{X}_i^- = \emptyset$。每个集合$\mathcal{X}^+ \cup \mathcal{X}_i^-$都由相应的半空间 H_i线性地 γ-分离。

根据上述构造，我们只假设并集（其中各子集可以彼此重叠）

$$\mathcal{X}^- = \mathcal{X}_1 \cup \ldots \cup \mathcal{X}_t$$

假设为 b　　固定的。我们将在本节中始终使用这一记法。（半空间的记法 H_i 并不固定，还可能进一　　光因语境不同而具有不同含义。）

满足上述性质的真实多面体可能有很多，我们只假设至少存在一个。需要再次强调的是，我们只在算法的理论分析中需要这一假想表示。

定义 2 如果一个集合 A 对于不同的 $i.$均满足 $A^- \subset \mathcal{X}_i^-.$，则称其为 i-正确（注意，A 可以是 i-正确的　　）

算法 2 分离算法

输入： X、 t 和 γ. 输出：如果存在一个 γ-分离的 t-多面体，则得到一个一致的 t-多面体。$\mathcal{A}$ ： $= \{X^+\} S(X^+, 0)$： $= \emptyset$ （搜索树根节点处的初始状态）当尚未找到多面体时执行 k ： $= k + 1$ 查找是否存在 A_1、 ……、 $A_{t'} \in \mathcal{A}$ ，使得 $t' \le t$ 满足 $F = \cap_{p \in [t']} H_{S(A_p, k-1)}$ 正确分类 X.若存在这样的 F 则返回 F 结束 if（分支步骤）设置 $S(A, k)$： $= S(A, k-1)$ 对所有 $A \in \mathcal{A}$.对于所有 (A, x) $\in \mathcal{A} \times X^-$ 执行 若 $x \in H_{S(A,k-1)}$ 则 $S(A \cup \{x\}, k)$： $=$ 线性规划算法 $(S(A, k-1)\ A \cup \{x\})$ 若线性规划算法未报告 $A \cup \{x\}$ 不可线性 γ-分离，则将 $A \cup \{x\}$ 添加到 $\mathcal{A}$.结束 if 结束 for 结束 while

此外，我们将集合 X 的一个 t-划分定义为由其子集组成的元组 $(X_1$、 ……、 $X_t)$，这些子集的并集为 X. 。在此定义中，元组中的集合允许为空。

令 $\mathcal{F}_i$、 $i \in [t]$，为 $\mathcal{H}$. 中的一些集合族。考虑以下条件：

条件 1 存在一个 t $X_1 \ldots X_t$ X $i \in [t]$ $X_i \neq \emptyset$ $H_i \in \mathcal{F}_i$-划分 (、、)，使得对于每个满足的，都存在一个使得 $H_i \cap X_i = \emptyset$.

就多面体分离而言，当 $\mathcal{F}_i$ 的每个元素都是半空间或整个空间 $\mathcal{H}$ 时，条件 1 有如下推论：若条件 1 成立，则存在集合 $H_i \in \mathcal{F}_i$、 $i \in [t]$，它们的交集不包含集合 X_i. 中的任何一个。在对给定样本 X 进行多面体分离时，我们关注这样的情形：集合族 $\mathcal{F}_i$ 中的每个集合都包含 X^+，且 $\tilde{X}$ 是负实例集合 X^-。此时，若条件 1 成立，相应的交集 $\cap_{i \in [t]} H_i$ 就构成一个 t'-多面体，其中 $t' \le t$，一致地划分 X.（根据我们的构造， H_i 可以是 $\mathcal{H}$ ，因此有 $t' \le t$ ）。

在每次迭代中，我们的分离算法（算法 2）都会尝试构造一个一致的 t'-多面体，其中包含 $t' \le t$. 。如果成功，它就返回以这种方式找到的多面体；否则，对于某些适当选取的族 $\mathcal{F}_i$ （算法中隐含使用，但稍后用于理论分析），条件 1 不成立，算法便继续运行。该算法的逻辑由以下引理启发；当条件 1 不成立时，该引理成立：

引理 2 如果对于 X，条件 *1* 不成立，那么对于 X 的每个 t-划分 $(X_1, \ldots, X_t)$，都存在 $j \in [t]$，满足 $X_j \neq \emptyset$，使得

$$\forall H \in \mathcal{F}_j : H \cap \tilde{X}_j \neq \emptyset. \tag{4}$$

证明。该引理可直接由条件 1 的否定得到。 ∎

为分析算法 2，我们引入搜索树的概念。搜索树的每一层对应于算法主循环的一次迭代，其中考虑某个集合族 $\mathcal{A}$ 。第 k 次迭代结束时的集合族 $\mathcal{A}$ 对应于构成搜索树第 k 层的节点集合。搜索树的每个节点

