Sharper bounds for online learning of smooth functions of a single variable

Jesse Geneson

June 1, 2021

Abstract

We investigate the generalization of the mistake-bound model to continuous real-valued single variable functions. Let $F_q$ be the class of absolutely continuous functions $f : [0, 1] \to \mathbb{R}$ with $\|f'\|_q \leq 1$, and define $\operatorname{opt}_p(F_q)$ as the best possible bound on the worst-case sum of the $p^{th}$ powers of the absolute prediction errors over any number of trials. Kimber and Long (Theoretical Computer Science, 1995) proved for $q \geq 2$ that $\operatorname{opt}_p(F_q) = 1$ when $p \geq 2$ and $\operatorname{opt}_p(F_q) = \infty$ when $p = 1$. For $1 < p < 2$ with $p = 1 + \epsilon$, the only known bound was $\operatorname{opt}_p(F_q) = O(\epsilon^{-1})$ from the same paper. We show for all $\epsilon \in (0, 1)$ and $q \geq 2$ that $\operatorname{opt}_{1+\epsilon}(F_q) = \Theta(\epsilon^{-\frac{1}{2}})$, where the constants in the bound do not depend on $q$. We also show that $\operatorname{opt}_{1+\epsilon}(F_{\infty}) = \Theta(\epsilon^{-\frac{1}{2}})$.

Keywords: online learning, smooth functions, single variable, general loss functions

1 Introduction

We consider a model of online learning previously studied in [1, 8, 9, 11, 7, 10] where an algorithm $A$ tries to learn a real-valued function $f$ from some class $F$. In each trial of this model, $A$ receives an input $x_t \in [0, 1]$, it must output some prediction $\hat{y}_t$ for $f(x_t)$, and then $A$ discovers the true value of $f(x_t)$.

For each $p \geq 1$ and $x = (x_0, \ldots, x_m) \in [0, 1]^{m+1}$, define $L_p(A, f, x) = \sum_{t=1}^{m} |\hat{y}_t - f(x_t)|^p$. Note that the summation starts on the second trial, since the guess on the first trial does not reflect the algorithm’s learning ability. Define $L_p(A, F) = \sup_{f \in F} L_p(A, f, x)$. In this paper, we study the optimum $\operatorname{opt}_p(F) = \inf_A L_p(A, F)$.

In particular we focus on the class of functions whose first derivatives have $q$-norms at most 1. For any real number $q \geq 1$, let $F_q$ be the family of absolutely continuous functions $f : [0, 1] \to \mathbb{R}$ for which $\int_0^1 |f'(x)|^q dx \leq 1$. Let $F_{\infty}$ be the family of absolutely continuous functions $f : [0, 1] \to \mathbb{R}$ for which $\sup_{x \in [0, 1]} |f'(x)| \leq 1$. By the definition of $F_{\infty}$ and Jensen’s inequality respectively, we have $F_{\infty} \subseteq F_r \subseteq F_q$ for any $1 \leq q \leq r$. Thus $\operatorname{opt}_p(F_{\infty}) \leq \operatorname{opt}_p(F_r) \leq \operatorname{opt}_p(F_q)$ for any $1 \leq q \leq r$. 

1
Hölder’s inequality to obtain a sharper bound of $\opt$, and a noisy version of this problem was studied in [3].

Kimber and Long [7] proved that $\opt_p(\mathcal{F}_1) = \infty$ for all $p \geq 1$. They also showed that $\opt_1(\mathcal{F}_q) = \opt_1(\mathcal{F}_\infty) = \infty$ for all $q \geq 1$. In contrast, they found that $\opt_p(\mathcal{F}_q) = \opt_p(\mathcal{F}_\infty) = 1$ for all $p \geq 2$ and $q \geq 2$. This was also proved by Faber and Mycielski [4] using a different algorithm, and a noisy version of this problem was studied in [3].

