Worst-case Error Bounds for Online Learning

Weian (Andrew) Xie
Mentor: Dr. Jesse Geneson

The Hotchkiss School

October 13, 2024
MIT PRIMES Conference

Motivation

- Imagine a situation where a learner makes daily predictions on the value of a hidden function at particular given inputs
  - For example, predicting tomorrow's temperature given known inputs such as location, today's temperature, humidity, etc.
  - Predicting stock price values given as another example
- After each prediction, the learner receives feedback—through obtaining the actual function value, for example.
- This model of learning is called **online learning**.

Motivation continued

- Using the repeated feedback, we expect the learner to eventually make better predictions.
- We are curious about how fast better predictions can be made.
  - Can we bound the maximum error that the best learner can produce in the worst-case scenario?

Definitions

We study a model first defined by Kimber and Long (1995).

- Let $\mathcal{F}$ be a family of real functions
- Let A be an algorithm attempting to learn a hidden function $f \in \mathcal{F}$.
- For every trial $t \geq 0$, the algorithm is given an input x_t and queried on the value of $f(x_t)$.
- A outputs its prediction $\hat{y}_t$.
- The value of $f(x_t)$ is then revealed to A.

Definitions

- For each trial t, let the **raw error** e_t denote the absolute value of the difference between the prediction $\hat{y}_t$ and the true value $f(x_t)$
  - Eventually, we expect e_t to approach zero
- For any real $p \geq 1$, define the *p*-**error function** to be the sum of p^{th} powers of the raw errors of each trial, for arbitrarily many trials:

$$\sum_{t \geq 1} e_t^p = \sum_{t \geq 1} |\hat{y}_t - f(x_t)|^p.$$

Definitions

We now define the worst-case learning error for the best possible algorithm in learning a function in $\mathcal{F}$.

Definition for $\text{opt}_p(\mathcal{F})$

For a real $p \geq 1$, define $\text{opt}_p(\mathcal{F})$ as the best upper bound on the p-**error function** for any algorithm attempting to learn a function $f \in \mathcal{F}$.

Functions satisfying "smoothness" constraints

- The basis of our assumption that a learner can eventually make better guesses relies on the fact that the function to be learned is sufficiently predictable (eg. "smooth" functions).
- Kimber and Long's model studies particular such families of functions $\mathcal{F}_q$, defined as follows.

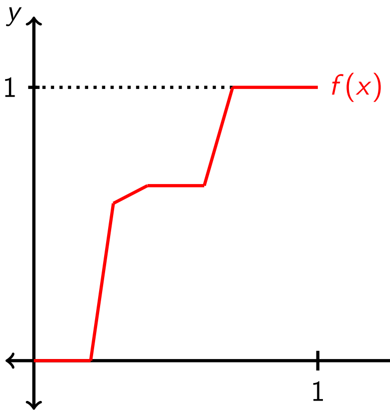
Definition

For any real $q \geq 1$, let $\mathcal{F}_q$ be the family of all absolutely continuous functions $f : [0, 1] \to \mathbb{R}$ such that $\int_0^1 |f'(x)|^q \mathrm{d}x \leq 1$.

Definition

Let $\mathcal{F}_\infty$ be the family of all absolutely continuous functions $f : [0, 1] \to \mathbb{R}$ such that $|f(x_1) - f(x_2)| \leq |x_1 - x_2|$ for all $x_1, x_2 \in [0, 1]$.

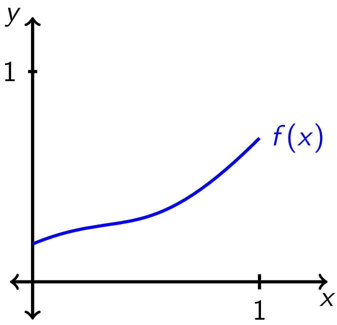
An example function f in $\mathcal{F}_1$



Note that $f \in \mathcal{F}_1$ as

$$\int_0^1 |f'(x)| \mathrm{d}x = \int_0^1 f'(x) \mathrm{d}x = f(1) - f(0) = 1.$$

An example function in $\mathcal{F}_\infty$



Let $f(x)$ be a function with $|f'(x)| \leq 1$ for all $x \in (0, 1)$. Then, by IVT, if there exists $0 \leq x_1 < x_2 \leq 1$ such that

$$|f(x_1) - f(x_2)| > |x_1 - x_2|,$$

there exists $x \in [x_1, x_2]$ with $|f'(x)| > 1$, contradiction; thus, $f \in \mathcal{F}_\infty$.

A Simple Example

To illustrate these concepts, we show a proof of a simple result.

Proposition (Kimber and Long, 1995)

For all $p, q \geq 1$, we have that $\text{opt}_p(\mathcal{F}_q) \geq 1$.