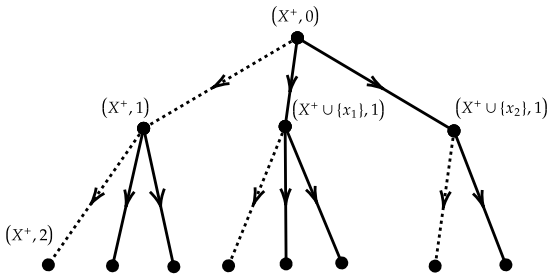

图 2：搜　　索树：实线弧表示有助于推进；对于每个父节点，如果在
分支步骤中有多条弧，那么其中一条不助于推进，其余弧都应有助于推进。

均由一对 $(A,\ k)$ 编码，其中在第 k 次迭代结束时，A 包含于 $\mathcal{A}$ 中；即 k 是搜索树中包含 $(A,\ k)$ 的层。相邻层的节点可以通过弧连接。当且仅当对某个 x.，$A = A'$ 或 $A' = A + \{x\}$ 时，存在一条从 $(A,\ k-1)$ 到 $(A',\ k)$ 的弧

当算法根据上一层的节点 $(A, k-1)$，决定是否将形如 $(A \cup \{x\}, k)$ 的新节点加入搜索树时，会求解相应的问题 $P(A \cup \{x\})$。如果线性规划算法找到一个能一致划分 $A \cup \{x\}$ 的半空间（即线性规划算法未报告 $A \cup \{x\}$ 不可线性 γ-分离），那么 $(A \cup \{x\}, k)$ 就会出现在搜索树中。算法会存储线性规划算法在搜索树相应节点处到达的状态 $S(A, k)$。概括而言，我们的算法如下：

- 给定一个样本 X，算法尝试构造一个由不超过 t 个不等式定义的分离多面体。为了“组装”这样的多面体，它尝试找到与当前层中存储的状态相关联的半空间的合适组合。

- 每次迭代时，算法都会通过添加一层来扩展搜索树。前一层的节点会保留在新的一层中。在分支步骤中，对于每个节点（$A,\ k-1$），算法都会生成一个新节点（$A \cup \{x\},\ k$），前提是 x 未被与父节点（$A,\ k-1$）相关联的半空间 $H_{S\ (A,\ k-1)}$ 分离。随后，将线性规划算法应用于问题 $P(A \cup \{x\})$，并从先前已为父节点到达的状态 $S(A,\ k-1)$开始。由于$H_{S\ (A,\ k-1)}$ 无法一致地分离 x，因此线性规划算法至少会执行一个贡献进展的步骤。我们将在算法运行时间的分析中利用这一性质。

- 搜索树具有这样的性质：与节点（$A,\ k$）关联的状态，与以（$A,\ k$）为根节点的子树中其他节点关联的状态相容。这使我们能够沿着搜索树中的每条路径追踪线性规划算法的进展。为此，我们引入“贡献进展的弧”这一概念，它对应于通过线性规划算法从一个状态转移到另一个状态，且该转移至少包含线性规划算法的一个贡献进展的步骤。

算法 2 构造的搜索树示例如图 2 所示。

回想一下，每个非空状态 S 都由一个关联对（$h_S,\ d_S$）构成，该关联对定义了一个半空间 H_S.。从 S，到达另一个状态 $S' \neq S$ 时，线性规划算法至少执行一个贡献进展的步骤，使$\|$（$h_S,\ d_S$）$\|^{-2}$ 增加一个由引理 1 的不等式 (2) 保证的特定值（$1/4$）。因此，自然可以引入以下势函数 π，并将其定义在搜索树的每个节点上：

$$\pi(A,k) := \begin{cases} \|(h_{S(A,k)}, d_{S(A,k)})\|^{-2}, & \text{if } S(A,k) \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

我们称弧 $((A,\ k-1),\ (A',\ k))$ 为贡献进展的步骤，当且仅当

$$\pi(A', k) \geq \pi(A, k-1) + 1/4.$$

当且仅当 LP 算法从 $S(A,\ k-1)$ 开始求解问题 $P(A')$时，至少执行了一次贡献进展的步骤时，该不等式成立。若 $A'=A$，则 $S(A',\ k):=S(A,\ k-1)$。因此，对于搜索树中的每条弧 $((A,\ k-1),\ (A',\ k))$，$\pi(A',\ k)\ \geq \pi(A,\ k-1)$。也就是说，$\pi$ 沿搜索树的每条路径都是非递减的。

另一方面，线性规划算法意味着

$$\pi(A,k)\leq 4\gamma^{-2}.$$

因此，搜索树中每条路径上的推进贡献弧数量不超过 $O(\gamma^{-2})$。

引理 3 在第 k次迭代结束时，$|\mathcal{A}|$ 有界于

$$((m+1)k)^{O(\gamma^{-2})}, \tag{5}$$

其中 $m=|X|$.

