arXiv:2301.01434v1 [cs.LG] 4 Jan 2023

光滑函数的在线学习

Jesse Geneson 和 Ethan Zhou

2023年1月5日

摘要

本文研究实值函数的在线学习，其中隐藏函数已知具有某些光滑性。具体而言，对于 $q \geq 1$，令 $\mathcal{F}_q$ 为绝对连续函数 $f\colon [0,1] \to \mathbb{R}$ 的类，使得 $\|f'\|_q \leq 1$。对于 $q \geq 1$ 和 $d \in \mathbb{Z}^+$，令 $\mathcal{F}_{q,d}$ 为函数 $f\colon [0,1]^d \to \mathbb{R}$ 的类，其中由 f 固定除一个参数以外的所有参数所形成的任意函数 $g\colon [0,1] \to \mathbb{R}$ 均属于 $\mathcal{F}_q$。对于任意实值函数类 $\mathcal{F}$ 和$p > 0$，令 $\text{opt}_p(\mathcal{F})$ 表示学习者在最坏情况下能够保证的绝对预测误差的 p^{th} 次幂之和的最佳上界。在单变量情形下，我们为 $\text{opt}_p(\mathcal{F}_q)$ 找到了新的界，其精确到一个常数因子。我们证明，对于所有 $\varepsilon \in (0,1)$，都有 $\text{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Theta(\varepsilon^{-\frac{1}{2}})$，且对于所有 $q \geq 2$，都有 $\text{opt}_{1+\varepsilon}(\mathcal{F}_q) = \Theta(\varepsilon^{-\frac{1}{2}})$。我们还证明，对于 $\varepsilon \in (0,1)$，有 $\text{opt}_2(\mathcal{F}_{1+\varepsilon}) = \Theta(\varepsilon^{-1})$。此外，我们通过证明当 $q \in (1,2)$ 且 $p \geq 2 + \frac{1}{q-1}$时 $\text{opt}_p(\mathcal{F}_q) = 1$，得到了新的精确结果。在多变量情形下，我们建立了 $\text{opt}_p(\mathcal{F}_{q,d})$ 与 $\text{opt}_p(\mathcal{F}_q)$ 之间的不等式，并证明当 $p < d$ 时 $\text{opt}_p(\mathcal{F}_{\infty,d})$ 为无穷大，而当 $p > d$ 时为有限值。对于试验次数有界的情形，我们还得到了关于学习 $\mathcal{F}_{\infty,d}$ （其中 $p < d$ ）的紧确界。

1 引言

设想有一位学习者，希望根据当天的温度范围、湿度、大气压、降水、风速、太阳辐射、地点和一年中的时间等输入，预测某一地点次日的温度范围。在我们的模型中，这位学习者每天接受测试。在某一天，学习者获得当天的输入，并据此输出对次日温度范围的预测；次日到来时，它看到正确的温度范围，然后利用这一反馈更新未来的预测。随着这一过程不断重复，学习者会积累信息，帮助自己做出更好的预测。于是自然会产生一个问题：学习者能否保证其预测会随着时间推移而变得更好？如果可以，速度有多快?

我们研究一种实值函数的在线学习模型，该模型此前已在 [9,12, 13, 1, 10, 11] 中进行过研究，其中算法 A 在若干次试验中从某个函数类 $\mathcal{F}$ 学习实值函数 f 。以往关于该模型的研究集中于单个输入的函数，例如仅根据一年中的时间预测温度范围。研究表明，只要函数足够光滑，学习者就能相当迅速地成为优秀的预测者。设$\mathcal{F}$由函数 f组成： $S \to \mathbb{R}$ 对于某个集合 S，并固定某个 $f \in \mathcal{F}$。在每次试验 $t = 0$、.....、 m中，A接收输入 $s_t \in S$，猜测 y_t作为 $f(s_t)$ 的值，并接收 $f(s_t)$ 的实际值。

按照 [9], ，我们关注一种误差函数，它衡量学习者在最坏情况下准确预测函数的难度。该误差函数取决于两个参数， p和 q，它们分别决定学习者因出错而受到惩罚的严厉程度，以及学习者可能遇到的函数类型。较小的 p 和 q 对学习者而言

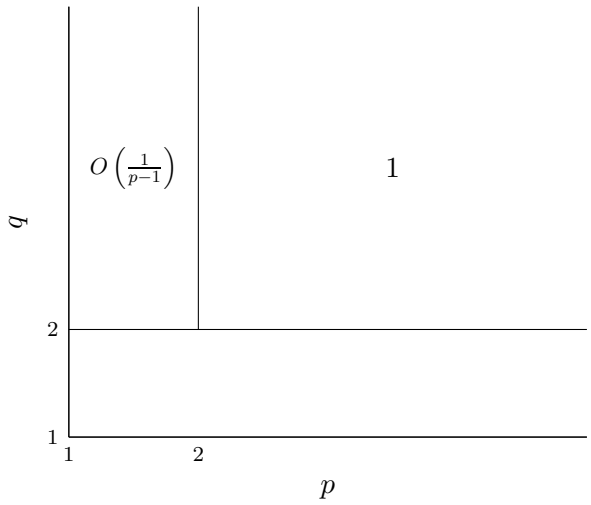

图1：精确值 以及本文结果之前 $\mathrm{opt}_p(\mathcal{F}_q)$ 的界，其中 $p, q > 1$ r

更难，从而导致误差函数的值更高。对于每个算法 A、 $p > 0$、 $f \in \mathcal{F}$ 和$\sigma = (s_0, \ldots\ldots, s_m) \in S^{m+1}$，定义

$$\mathscr{L}_p(A, f, \sigma) = \sum_{t=1}^{m} |\hat{y}_t - f(s_t)|^p.$$

当 f 和 σ 的含义可由上下文明确时，我们将 $\mathscr{L}_p(A, f, \sigma)$ 称为 A 的总 p 误差。定义

$$\mathscr{L}_p(A, \mathcal{F}) = \sup_{f \in \mathcal{F}, \sigma \in \cup_{m \in \mathbb{Z}^+} S^m} \mathscr{L}_p(A, f, \sigma)$$

并且 $\mathrm{opt}_p(\mathcal{F}) = \inf_A \mathscr{L}_p(A, \mathcal{F})$。请注意，与 [9, 12, 5] 中给出的 opt 定义不同， $\mathcal{F}$ 可以由定义在任意定义域上的实值函数组成，而不只是从 $[0,1]$ 到 $\mathbb{R}$ 的函数。

对于 $\mathcal{F}$ 包含函数 $f: [0,1] \to \mathbb{R}$ ，且其导数具有各种有界范数的情形， [9, 12, 5] 对此进行了研究。对于 $q \geq 1$, 令 $\mathcal{F}_q$ 为满足 $\int_0^1 |f'(x)|^q \mathrm{d}x \leq 1$ 的绝对连续函数$f: [0,1] \to \mathbb{R}$ 类，并令 $\mathcal{F}_\infty$ 为满足 $\sup |f'(x)| \leq 1$ 的绝对连续函数 $f: [0,1] \to \mathbb{R}$ 类。正如 [12] 中所指出的， $\mathcal{F}_\infty$ 恰好包含那些 $f: [0,1] \to \mathbb{R}x \in (0,1)$

使得 $|f(x) - f(y)| \leq |x - y|$ 对所有 x、 $y \in [0,1]$均成立。此外，根据詹森不等式， $\mathcal{F}_\infty \subseteq \mathcal{F}_q \subseteq \mathcal{F}_r$对所有 $q \geq r \geq 1$均成立。因此，对于所有 $p \geq 1$ 和 $q \geq r \geq 1$，$\mathrm{opt}_p(\mathcal{F}_\infty) \leq \mathrm{opt}_p(\mathcal{F}_q) \leq \mathrm{opt}_p(\mathcal{F}_r)$。此前的论文确定了当 $p = 1$、 $q = 1$以及 p、 $q \geq 2$时 $\mathrm{opt}_p(\mathcal{F}_q)$ 的精确值，并给出了当 $p \in (1,2)$ 和 $q \geq 2$时 $\mathrm{opt}_p(\mathcal{F}_q)$ 的界。

该论文 [9] 证明了 $\mathrm{opt}_p(\mathcal{F}_1) = \infty$ 对所有 $p \geq 1$均成立。他们还证明了，对于所有 $q \geq 1$，$\mathrm{opt}_1(\mathcal{F}_q) = \mathrm{opt}_1(\mathcal{F}_\infty) = \infty$。相比之下，他们发现，对于所有 $p \geq 2$和 $q \geq 2$，$\mathrm{opt}_p(\mathcal{F}_q) = \mathrm{opt}_p(\mathcal{F}_\infty) = 1$。[4] 中也使用基于 Widrow-Hoff 算法 [8, 14], 的推广的另一种算法证明了这一点，并且在 [3]中研究了该问题的一个带噪声版本。在本文中，我们扩展了已知 $\mathrm{opt}_p(\mathcal{F}_q) = 1$成立时 p、 q 的取值区域。

定理 1.1. 对任意实数 $q > 1$ 和 $p \geq 2 + \frac{1}{q-1}$，都有 $\mathrm{opt}_p(\mathcal{F}_q) = 1$。

对于 $p = 1 + \varepsilon$ 且 $\varepsilon \in (0, 1)$，论文 [9] 证明了，对于所有 $q \geq 2$，$\mathrm{opt}_p(\mathcal{F}_q) = O(\varepsilon^{-1})$，这意味着 $\mathrm{opt}_p(\mathcal{F}_\infty) = O(\varepsilon^{-1})$。然而，这些界并不紧。在本文中，对于所有 $\varepsilon \in (0, 1)$ 和 $q \geq 2$，我们将 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q)$ 确定到一个常数因子以内。

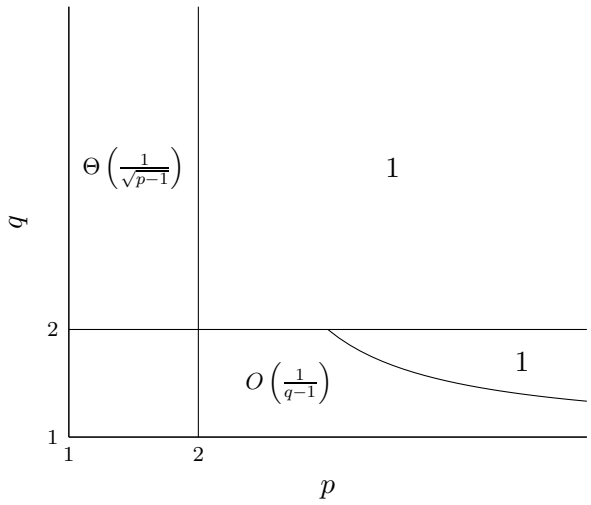

图2：opt 的确切值以及 $_p(\mathcal{F}_q)$ 的界，其中 p、$q > 1$ 包括本文的结果 er

定理 1.2. 对于所有 $\varepsilon \in (0, 1)$，我们有 $\text{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Theta(\varepsilon^{-\frac{1}{2}})$，并且对于所有$q \geq 2$，$\text{opt}_{1+\varepsilon}(\mathcal{F}_q) = \Theta(\varepsilon^{-\frac{1}{2}})$，其中界中的常数不依赖于 q。

定理 1.2 的证明分为上界和下界两部分。对于上界，我们结合 [9] 中的结果使用 H¨older 不等式。对于下界，我们修改了 [12], 中使用的一种构造；该构造获得了 $\text{opt}_1(\mathcal{F}_q)$ 的有限变体的界，而该变体取决于试验次数 m。

[9] 和 [12] 的结果留下了确定 $_p(F_q)$（ $q \in$ 中的 (1,2)）这一问题。即使精确到一个常数因子，也无人知晓其结果。我们通过将 $_2(\mathcal{F}_{1+\varepsilon})$ 在 $\varepsilon \in$ 中的结果确定到一个常数因子，推进了这一问题的研究。图1展示了在本文结果之前， p、$q > 1$ 已知的界和确切值；而图2展示了包括本文结果在内， p、$q > 1$ 已知的界和确切值。

定理 1.3。对于 $\varepsilon \in (0, 1)$，我们有 $\text{opt}_2(\mathcal{F}_{1+\varepsilon}) = \Theta(\varepsilon^{-1})$。

论文 [9] 还讨论了多个变量的光滑函数的在线学习问题。以往关于多变量函数学习的研究 [2, 6, 7] 关注的是期望损失，而非最坏情况损失，并采用输入 x_i 由某个概率分布决定的模型。