For $p = 1 + \epsilon$ with $\epsilon \in (0, 1)$, Kimber and Long proved that $\opt_p(\mathcal{F}_q) = O(\epsilon^{-1})$ for all $q \geq 2$, which implies that $\opt_p(\mathcal{F}_\infty) = O(\epsilon^{-1})$. However, these bounds are not sharp. In this paper, we determine $\opt_{1+\epsilon}(\mathcal{F}_q)$ up to a constant factor for all $\epsilon \in (0, 1)$ and $q \geq 2$.

**Theorem 1.1.** For all $\epsilon \in (0, 1)$, we have $\opt_{1+\epsilon}(\mathcal{F}_\infty) = \Theta(\epsilon^{-\frac{1}{2}})$ and $\opt_{1+\epsilon}(\mathcal{F}_q) = \Theta(\epsilon^{-\frac{1}{2}})$ for all $q \geq 2$, where the constants in the bound do not depend on $q$.

The proof splits into an upper bound and a lower bound. For the upper bound, we use Hölder’s inequality combined with past results of Kimber and Long. For the lower bound, we modify a construction used in [10], where Long obtained bounds on a finite variant of $\opt_1(\mathcal{F}_q)$ that depends on the number of trials $m$.

## 2 Sharp bounds for general loss functions

In this section, we derive an upper bound and a lower bound to prove Theorem 1.1. We introduce some terminology from [7] and [10] that we use for the proofs in this section. For a function $f : [0, 1] \to \mathbb{R}$, we define $J[f] = \int_0^1 f'(x)^2\,dx$. Given a finite subset $S \subseteq [0, 1] \times \mathbb{R}$ with $S = \{(u_i, v_i) : 1 \leq i \leq m\}$ and $u_1 < u_2 < \cdots < u_m$, we define $f_S : [0, 1] \to \mathbb{R}$ as follows.

Let $f_\emptyset(x) = 0$ for all $x$, and for each nonempty $S$ let $f_S$ be the piecewise function defined by $f_S(x) = v_1$ for $x \leq u_1$, $f_S(x) = v_i + \frac{(x-u_{i-1})(v_{i+1}-v_i)}{u_i-u_{i-1}}$ for $x \in (u_i, u_{i+1}]$, and $f_S(x) = v_m$ for $x > u_m$.

For the upper bound, we use the LININT learning algorithm which is defined using $f_S$. Specifically we define LININT($\emptyset, x_1$) = 0 and LININT(((x_1, y_1), \ldots, (x_{t-1}, y_{t-1}), x_t) = \int_0^1 f_((x_1, y_1), \ldots, (x_{t-1}, y_{t-1}))\,dx_t$. The beginning of our proof is similar to Kimber and Long’s proof from [7] that $\opt_{1+\epsilon}(\mathcal{F}_2) = O(\epsilon^{-1})$ for $\epsilon \in (0, 1)$, but we change the end of the proof by using Hölder’s inequality to obtain a sharper bound of $\opt_{1+\epsilon}(\mathcal{F}_2) = O(\epsilon^{-\frac{1}{2}})$.

**Theorem 2.1.** If $\epsilon \in (0, 1)$, then $\opt_{1+\epsilon}(\mathcal{F}_2) = O(\epsilon^{-\frac{1}{2}})$.

**Proof.** Fix $\epsilon \in (0, 1)$ and let $p = 1 + \epsilon$. Let $x_0, \ldots, x_m$ be any sequence of elements of $[0, 1]$, and let $f \in \mathcal{F}_2$. Let $y_1, \ldots, y_m$ be LININT’s predictions on trials $1, \ldots, m$. For each $i > 1$, let $d_i = \min_{j<i} |x_j - x_i|$ and let $e_i = |y_i - f(x_i)|$. Kimber and Long proved in Theorem 17 of [7] that $\sum_{i=1}^m \frac{e_i^2}{d_i^2} \leq 1$. In the same theorem, they also proved that $\sum_{i=1}^m d_i^p \leq 1 + \frac{1}{2p-2}$ for $p > 1$. This is where our proofs diverge.