证明。由于搜索树中的每条路径至多包含 $O(\gamma^{-2})$条推进贡献弧，因此，从根节点到树的第 k层的每条路径都可以唯一地编码为一个 k维向量，其分量取值为 0 到 m 之间的整数，且非零分量的数量不超过 $O(\gamma^{-2})$。式 (5) 是这类 k维向量总数的上界。同时，这也是搜索树第 k层节点数量的上界，即第 k次迭代结束时 $|\mathcal{A}|$的上界。 ■

现在我们回到 i-正确性这一概念（定义 2），并将其扩展为路径的 i-正确性概念：进一步地，如果该路径上的每个节点（$A,\ k$）都满足 A 为 i-正确，则称搜索树中的这条路径为 i-正确。

引理 4 若(A,k) 是搜索树中的一个节点，且 A 是 i-正确的，那么连接根节点与(A,k) 的路径是i-正确的。

证明。从根节点出发的路径形式如下

$$(A_0,0)=(X^+,0),(A_1,1),\ldots,(A_k,k)=(A,k).$$

注意，$A_l\subseteq A_{l+1},\ l=0,\ \ldots\ldots,\ k-1$. 这意味着，如果某个 A_l 不是 i-正确的，那么 A_k 就不是 i-正确的，因为A_l 包含于 A_k，这与 A 是 i-正确的假设相矛盾。 ■

定理 1 如果存在一个一致的 γ-分离 t-多面体，那么算法 2 的运行时间有如下上界

$$m^{O(t\gamma^{-2}\lceil\log_m(\gamma^{-1})\rceil)}. \tag{6}$$

证明。考虑第 k. 次迭代。令 $X=X^-.$ t $X_1\ \ldots\ X_t$ X X_i。考虑一个 -划分 (,), 使得每个非空的都是 i-正确的。对于每个满足 $X_i\neq\emptyset$ 的 $i\in[t]$，令 $\mathcal{F}_i$ 为与状态 $S(A,k-1)$ 相关联的所有半空间的族，这些状态对应于 $\mathcal{A}$. 中的 i-正确集合 A。对于每个满足 $X_i=\emptyset$ 的 i，令 $\mathcal{F}_i=\{\mathcal{H}\}$.

令 $\mathcal{A}_i$ 表示 $\mathcal{A}$. 中所有 i-正确集合构成的族。注意，对于每个 $i\in[t]$，族 $\mathcal{A}_i$ 都非空，因为 e it 至少包含 X^+.

假设该算法在当前迭代中未能找到具有 $t'\leq t$ 的一致 t'-多面体。则无法由族 $\mathcal{F}_i,\ i\in[t]$. 组装出任何一致的 t-多面体。因此，条件 1 不满足。于是，引理 2 表明，存在 $j\in[t]$，使得对于所有 j-正确集合 $A\in\mathcal{A}.$，都有 $X_j\cap H_{S(A,k-1)}\neq\emptyset$。这意味着，对于每个 $A\in\mathcal{A}_j$，当前迭代 k 应构造一条贡献进展的弧，将$(A,\ k-1)$与$(A',\ k)$连接起来，其中 $A'=A\cup\{x\},\ x\in X_j.$。也就是说，$A'$也是 j-正确的。将此 j 记为 j_k，其中 k 是当前迭代的编号。

现在考虑算法所有迭代中的指标序列 $j_1,\ j_2,\ \ldots\ldots$。考虑其中等于某个 $r\in[t]$ 的指标构成的子序列：$j_{k_1}=r,\ j_{k_2}=r,\ \ldots\ldots$。这里，$k_1,\ k_2,\ \ldots\ldots$分别是搜索树的相应层。对于每个 k_s，搜索树中的前一层 k_s-1 都包含一个节点（$A,\ k_s-1$），其中 A 是 r-正确的。引理 4

算法 3 非适当分离算法

输入： X 和 γ.输出：一个与 X 一致的多面体，前提是存在一个 γ-分离多面体。$\mathcal{A} := \{X^+\}S(X^+, 0)$： $= \emptyset$ （搜索树根节点处的初始状态）当未找到多面体时执行$k := k + 1F := \cap_{A\in\mathcal{A}} H_{S(A,k-1)}$如果 F 正确地分离了 X^- 并且 X^+，则返回 F结束如果任意选取一个被 F.错误分类的 $x \in X$ （$x \in X^-$，因为 $X^+ \subset F$.（分支步骤）设置 $S(A, k)$： $= S(A, k-1)$ 对所有 $A \in \mathcal{A}$.对所有 $A \in \mathcal{A}$ 执行$S(A \cup \{x\}, k)$： $=$ LP$(S(A, k-1)\ A \cup \{x\})$ 如果线性规划算法未报告 $A \cup \{x\}$ 不能线性地 γ-可分，则将 $A \cup \{x\}$ 添加到 $\mathcal{A}$.结束如果结束循环

这意味着从根节点到 $(A, k_s - 1)$ 的路径是 r-正确的。我们的构造意味着，从根节点到层 $k_s - 1$ 的每条 r-正确路径，都通过一组弧与层 k_s 相连，其中至少有一条弧是贡献进展的，且形式为 $((A, k_s - 1), (A', k_s))$，其中对于每个 $s.$， A' 都是 r-正确的。因此，在搜索树的所有层中，存在一条 r-正确的路径，其中包含的推进贡献弧至少与该子序列的元素数量相同。