我们将 [9] 单变量情形自然地推广到多变量函数。具体而言，对于 $q \geq 1$ 和 $d \in \mathbb{Z}^+$，令 $\mathcal{F}_{q,d}$ 为函数类 $f : [0,1]^d \to \mathbb{R}$，使得对于任意的 $(d-1)$ 元组 $(x_1, \ldots, x_{d-1}) \in [0,1]^{d-1}$ 以及满足 $1 \leq i \leq d$ 的整数 i，由 $g(x) = f(\mathbf{v}_{i,x})$ 给出的函数 $g : [0,1] \to \mathbb{R}$属于 $\mathcal{F}_q$，其中 $\mathbf{v}_{i,x} \in [0,1]^d$是将 x 插入到 $(x_1, \ldots, x_{d-1})$ 的第i^{th} 个位置后形成的向量。

关于 $\text{opt}_p(\mathcal{F}_{q,d})$，最根本的问题之一是确定它何时有限、何时无限。当 $q = \infty$ 时，我们几乎完全解答了这一问题。

定理 1.4. 对任意正整数 d，当 $p > d$ 时，$\text{opt}_p(\mathcal{F}_{\infty,d})$ 为有限值，而当 $0 < p < d$ 时为无限值。

作为推论，对于 $0 < p < d$，立即可得 $\text{opt}_p(\mathcal{F}_{q,d}) = \infty$ 对所有 $q \geq 1$ 均成立。此外，很容易看出，$\text{opt}_p(\mathcal{F}_{1,d}) = \infty$ 对所有正整数 d 和 p 均成立。

论文 [9] 和 [12] 还研究了试验次数有上限时，在线学习光滑函数的最坏情况错误界。具体而言，沿用本节第一段中的记号，定义

$$\mathscr{L}_p(A, f, \sigma, m) = \sum_{t=1}^{m} |\hat{y}_t - f(s_t)|^p.$$

此外，定义

$$\mathscr{L}_p(A, \mathcal{F}, m) = \sup_{f\in\mathcal{F}, \sigma\in S^{m+1}} \mathscr{L}_p(A, f, \sigma, m)$$

以及 $\mathrm{opt}_p(\mathcal{F}, m) = \inf_A \mathscr{L}_p(A, \mathcal{F}, m)$。

论文 [9] 证明了，对于所有 $q \geq 2$，$\mathrm{opt}_1(\mathcal{F}_q, m) = O(\log(m))$，且 $\mathrm{opt}_1(\mathcal{F}_2, m) = \Omega(\sqrt{\log(m)})$。论文 [12] 通过证明，对于所有 $q \geq 2$，$\mathrm{opt}_1(\mathcal{F}_q, m) = \Theta(\sqrt{\log(m)})$，且 $\mathrm{opt}_1(\mathcal{F}_2, m) = \sqrt{\log_2(m)}/2 \pm O(1)$，改进了这些界。当 $0 < p < d$ 时，我们得到了试验次数有界时光滑函数在线学习的紧确界。特别是，这些紧确界在单变量情形下也是新的。

定理 1.5. 对于任意正整数 d 和满足 $0 < p < d$ 的实数 p，有 $\mathrm{opt}_p(\mathcal{F}_{\infty,d}, m) = \Theta(m^{1-\frac{p}{d}})$，其中界中的常数取决于 p 和 d。

在第 2 节中，我们聚焦于单变量情形。我们在第 2.1 小节和第 2.2 小节中证明定理 1.3。第 2.1 小节给出了下界，而第 2.2 小节给出了上界以及若干有用的引理。第 2.3 小节聚焦于证明定理 1.1。在第 2.4 小节中，我们证明定理 1.2。在第 3 节中，我们聚焦于多变量情形，并给出 $\mathrm{opt}_p(\mathcal{F}_{q,d})$ 的各种界。最后，在第 4 节中，我们讨论开放问题。

2 单变量情形下 $q \in (1, 2)$ 的结果

首先，我们采用 [9] 中的一些记号。对于 $f\colon [0,1] \to \mathbb{R}$，将 f 的 q-作用记作 $J_q[f]$，定义为

$$J_q[f] = \int_0^1 |f'(x)|^q \mathrm{d}x,$$

从而 $\mathcal{F}_q$ 恰好是满足 $J_q[f] \leq 1$ 的绝对连续 $f\colon [0,1] \to \mathbb{R}$.

此外，对于 $[0,1]\times\mathbb{R}$ 中满足 $u_1 < \ldots < u_m$ 的非空点集 $S = \{(u_i, v_i) : 1 \leq i \leq m\}$，定义

$$f_S(x) = \begin{cases} v_1 & x \leq u_1 \\ v_i + \frac{(x-u_i)(v_{i+1}-v_i)}{u_{i+1}-u_i} & x \in (u_i, u_{i+1}] \\ v_m & x > u_m \end{cases}$$

并令 $f_\emptyset(x) \equiv 0$。

最后，将学习算法 LININT 定义如下：在试验 0 中，LININT 猜测 $\hat{}y_0 = 0$；在试验 $i > 0$ 中，在点 $S = \{(x_0, f(x_0)) \ldots, (x_{i-1}, f(x_{i-1}))\}$ 已揭示且给定 x_i 后，LININT 猜测 $\hat{}y_i = f_S(x_i)$。

2.1 $\mathrm{opt}_p(\mathcal{F}_q)$ 的下界

首先，对于所有 $p, q > 1$，$\mathrm{opt}_p(\mathcal{F}_q)$ 都有一个显然的下界。

命题 2.1. 对于 $p, q > 1$，我们有 $\mathrm{opt}_p(\mathcal{F}_q) \geq 1$。

论文 [9] 证明了当 $p, q \geq 2$ 时等号成立。正如我们将看到的，当 $q \in (1,2)$ 时，对于足够大的 p，等号也成立。

对于 $q \in (1,2)$ 和 $p > 1$，我们还利用与 [9] 中定理 8 类似的对手策略，证明了 $\mathrm{opt}_2(\mathcal{F}_q)$ 的一个下界。

定理 2.2. 对于 $q \in (1, 2)$，我们有 $\mathrm{opt}_p(\mathcal{F}_q) \geq \frac{q}{(p2^p e \ln 2)(q-1)}$。

证明. 固定 $q \in (1, 2)$ 和一个用于学习 $\mathcal{F}_q$ 的算法 A。考虑以下对手策略族，它们依赖于参数 $b \in (0, 1)$。对手选取 $x_0 = 0$ 并揭示 $f(x_0) = 0$，然后选取 $x_1 = 1$ 并揭示 $f(x_1) = \pm b$，使得 $|y_1 - f(x_1)| \geq b$；不失一般性，假设 $f(x_1) = b$。接下来的 $k = \left\lfloor -q \log_{q-1}^{-2} b \right\rfloor$ 次试验中，对手递归地选取 x_i 和 $f(x_i)$，具体如下。在试验 $2 \leq i < k+2$ 中，对手将 l_i 设为满足 $f(x) = 0$ 已被先前揭示的最大实数 $x \in [0,1]$，并类似地将 r_i 设为满足 $f(x) = b$ 已被先前揭示的最小实数 $x \in [0,1]$，然后设置 $x_i = \frac{l_i + r_i}{2}$。收到 A 的猜测 $\hat{y}_i$ 后，对手揭示 $f(x_i) = 0$ 或 $f(x_i) = b$，使得 $|f(x_i) - \hat{y}_i| \geq \frac{b}{2}$。

要证明该策略定义良好，注意到所有 x_i 都互不相同，因此只需证明存在一个函数 $f \in \mathcal{F}_q$，使其与所有 $(x_i, f(x_i))$ 相一致。事实上，取 $f = f_{\{(x_0, f(x_0)), \ldots, (x_{k+1}, f(x_{k+1}))\}}$，使其在所有点 $(x_i, f(x_i))$ 之间进行线性插值；那么 f 只有一条斜率非零的线段，其

$$J_q[f] = 2^{-k} \left(\frac{b}{2^{-k}} \right)^q = 2^{k(q-1)} b^q \leq 2^{-\frac{q \log_2 b}{q-1} \cdot (q-1)} b^q = 1.$$

因此，对手保证误差至少为

$$\mathrm{opt}_p(\mathcal{F}_q) \geq \sum_{i=1}^{k+1} |f(x_i) - \hat{y}_i|^p \geq b^p + k \left(\frac{b}{2} \right)^p \geq -\frac{b^p q \log_2 b}{2^p (q-1)}.$$

选取 $b = e^{-\frac{1}{p}}$，得到 $\mathrm{opt}_p(\mathcal{F}_q) \geq \frac{q}{(p2^p e \ln 2)(q-1)}$。 ■

特别地，当 $p = 2$ 时，我们得到如下结果：

推论 2.3. 对于 $q \in (1, 2)$，我们有

$$\mathrm{opt}_2(\mathcal{F}_q) \geq \frac{q}{(8e \ln 2)(q-1)} > \frac{1}{(8e \ln 2)(q-1)}.$$

2.2 $\mathrm{opt}_2(\mathcal{F}_q)$ 的界

本节的主要结果是，对于 $\varepsilon \in (0, 1)$，$\mathrm{opt}_2(\mathcal{F}_{1+\varepsilon}) = \Theta(\varepsilon^{-1})$。对于 $q \in (1, 2)$，推论 2.3 给出了 $\mathrm{opt}_2(\mathcal{F}_q)$ 的下界；现在我们证明 $\mathrm{opt}_2(\mathcal{F}_q)$ 的上界，并利用这一点推导出所需结果。首先，我们证明 [9] 中与引理 9 类似的事实成立。

引理 2.4. 设 $u_1 < \ldots < u_m$ 是 $[0,1]$ 和 $v_1, \ldots, v_m$ 中的实数，并定义 $S = \{(u_1, v_1) \ldots, (u_m, v_m)\}$. 那么对于任意 $q \in (1, 2)$ 以及绝对连续的 $f\colon [0,1] \to \mathbb{R}$，若满足 $f(u_i) = v_i$（对 $1 \leq i \leq m$），则有 $J_q[f] \geq J_q[f_S]$。

证明. 若 $m=1$，则 $J_q[f_S]=0$，结论显然成立。否则，固定一个与 (u_i,v_i) 一致的绝对连续函数f: $[0,1]\to\mathbb{R}$，并固定 $1\le i<m$。由詹森不等式，

$$\frac{\int_{u_i}^{u_{i+1}}|f'(x)|^q\mathrm{d}x}{u_{i+1}-u_i}\ge\left(\frac{\int_{u_i}^{u_{i+1}}|f'(x)|\mathrm{d}x}{u_{i+1}-u_i}\right)^q\ge\left(\frac{\left|\int_{u_i}^{u_{i+1}}f'(x)\mathrm{d}x\right|}{u_{i+1}-u_i}\right)^q=\left|\frac{v_{i+1}-v_i}{u_{i+1}-u_i}\right|^q.$$

由此可得

$$J_q[f]\ge\int_{u_1}^{u_m}|f'(x)|^q\mathrm{d}x\ge\sum_{i=1}^{m-1}(u_{i+1}-u_i)\left|\frac{v_{i+1}-v_i}{u_{i+1}-u_i}\right|^q=J_q[f_S]$$

对所有 $1\le i<m$ 求和。 ■

由此得到以下有用的事实。

引理 2.5. 对任意 $q>1$、目标函数 $f\in\mathcal{F}_q$、整数 $m\ge1$和输入序列$x_0,\dots,x_m\in[0,1]$，LININT 在任何试验 $i\ge1$中都不会产生误差 $|\hat{}y_i-f(x_i)|>1$。

证明. 反设 LININT 对 $i\ge1$产生误差 $|\hat{}y_i-f(x_i)|>1$；那么存在 $0\le j<i$ 使得 $|f(x_i)-f(x_j)|>1$。令 $S=\{(x_0,f(x_0))\dots,(x_i,f(x_i))\}$，由引理 2.4

$$J_q[f]\ge J_q[f_S]\ge|x_i-x_j|\left|\frac{f(x_i)-f(x_j)}{x_i-x_j}\right|^q\ge|f(x_i)-f(x_j)|>1,$$

根据该结论，$f\notin\mathcal{F}_q$，矛盾。 ■

推论 2.6. 对于任意 $q>1$和 $p'>p>1$，我们有 $\mathscr{L}_{p'}(\text{LININT }\mathcal{F}_q)\le\mathscr{L}_p(\text{LININT }\mathcal{F}_q)$。

证明。在每次试验 $i\ge1$中，LININT 产生误差 $|\hat{}y_i-f(x_i)|\le1$，因此对于所有 $i\ge1$，$|\hat{}y_i-f(x_i)|^{p'}\le|\hat{}y_i-f(x_i)|^p$。 ■

由此，证明过程与 [9]中的定理 11 的证明类似。具体而言，我们将比较随着新点加入 S，$J_q[f_S]$ 的变化与 LININT 产生的平方误差 $(\hat{}y-f_S(x))^2$，以界定 $\mathscr{L}_2(\text{LININT }\mathcal{F}_q)$。这需要以下不等式。