First, note that $\sum_{i=1}^m e_i^p = \sum_{i=1}^m \frac{e_i^p}{d_i^2} \cdot d_i^2$. By Hölder’s inequality, we have

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

$$\left(\sum_{i=1}^m \frac{e_i^p}{d_i^2}\right)^{\frac{2}{p}} \cdot \left(\sum_{i=1}^m (d_i^2)^{\frac{2}{2p}}\right)^{1-\frac{2}{p}}.$$
Note that $\sum_{i=1}^m (\frac{e^i}{d_i^2})^{\frac{2}{p}} = \sum_{i=1}^m \frac{e^i}{d_i^2} \leq 1$ by the first bound from [7] that we cited in the first paragraph. By the second bound from [7] that we cited in the first paragraph, we have $\sum_{i=1}^m (d_i^2)^{\frac{2}{p}} = \sum_{i=1}^m d_i^{2/p} \leq 1 + \frac{1}{2^{2/p-2}}$ since $\frac{p}{2-p} > 1$.

Thus $(\sum_{i=1}^m (d_i^2)^{\frac{2}{p}})^{1-\frac{2}{p}} \leq (1 + \frac{1}{2^{2/p-2}})^{1-\frac{2}{p}} = (1 + \frac{1}{2^{2/p-2}})^{1-\frac{2}{p}}$.

Let $\delta = \frac{1+1}{4\epsilon} - 1$, and note that $\delta = \frac{2\epsilon}{1+\epsilon} \geq 2\epsilon$ since $\epsilon \in (0, 1)$. Thus $(1 + \frac{1}{2^{2/p-2}})^{1-\frac{2}{p}} = (1 + \frac{1}{2^{2/p-2}})^{1-\frac{2}{p}} = O(\delta^{-\frac{2}{p}})$ since $e^{\delta \ln 2} \geq 1 + \delta \ln 2$. Moreover $\delta^{-\frac{2}{p}} = O(\epsilon^{-\frac{2}{2\epsilon}}) = O(\epsilon^{-\frac{2}{p}})$ since $\delta \geq 2\epsilon$ and $\epsilon^p = \Theta(1)$ for $\epsilon \in (0, 1)$.

Thus we have proved that $\sum_{i=1}^m e_i^p = O(\epsilon^{-\frac{1}{2}})$, so $\text{opt}_{1+\epsilon}(F_2) = O(\epsilon^{-\frac{1}{2}})$.

We obtain the next corollary from Theorem [2.1] since $\text{opt}_p(F_\infty) \leq \text{opt}_p(F_r) \leq \text{opt}_p(F_q)$ whenever $1 \leq q \leq r$.

**Corollary 2.2.** If $\epsilon \in (0, 1)$, then $\text{opt}_{1+\epsilon}(F_\infty) = O(\epsilon^{-\frac{1}{2}})$ and $\text{opt}_{1+\epsilon}(F_q) = O(\epsilon^{-\frac{1}{2}})$ for all $q \geq 2$, where the constant does not depend on $q$.

In order to show that the last theorem is sharp up to a constant factor, we construct a family of functions in $F_\infty$. Our proof uses the following lemma from [7] which was also used in [10].

**Lemma 2.3.** Let $S \subseteq [0, 1] \times \mathbb{R}$ with $S = \{(u_i, v_i) : 1 \leq i \leq m\}$ and $u_1 < u_2 < \cdots < u_m$. If $(x, y) \in [0, 1] \times \mathbb{R}$ and there exists $1 \leq j \leq m$ such that $|x - u_j| = \min_i |x - u_i|$, then $J[f_{S \cup \{(x, y)\}}] = J[f_S] + \frac{2(y - f_S(x))^2}{\min_i |x - u_i|}$.

The construction and method in the following proof is very similar to one used by Long in [10] to obtain bounds for a finite variant of $\text{opt}_1(F_q)$ for $q \geq 2$ that depends on the number of trials $m$. The proofs differ in that we have adjusted the parameters of the construction to give the desired lower bound in terms of $\epsilon$, and some summations that were finite in Long’s proof are infinite in the next proof.