假设迭代次数为无限，或超过 $4t\gamma^{-2}.$。那么，对于某个 $l,$，等于 l 的索引子序列包含的元素多于 $4\gamma^{-2}$。这会导致矛盾，因为在这种情况下，搜索树应当包含一条 l-正确的路径，且其中有超过 $4\gamma^{-2}$ 条推进贡献弧。因此，迭代次数不超过 $4t\gamma^{-2}.$。算法停止时，会返回一个一致的 t'-多面体，其中包含 $t' \le t$，因为这是它唯一可能停止的条件。

每当线性规划算法由算法 2 调用时，所需的基本运算次数上界为 $O(m\gamma^{-2})$，其中样本量 $m = |X|$ 出现的原因是，在每次迭代中，线性规划算法都需要找到如下形式的问题 $P(A \cup \{x\})$ 中一个违反的约束，其中 $(A \cup \{x\}) \subseteq X.$ 在每次迭代中，构造 F 或确认其不存在所需的时间上界为 $O(|\mathcal{A}|^t)$。由于算法2的迭代次数上界为 $O(t\gamma^{-2})$，由引理 3 可知，将 (5) 中的 k 替换为 $t\gamma^{-2}.$ 后， $|\mathcal{A}|$ 的上界为 (5)。将γ^{-2} 替换为 $m^{\log_m(\gamma^{-2})}$ 并重新整理各项，即可得到复杂度界 (6)。 ∎

现在，我们基于类似的思路构造另一种算法。与算法 2 不同，它将当前迭代中所有可用半空间的交集构造为 F。如果 F 与给定样本 X 不一致，那么它会任意选择 X 中的一个元素 x，并由 F 为其赋予错误标签。由于所有半空间都包含 X^+，因此可得 $x \in X^-.$。由于 $x \in F,$ 可得，对于当前迭代开始时的所有 $A \in \mathcal{A}$， $x \in H_{S(A,k-1)}$。这一观察使我们能够证明，算法 2 的迭代次数上界同样适用于算法3。

定理 2 如果存在一个能够正确分类实例空间的 γ-分离 t-多面体，那么算法 *3* 会在以下时间内找到一个一致多面体

$$O\left(m \cdot (2t\gamma^{-2})^{4\gamma^{-2}}\right) \tag{7}$$

定义该多面体的半空间数量上界为

$$O((8t\gamma^{-2})^{4\gamma^{-2}}). \tag{8}$$

证明。 迭代次数的上界与算法 2 相同。这可以通过与定理 1 证明中所用的相同论证得出。唯一的区别在于多面体 F 的构造方式，以及找不到多面体 F 时得出的结论；在算法 3 中，我们只需检查由前一层到达的所有状态所对应的半空间交集构成的多面体 F，是否是一致多面体。如果不是，那么每条弧 ($(A,\ k-1)$，$(A\cup\{x\},\ k)$) 都是贡献进展的步骤，因为 $x\in X^-$ 包含于每个半空间 $H_{S(A,k-1)}$，这意味着在线性规划算法从 $S(A,\ k-1)$ 转移到 $S(A\cup\{x\},\ k)$ 的过程中，至少执行了一次贡献进展的步骤。 现在，我们可以用与定理 1 证明中相同的方式推导出迭代次数的上界 $4t\gamma^{-2}$。为估计每次迭代结束时的 $|\mathcal{A}|$，我们使用 (5)，其中应将 k 替换为 $4t\gamma^{-2}$，将 $m+1$ 替换为 2，因为算法 3 构造的搜索树是二叉树，这意味着路径可以用二进制向量编码。 ∎

注。若不存在一致的、 γ-分离的 t-多面体，则两种算法的迭代次数都将是无限的。若存在这样的多面体，则两种算法的迭代次数均由 $4t\gamma^{-2}$. 限定。因此，我们可以增加一步来验证 $k\le 4t\gamma^{-2}$. 是否成立。若不成立，则不存在一致的、 γ-分离的 t-多面体。

5 PAC 多面体分类

若 $\mathcal{H}$ 是有限维的，则算法 2 返回的 t-多面体的 VC 维数由 $O(dt\log\ t)$。为了获得算法 3 返回的多面体的 VC 维数，应将上述估计中的 t 替换为相应的界 (8)。因此，我们得到以下推论：

推论 1 设 $\mathcal{H}$ 是一个维数为 $d<\infty$. 的空间。则以下陈述成立：

- 算法 *2* 可在时间 $m^{O(t\gamma^{-2})}$ 内，从 *VC* 维数 D 由 $O(dt\log\ t)$ 限定的概念类中构造一个 *PAC* 多面体。
- 算法 *3* 可在时间 *(7)* 内，从 *VC* 维数被$O((\gamma^{-2}+\log t)(t\gamma^{-2})^{4\gamma^{-2}})$界定的概念类构造一个 *PAC* 多面体。

因此，上述 PAC 学习可通过样本复杂度为 $O(\varepsilon^{-1}(D+\log(\delta^{-1}))$，来实现，其中 D 是相应的 VC 维数。

对于无限维情形，我们证明如下结论：

推论 2 设 $\mathcal{H}$ 是一个具有连续核 K. 的再生核希尔伯特空间。设 $\mathcal{X}$ 是紧集 Ω 在 $\mathbb{R}^s$ 下经该再生核希尔伯特空间的特征映射 φ 所得的像。算法 *2* 和算法 *3* 均可在与前述相同的假设下实现，以生成 *PAC* 多面体。样本复杂度可由包含Ω γ^{-1}的空间维数 s 以多项式方式界定，且 t.