引理 2.7. 对于实数 $a>0$、$b>0$、$q\in(1,2)$ 和 $x\in(-a,b)$，我们有

$$a\left(1+\frac{x}{a}\right)^q+b\left(1-\frac{x}{b}\right)^q-(a+b)\ge\frac{2q(q-1)}{a+b}\cdot x^2.$$

证明。固定 a、b、q，并定义函数

$$f(x)=a\left(1+\frac{x}{a}\right)^q+b\left(1-\frac{x}{b}\right)^q-(a+b)-\frac{2q(q-1)}{a+b}\cdot x^2$$

对于 $x\in(-a,b)$，因此我们希望证明对所有 $x\in(-a,b)$，$f(x)\ge0$。计算

$$f'(x)=q\left[\left(1+\frac{x}{a}\right)^{q-1}-\left(1-\frac{x}{b}\right)^{q-1}-\frac{4(q-1)}{a+b}x\right]$$

$$f''(x)=q(q-1)\left[\frac{1}{a}\left(1+\frac{x}{a}\right)^{q-2}+\frac{1}{b}\left(1-\frac{x}{b}\right)^{q-2}-\frac{4}{a+b}\right]$$

$$f^{(3)}(x)=q(q-1)(q-2)\left[\frac{1}{a^2}\left(1+\frac{x}{a}\right)^{q-3}-\frac{1}{b^2}\left(1-\frac{x}{b}\right)^{q-3}\right]$$

$$f^{(4)}(x)=q(q-1)(q-2)(q-3)\left[\frac{1}{a^3}\left(1+\frac{x}{a}\right)^{q-4}+\frac{1}{b^3}\left(1-\frac{x}{b}\right)^{q-4}\right].$$

首先，我们证明对所有 $x \in (-a, b)$，$f''(x) \geq 0$。注意，对于所有 x，$f_{(4)}(x) > 0$，并且

$$\lim_{x \to -a^+} f''(x) = \lim_{x \to b-} f''(x) = \infty,$$

因此，$f_{(3)}(x)$ 在 $(-a, b)$ 上递增，只需检查 $f''(x)$ 在 $f^{(3)}(x) = 0$ 时的取值是否满足 ≥ 0。解出这个 x，得到

$$\begin{aligned} f^{(3)}(x) = 0 &\iff \frac{1}{a^2}\left(1 + \frac{x}{a}\right)^{q-3} = \frac{1}{b^2}\left(1 - \frac{x}{b}\right)^{q-3} \\ &\iff \left(\frac{1 + \frac{x}{a}}{1 - \frac{x}{b}}\right)^{3-q} = \frac{b^2}{a^2} \iff \frac{1 + \frac{x}{a}}{1 - \frac{x}{b}} = \frac{b^{\frac{2}{3-q}}}{a^{\frac{2}{3-q}}} \\ &\iff x = \frac{-a^{\frac{2}{3-q}} + b^{\frac{2}{3-q}}}{a^{\frac{q-1}{3-q}} + b^{\frac{q-1}{3-q}}}. \end{aligned}$$

在这个 x 处，

$$\begin{aligned} f''(x) &= q(q-1)\left[\frac{1}{a}\left(1 + \frac{-a^{\frac{2}{3-q}} + b^{\frac{2}{3-q}}}{a^{\frac{2}{3-q}} + ab^{\frac{q-1}{3-q}}}\right)^{q-2} + \frac{1}{b}\left(1 - \frac{-a^{\frac{2}{3-q}} + b^{\frac{2}{3-q}}}{a^{\frac{q-1}{3-q}}b + b^{\frac{2}{3-q}}}\right)^{q-2} - \frac{4}{a+b}\right] \\ &= q(q-1)\left[\frac{1}{a}\left(\frac{(a+b)b^{\frac{q-1}{3-q}}}{a\left(a^{\frac{q-1}{3-q}} + b^{\frac{q-1}{3-q}}\right)}\right)^{q-2} + \frac{1}{b}\left(\frac{(a+b)a^{\frac{q-1}{3-q}}}{b\left(a^{\frac{q-1}{3-q}} + b^{\frac{q-1}{3-q}}\right)}\right)^{q-2} - \frac{4}{a+b}\right] \\ &= q(q-1)\left[\frac{\left(a^{\frac{q-1}{3-q}} + b^{\frac{q-1}{3-q}}\right)^{2-q}\left(a^{1-q}b^{\frac{(q-1)(q-2)}{3-q}} + a^{\frac{(q-1)(q-2)}{3-q}}b^{1-q}\right)}{(a+b)^{2-q}} - \frac{4}{a+b}\right] \\ &\geq q(q-1)\left[\frac{2^{2-q}(ab)^{\frac{(q-1)(2-q)}{2(3-q)}} \cdot 2(ab)^{\frac{(q-1)(2q-5)}{2(3-q)}}}{(a+b)^{2-q}} - \frac{4}{a+b}\right] \\ &= 4q(q-1)\left[\frac{1}{(a+b)^{2-q}(4ab)^{\frac{q-1}{2}}} - \frac{1}{a+b}\right] \\ &\geq 4q(q-1)\left[\frac{1}{(a+b)^{2-q}(a+b)^{q-1}} - \frac{1}{a+b}\right] = 0, \end{aligned}$$

其中，所有不等式都可由不等式 $(u+v)^2 \geq 4uv \iff (u-v)^2 \geq 0$ 对所有实数 u, v 成立推出。由于 $f''(x)$ 在此处取最小值，因此对所有 $x \in (-a, b)$，都有 $f''(x) \geq 0$。

由于 $f'(0) = 0$，且对所有 $x \in (-a, b)$，都有 $f''(x) \geq 0$，因此对于 $x < 0$，有 $f'(x) \leq 0$；对于 $x > 0$，有 $f'(x) \geq 0$，所以对所有 $x \in (-a, b)$，都有 $f(x) \geq f(0) = 0$。■

引理 2.8. 对于实数 $a, b \in (0,1)$、$q \in (1, 2)$ 和 $x \notin (-a, b)$，我们有

$$a\left|\frac{x}{a} + 1\right|^q + b\left|\frac{x}{b} - 1\right|^q - (a+b) \geq \frac{(q-1)|x|^q}{(a+b)^{q-1}}.$$

证明。固定 a, b, q；根据对称性，只需考虑 $x \geq b$。定义函数 n

$$f(x) = a\left(\frac{x}{a} + 1\right)^q + b\left(\frac{x}{b} - 1\right)^q - (a+b) - \frac{(q-1)x^q}{(a+b)^{q-1}}$$

对于 $x \geq b$，因此我们希望证明对所有 $x \geq b$，$f(x) \geq 0$。由于

$$\begin{aligned} f'(x) &= q\left(\frac{x}{a}+1\right)^{q-1} + q\left(\frac{x}{b}-1\right)^{q-1} - \frac{q(q-1)x^{q-1}}{(a+b)^{q-1}} \\ &\geq q\left(\frac{x}{a+b}\right)^{q-1} + q\left(\frac{x}{b}-1\right)^{q-1} - \frac{qx^{q-1}}{(a+b)^{q-1}} > 0 \end{aligned}$$

对所有 $x > b$，f 都是递增的，因此只需证明

$$f(b) = a\left(\frac{a+b}{a}\right)^{q} - (a+b) - \frac{(q-1)b^q}{(a+b)^{q-1}} \geq 0.$$

两边除以 $a+b$ 并代入 $r = \frac{a}{a+b}$，可知这等价于

$$g(r) = \frac{1}{r^{q-1}} - 1 - (q-1)(1-r)^q \geq 0$$

对于 $r \in (0,1)$。由于 $g(1) > 0$，只需证明

$$\begin{aligned} & g'(r) = r^{-q}(q-1)\left(qr^q(1-r)^{q-1} - 1\right) < 0 \\ \iff & qr^{q-1}(1-r)^{q-1} < \frac{1}{r}. \end{aligned}$$

对于 $r \in (0,1)$，量 $r^{q-1}(1-r)^{q-1}$ 在 $r = \frac{1}{2}$ 时取最大值，因此只需证明 $q\left(\frac{1}{2}\right)^{2q-2} < 1$ 对 $q \in (1,2)$ 成立。定义 $h(q) = q\left(\frac{1}{2}\right)^{2q-2}$，则对于 $q \in (1,2)$，有 $h'(q) = 4^{1-q}(1 - q\ln 4) < 0$。由于 $h(1) = 1$，对于 $q \in (1,2)$，有 $h(q) < 1$。■

推论 2.9。对于满足 $a+b \leq 1$，$q \in (1,2)$ 的实数 $a, b \in (0,1)$，以及 $x \notin (-a, b)$，我们有

$$a\left|\frac{x}{a}+1\right|^q + b\left|\frac{x}{b}-1\right|^q - (a+b) \geq (q-1)|x|^q.$$

结合以上结果可得以下关键结论。

引理 2.10。固定 $q \in (1,2)$，一个非空集合 $S = \{(u_1, v_1) \ldots, (u_k, v_k)\}$，它是 $[0,1] \times \mathbb{R}$ 中的点的集合，并且 $(x, y) \in [0,1] \times \mathbb{R}$ 满足：对任意 $1 \leq i \leq k$，均有 $u_1 < \ldots < u_k$，$x \neq u_i$，且 $J_q\left[f_{S \cup \{(x,y)\}}\right] \leq 1$。那么

$$J_q\left[f_{S\cup\{(x,y)\}}\right] - J_q[f_S] \geq (q-1)(y - f_S(x))^2.$$

证明。首先，假设 $x < u_1$。那么，与 f_S 相比，函数 $f_{S\cup\{(x,y)\}}$ 在 (x,y) 和 (u_1, v_1) 之间包含一条新的（可能具有非零斜率的）线段，因此，由于 $|y - f_S(x)| = |v_1 - y| \leq 1$（根据引理 2.5）且 $|u_1 - x| \leq 1$，

$$J_q\left[f_{S\cup\{(x,y)\}}\right] - J_q[f_S] = (u_1 - x)\left|\frac{v_1 - y}{u_1 - x}\right|^q \geq (v_1 - y)^2 = (y - f_S(x))^2 \geq (q-1)(y - f_S(x))^2.$$

情形 $x > u_k$ 类似。

现在假设存在一个整数 $1 \leq i < k$，使得 $u_i < x < u_{i+1}$。在这种情况下，

$$J_q\left[f_{S\cup\{(x,y)\}}\right] - J_q[f_S] = (x - u_i)\left|\frac{y - v_i}{x - u_i}\right|^q + (u_{i+1} - x)\left|\frac{v_{i+1} - y}{u_{i+1} - x}\right|^q - (u_{i+1} - u_i)\left|\frac{v_{i+1} - v_i}{u_{i+1} - u_i}\right|^q.$$

代入 $a = x - u_i$、$b = u_{i+1} - x$、$d = y - f_S(x)$ 和 $m = \frac{v_{i+1}-v_i}{u_{i+1}-u_i} = \frac{f_S(x) - v_i}{a} = \frac{v_{i+1} - f_S(x)}{b}$，我们可以将上式改写为

$$\begin{aligned} J_q\left[f_{S\cup\{(x,y)\}}\right] - J_q[f_S] &= a\left|m + \frac{d}{a}\right|^q + b\left|m - \frac{d}{b}\right|^q - (a+b)|m|^q \\ &= |m|^q\left(a\left|1 + \frac{d}{ma}\right|^q + b\left|1 - \frac{d}{mb}\right|^q - (a+b)\right). \end{aligned}$$

然后应用 引理 2.7 或推论 2.9（取决于 $\frac{d}{m} \in (-a, b)$）可得

$$\begin{aligned} J_q\left[f_{S\cup\{(x,y)\}}\right] - J_q[f_S] &\geq |m|^q \min\left\{\frac{2q(q-1)}{a+b}\cdot\left(\frac{d}{m}\right)^2, (q-1)\left|\frac{d}{m}\right|^q\right\} \\ &= \min\left\{\frac{2q(q-1)}{|m|^{2-q}(a+b)}\cdot d^2, (q-1)|d|^q\right\}. \end{aligned}$$

如果 $0 < |m| \leq 1$，那么 $a + b = u_{i+1} - u_i \leq 1 \implies |m|^{2-q}(a+b) \leq 1$；而如果 $|m| \geq 1$，那么

$$|m|^{2-q}(a+b) \leq |m|^q(a+b) \leq J_q[f_S] \leq J_q[f_{S\cup\{(x,y)\}}] \leq 1$$

根据引理 2.4，因此无论哪种情况

$$\frac{2q(q-1)}{|m|^{2-q}(a+b)}\cdot d^2 \geq 2q(q-1)d^2 \geq (q-1)d^2.$$

此外，由于 $J_q\left[f_{S\cup\{(x,y)\}}\right] \leq 1$，根据引理 2.5 $|d| \leq 1$. 因此 $(q-1)|d|^q \geq (q-1)d^2$ 同样成立。■