**Theorem 2.4.** If $\epsilon \in (0, 1)$, then $\text{opt}_{1+\epsilon}(F_\infty) = \Omega(\epsilon^{-\frac{1}{2}})$.

**Proof.** Since $\text{opt}_{1+\epsilon}(F_\infty) \geq 1$ for all $\epsilon \in (0, 1)$, it suffices to prove the theorem for $\epsilon \in (0, \frac{1}{2})$. Define $x_0 = 1$ and $y_0 = 0$. For natural numbers $i, j$ with $0 \leq j < 2^{i-1}$, define $x_{2^{i-1} + j} = \frac{1}{2^i} + \frac{j}{2^{i-1}}$. For each $i = 1, 2, \ldots$, we consider the trials for $x_{2^{i-1}}, \ldots, x_{2^i-1}$ to be part of stage $i$, so that $x_1 = \frac{1}{2}$ is in stage 1, $x_2 = \frac{1}{4}$ and $x_3 = \frac{3}{4}$ are in stage 2, and so on.

Let $A$ be any algorithm for learning $F_\infty$. Using $A$, we construct an infinite sequence of functions $f_0, f_1, \cdots \in F_\infty$ and an infinite sequence of real numbers $y_0, y_1, \ldots$ for which $f_i$ is consistent with the $x_k$ and $y_k$ values for $k \leq i$ and $A$ has total $(1 + \epsilon)$-error at least $\sum_{i=1}^{\infty} 2^{k-2}(\frac{\sqrt{(1+\epsilon)^2}}{2k+1})^{1+\epsilon}$ when $f_i$ is the target function. This will imply that $\text{opt}_{1+\epsilon}(F_\infty) \geq \sum_{k=1}^{\infty} 2^{k-2}(\frac{\sqrt{(1+\epsilon)^2}}{2k+1})^{1+\epsilon}$.

For the proof, we will also define another infinite sequence of functions $g_{i,j}$ with $0 \leq j \leq 2^{i-1}$ and another infinite sequence of real numbers $v_1, v_2, \ldots$. We start by letting $f_0$ be the 0-function.

3
Fix a stage $i$, and let $g_{t,0} = f_{2i-1}$. Let $t$ be a trial in stage $i$, and let $v_i$ be whichever of $f_{t-1}(x_t) \pm \frac{\sqrt{1-\epsilon}}{2^{i+1}}$ is furthest from $y_t$. Let $g_{i,t-2i+1+1}$ be the function which linearly interpolates $\{(0,0), (1,0)\} \cup \{(x_s, y_s) : s < 2^{i-1}\} \cup \{(x_s, y_s) : 2^{i-1} \leq s \leq t\}$.

For any $t \geq 1$, let $L_t$ and $R_t$ be the elements of $\{0,1\} \cup \{x_s : s < t\}$ that are closest to $x_t$ on the left and right respectively. If both $|v_t - f_{t-1}(L_t)| \leq 2^{-i}$ and $|v_t - f_{t-1}(R_t)| \leq 2^{-i}$, then let $y_t = v_t$. Otherwise we let $y_t = f_{t-1}(x_t)$. Finally, we define $f_t$ to be the function which linearly interpolates $\{(0,0), (1,0)\} \cup \{(x_s, y_s) : s \leq t\}$.

By definition, we have $f_t \in F_{\infty}$ for each $t \geq 0$. We will prove next that for all $i, j$ we have $J[g_{i,j}] \leq \frac{1}{4}$. The proof will use double induction, first on $i$ and then on $j$, and we will prove a slightly stronger statement.

In order to prove that $J[g_{i,j}] \leq \frac{1}{4}$ for all $i$ and $j$, we will prove that $J[f_{2i-1} - \frac{1}{4}] \leq \frac{1}{4} \sum_{k=0}^{\lfloor i/2 \rfloor} (1-\epsilon)^k$ for all $i \geq 1$. Note that this is equivalent to proving that $J[g_{i,0}] \leq \frac{1}{4} \sum_{k=0}^{i-1} (1-\epsilon)^k$ for all $i \geq 1$. Clearly this is true for $i = 1$, which is the base case of the induction on $i$.