证明。 主要思路是采用众所周知的离散化技巧，即用一个具有足够小步长的有限集$\Omega^{\#}$ 替换 Ω。随后修改算法，将算法运行过程中出现的 $\mathcal{H}$ 中的元素 x 替换为其在像 $\varphi(\Omega^{\#})$中的近似。 当然，由于我们只需要 K 计算内积，因此不会显式计算 φ。更多细节请参见附录。 ∎

6 总结

我们提出了一种用于希尔伯特空间（例如再生核希尔伯特空间）的 PAC 学习算法。该算法学习多面体概念，因此具有普适适用性，因为总能显式选取一个合适的核，使多面体概念在相应的再生核希尔伯特空间中可实现。

参考文献

Martin Anthony。阈值决策列表的泛化误差界。机器学习研究杂志，5：189–217，2004年。Martin Anthony 和 Joel Ratsaby。数值数据逻辑分析中的稳健切分点。离散应用数学，160(4-5)：355–364，2012年。Kristin P. Bennett 和 Erin J. Bredensteiner。SVM 分类器中的对偶性与几何。载于国际机器学习大会，第 57–64 页，2000年。Anselm Blumer、Andrzej Ehrenfeucht、David Haussler 和 Manfred K. Warmuth。奥卡姆剃刀。信息处理快报，24(6)：377–380，1987年。doi：10.1016/0020-0190(87)90114-1。Anselm Blumer、Andrzej Ehrenfeucht、David Haussler 和 Manfred K. Warmuth。可学习性与 Vapnik–Chervonenkis 维数。*J. ACM*，36(4)：929–965，1989年。doi：10.1145/76359.76371。Arkadev Chattopadhyay、Meena Mahajan、Nikhil S. Mande 和 Nitin Saurabh。线性决策列表的下界。计算研究库，abs/1901.05911，2019年。Sergei Chubanov。用于线性可行性问题的多项式投影算法。数学规划，153：687–713，2015年。G. Dantzig。一个与问题规模无关、经 $1/\epsilon^2$ 次迭代即可获得的满足凸性约束的线性规划 ϵ-精确可行解。技术报告 *SOL 92-5*，1992年。Ricky Der 和 Daniel Lee。巴拿赫空间中的大间隔分类。载于第 *11* 届国际人工智能与统计会议（人工智能与统计会议 *2007*）论文集，第 91–98 页，2007年。Lee-Ad Gottlieb、Eran Kaufman、Aryeh Kontorovich 和 Gabriel Nivasch。带间隔的凸多胞体学习。载于神经信息处理系统进展第 *31* 卷：*2018* 年神经信息处理系统年度会议，第 5711–5721 页，2018年。Lee-Ad Gottlieb、Eran Kaufman、Aryeh Kontorovich 和 Gabriel Nivasch。带间隔的凸多面体学习。*IEEE* 信息论汇刊，68：1976–1984，2022年。Robert G. Jeroslow。关于用线性不等式定义超立方体顶点集合。离散数学，11(2)：119–124，1975年。Alex Kantchelian、Michael C Tschantz、Ling Huang、Peter L Bartlett、Anthony D Joseph 和 J. D. Tygar。大间隔凸多胞体机。载于*2014* 年神经信息处理系统大会论文集，第 3248–3256 页。2014年。Adam R. Klivans、Ryan O’Donnell 和 Rocco A. Servedio。通过高斯表面积学习几何概念。载于*2008* 年第 *49* 届 *IEEE* 计算机科学基础研讨会，第 541–550 页，2008年。doi：10.1109/ FOCS.2008.64。Philip M. Long 和 Rocco A. Servedio。在非集中分布下对决策列表和线性阈值函数进行属性高效学习。载于 *2006* 年第 *20* 届神经信息处理系统年度会议论文集，第 921–928 页，2006年。O. L. Mangasarian。模式分离的多曲面方法。*IEEE* 信息论汇刊，14(6)：801–807，1968年。O.L. Mangasarian、R. Setiono 和 W. H. Wolberg。通过线性规划进行模式识别：理论及其在医学诊断中的应用，1990年。

Naresh Manwani 和 P. S. Sastry。使用逻辑函数学习多面体分类器。载于*ACML*，*JMLR Proceedings*第 13 卷，第 17–30 页。JMLR.org，2010年。Michael E. Mavroforakis 和 Sergios Theodoridis。支持向量机（SVM）分类的几何方法。*IEEE Trans. Neural Networks*，17(3)：671–682，2006年。N. Megiddo。关于多面体可分性的复杂度。技术报告 RJ 5252，IBM Almaden Research Center，1996年。Fabrice Rossi 和 Nathalie Villa。用于函数数据分类的支持向量机。*Neurocomputing*，69(7-9)：730–742，2006年。Santosh S. Vempala。一种用于学习半空间交集的基于随机采样的算法。*J.ACM*，57(6)，2010年11月。ISSN 0004-5411。doi：10.1145/1857914.1857916。URL https://doi.org/10.1145/1857914.1857916。