这直接给出了所需的上界。

定理 2.11。对于 $q \in (1, 2)$，我们有 $\mathscr{L}_2(\text{LININT}\ \mathcal{F}_q) \leq \frac{1}{q-1}$。

证明。固定一个目标函数 $f \in \mathcal{F}_q$、一个整数 $m \geq 1$ 以及一个输入序列 $\sigma = (x_0, \ldots, x_m) \in [0,1]^{m+1}$。不妨设所有 x_i 均不相同。对于 $0 \leq i \leq m$，定义 $S_i = \{(x_0, f(x_0)) \ldots, (x_i, f(x_i))\}$，并假设 LININT 生成猜测 $\hat{y}_0, \ldots, \hat{y}_m \in \mathbb{R}$。根据引理 2.4 和引理 2.10，

$$1 \geq J_q[f] \geq J_q[f_{S_m}] = \sum_{i=1}^{m}\left(J_q[f_{S_i}] - J_q\left[f_{S_{i-1}}\right]\right) \geq (q-1)\sum_{i=1}^{m}(\hat{y}_i - f(x_i))^2,$$

so

$$\mathscr{L}_2(\text{LININT}, f, \sigma) = \sum_{i=1}^{m}(\hat{y}_i - f(x_i))^2 \leq \frac{1}{q-1}$$

对于任意 $f \in \mathcal{F}_q$、整数 $m \geq 1$ 和 $\sigma \in [0,1]^{m+1}$。因此 $\mathscr{L}_2(\text{LININT}\ \mathcal{F}_q) \leq \frac{1}{q-1}$。 ■

最后，将上述结果与推论 2.3 中的下界结合起来，可得以下结果。

定理 2.12. 对于 $\varepsilon \in (0, 1)$，我们有 $\text{opt}_2(\mathcal{F}_{1+\varepsilon}) = \Theta(\varepsilon^{-1})$。

证明. 结合推论 2.3 和定理 2.11，

$$\frac{\varepsilon^{-1}}{8e\ln 2} < \frac{1+\varepsilon}{(8e\ln 2)\varepsilon} \leq \text{opt}_2(\mathcal{F}_{1+\varepsilon}) \leq \varepsilon^{-1}.$$

因此，$\text{opt}_2(\mathcal{F}_{1+\varepsilon}) = \Theta(\varepsilon^{-1})$。 ■

将定理 2.11 中的上界推广到所有 $p \geq 2$ 很简单。

推论 2.13。对于 $\varepsilon \in (0, 1)$ 和 $p \geq 2$，我们有 $\mathrm{opt}_p(\mathcal{F}_{1+\varepsilon}) = O(\varepsilon^{-1})$。

证明。根据引理 2.6，$\mathrm{opt}_p(\mathcal{F}_{1+\varepsilon}) \leq \mathscr{L}_p(\mathrm{LININT}\ \mathcal{F}_{1+\varepsilon}) \leq \mathscr{L}_2(\mathrm{LININT}\ \mathcal{F}_{1+\varepsilon}) = O(\varepsilon^{-1})$。 ■

2.3 关于大 p 的精确结果

本节将证明，对于 $q \in (1, 2)$ 和 $p \geq 2 + \frac{1}{q-1}$，$\mathrm{opt}_p(\mathcal{F}_q) = 1$。首先需要以下引理。

引理 2.14。对于实数 $q \in (1, 2)$、$a \in (0, 1)$ 和 u，满足 v satisfying $|u - v| \geq \frac{(q-1\ q-1}{a\ 1-a})()$，我们有

$$a|u|^q + (1-a)|v|^q > 1.$$

证明。不失一般性，假设 $u > v$，从而 $u \geq v + \frac{(q-1\ q-1}{a\ 1-a})()$。

首先，假设 $v < 0 < u$；那么 $|u| + |v| \geq \frac{(q-1\ q-1}{a\ 1-a})()$。根据加权幂平均不等式，

$$\begin{aligned}
\frac{|u|(a|u|^{q-1}) + |v|((1-a)|v|^{q-1})}{|u| + |v|} &\geq \left(\frac{a^{-\frac{1}{q-1}} + (1-a)^{-\frac{1}{q-1}}}{|u| + |v|}\right)^{-(q-1)} \\
\implies a|u|^q + (1-a)|v|^q &\geq \frac{(|u| + |v|)^q}{\left(a^{-\frac{1}{q-1}} + (1-a)^{-\frac{1}{q-1}}\right)^{q-1}} \\
&\geq \frac{(q-1)^{q(q-1)}}{a^q(1-a)^q\left(a^{-\frac{1}{q-1}} + (1-a)^{-\frac{1}{q-1}}\right)^{q-1}} \\
&\geq \frac{(q-1)^{q(q-1)}}{2^{q-1}a^q(1-a)^q \max\{a^{-1}, (1-a)^{-1}\}} \\
&= \frac{(q-1)^{q(q-1)}}{2^{q-1}\max\{a^{q-1}(1-a)^q, a^q(1-a)^{q-1}\}}.
\end{aligned}$$

By t加权算术-几何平均不等式，对于 $r \in (0, 1)$，我们有

$$\begin{aligned}
r^q(1-r)^{q-1} &= \frac{q^q}{(q-1)^q}\left(\frac{(q-1)r}{q}\right)^q(1-r)^{q-1} \\
&\leq \frac{q^q}{(q-1)^q}\left(\frac{q \cdot \frac{(q-1)r}{q} + (q-1)(1-r)}{(q-1)+q}\right)^{(q-1)+q} \\
&= \frac{q^q}{(q-1)^q}\left(\frac{q-1}{2q-1}\right)^{2q-1}.
\end{aligned}$$

Thus

$$\max\left\{a^{q-1}(1-a)^q, a^q(1-a)^{q-1}\right\} \leq \frac{q^q}{(q-1)^q}\left(\frac{q-1}{2q-1}\right)^{2q-1} = \frac{q^q(q-1)^{q-1}}{(2q-1)^{2q-1}},$$

so

$$a|u|^q + (1-a)|v|^q \geq \frac{(q-1)^{(q-1)^2}(2q-1)^{2q-1}}{2^{q-1}q^q}.$$

考虑

$$f(q)=(q-1)^2\ln(q-1)+(2q-1)\ln(2q-1)-(q-1)\ln 2-q\ln q$$

在 $q\in(1,2)$ 上。注意

$$\begin{aligned}f'(q)&=(2(q-1)\ln(q-1)+(q-1))+(2\ln(2q-1)+2)-\ln 2-(\ln q+1)\\&=1-\ln 2+2(q-1)\ln(q-1)+(q-1-\ln q)+2\ln(2q-1)\\&\geq 1-\ln 2+2(q-1)\ln(q-1)+2\ln(2q-1),\end{aligned}$$

由于 $e^x\geq 1+x$，可知对于 $x>-1$，$x\geq\ln(1+x)$。由于 $x\ln x$ 在 $\left(0,\frac{1}{e}\right)$ 上递减，在 $\left(\frac{1}{e},\infty\right)$ 上递增（因此特别地，对于 $x>0$，$x\ln x\geq-\frac{1}{e}$），

$$\begin{aligned}q\in(1,1.004]&\implies f'(q)\geq 1-\ln 2+0.008\ln 0.004>0\\q\in[1.004,1.055]&\implies f'(q)\geq 1-\ln 2+0.11\ln 0.055+2\ln 1.008>0\\q\in[1.055,1.12]&\implies f'(q)\geq 1-\ln 2+0.24\ln 0.12+2\ln 1.11>0\\q\in[1.12,2)&\implies f'(q)\geq 1-\ln 2-\frac{2}{e}+2\ln 1.24>0,\end{aligned}$$

因此，对于所有 $q\in(1,2)$，$f'(q)>0$。由于 $\lim\limits_{q\to1^+}f(q)=0$，可得对于 $q\in(1,2)$，$f(q)>0$，因此只要 $v<0<u$，就有 $a|u|^q+(1-a)|v|^q\geq e^{f(q)}>1$。

现在假设 $v\geq 0$。由于 $|x|^q$ 在 $x\geq 0$ 上递增，

$$a|u|^q+(1-a)|v|^q\geq a\left(\frac{(q-1)^{q-1}}{a(1-a)}\right)^q=\frac{(q-1)^{q(q-1)}}{a^{q-1}(1-a)^q}.$$

利用以上推导，

$$\frac{(q-1)^{q(q-1)}}{a^{q-1}(1-a)^q}\geq\frac{(q-1)^{q(q-1)}}{\max\left\{a^{q-1}(1-a)^q,a^q(1-a)^{q-1}\right\}}>2^{q-1}>1,$$

因此，当 $v\geq 0$ 时，不等式成立。$u\leq 0$ 的情形完全相同，证明至此完成。■

由此，我们得到以下关键结果。

引理 2.15. 固定 $q\in(1,2)$，取 $[0,1]\times\mathbb{R}$ 中的非空点集 $S=\{(u_1,v_1)\ldots,(u_k,v_k)\}$，以及 $(x,y)\in[0,1]\times\mathbb{R}$，使得对任意 $1\leq i\leq k$，均有 $u_1<\ldots<u_k$，$x\neq u_i$，且 $J_q\left[f_{S\cup\{(x,y)\}}\right]\leq 1$。令 $p=2+\frac{1}{q-1}$。则

$$J_q\left[f_{S\cup\{(x,y)\}}\right]-J_q[f_S]\geq|y-f_S(x)|^p.$$

证明. 我们首先证明，$|y-f_S(x)|>(q-1)^{q-1}$ 与 $x\in(u_1,u_k)$ 不可能同时成立。反设二者同时成立，则存在一个整数 $1\leq i<k$，使得 $u_i<x<u_{i+1}$。我们将通过证明 $J_q\left[f_{S\cup\{(x,y)\}}\right]>1$ 导出矛盾。显然

$$J_q\left[f_{S\cup\{(x,y)\}}\right]\geq(x-u_i)\left|\frac{y-v_i}{x-u_i}\right|^q+(u_{i+1}-x)\left|\frac{v_{i+1}-y}{u_{i+1}-x}\right|^q.$$

将引理 2.10 中的 $a=x-u_i$、$b=u_{i+1}-x$、$d=y-f_S(x)$ 和 $m=\frac{v_{i+1}-v_i}{u_{i+1}-u_i}$ 代入后，可改写为

$$J_q\left[f_{S\cup\{(x,y)\}}\right]\geq a\left|m+\frac{d}{a}\right|^q+b\left|m-\frac{d}{b}\right|^q.$$

由于 $a+b=u_{i+1}-u_i\le 1$ 和 $q>1$,

$$J_q\left[f_{S\cup\{(x,y)\}}\right]\ge\frac{a}{a+b}\left|(a+b)m+\frac{d(a+b)}{a}\right|^q+\frac{b}{a+b}\left|(a+b)m-\frac{d(a+b)}{b}\right|^q,$$

并且由于

$$|d|>(q-1)^{q-1}\implies\left|d(a+b)\left(\frac{1}{a}+\frac{1}{b}\right)\right|=\frac{|d|(a+b)^2}{ab}\ge\frac{(q-1)^{q-1}}{\frac{a}{a+b}\cdot\frac{b}{a+b}},$$

应用引理 2.14 得到 $J_q\left[f_{S\cup\{(x,y)\}}\right]>1$，矛盾。

因此，$|y-f_S(x)|\le(q-1)^{q-1}$ 和 $x\notin(u_1,u_k)$ 中至少有一个成立。如果

$$|y-f_S(x)|\le(q-1)^{q-1}\implies(q-1)(y-f_S(x))^2\ge|y-f_S(x)|^p,$$

该结果可由引理 2.10 得出。否则，不妨假设 $x<u_1$ （ $x>u_k$ 的情形类似）；于是

$$J_q\left[f_{S\cup\{(x,y)\}}\right]-J_q[f_S]=(u_1-x)\left|\frac{v_1-y}{u_1-x}\right|^q\ge|v_1-y|^q\ge|v_1-y|^p=|y-f_S(x)|^p$$

根据引理 2.5（因为 $q<2<p$），该结果在此情形下同样成立。 ■

由此立即得到以下结论。

定理 2.16. 对任意实数 $q>1$ 和 $p\ge 2+\frac{1}{q-1}$，我们有 $\mathrm{opt}_p(\mathcal{F}_q)=1$。

证明。由命题 2.1 和推论 2.6，只需证明对于 $q\in(1,2)$ 和 $p=2+\frac{1}{q-1}$，$\mathscr{L}_p(\mathrm{LININT}\ \mathcal{F}_q)\le 1$。固定 $p=2+\frac{1}{q-1}$、一个目标函数 $f\in\mathcal{F}_q$、一个整数 $m\ge 1$，以及一个输入序列 $\sigma=(x_0,\dots,x_m)\in[0,1]^{m+1}$。不失一般性，假设所有 x_i 互不相同。对于 $0\le i\le m$，定义 $S_i=\{(x_0,f(x_0))\dots,(x_i,f(x_i))\}$，并假设 LININT 产生了猜测 $\hat{y}_0,\dots,\hat{y}_m\in\mathbb{R}$。由引理 2.4 和引理 2.15,

$$1\ge J_q[f]\ge J_q\left[f_{S_m}\right]=\sum_{i=1}^{m}\left(J_q\left[f_{S_i}\right]-J_q\left[f_{S_{i-1}}\right]\right)\ge\sum_{i=1}^{m}|\hat{y}_i-f(x_i)|^p,$$