Fix some integer $i \geq 1$. We will assume that $J[f_{2i-1} - \frac{1}{4}] \leq \frac{1}{4} \sum_{k=0}^{i-1} (1-\epsilon)^k$, and use this to prove that $J[f_{2i-1}] \leq \frac{1}{4} \sum_{k=0}^{i} (1-\epsilon)^k$. In order to prove that $J[f_{2i-1}] \leq \frac{1}{4} \sum_{k=0}^{i} (1-\epsilon)^k$, we will prove the stronger claim that $J[g_{i,j}] \leq \left(\frac{1}{4} \sum_{k=0}^{i-1} (1-\epsilon)^k + \frac{j(1-\epsilon)^i}{2^{i+1}}\right)$. This follows from the inductive hypothesis for $i$ and the definition of $g_{i,0}$ when $j = 0$, which is the base case of the induction on $j$.

Fix some integer $j$ with $0 \leq j \leq 2^{i-1} - 1$ and assume that $J[g_{i,j}] \leq \left(\frac{1}{4} \sum_{k=0}^{i-1} (1-\epsilon)^k + \frac{j(1-\epsilon)^i}{2^{i+1}}\right)$. By Lemma 2.3 we have $J[g_{i,j+1}] = J[g_{i,j}] + \frac{2j(\sqrt{1-\epsilon})^2}{2^{i+1}} = J[g_{i,j}] + \frac{(1-\epsilon)^i}{2^{i+1}}$.

By the inductive hypothesis for $j$, we obtain $J[g_{i,j+1}] \leq \left(\frac{1}{4} \sum_{k=0}^{i-1} (1-\epsilon)^k + \frac{j(1-\epsilon)^i}{2^{i+1}}\right)^2$, which completes the inductive step for $j$. Substituting $j = 2^{i-1}$, we obtain $J[g_{i,2^{i-1}-1}] \leq \frac{1}{4} \sum_{k=0}^{i} (1-\epsilon)^k$. Note that Lemma 2.3 implies that $J[f_{2^{i-1}-1+j}] \leq J[g_{i,j}]$ for all $j = 0, \ldots, 2^{i-1}$, so we obtain $J[f_{2^{i-1}}] \leq \frac{1}{4} \sum_{k=0}^{i} (1-\epsilon)^k$ using $j = 2^{i-1}$, which completes the inductive step for $i$.

Since $J[g_{i,j}] \leq \left(\frac{1}{4} \sum_{k=0}^{i-1} (1-\epsilon)^k + \frac{j(1-\epsilon)^i}{2^{i+1}}\right)$, we obtain $J[g_{i,j}] \leq \left(\frac{1}{4} \sum_{k=0}^{i-1} (1-\epsilon)^k + \frac{(1-\epsilon)^i}{4}\right) = \frac{1}{4} \sum_{k=0}^{i} (1-\epsilon)^k$ for all $j$ with $0 \leq j \leq 2^{i-1}$. Note that $\frac{1}{4} \sum_{k=0}^{i} (1-\epsilon)^k < \frac{1}{4} \sum_{k=0}^{\infty} (1-\epsilon)^k = \frac{1}{4}$.

For each $i \geq 1$, we claim that $y_t = f_{t-1}(x_t)$ for at most half of the intervals $t$ in stage $i$. For each trial $t$ with $y_t = f_{t-1}(x_t)$, note that the absolute value of the slope of $g_{i,t-2^{i-1}+1}$ must exceed 1 in at least one of the intervals of length $2^{-i}$ on either side of $x_t$. If $y_t = f_{t-1}(x_t)$ for at least $b$ of the trials in stage $i$, then restricting to intervals of slope at least 1 implies that $J[g_{i,2^{i-1}-1}] \geq b2^{-i}$.