7 附录

7.1 多面体分类的普适性

我们首先考虑这样一种特殊情况：$\mathcal{X}$是布尔立方体 $\{0,1\}^n$.的子集。$\mathcal{X}$划分为 $\mathcal{X}^-$ 和 $\mathcal{X}^+$的任意划分，都可以通过适当的合取范式（CNF）进行正确分类；该合取范式可以表示为一个有限的线性不等式系统，其解集包含 $\mathcal{X}^+$，且不包含 $\mathcal{X}^-$. 中的任何实例。设 $Ax \leq b$是这样一个具有 t 个不等式的系统，其中 A 是系数矩阵，$b \in \mathbb{R}^t$. 。于是，$\mathcal{X}^-$ 可以表示为$\mathcal{X}^- = \mathcal{X}_1 \cup \cdots \cup \mathcal{X}_t$，其中 $\mathcal{X}_i$由 $\mathcal{X}^-$ 中违反第 i个约束的点组成。设 $2\gamma_i$为 $\mathcal{X}_i^-$ 的凸包与 $\mathcal{X}^+$. 的凸包之间的距离。令 $\gamma = \min_i \gamma_i$. 。则 $\mathcal{X}^-$ 可以通过由 $Ax \leq b$.定义的 t-多面体与 $\mathcal{X}^+$ 实现 γ-分离。

因此，如果我们处理这样一个实例空间，其中实例由长度不超过某个 d，的二进制向量编码，那么上述二进制情形表明，此类数据可以通过二进制编码空间中的适当多面体进行正确分类。

现在假设我们需要在某个维数为 d. 的欧几里得空间中，将以原点为中心的单位球内的某个子集 Ω 一致地分离出来。我们还假设，相反类别 Ω^- 和 Ω^+ 是可区分的，即存在 $\rho > 0$，使得对于所有 $\omega_0 \in \Omega^-$ 和 $\omega_1 \in \Omega^+$. ，都有 $\|\omega_0 - \omega_1\| \geq \rho$ 。令 $\mathcal{X}$ 由向量 $(\omega, \sqrt{1 - \|\omega\|^2})$ $\omega \in \Omega$ 组成。因此，在相应的映射下，$\mathcal{X}$ 是 Ω 在球面上的像，位于空间 $\mathbb{R}^{d+1}$. 中。集合 $\mathcal{X}^-$ 可以由 $\mathbb{R}^{d+1}$ 中有限个彼此不相交的较小球体 B_i 覆盖。记这些球体的数量 B_i 为 t. 。每个球体 B_i 与球面的交集都可以与 $\mathcal{X}^+$，线性 γ_i-分离，其中 $\gamma_i > 0$. 为某个值。令 γ 为所有 γ_i，$i \in [t]$. 中的最小值。于是，$\mathcal{X}^+$ 包含于一个 t-多面体中，该多面体将 $\mathcal{X}^-$ 与 $\mathcal{X}^+$. γ-分离。

上述构造也适用于希尔伯特空间。下面我们给出另一种构造，它将整个空间映射到另一个希尔伯特空间中的一个球面上。

Let $\tilde{\mathcal{H}}$ 表示正定核的再生核希尔伯特空间

现在我们希望对空间 $\mathcal{H}$ 进行同胚变换，使其像属于另一个空间中的某个球面。为此，我们引入一个新核 K，定义如下：

$$K(x, x') = \frac{\tilde{K}(x, x') + 1}{\sqrt{\tilde{K}(x, x) + 1}\sqrt{\tilde{K}(x', x') + 1}}.$$

新的核 K 也是正定的。按此方式定义 K，不会破坏原有 再生核希尔伯特空间 $\mathcal{H}$. 的拓扑。因此，令 $\mathcal{H}$ 为 K. 的再生核希尔伯特空间。若 $\mathcal{X}$ 是 $\mathcal{H}$中的一个实例空间，那么它在 $\mathcal{H}$ 中的像 $\mathcal{X}$ 位于以原点为中心的单位球面 上，因为 $x \in \mathcal{X}$ 的像的范数平方为 $K(x, x) = 1$. 。若 $\mathcal{X}$ 是欧几里得空间中某个紧集的像，且相反类别可以区分，那么与之前类似，我们可以得出结论：$\mathcal{X}^-$ 和 $\mathcal{X}^+$ 可以被 γ-分离，其中分离边界是 $\mathcal{H}$中的一个多面体，对于某个 $\gamma > 0$.