因此，对于任意 $f\in\mathcal{F}_q$ 和 σ，$\mathscr{L}_p(\mathrm{LININT}\ f,\sigma)\le 1$。由此可得 $\mathscr{L}_p(\mathrm{LININT}\ \mathcal{F}_q)\le 1$。 ■

2.4 $p\in\in(1,2)$ 时的紧确界

论文 [9] 证明了，对于所有 $\varepsilon\in\in(0,1)$ 和 $q\ge 2$，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q)=O(\varepsilon^{-1})$。本节中，我们首先通过证明对于所有 $\varepsilon\in\in(0,1)$ 和 $q\ge 2$，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q)=O(\varepsilon^{-\frac{1}{2}})$，改进他们给出的上界。随后，我们通过证明对于所有 $q\ge 1$，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q)=\Omega(\varepsilon^{-\frac{1}{2}})$，说明该界是紧确的。为了证明这个上界，我们将用到文献 [9] 中的两个引理。为陈述这些引理并证明我们的上界，我们使用以下记号。设 $x_0,\dots,x_m$ 是 $[0,1]$, 中任意一列互不相同的元素，并设 $f\in\mathcal{F}_2$。设 $\hat{y}_1,\dots,\hat{y}_m$ 为 LININT 在试验 $1,\dots,m$ 上的预测。对每个 $i>1$，令 $d_i=\min_{j<i}|x_j-x_i|$，并令 $e_i=|\hat{y}_i-f(x_i)|$。

引理 2.17（[9]）。对于所有正整数 m，我们有 $\sum_{i=1}^{m}\frac{e_i^2}{d_i}\le 1$。

引理 2.18 ([9])　　。对于所有正整数 m 和实数 $x > 1$，我们有 $\sum_{i=1}^{m} d_i^x \le 1 + \frac{1}{2^x - 2}$　　.

结合引理 2.17 和引理 2.18，并应用 H¨older 不等式，我们得到以下紧上界。

定理 2.19. 若 $p = 1 + \varepsilon \in (1, 2)$，则 $\mathrm{opt}_p(\mathcal{F}_2) = O(\varepsilon^{-\frac{1}{2}})$。

证明。首先，注意到

$$\sum_{i=1}^{m} e_i^p = \sum_{i=1}^{m} \frac{e_i^p}{d_i^{\frac{p}{2}}} \cdot d_i^{\frac{p}{2}}.$$

根据 H¨older 不等式，我们有

$$\sum_{i=1}^{m} \frac{e_i^p}{d_i^{\frac{p}{2}}} \cdot d_i^{\frac{p}{2}} \le \left(\sum_{i=1}^{m} \frac{e_i^2}{d_i} \right)^{\frac{p}{2}} \left(\sum_{i=1}^{m} d_i^{\frac{p}{2-p}} \right)^{1-\frac{p}{2}}.$$

注意，由引理 2.17 可知 $\sum_{i=1}^{m} \frac{e_i^2}{d_i} \le 1$，并且

$$\sum_{i=1}^{m} d_i^{\frac{p}{2-p}} \le 1 + \frac{1}{2^{\frac{p}{2-p}} - 2}$$

根据引理 2.18，由于 $p > 1$ 蕴含 $\frac{p}{2-p} > 1$。因此

$$\left(\sum_{i=1}^{m} d_i^{\frac{p}{2-p}} \right)^{1-\frac{p}{2}} \le \left(1 + \frac{1}{2^{\frac{p}{2-p}} - 2} \right)^{1-\frac{p}{2}}.$$

令 $\delta = \frac{p}{2-p} - 1$，并注意到 $\frac{1}{\delta} = \frac{2-p}{2p-2}$。因此

$$\left(1 + \frac{1}{2^{\frac{p}{2-p}} - 2} \right)^{1-\frac{p}{2}} = \left(1 + \frac{1}{2^{1+\delta} - 2} \right)^{1-\frac{p}{2}} = O\left(\left(1 + \frac{1}{\delta} \right)^{\frac{2-p}{2}} \right) = O\left(\left(\frac{p}{2p-2} \right)^{\frac{2-p}{2}} \right),$$

其中上界　　由以下事实可得：$e^{\delta \ln^2} \ge 1 + \delta \ln 2$。因此我们已证明 tha　　t

$$\sum_{i=1}^{m} e_i^p = O\left(\left(\frac{p}{2p-2} \right)^{\frac{2-p}{2}} \right),$$

因此 $\mathrm{opt}_p(\mathcal{F}_2) = O\left((2p-2)^{-\frac{2-p}{2}} \right)$，其中我们利用以下事实得到最后一个界：$p^{\frac{2-p}{2}} = \Theta(1)$ 对于 $p \in (1, 2)$。由于 $p = 1 + \varepsilon$，我们有

$$\mathrm{opt}_p(\mathcal{F}_2) = O\left((2p-2)^{-\frac{2-p}{2}} \right) = O\left(\varepsilon^{-\frac{1-\varepsilon}{2}} \right) = O(\varepsilon^{-\frac{1}{2}}),$$

其中我们利用以下事实得到最后一个界：$\varepsilon^{\varepsilon} = \Theta(1)$ 对于 $\varepsilon \in (0, 1)$。 ■

我们得到下一个推论，因为当 $1 \le q \le r$ 时，$\mathrm{opt}_p(\mathcal{F}_\infty) \le \mathrm{opt}_p(\mathcal{F}_r) \le \mathrm{opt}_p(\mathcal{F}_q)$。

推论 2.20. 如果 $\varepsilon \in (0, 1)$，那么对于所有 $q \ge 2$，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = O(\varepsilon^{-\frac{1}{2}})$ 且 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q) = O(\varepsilon^{-\frac{1}{2}})$，其中该常数不依赖于 q。

为了说明上一个推论在常数因子的意义下是紧的，我们在 $\mathcal{F}_\infty$ 中构造了一个函数族。我们的证明使用了来自 [9] 的以下引理，该引理也曾用于 [12]。

引理 2.21 ([9])。设 $S \subseteq [0,1] \times \mathbb{R}$ with $S = \{(u_i, v_i) : 1 \le i \le m\}$ and $u_1 < u_2 < \cdots < u_m$。若 $(x, y) \in [0,1] \times \mathbb{R}$，且存在 $1 \le j \le m$ 使得 $|x - u_j| = |x - u_{j+1}| = \min_i |x - u_i|$，那么 $J_2[f_{S \cup \{(x,y)\}}] = J_2[f_S] + \frac{1}{2(y - f_S(x))^2} \min_i |x - u_i|$。

以下证明中的方法类似于 [12] 中使用的方法，后者用于获得 $\mathrm{opt}_1(\mathcal{F}_q)$ 的界，这是 $q \ge 2$ 的一个有限变体，且该变体依赖于试验次数 m。

定理 2.22。若 $\varepsilon \in (0, 1)$，则 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Omega(\varepsilon^{-\frac{1}{2}})$。

证明。由于对于所有 $\varepsilon \in (0, 1)$，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) \ge 1$，因此只需证明定理对 $\varepsilon \in \left(0, \frac{1}{2}\right)$ 成立。定义 $x_0 = 1$ 和 $y_0 = 0$。对于满足 $0 \le j < 2^{i-1}$ 的自然数 i, j，定义 $x_{2^{i-1}+j} = \frac{1}{2^i} + \frac{j}{2^{i-1}}$。对于每个 $i = 1, 2, \ldots$，我们将 $x_{2^{i-1}}, \ldots, x_{2^i - 1}$ 的试验视为阶段 i 的一部分，因此 $x_1 = \frac{1}{2}$ 处于阶段 1，$x_2 = \frac{1}{4}$ 和 $x_3 = \frac{3}{4}$ 处于阶段 2，依此类推。

令 A 为任意一个用于学习 $\mathcal{F}_\infty$ 的算法。利用 A，我们构造一个无限的分段线性函数序列 $f_0, f_1, \ldots \in \mathcal{F}_\infty$，以及一个无限的数列 $y_0, y_1, \ldots \in \mathbb{R}$，使得 f_t 与 $k \le t$ 的 x_k 和 y_k 值一致，并且 A 的总 $(1+\varepsilon)$ 误差至少为

$$\sum_{k=1}^{i} 2^{k-2} \left(\frac{\sqrt{\varepsilon}(1-\varepsilon)^{\frac{k}{2}}}{2^{k+1}} \right)^{1+\varepsilon}$$

经过 i 个阶段后。这意味着

$$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) \ge \sum_{k=1}^{\infty} 2^{k-2} \left(\frac{\sqrt{\varepsilon}(1-\varepsilon)^{\frac{k}{2}}}{2^{k+1}} \right)^{1+\varepsilon}.$$

为了分析函数 f_i，我们还将定义并分析另一个由分段线性函数组成的无限序列 $g_{i,\,j}$ 其中 $0 \le j \le 2^{i-1}$，以及另一个无限数列 v_1、v_2、$\ldots \in \mathbb{R}$。我们先令 f_0 为零函数。接下来，我们通过归纳定义这两个分段线性函数序列。

固定一个阶段 i，并令 $g_{i,\,0} = f_{2^{i-1}-1}$。令 t 为阶段 i 中的一次试验，并令 v_t 为 $f_{t-1}(x_t) \pm \frac{\sqrt{\varepsilon}(1-\varepsilon)^{\frac{i}{2}}}{2^{i+1}}$ 中距离 y_t 最远的那个。令 $g_{i,\,t-2^{i-1}+1}$ 为经过以下各点的函数：$\{(0, 0), (1, 0)\} \cup \{(x_s, y_s) : s < 2^{i-1}\} \cup \{(x_s, v_s) : 2^{i-1} \le s \le t\}$，并对这些点进行线性插值。

对于任意 $t \ge 1$，令 L_t 和 R_t 分别为 $\{0, 1\} \cup \{x_s : s < t\}$ 中最接近 x_t 左侧和右侧的元素。如果 $|v_t - f_{t-1}(L_t)| \le 2^{-i}$ 和 $|v_t - f_{t-1}(R_t)| \le 2^{-i}$ 均成立，则令 $y_t = v_t$。否则，令 $y_t = f_{t-1}(x_t)$。最后，我们将 f_t 定义为在 $\{(0, 0), (1, 0)\} \cup \{(x_s, y_s) : s \le t\}$ 上进行线性插值的函数。

根据定义，对于每个 $t \ge 0$，我们有 $f_t \in \mathcal{F}_\infty$。接下来，我们将证明对于所有 i，j 都有 $J_2[g_{i,j}] \le \frac{1}{4}$，然后用它来证明，在阶段 i 中，至多一半的试验 t 满足 $y_t = f_{t-1}(x_t)$。证明将使用双重归纳，先对 i，再对 j 进行归纳，并证明以下更强的命题：

$$J_2[g_{i,j}] \le \frac{\varepsilon}{4} \sum_{k=0}^{i-1} (1-\varepsilon)^k + \frac{j\varepsilon(1-\varepsilon)^i}{2^{i+1}}. \tag{1}$$

为了证明这一命题，我们还将证明：

$$J_2[f_{2^{i-1}-1}] \leq \frac{\varepsilon}{4}\sum_{k=0}^{i-1}(1-\varepsilon)^k \tag{2}$$

对所有 $i \geq 1$ 均成立。注意，这等价于证明

$$J_2[g_{i,0}] \leq \frac{\varepsilon}{4}\sum_{k=0}^{i-1}(1-\varepsilon)^k$$

对所有 $i \geq 1$ 均成立。显然，当 $i = 1$ 时结论成立，这就是关于 i 归纳法的基例。固定某个阶段 $i \geq 1$。我们假设不等式 2 对这个固定的 i 成立，并以此证明

$$J_2[f_{2^i-1}] \leq \frac{\varepsilon}{4}\sum_{k=0}^{i}(1-\varepsilon)^k. \tag{3}$$

为证明不等式 3，我们将证明不等式 1 对所有 $0 \leq j \leq 2^{i-1}$ 均成立。这由关于 i 的归纳假设以及 $g_{i,0}$ 的定义可得；当 $j = 0$ 时结论成立，这就是关于 j 归纳法的基例。固定一个满足 $0 \leq j \leq 2^{i-1} - 1$ 的整数 j，并假设不等式 1 对这个固定的 j 成立。根据引理 2.21，我们有

$$J_2[g_{i,j+1}] = J_2[g_{i,j}] + \frac{2\left(\frac{\sqrt{\varepsilon}(1-\varepsilon)^{\frac{i}{2}}}{2^{i+1}}\right)^2}{2^{-i}} = J_2[g_{i,j}] + \frac{\varepsilon(1-\varepsilon)^i}{2^{i+1}}.$$