Since $J[g_{i,2^{i-1}-1}] \leq \frac{1}{4}$, we must have $b \leq 2^{i-2}$. Thus during stage $i$, there are at most $2^{i-2}$ trials $t$ with $y_t = f_{t-1}(x_t)$, which implies that there are at least $2^{i-2}$ trials with $y_t = v_t$. In each of those trials, $A$ was off by at least $\frac{\sqrt{1-\epsilon}}{2^{i+1}}$, so the total $(1+\epsilon)$-error of $A$ after $i$ stages is at least $\sum_{k=1}^{i} 2^{k-2} \left(\frac{\sqrt{1-\epsilon}}{2^{k+1}}\right)^{1+\epsilon}$. Thus
\[
\text{opt}_{1+\epsilon}(\mathcal{F}_\infty) \geq \sum_{k=1}^{\infty} 2^{k-2} \left( \frac{\sqrt{\epsilon}(1-\epsilon)^{\frac{k}{2}}}{2^{k+1}} \right)_{1+\epsilon} = \\
\frac{\frac{1}{2} \left( 1 - \epsilon \right)^{1+\epsilon}}{1 - 2 \left( \frac{\sqrt{1-\epsilon}}{2} \right)^{1+\epsilon}} = \\
\Omega \left( \frac{\epsilon(1-\epsilon)^{\frac{1+\epsilon}{2}}}{1 - 2 \left( \frac{\sqrt{1-\epsilon}}{2} \right)^{1+\epsilon}} \right).
\]

Since \( \epsilon^\epsilon = \Theta(1) \) and \((1-\epsilon)^{1+\epsilon} = \Theta(1)\) for \( \epsilon \in (0, \frac{1}{2}) \), we have \( \text{opt}_{1+\epsilon}(\mathcal{F}_\infty) = \Omega \left( \frac{\sqrt{\epsilon}}{1 - 2 \left( \frac{\sqrt{1-\epsilon}}{2} \right)^{1+\epsilon}} \right) \).

Since \( 2^\epsilon = \Theta(1) \) for \( \epsilon \in (0, \frac{1}{2}) \), we have \( \text{opt}_{1+\epsilon}(\mathcal{F}_\infty) = \Omega \left( \frac{\sqrt{\epsilon}}{2^{\epsilon} - \sqrt{1-\epsilon}} \right) \).

Note that \((1-\epsilon)^{\frac{1+\epsilon}{2}} \geq 1 - \epsilon(1+\epsilon)\) for \( \epsilon \in (0, \frac{1}{2}) \). To check this, note that it is true when \( \epsilon = 0 \), and the derivative of \((1-\epsilon)^{\frac{1+\epsilon}{2}} - (1-\epsilon(1+\epsilon))\) is \( 2\epsilon + 1 + (1-\epsilon)^{\frac{1+\epsilon}{2}} \left( \frac{1}{2} \ln(1-\epsilon) - \frac{1}{2} - \frac{\epsilon}{1-\epsilon} \right) \).

For \( \epsilon \in (0, \frac{1}{2}) \), we have

\[
2\epsilon + 1 + (1-\epsilon)^{\frac{1+\epsilon}{2}} \left( \frac{1}{2} \ln(1-\epsilon) - \frac{1}{2} - \frac{\epsilon}{1-\epsilon} \right) > \\
2\epsilon + 1 + \frac{1}{2} \ln(1-\epsilon) - \frac{1}{2} - \frac{\epsilon}{1-\epsilon} > \\
2\epsilon + \frac{1}{2} \ln(1-\epsilon) - \frac{1}{2} - \frac{\epsilon}{1-\epsilon} > \\
2\epsilon + 1 - \frac{1}{2} \ln(2) - \frac{1}{2} - 2\epsilon > 0.
\]

Thus, \( \text{opt}_{1+\epsilon}(\mathcal{F}_\infty) = \Omega \left( \frac{\sqrt{\epsilon}}{2^{\epsilon} - \sqrt{1-\epsilon}} \right) \).