7.2 PAC 学习

线性规划算法从 $\mathcal{H}$. 的一个有限维子集中隐式地进行学习。设 $\mathcal{X}$ 是线性 γ-可分的。现在，对给定样本 $X.$，从空状态开始运行线性规划算法。令 S 为到达的状态。注意，该状态与整个实例空间 $\mathcal{X}$. 相容。假设我们有一个假想的分离预言机，能够返回由所有可能实例构成的系统 $P(\mathcal{X})$ 中违反的约束。再次运行线性规划算法，从 S开始，针对 $\mathcal{X}.$。从空状态开始，这两次线性规划算法的运行经过状态 S，最终到达状态 S'，总共进行了 $O(\gamma^{-2})$ 次迭代。每次迭代中，线性规划算法都会将旧的 $(h,\ d)$ 与 $y(x)(x,\ 1)$，的凸组合构造成一个新的 $(h,\ d)$，其中 $x \in \mathcal{X}.$。由此可知，$h_{S'}$ 是 $\mathcal{X}$的一个有限子集的凸组合，该子集的基数不超过 $O(\gamma^{-2})$。换言之，$h_{S'}$ 位于一个子空间中，其维数不超过$O(\gamma^{-2})$。因此，对于任何给定样本，线性规划算法都能从一类具有有限 VC 维数的适当线性概念中学习。当 $\mathcal{X}$线性 γ-可分时，这一类包含一个真实线性概念，能够正确划分整个实例空间 $\mathcal{X}$.

与.....不同 有限维情形，我们无法直接得出结论，认为我们的算法构造了一个多面体 ral 分类器，属于 一个 VC 维数有限的概念类。因此，我们需要一些额外的假设。

假设 $\mathcal{H}$ 是一个核为 K 的再生核希尔伯特空间，并且 $K\colon \mathbb{R}^s \times \mathbb{R}^s \longrightarrow \mathbb{R}$ 连续可微，且 Ω 是立方体 $[0,1]^s$ 在 $\mathbb{R}^s$. 中的子集。在这种情况下，K 在 Ω 上是利普希茨连续的。

因此，令 φ 为 $\mathcal{H}$. 的特征映射。它在 Ω 上是利普希茨连续的。于是，存在某个常数 L,

$$\|\varphi(\omega) - \varphi(\omega + \Delta\omega)\| \leq L\|\Delta\omega\|.$$

对于 ω，令 $\omega^{\#}$ 表示由形如 βz 的点所构成的格中与其最近的元素，其中 $z \in \mathbb{Z}^s$，而 β 是一个足够小的常数，我们将在下文计算其值。

注意到

$$\|\omega - \omega^{\#}\| \leq \sqrt{s}\beta.$$

于是，选择

$$\beta \leq \gamma/(2\sqrt{s}L),$$

我们得到

$$\|\varphi(\omega) - \varphi(\omega^{\#})\| \leq \gamma/2.$$

令 B 为以原点为中心的单位球，并令 $proj_B$ 表示到 $B.$ 上的正交投影。于是,

$$\|\varphi(\omega) - proj_B(\varphi(\omega^{\#}))\| \leq \gamma/2,$$

因为 $\varphi(\omega)$ 属于 B，根据我们的假设，$\mathcal{X}$ 位于 B，且投影到 B 不会增大到 $\varphi(\omega)$ 的距离。

由此可知，定义 γ-分离 t-多面体的半空间的所有 $\gamma/2$-间隔中，没有一个包含 $proj_B(\varphi(\omega^{\#}))$. 因此，我们可以始终假设 $proj_B(\varphi(\omega^{\#}))$ 的标签是 $y(\varphi(\omega))$。我们的离散化实例空间是 $\mathcal{X}^{\#} = proj_B((\varphi(\Omega^{\#}))$，其中 $\Omega^{\#} = \{\omega^{\#} : \omega \in \Omega\}$. $\mathcal{X}^{\#}$ 的相反类别可以由一个 t- 多面体进行 $\gamma/2$- 分离。

设 X 是给定的样本，$X^{\#}$ 表示所有 $proj_B(\varphi(\omega^{\#}))$ 构成的集合，其中 ω 遍历 Ω 中编码 $X.$ 各元素的元素。注意，我们不需要显式给出 φ。所有计算都可以仅使用内积（因此也使用 K）以及相应元素 $X.$ 的编码 ω 来完成。

如果 X 是线性 γ-可分的，那么 $X^{\#}$ 是 $\gamma/2$-可分的。如果我们检测到它不可分，那么原始样本 X 就不可分。我们在下文提出的线性规划算法的修改思路是：在训练过程中，只考虑 $X^{\#}$ 中的元素，直到构造出一个分离超平面（此时，由于 $\mathcal{X}^{\#}$是有限的且系数经过了适当舍入，我们是从一个有限概念类中学习），或者 $(h,\ d)$ 的长度足够小，从而可以断定 $X^{\#}$ 不是线性 $\gamma/2$-可分的，因此 X 不是 γ-线性可分的。

线性规划算法中贡献进展的步骤应修改如下：

- 令 $\hat{}x \in X$ 对应于 $P(X)$ 中被 $(h,\ d)$ 违反的一项约束：$\hat{}x$ 由其编码 $\omega \in \Omega$。表示
- 设 $x = proj_B(\varphi(\omega\#))$.