根据关于 j 的归纳假设，我们得到

$$J_2[g_{i,j+1}] \leq \frac{\varepsilon}{4}\sum_{k=0}^{i-1}(1-\varepsilon)^k + \frac{j\varepsilon(1-\varepsilon)^i}{2^{i+1}} + \frac{\varepsilon(1-\varepsilon)^i}{2^{i+1}},$$

由此完成了关于 j 的归纳步骤。代入 $j = 2^{i-1}$，我们得到

$$J_2[g_{i,2^{i-1}}] \leq \frac{\varepsilon}{4}\sum_{k=0}^{i}(1-\varepsilon)^k.$$

注意，引理 2.21 意味着

$$J_2[f_{2^{i-1}-1+j}] \leq J_2[g_{i,j}]$$

对所有 $j = 0, \ldots, 2^{i-1}$，因此我们得到不等式 3，从而完成了 i 的归纳步骤。由不等式 1，我们得到

$$J_2[g_{i,j}] \leq \frac{\varepsilon}{4}\sum_{k=0}^{i-1}(1-\varepsilon)^k + \frac{\varepsilon(1-\varepsilon)^i}{4} = \frac{\varepsilon}{4}\sum_{k=0}^{i}(1-\varepsilon)^k$$

对于所有满足 $0 \leq j \leq 2^{i-1}$ 的 j。注意，

$$\frac{\varepsilon}{4}\sum_{k=0}^{i}(1-\varepsilon)^k < \frac{\varepsilon}{4}\sum_{k=0}^{\infty}(1-\varepsilon)^k = \frac{1}{4}.$$

既然我们已经证明了 $J_2[g_{i,j}] \le \frac{1}{4}$，现在就可以证明：对于每个 $i \ge 1$，在阶段 i 的试验 t 中，至多有一半满足 $y_t = f_{t-1}(x_t)$。对于每次满足 $y_t = f_{t-1}(x_t)$ 的试验 t，注意，$g_{i,t-2^{i-1}+1}$ 的斜率绝对值在 x_t 两侧长度为 2^{-i} 的区间中，至少有一个区间内必须超过 1。如果阶段 i 中至少有 b 的试验满足 $y_t = f_{t-1}(x_t)$，那么只考虑斜率至少为 1 的区间可得 $J_2[g_{i,2^{i-1}}] \ge b2^{-i}$。由于 $J_2[g_{i,2^{i-1}}] \le \frac{1}{4}$，我们必有 $b \le 2^{i-2}$。因此，在阶段 i 中，满足 $y_t = f_{t-1}(x_t)$ 的试验 t 至多有 2^{i-2} 次，这意味着至少有 2^{i-2} 次试验满足 $y_t = v_t$。在这些试验中，每次 A 的误差至少为 $\frac{\sqrt{\varepsilon}(1-\varepsilon)^{\frac{i}{2}}}{2^{i+1}}$，所以经过 i 个阶段后，A 的总 $(1+\varepsilon)$ 误差至少为 $\sum_{k=1}^{i} 2^{k-2}\left(\frac{\sqrt{\varepsilon}(1-\varepsilon)^{\frac{k}{2}}}{2^{k+1}}\right)^{1+\varepsilon}$。因此

$$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) \ge \sum_{k=1}^{\infty} 2^{k-2}\left(\frac{\sqrt{\varepsilon}(1-\varepsilon)^{\frac{k}{2}}}{2^{k+1}}\right)^{1+\varepsilon} = \frac{\frac{1}{2}\left(\frac{\sqrt{\varepsilon(1-\varepsilon)}}{4}\right)^{1+\varepsilon}}{1-2\left(\frac{\sqrt{1-\varepsilon}}{2}\right)^{1+\varepsilon}} = \Omega\left(\frac{(\varepsilon(1-\varepsilon))^{\frac{1+\varepsilon}{2}}}{1-2\left(\frac{\sqrt{1-\varepsilon}}{2}\right)^{1+\varepsilon}}\right).$$

由于 $\varepsilon^\varepsilon = \Theta(1)$ 且 $(1-\varepsilon)^{1+\varepsilon} = \Theta(1)$ 对于 $\varepsilon \in \left(0, \frac{1}{2}\right)$，我们有 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Omega\left(\frac{\sqrt{\varepsilon}}{1-2\left(\frac{\sqrt{1-\varepsilon}}{2}\right)^{1+\varepsilon}}\right)$。由于 $2^\varepsilon = \Theta(1)$ 对于 $\varepsilon \in \left(0, \frac{1}{2}\right)$，我们有 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Omega\left(\frac{\sqrt{\varepsilon}}{2^\varepsilon - \sqrt{1-\varepsilon}^{1+\varepsilon}}\right)$。注意，对于 $\varepsilon \in \left(0, \frac{1}{2}\right)$，有 $(1-\varepsilon)^{\frac{1+\varepsilon}{2}} \ge 1-\varepsilon(1+\varepsilon)$。为验证这一点，注意当 $\varepsilon = 0$ 时它成立，而 $(1-\varepsilon)^{\frac{1+\varepsilon}{2}} - (1-\varepsilon(1+\varepsilon))$ 的导数为

$$2\varepsilon + 1 + (1-\varepsilon)^{\frac{1+\varepsilon}{2}}\left(\frac{1}{2}\ln(1-\varepsilon) - \frac{1}{2} - \frac{\varepsilon}{1-\varepsilon}\right) > 0$$

对于 $\varepsilon \in \left(0, \frac{1}{2}\right)$。因此，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Omega\left(\frac{\sqrt{\varepsilon}}{2^\varepsilon - 1 + \varepsilon(1+\varepsilon)}\right)$。还要注意，对于 $\varepsilon \in (0, 1)$，$2^\varepsilon \le 1 + \varepsilon$。当 $\varepsilon = 0$ 和 $\varepsilon = 1$ 时等号成立，而 $1 + \varepsilon - 2^\varepsilon$ 的导数为 $1 - 2^\varepsilon \ln 2$；当 $\varepsilon \in \left(0, -\ln\ln 2^2\right)\ln$ 时，该导数为正；当 $\varepsilon \in \left(-\ln\ln 2\ln 2, 1\right)$ 时，该导数为负。因此 $2^\varepsilon - 1 + \varepsilon(1+\varepsilon) < 3\varepsilon$，所以 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Omega(\varepsilon^{-\frac{1}{2}})$。■

下一个推论由定理 2.22 得出，同样利用了这样一个事实：当 $1 \le q \le r$ 时，$\mathrm{opt}_p(\mathcal{F}_\infty) \le \mathrm{opt}_p(\mathcal{F}_r) \le \mathrm{opt}_p(\mathcal{F}_q)$。

推论 2.23. 若 $\varepsilon \in (0, 1)$，则对所有 $q \ge 1$，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q) = \Omega(\varepsilon^{-\frac{1}{2}})$，其中常数与 q 无关。

结合推论 2.20 和推论 2.23，可得以下定理。

定理 2.24. 若 $\varepsilon \in (0, 1)$，则对所有 $q \ge 2$，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_\infty) = \Theta(\varepsilon^{-\frac{1}{2}})$，且 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q) = \Theta(\varepsilon^{-\frac{1}{2}})$，其中界中的常数与 q 无关。

3 多变量推广

本节中，我们证明关于 $\mathrm{opt}_p(\mathcal{F}_{q,\ d})$ 的若干结果。首先，我们根据 $\mathrm{opt}_p(\mathcal{F}_q)$，证明 $\mathrm{opt}_p(\mathcal{F}_{q,\ d})$ 的一个简单下界。

命题 3.1. 对于任意正整数 d、实数 $p > 0$ 和 $q \in [1, \infty) \cup \{\infty\}$，我们有

$$\mathrm{opt}_p(\mathcal{F}_{q,d}) \ge d^p \cdot \mathrm{opt}_p(\mathcal{F}_q).$$

证明。如果 $\mathrm{opt}_p(\mathcal{F}_{q,\ d}) = \infty$，则无需证明；如果 $\mathrm{opt}_p(\mathcal{F}_q) = \infty$，显然，只需将输入 $\mathbf{x}_i$ 限制在集合 $\{c\mathbf{e}_1$ 上：$c \in [0,1]\} \subset [0,1]^d$（其中 $\mathbf{e}_1 \in [0,1]^d$ 的第一个分量为 1，其余分量为 0），即可得到 $\mathrm{opt}_p(\mathcal{F}_{q,\ d}) = \infty$。

现在假设 $\mathrm{opt}_p(\mathcal{F}_{q,\ d})$ 和 $\mathrm{opt}_p(\mathcal{F}_q)$ 都是有限的。固定任意一个用于学习 $\mathcal{F}_{q,\ d}$ 的算法 A。令 $\mathbf{1}$ 表示全 1 的 d 元组，并令 $a(x_i : (x_0, z_0) \ldots, (x_{i-1}, z_{i-1}))$ 表示 A 在学习了 $j < i$ 对 $(x_j\mathbf{1}, z_j)$ 后，给定输入 $x_i\mathbf{1}$ 时的输出；假设存在一个属于 $\mathcal{F}_{q,\ d}$ 的函数，它经过点 $(x_j\mathbf{1}, z_j)$（$j < i$）。接着令 A' 为学习 $\mathcal{F}_q$ 的算法；在学习了 $j < i$ 对 (x_j, w_j) 后，给定输入 x_i 时，该算法返回输出

$$a'(x_i : (x_0, w_0), \ldots, (x_{i-1}, w_{i-1})) = \frac{a(x_i : (x_0, dw_0), \ldots, (x_{i-1}, dw_{i-1}))}{d},$$

给定 t 在 $\mathcal{F}_q$ 中存在一个函数，它经过点 (x_j, w_j)（$j < i$）。

固定 $\varepsilon > 0$。那么存在 $f \in \mathcal{F}_q$ 和一个输入序列 $x_0, x_1, \ldots, x_M$，使得 t

$$\sum_{i=1}^{M} |a'(x_i : (x_0, f(x_0)), \ldots, (x_{i-1}, f(x_{i-1}))) - f(x_i)|^p \geq \mathrm{opt}_p(\mathcal{F}_q) - \varepsilon.$$

面对 A，对手使用函数

$$\gamma(a_1, \ldots, a_d) = \sum_{i=1}^{d} f(a_i)$$

其输入为 $x_0\mathbf{1}, x_1\mathbf{1}, \ldots, x_M\mathbf{1}$。首先，假设 $q \in [1, \infty)$。注意，对于任意 $1 \leq k \leq d$ 和 $d-1$ 个实数 $x_i \in [0,1]$，其中 $1 \leq i \leq d$，但 $i \neq k$，

$$\int_0^1 \left| \frac{\mathrm{d}\gamma}{\mathrm{d}x_k} \right|^q \mathrm{d}x_k = \int_0^1 |f'(x)|^q \mathrm{d}x \leq 1$$

因为 $f \in \mathcal{F}_q$；因此，$\gamma \in \mathcal{F}_{q,\ d}$。接下来，假设 $q = \infty$。注意，对于所有 $x_k \in [0,1]$，有 $\left| \frac{\mathrm{d}\gamma}{\mathrm{d}x_k} \right| = |f'_{\mathrm{d}}(x_k)| \leq 1$，因为 $f \in \mathcal{F}_q$；因此，在这种情况下我们也有 $\gamma \in \mathcal{F}_{q,\ d}$。为了完成证明，令

$$e_i = |a'(x_i : (x_0, f(x_0)), \ldots, (x_{i-1}, f(x_{i-1}))) - f(x_i)|$$

and

$$k_i = |a(x_i : (x_0, \gamma(x_0\mathbf{1})), \ldots, (x_{i-1}, \gamma(x_{i-1}\mathbf{1}))) - \gamma(x_i\mathbf{1})|$$

对每个 i。因此，

$$k_i = |da'(x_i : (x_0, f(x_0)), \ldots, (x_{i-1}, f(x_{i-1}))) - df(x_i)| = de_i$$

and

$$\sum_{i=1}^{M} e_i^p \geq \mathrm{opt}_p(\mathcal{F}_q) - \varepsilon.$$

因此，

$$\sum_{i=1}^{M} k_i^p \geq d^p(\mathrm{opt}_p(\mathcal{F}_q) - \varepsilon).$$

取 $\varepsilon \to 0$ 即可完成证明。 ■

下一个推论由命题 3.1 得出，因为 $\text{opt}_p(\mathcal{F}_1) = \infty$ [9]。

推论 3 .2. 对于任意正整数 d 和实数 $p > 0$，我们有 $\text{opt}_p(\mathcal{F}_{1,d}) = \infty$.