Also note that \( 2^\epsilon \leq 1 + \epsilon \) for \( \epsilon \in (0, 1) \). There is equality at \( \epsilon = 0 \) and \( \epsilon = 1 \), and the derivative of \( 1 + \epsilon - 2^\epsilon \) is \( 1 - 2^\epsilon \ln 2 \), which is positive for \( \epsilon \in (0, \frac{\ln \ln 2}{\ln 2}) \) and negative for \( \epsilon \in \left( \frac{-\ln \ln 2}{\ln 2}, 1 \right) \). Thus \( 2^\epsilon - 1 + \epsilon(1+\epsilon) < 3\epsilon \), so \( \text{opt}_{1+\epsilon}(\mathcal{F}_\infty) = \Omega(\epsilon^{\frac{3}{2}}) \).

The next corollary follows from Theorem 2.4 again using the fact that \( \text{opt}_p(\mathcal{F}_\infty) \leq \text{opt}_p(\mathcal{F}_r) \leq \text{opt}_p(\mathcal{F}_q) \) whenever \( 1 \leq q \leq r \).

**Corollary 2.5.** If \( \epsilon \in (0, 1) \), then \( \text{opt}_{1+\epsilon}(\mathcal{F}_q) = \Omega(\epsilon^{\frac{3}{2}}) \) for all \( q \geq 1 \), where the constant does not depend on \( q \).

### 3 Discussion

With the results in this paper, the value of \( \text{opt}_p(\mathcal{F}_q) \) is now known up to a constant factor for all \( p, q \geq 1 \) except for \((p, q)\) with \( p \in (1, \infty) \) and \( q \in (1, 2) \). It remains to investigate
$\text{opt}_p(\mathcal{F}_q)$ when $p \in (1, \infty)$ and $q \in (1, 2)$, and to narrow the constant gap between the upper and lower bounds for $\text{opt}_{1+\epsilon}(\mathcal{F}_q) = \Theta(\epsilon^{-\frac{1}{2}})$ when $\epsilon \in (0, 1)$ and $q \in [2, \infty) \cup \{\infty\}$.

A possible extension of this research is to investigate analogues of these problems for multivariable functions. Previous research on learning multivariable functions [2, 5, 6] has focused on expected loss rather than worst-case loss, using models where the inputs $x_i$ are determined by a probability distribution.

In [10], Long investigated a finite variant of $\text{opt}_1(\mathcal{F}_q)$ for $q \geq 2$ that depends on the number of trials $m$. It seems interesting to extend this variant to $\text{opt}_p(\mathcal{F}_q)$ for $p = 1 + \epsilon$ with $0 < \epsilon < 1$ and $q \geq 1$, since $\text{opt}_{1+\epsilon}(\mathcal{F}_q)$ can grow arbitrarily large as $\epsilon \to 0$.

References

[1] D. Angluin. Queries and concept learning. Machine Learning 2 (1988) 319-342.
[2] A. Barron. Approximation and estimation bounds for artificial neural networks. Workshop on Computational Learning Theory (1991)
[3] N. Cesa-Bianchi, P.M. Long, and M.K. Warmuth. Worst-case quadratic loss bounds for prediction using linear functions and gradient descent. IEEE Transactions on Neural Networks 7 (1996) 604-619.
[4] V. Faber and J. Mycielski. Applications of learning theorems. Fundamenta Informaticae 15 (1991) 145-167.
[5] W. Hardle. Smoothing techniques. Springer Verlag (1991)
[6] D. Haussler. Generalizing the PAC model: sample size bounds from metric dimension-based uniform convergence results. Proceedings of the 30th Annual Symposium on the Foundations of Computer Science (1989)
[7] D. Kimber and P. M. Long. On-line learning of smooth functions of a single variable. Theoretical Computer Science 148 (1995) 141-156.
[8] N. Littlestone. Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm. Machine Learning 2 (1988) 285-318.
[9] N. Littlestone and M.K. Warmuth. The weighted majority algorithm. Proceedings of the 30th Annual Symposium on the Foundations of Computer Science (1989)
[10] P. M. Long. Improved bounds about on-line learning of smooth functions of a single variable. Theoretical Computer Science, 241 (2000) 25-35.
[11] J. Mycielski. A learning algorithm for linear operators. Proceedings of the American Mathematical Society 103 (1988) 547-550.