- 令 (ˆ ˆh， d) 是 (0， 0) 在线段 [(h， d) y(x)(x， 1)]. 上的投影。它由以下公式计算 la

$$(\hat{h},\hat{d}) = \alpha(h,d) + (1-\alpha)y(x)(x,1)$$

其中

$$\alpha = \frac{\langle y(x)(x,1), y(x)(x,1)-(h,d)\rangle}{\|y(x)(x,1)-(h,d)\|^2}.$$

- 如果 $\|$(ˆh， ˆd)$\| \le \gamma/4$，那么返回：由于在此情况下， $X_\#$ 不可线性 $\gamma/2$-分离，因此 X 不可线性 γ-分离。

- 通过舍入 α 计算 (h'， d')：

$$(h',d') := \beta\lfloor\alpha/\beta\rfloor(h,d) + (1-\beta\lfloor\alpha/\beta\rfloor)y(x)(x,1).$$

请注意， x 不一定违反与 ˆx. 相关的约束。只要势函数 π 意义上的相应进展得以保持即可。

将相应变体的线性规划算法称为修改后的线性规划算法。
还需要确定一个 β 的值，以保证

$$\|(h',d')\|^{-2} \ge \|(h,d)\|^{-2} + const, \tag{9}$$

其中 $const$ 是某个正常数。
注意，$\|$(h'， d')$\| \ge \|$(ˆh， ˆd)$\|$，因为(h， d)是(0， ˆ 0)在相应线段上的正交投影，而(h'， d')位于同一线段上。

首先，我们估计非负差值 $\Delta = \|$(h'， d')$\|^2 - \|$(ˆh， ˆd)$\|^2$ 需要多小，才能确保ˆ ˆ (9)。我们将利用以下事实：在计算(h'， d')的阶段，我们有 $\|$(h， d)$\| \ge \gamma/2$：

$$\begin{aligned}
\|(h',d')\|^{-2} &= (\|(\hat{h},\hat{d})\|^2 + \Delta)^{-1} \\
&= (1+\Delta\|(\hat{h},\hat{d})\|^{-2})^{-1}\cdot\|(\hat{h},\hat{d})\|^{-2} \\
&\ge (1+4\Delta\gamma^{-2})^{-1}\cdot\|(\hat{h},\hat{d})\|^{-2} \\
&= \left(1-\frac{4\Delta\gamma^{-2}}{1+4\Delta\gamma^{-2}}\right)\|(\hat{h},\hat{d})\|^{-2} \\
&\ge (1-4\Delta\gamma^{-2})\|(\hat{h},\hat{d})\|^{-2} \\
&\ge (1-4\Delta\gamma^{-2})(\|(h,d)\|^{-2}+1/4) \\
&= \|(h,d)\|^{-2} + 1/4 - 4\Delta\gamma^{-2}(\|(h,d)\|^{-2}+1/4) \\
&= \|(h,d)\|^{-2} + 1/4 - \Delta(4\gamma^{-2}\|(h,d)\|^{-2}+\gamma^{-2}) \\
&\ge \|(h,d)\|^{-2} + 1/4 - \Delta(16\gamma^{-2}+\gamma^{-2}).
\end{aligned}$$

因此，只需选择 β，使得 $\Delta \le \gamma^2/(8\cdot 17)$，在这种情况下，(9) 对 $const = 1/8$. 成立。考虑到 $\|$(x， 1)$\|^2 \le 2$，我们注意到 $\Delta \le 4\beta^2$. 。因此，我们可以选择 $\beta \le \gamma^2/\sqrt{4\cdot 8\cdot 17}$. 。结合前面的界，我们可以设定

$$\beta = \min\{\gamma^2/\sqrt{4\cdot 8\cdot 17}, \gamma/(2\sqrt{s}L)\}.$$

修改后的线性规划算法最多执行 $O(\gamma^{-2})$ 次迭代。相应的线性半空间类是有限的。其中每个半空间都由不超过 $O(\gamma^{-2})$ 个形如 ($proj_B$($\omega^\#$)， 1)$\omega \in \Omega$。的点的线性组合定义。

现在不难估计此类组合的数量，同时考虑系数量化 α以及 β.的选择

注意， $|\Omega^{\#}|$ 有界于 $O(\beta^{-s})$，因为Ω是 $[0,1]^s$.的子集

系数 $\beta\lfloor\alpha/\beta\rfloor$ 有 $O(\beta^{-1})$ 种变体， $|\Omega^{\#}|$ 有 x. 种可能的选择。因此，无论所考虑的样本为何，修改后的线性规划算法都会从 $(\beta^{-1})^{O(s\gamma^{-2})}$ 种变体中选择线性组合。因此，相应半空间数量的对数上界为 $O(\gamma^{-2}s\log^2(\beta^{-1}))$. 这给出了修改后的线性规划算法可能构造的半空间类的 VC 维数上界。要获得基于该半空间类构造的 t-多面体类的 VC 维数，还应将此上界乘以 t.随后，将“PAC 多面体分类”一节中的 D 替换为上述 VC 维数上界，即可得到样本复杂度。对于非适当分离算法，我们使用 $t^{O(\gamma^{-2})}$ 代替 t.