现在我们直接证明一些关于 $\text{opt}_p(\mathcal{F}_{\infty,d})$ 的结果，这取决于 $p < d$ 还是 $p > d$。主要的否定结果如下。

定理 3.3. 设 $d > 0$ 为整数， p 为实数，且满足 $0 < p < d$。则 $\text{opt}_p(\mathcal{F}_{\infty,d}) = \infty$。

证明。固定任意一个用于学习 $\mathcal{F}_{\infty,d}$ 的算法 A。然后任取整数 $n \geq 1$，并令 S 为满足 $2nr$ 是奇整数的实数集合 $0 < r < 1$ （因此 $|S| = n$）。对手首先揭示 $f(0, \ldots, 0) = 0$，然后按字典序依次选择 $\mathbf{x}_i$，遍历 S^d 中的所有元素，从 A 接收输入 $\hat{y}_i$，并揭示与 $\hat{y}_i$ 相距更远的那个 $f(\mathbf{x}_i) = \pm\frac{1}{2n}$。

令 $\{x\} = x - \lfloor x \rfloor$ 表示 x 的小数部分。在进行完 $n^d + 1$ 次试验后，算法所揭示的 f 值与如下函数 $f\colon [0,1]^d \to \mathbb{R}$ 一致：

$$f(x_1, \ldots, x_d) = \pm\frac{1}{n} \min_{1 \leq i \leq d} \left(\min\left(\{nx_i\}, \{-nx_i\}\right)\right),$$

其中，符号 $\pm$ 的选取满足：对于任意 $\mathbf{x} = (x_1, \ldots, x_d) \in S^d$， $f(\mathbf{x})$ 都与对手的输出一致，并且 f 在任意区域内都具有恒定符号

$$\left(\frac{n_1}{n}, \frac{n_1+1}{n}\right) \times \ldots \times \left(\frac{n_d}{n}, \frac{n_d+1}{n}\right) \subset [0,1]^d$$

对于整数 $\mathbf{s}$ $0 \leq n_i < n$。一致性成立是因为对于所有 $x_i \in S$， $\{nx_i\} = \{-nx_i\} = \frac{1}{2}$.

首先，我们证明 $f \in \mathcal{F}_{\infty,d}$。任取 $1 \leq i \leq d$ 和 $\mathbf{x} = (x_1, \ldots, x_{d-1}) \in [0,1]^{d-1}$，并考虑如下函数 $g\colon [0,1] \to \mathbb{R}$： $g(x) = f(\mathbf{x}')$，其中 $\mathbf{x}' \in [0,1]^d$ 是通过将 x 插入 $\mathbf{x}$ 的第 i^{th} 个位置得到的。于是， g 为

$$g(x) = \pm\frac{1}{n} \min\left(\min\left(\{nx\}, \{-nx\}\right), M\right),$$

其中

$$M = \min_{1 \leq i \leq d-1} \left(\min\left(\{nx_i\}, \{-nx_i\}\right)\right).$$

显然， g 是分段线性函数，只有有限个点处 g' 未定义；根据 g 的定义，在其他所有地方都有 $|g'(x)| = 1$ 或 $g'(x) = 0$ ；此外，由于函数 $\min(\{nx\}, \{-nx\})$ 是连续的，因此 g 也是连续的。因此， $g \in \mathcal{F}_\infty$。由于这对 $1 \leq i \leq d$ 和 $\mathbf{x} \in [0,1]^d$ 的任意选择都成立，因此 $f \in \mathcal{F}_{\infty,\ d}$。

现在我们来求对手能够保证的误差下界。首次试验之后还有 n^d 次试验，每次都有 $\lceil y_i - f(\mathbf{x}_i)| \geq \frac{1}{2n}$；因此，对手可以保证

$$\sum_{i>0} |\hat{y}_i - f(\mathbf{x}_i)|^p \geq \frac{n^d}{(2n)^p} = \frac{1}{2^p} \cdot n^{d-p}.$$

Because $p < d$，当 n 增大时，这会变得任意大；因此 $\text{opt}_p(\mathcal{F}_{\infty,\ d}) = \infty$. ■

由于 $\mathcal{F}_\infty \subseteq \mathcal{F}_q$ 意味着，对于任意 $q \geq 1$， $\mathcal{F}_{\infty,d} \subseteq \mathcal{F}_{q,d}$ ，因此该界可推广至 $q \neq \infty$。

推论 3.4。设 $d > 0$ 为整数， p 为满足 $0 < p < d$的实数。对于任意 $q \geq 1$，我们有 $\mathrm{opt}_p(\mathcal{F}_{q,d}) = \infty$。

为了建立 $\mathrm{opt}_p(\mathcal{F}_{\infty,d})$ 的上界，我们证明以下引理。

引理 3.5。 对于 $f \in \mathcal{F}_{\infty,d}$，以及 $\mathbf{x}_1 = (x_{1,1}, \ldots, x_{d,1})$ 和 $\mathbf{x}_2 = (x_{1,2}, \ldots, x_{d,2}) \in [0,1]^d$，我们有

$$|f(\mathbf{x}_1) - f(\mathbf{x}_2)| \leq \sum_{i=1}^{d} |x_{i,1} - x_{i,2}|.$$

证明。固定这样的 f、$\mathbf{x}_1, \mathbf{x}_2$。定义一个由 $\mathbf{x}'_i \in [0,1]^d$ 组成的序列，其中 $0 \leq i \leq d$，使得 $\mathbf{x}'_i$ 的前i 个分量与 $\mathbf{x}_2$ 的前 i 个分量相同，而最后 $d-i$ 个分量与 $\mathbf{x}_1$ 的最后$d-i$ 个分量相同（因此 $\mathbf{x}'_0 = \mathbf{x}_1$ 且 $\mathbf{x}'_d = \mathbf{x}_2$）。根据三角不等式，

$$|f(\mathbf{x}_1) - f(\mathbf{x}_2)| = \left| \sum_{i=1}^{d} \left(f(\mathbf{x}'_{i-1}) - f(\mathbf{x}'_i) \right) \right| \leq \sum_{i=1}^{d} \left| f(\mathbf{x}'_{i-1}) - f(\mathbf{x}'_i) \right|.$$

现在考虑任意 $1 \leq i \leq d$。注意，$\mathbf{x}_{i-1}$ 和 $\mathbf{x}'_i$ 仅在其 i^{th} 个分量上不同，其中一个为 $x_{i,1}$，另一个为 $x_{i,2}$。那么，根据 $\mathcal{F}_{\infty,d}$ 的定义，并利用对于 $g \in \mathcal{F}_\infty$ 和 x_1， $x_2 \in [0,1]$，$|g(x_1) - g(x_2)| \leq |x_1 - x_2|$ 这一事实，可得 $\left| f(\mathbf{x}'_{i-1}) - f(\mathbf{x}'_i) \right| \leq |x_{i,1} - x_{i,2}|$。对 $1 \leq i \leq d$ 求和即可得到所需结果。■

引理 3.5 使得类 $\mathcal{F}_{\infty,d}$ 特别容易处理。通过使用最近邻算法，我们得到了以下上界。

定理 3.6. 假设 $p > d$。那么 $\mathrm{opt}_p(\mathcal{F}_{\infty,d}) \leq (2^d - 1)^{\frac{d^p}{1 - \frac{2^d}{2^p}}}$。

证明。考虑算法 A：它在第一个输入上猜测 0，并在试验 i （接收输入 $\mathbf{x}_0, \ldots, \mathbf{x}_{i-1}$ 后）选择使 $\mathbf{x}_j$ 与 $\mathbf{x}_i$ 之间的 L_1 距离最小的最小索引 $0 \leq j < i$，然后猜测 $\hat{y}_i = f(\mathbf{x}_j)$。我们将证明 $\mathscr{L}_p(A, \mathcal{F}_{\infty,d}) \leq (2^d - 1)^{\frac{d^p}{1 - \frac{2^d}{2^p}}}$。

固定 $f \in \mathcal{F}_{\infty,d}$，以及由 $\mathbf{x}_i \in [0,1]^d$ 中的元素组成的序列 $\mathbf{x}_0, \ldots, \mathbf{x}_m$。假设所有 $\mathbf{x}_i$ 互不相同。那么对于每个 $1 \leq i \leq m$，都存在最小整数 k_i，使得若将 $[0,1]^d$ 划分为 $2^{k_i d}$ 个区域，划分方式如下

$$\{(x_1, \ldots, x_d) \in [0,1]^d : n_i \leq 2^{k_i} x_i \leq n_i + 1\}$$

遍历所有由整数 $0 \leq n_i < 2^{k_i}$ 构成的 d 元组 $(n_1, \ldots, n_d)$ 时，$\mathbf{x}_i$ 不在与 $\mathbf{x}_0, \ldots, \mathbf{x}_{i-1}$ 相同的区域内。注意，由于对于任意 $1 \leq i \leq m$，$\mathbf{x}_0$ 和 $\mathbf{x}_i$ 都在 $[0,1]^d$ 中，因此所有 k_i 都至少为 1。对于每个整数 $k \geq 1$，令 c_k 表示满足 $k_i = k$ 的整数 $1 \leq i \leq m$ 的个数。根据鸽巢原理，对于任意固定的整数 $k \geq 0$，满足 $k_i \leq k$ 的指标 $1 \leq i \leq m$ 至多有 $2^{kd} - 1$ 个；否则，至少有 $2^{kd} + 1$ 个 $\mathbf{x}_i$ （包括 $\mathbf{x}_0$）会成为其所在长度为 2^{-k} 的超立方体区域中的第一个。因此

$$\sum_{k=1}^{K} c_k \leq 2^{Kd} - 1 \tag{4}$$

对于任意整数 $K \geq 1$。此外，对于任意 $1 \leq i \leq m$，$\mathbf{x}_i$ 与 $\mathbf{x}_0$、...、$\mathbf{x}_{i-1}$ 中的一个点位于同一个长度为 $2^{-(k_i - 1)}$ 的超立方体中，而该超立方体的两个对顶点之间的距离为 L_1，因此根据引理 3.5，

$$|\hat{y}_i - f(\mathbf{x}_i)| \leq \frac{d}{2^{k_i - 1}}. \tag{5}$$

综合上述结果,

$$\begin{aligned}\sum_{i=1}^{m}|\hat{y}_i - f(\mathbf{x}_i)|^p &\le \sum_{k\ge 1} c_k \left(\frac{d}{2^{k-1}}\right)^p = \sum_{K\ge 1}\left[\left(\sum_{k=1}^{K} c_k\right)\left(\left(\frac{d}{2^{K-1}}\right)^p - \left(\frac{d}{2^K}\right)^p\right)\right] \\ &\le d^p \sum_{K\ge 1}(2^{Kd}-1)(2^{-p(K-1)} - 2^{-pK}) = d^p(2^p-1)\sum_{K\ge 1} 2^{-pK}(2^{Kd}-1) \\ &= d^p(2^p-1)\left(\frac{2^{d-p}}{1-2^{d-p}} - \frac{2^{-p}}{1-2^{-p}}\right) = \frac{(2^d-1)d^p}{1-\frac{2^d}{2^p}}.\end{aligned}$$

这对所有 $f \in \mathcal{F}_{\infty,\,d}$ 以及 $\mathbf{x}_i$ 的序列都成立；因此 $\mathscr{L}_p(A, \mathcal{F}_{\infty,\,d}) \le \frac{(2^d-1)\,d^p}{1-\frac{2^d}{2^p}}$。 ■

由于对所有$p \ge 2$都有 $\mathrm{opt}_p(\mathcal{F}_\infty) = 1$ ，下一个推论可由命题 3.1 和定理 3.6 得出。

推论 3.7。对于任意固定的正整数 d 和实数 $p \ge d+1$，我们有 $\mathrm{opt}_p(\mathcal{F}_{\infty,\,d}) = \Theta(d^p)$，其中上界中的常数仅取决于 d。

当试验次数有界时，我们还可以利用定理 3.3 和定理 3.6，得到学习 $\mathcal{F}_{\infty,\,d}$ 的最坏情况误差的紧确界。

推论 3.8。设 $d > 0$ 为整数， p 为满足 $0 < p < d$ 的实数。那么 $\mathrm{opt}_p(\mathcal{F}_{\infty,\,d}, m) = \Theta(m^{1-\frac{p}{d}})$，其中界中的常数取决于 p 和 d。

证明。根据定理 3.3，对于 $n = \lfloor m^{1/d} \rfloor$，我们有 $\mathrm{opt}_p(\mathcal{F}_{\infty,\,d}, m) \ge \frac{1}{2^p} n^{d-p}$，因此得到下界 $\mathrm{opt}_p(\mathcal{F}_{\infty,\,d}, m) \ge \frac{1}{2^p} m^{\frac{d-p}{d}}(1 - o(1))$。对于上界，我们使用定理 3.6 中的算法和记号，并取 $K = \left\lceil \log_2\left(\frac{m+1}{d}\right) \right\rceil$，得到

$$\begin{aligned}\sum_{i=1}^{m}|\hat{y}_i - f(\mathbf{x}_i)|^p &\le \sum_{i=1}^{m}\left(\frac{d}{2^{k_i-1}}\right)^p \le \sum_{k=1}^{K}(2^{kd} - 2^{(k-1)d})\left(\frac{d}{2^{k-1}}\right)^p \\ &= d^p(2^d-1)\sum_{k=1}^{K} 2^{(k-1)(d-p)} = d^p(2^d-1)\frac{2^{K(d-p)}-1}{2^{d-p}-1} \\ &< \frac{d^p(2^d-1)2^{d-p}}{2^{d-p}-1} m^{\frac{d-p}{d}}(1+o(1)),\end{aligned}$$

其中，第一个不等式由不等式 5 得出，第二个不等式由不等式 4 得出，因为 $\left(\frac{d}{2^{k-1}}\right)^p$ 随 k 递减。因此

$$\mathrm{opt}_p(\mathcal{F}_{\infty,d}, m) \le \frac{d^p(2^d-1)2^{d-p}}{2^{d-p}-1} m^{\frac{d-p}{d}}(1+o(1)).$$

■

4 讨论与开放问题

借助本文的结果，对于所有 p、 $q \geq 1$，$\mathrm{opt}_p(\mathcal{F}_q)$ 的值现在都被限定在相差一个常数因子的范围内；但 $q \in (1,2)$ 且 $p \in (1,2) \cup (2, 2+\frac{1}{q-1})$ 的情况除外。具体而言，将本文的结果与 [9], 中的结果结合起来，我们现在知道，对于位于以下区域内的所有 (p,q)，$\mathrm{opt}_p(\mathcal{F}_q) = 1$。

- $p, q \geq 2$
- $q \in (1,2)$ 和 $p \geq 2 + \frac{1}{q-1}$

除了研究 $\mathrm{opt}_p(\mathcal{F}_q)$ 不受常数因子限制的区域外，当 $\varepsilon \in (0,1)$ 且 $q \in [2,\infty) \cup \{\infty\}$ 时，仍需缩小 $\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q) = \Theta(\varepsilon^{-\frac{1}{2}})$ 的上界与下界之间的常数差距。另一个类似问题是，当 $q \in (1,2)$ 时，缩小 $\mathrm{opt}_2(\mathcal{F}_q) = \Theta(\varepsilon^{-1})$ 的上界与下界之间的常数差距。

本文的结果还有助于刻画使 $\mathrm{opt}_p(\mathcal{F}_q)$ 有限的 (p,q) 的取值。在本文发表之前，人们仅知道当 $p > 1$ 且 $q \geq 2$ 时 $\mathrm{opt}_2(\mathcal{F}_q)$ 有限，以及当 $p = 1$ 或 $q = 1$ 时 $\mathrm{opt}_p(\mathcal{F}_q) = \infty$。借助我们的新结果，我们现在知道，当 $p \geq 2$且 $q > 1$ 时，$\mathrm{opt}_p(\mathcal{F}_q)$ 也有限。我们对这个问题提出如下猜想。

猜想 4.1. 对所有 $p > 1$ 和 $q > 1$，$\mathrm{opt}_p(\mathcal{F}_q)$ 都是有限的。

除了关于单变量光滑函数的新结果外，我们还将该模型推广到了多变量函数，并为这种多变量在线学习情形找到了若干界。我们证明，当 $0 < p < d$ 时，$\mathrm{opt}_p(\mathcal{F}_{\infty,d})$ 是无限的；当 $p > d$ 时，它是有限的，但对于 $d > 1$，$\mathrm{opt}_d(\mathcal{F}_{\infty,d})$ 是否有限仍有待确定。对于有限的 $q \geq 1$ 和 $0 < p < d$，我们还知道 $\mathrm{opt}_p(\mathcal{F}_{q,d})$ 是无限的，但对于 $p \geq d$ 和 $q \in (1,\infty)$，$\mathrm{opt}_p(\mathcal{F}_{q,d})$ 是否有限仍有待确定。此外，我们证明，对于任意固定的正整数 d，当 $p \geq d+1$ 时，$\mathrm{opt}_p(\mathcal{F}_{\infty,d}) = \Theta(d^p)$，其中上界中的常数取决于 d。上界与下界之间的乘法差距为 $2^{d+1} - 2$。我们猜想，对于相对于 d 和 q 足够大的 p，下界是紧的。

猜想 4.2. 对所有 $q \in [1,\infty) \cup \{\infty\}$、所有正整数 d，以及所有相对于 q 和 d充分大的实数 p，我们有 $\mathrm{opt}_p(\mathcal{F}_{q,\ d}) = d^p$。

论文 [9] 和 [12] 研究了当 $q \geq 2$时的 $\mathrm{opt}_1(\mathcal{F}_q, m)$，其中 m 是试验次数。研究满足 $0 < \varepsilon < 1$ 和 $q \geq 1$的 $p = 1+\varepsilon$下的 $\mathrm{opt}_p(\mathcal{F}_q, m)$ 是很自然的，因为当 $\varepsilon \to 0$时，$\mathrm{opt}_{1+\varepsilon}(\mathcal{F}_q)$ 可以任意增大。对于任意固定的正整数 d 和满足 $0 < p < d$的固定实数 p，我们将 $\mathrm{opt}_p(\mathcal{F}_{\infty,\ d}, m)$ 的界限制在相差一个常数因子的范围内，但界中的常数取决于 p 和 d。仍需缩小 $\frac{d^p}{(2^d-1\ 2^{d-p}}m^{\frac{d-p}{d}})(1+o(1))$ $2^{d-p}-1$ 的上界与 $\frac{1}{2^p}m^{\frac{d-p}{d}}(1-o(1))$ 的下界之间的差距。研究 q取有限值时的 $\mathrm{opt}_p(\mathcal{F}_{q,\ d}, m)$ 也会很有意思。

另一个可能的方向是研究具有额外限制条件的光滑函数族。例如，设 $\mathcal{E}_q \subseteq \mathcal{F}_q$ 为指数函数族 $f(x) = e^{ax+b}$，其中满足 $f \in \mathcal{F}_q$。

命题 4.3。对于所有 $p > 0$ 和 $q \geq 1$，我们有 $\mathrm{opt}_p(\mathcal{E}_q) = 1$。

证明。上界 $\mathrm{opt}_p(\mathcal{E}_q) \le 1$ 由引理 2.5 得出，因为学习者在输入不同的两轮之后就知道该函数，且第一轮不计入总误差。对于下界，考虑一个对手：其选择某个 $\varepsilon \in (0,1)$，定义 $\varrho = 1 - \sqrt[1-\varepsilon]{1-\varepsilon}$，并揭示 $f(0) = -(1-\varepsilon)\ln(1-\varepsilon)$。在第二轮中，对手会揭示 $f(1) = -\frac{1}{\ln(1-\varepsilon)}$ 或 $f(1) = -\frac{1-\varrho}{\ln(1-\varrho)}$，选择其中使学习者猜测所产生的误差最大者。

若 $f(1) = -\frac{1}{\ln(1-\varepsilon)}$，则 $f(x) = e^{ax+b}$ 满足 $a = -\ln(1-\varepsilon)$，且 $b = \ln(1-\varepsilon) - \ln(-\ln(1-\varepsilon))$。注意，对所有 $x \in [0,1]$，$f'(x) = ae^{ax+b} \in [1-\varepsilon, 1]$，因此对所有 $q \ge 1$，~~$f \in \mathcal{E}_q$~~。若 $f(1) = -(1-\varrho)\ln(1-\varrho)$，则 $f(x) = e^{ax+b}$ 满足 $a = \ln(1-\varrho)$，且 $b = -\ln(-\ln(1-\varrho))$。注意，对所有 $x \in [0,1]$，$f'(x) = ae^{ax+b} \in [-1, -1+\varrho]$，因此对所有 $q \ge 1$，$f \in \mathcal{E}_q$。此外，注意

$$\lim_{\varepsilon \to 0} \left(-\frac{1}{\ln(1-\varepsilon)} + \frac{1-\varrho}{\ln(1-\varrho)} \right) = \lim_{\varepsilon \to 0} \left(\frac{-1 + (1-\varepsilon)\sqrt[1-\varepsilon]{1-\varepsilon}}{\ln(1-\varepsilon)} \right) = 2,$$

根据洛必达法则。因此 $\mathrm{opt}_p(\mathcal{E}_q) \ge 1$。 ■

令 $\mathcal{P}_{q,\,m} \subseteq \mathcal{F}_q$ 为次数至多为 m 的多项式函数族 $f \in \mathcal{F}_q$。不难看出，对所有 $p > 0$ 和 $q \ge 1$，$\mathrm{opt}_p(\mathcal{P}_{q,\,1}) = 1$；但研究 $m > 1$ 时的 $\mathrm{opt}_p(\mathcal{P}_{q,\,m})$ 会很有意思。注意，对所有 $p > 0$、$q \ge 1$ 和 $m \ge 1$，我们有 $\mathrm{opt}_p(\mathcal{P}_{q,\,m}) \le \mathrm{opt}_p(\mathcal{F}_q, m)$，因为学习者在接受 $m+1$ 个不同输入的测试后，就能确定地知道 $f \in \mathcal{P}_{q,\,m}$。令 $\mathcal{P}_q \subseteq \mathcal{F}_q$ 为所有多项式函数 $f \in \mathcal{F}_q$ 组成的族。我们针对这个函数族提出以下猜想。

猜想 4.4. 对所有 $p > 0$ 和 $q \ge 1$，我们有 $\mathrm{opt}_p(\mathcal{P}_q) = \mathrm{opt}_p(\mathcal{F}_q)$。

还可以研究 $\mathcal{F}_q$ 的其他一些可能子集，例如最多包含 k 个片段的分段函数（其中每个片段都是次数至多为 m 的多项式）、指数函数之和，以及三角函数之和。

最后，我们回到引言中的问题：预测给定地点次日的温度范围。具体而言，考虑这样一个单变量问题：仅根据一年中的时间来预测次日的温度范围。将 [9] 中的模型用于这一温度预测问题时，一个问题在于，一年中的时间这一输入相同，在不同年份可能对应不同的温度范围输出。对此，更现实的建模方式是根据输入从概率分布中选择输出。为了使学习者能够保证最坏情况误差存在有限上界，需要限制试验次数，并对概率分布施加约束。例如，可以要求该概率分布的密度函数具有与 [9], 中的函数类似的光滑性，并要求密度函数的支撑集是某个 $r > 0$ 的 $[0, r]$ 子集。研究这样的模型将是未来研究的一个有趣方向。请注意，当支撑集只包含一个点时，该模型就退化为 [9] 中的模型。

5 致谢

这项研究的大部分工作是在 PRIMES 2022 期间完成的。我们感谢组织者提供了这一研究机会。我们的论文涵盖了证明定理 1.2 的工作 [5], 。我们也感谢匿名审稿人提出的宝贵意见，这些意见提升了本文结果的清晰度和呈现效果。

参考文献

[1] D. Angluin，查询与概念学习。《机器学习》2（1988）319–342。[2] A. Barron，人工神经网络的逼近与估计界。计算学习理论研讨会（1991）[3] N. Cesa-Bianchi、P.M. Long 和 M.K. Warmuth，使用线性函数和梯度下降进行预测时的最坏情况二次损失界。《IEEE神经网络汇刊》7（1996）604–619。[4] V. Faber 和 J. Mycielski，学习定理的应用。《Fundamenta Informaticae》15（1991）145–167。[5] J. Geneson，单变量光滑函数在线学习的更紧界。CoRR abs/2105.14648（2021）[6] W. Hardle，平滑技术。施普林格出版社（1991）[7] D. Haussler，推广 PAC 模型：基于度量维数的一致收敛结果所给出的样本量界。第30届年度计算机科学基础研讨会论文集（1989）[8] S. Kaczmarz，线性方程组的近似解。波兰科学院学报A辑 35（1937）355–357。[9] D. Kimber 和 P. M. Long，单变量光滑函数的在线学习。《理论计算机科学》148（1995）141–156。[10] N. Littlestone，当无关属性大量存在时如何快速学习：一种新的线性阈值算法。《机器学习》2（1988）285–318。[11] N. Littlestone 和 M.K. Warmuth，加权多数算法。第30届年度计算机科学基础研讨会论文集（1989）[12] P. M. Long，关于单变量光滑函数在线学习的改进界。《理论计算机科学》241（2000）25–35。[13] J. Mycielski，线性算子的学习算法。美国数学会会刊 103（1988）547–550。[14] B. Widrow 和 M.E. Hoff，自适应开关电路。1960年 IRE WESCON 会议记录（1960）96–104。