Proof:

- Let $f \in \mathcal{F}_q$ be the hidden function, which we don't fix yet.
- Consider the perspective of an adversary, working against the learner to maximize error by adapting f.
- For trial 0, let $x_0 = 0$, and reveal $f(x_0) = 0$.
- Note that the functions $g(x) \equiv x$ and $h(x) \equiv -x$ are both in $\mathcal{F}_q$, for any $q \geq 1$, as

$$\int_0^1 |g'(x)|^q \mathrm{d}x = \int_0^1 |h'(x)|^q \mathrm{d}x = \int_0^1 1 \mathrm{d}x \leq 1.$$

A Simple Example Continued

Proposition (Kimber and Long, 1995)

For all $p, q \geq 1$, we have that $\text{opt}_p(\mathcal{F}_q) \geq 1$.

Proof (continued):

- Now, for trial 1, let $x_1 = 1$, and query the learner on the value of $f(x_1)$
- If the prediction $\hat{y}_1 \geq 0$, we reveal $f(1) = -1$, and from now on fix $f(x) \equiv h(x) = -x$.
- If the prediction $\hat{y}_1 < 0$, we reveal $f(1) = 1$, and from now on fix $f(x) \equiv g(x) = x$.
- In any case, we (the adversary) guarantee a raw error $|\hat{y}_t - f(1)| \geq 1$.
- Thus, the *p*-**error function** will always be at least $|\hat{y}_t - f(1)|^q \geq 1$.
- Therefore, the best upper bound on the *p*-**error function**, which is $\text{opt}_p(\mathcal{F}_q)$, is at least 1.

Research Question

- What is the value of $\text{opt}_p(\mathcal{F}_q)$ for all choices of $p, q \geq 1$?
  - When is it finite?
  - Can we obtain precise values?
  - Can we obtain upper and lower bounds that are close?

Kimber and Long (1995)

Kimber and Long (1995) proved the following negative results.

Theorem (1995)

For any $p \geq 1$, we have $\text{opt}_p(\mathcal{F}_1) = \infty$.

Theorem (1995)

For any $q \geq 1$ or $q = \infty$, we have $\text{opt}_1(\mathcal{F}_q) = \infty$.

In other words, whenever $p = 1$ or $q = 1$, finite learning error cannot be guaranteed.

Kimber and Long (1995)

Kimber and Long (1995) also established the following positive results, including a precise equality when $p, q \geq 2$.

Theorem (1995)

For all $p, q \geq 2$, we have $\mathrm{opt}_p(\mathcal{F}_q) = 1$.

Theorem (1995)

For all $p \in (1, 2)$ and all $q \geq 2$, we have $\mathrm{opt}_p(\mathcal{F}_q) = O\left(\frac{1}{p-1}\right)$.

Geneson and Zhou (2023)

Geneson and Zhou (2023) improved the previous bound and proved a matching lower bound differing by a constant factor.

Theorem (2023)

For all $p \in (1, 2)$ and all $q \geq 2$, we have $\mathrm{opt}_p(\mathcal{F}_q) = \Theta\left(\frac{1}{\sqrt{p-1}}\right)$.

They also established another upper bound.

Theorem (2023)

For all $p \geq 2$ and $q \in (1, 2)$, we have $\mathrm{opt}_p(\mathcal{F}_q) = O\left(\frac{1}{q-1}\right)$.

In addition, they extended the region of (p, q) where $\mathrm{opt}_p(\mathcal{F}_q) = 1$.

Theorem (2023)

For any $q > 1$ and $p \geq 2 + \frac{1}{q-1}$, we have $\mathrm{opt}_p(\mathcal{F}_q) = 1$.

Previous Bounds

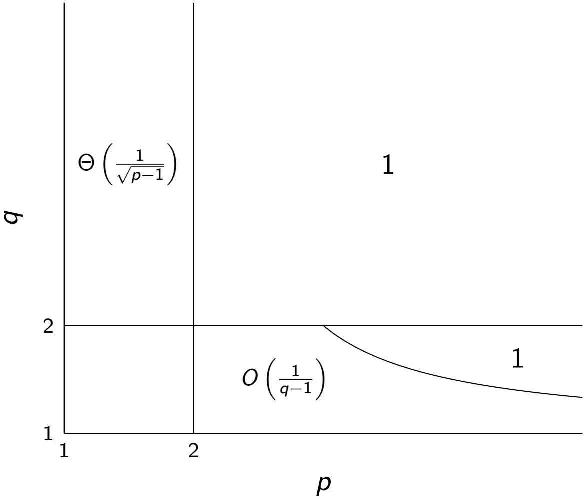


Figure 1: All previous bounds/equalities on $\mathrm{opt}_p(\mathcal{F}_q)$ for $p, q > 1$

New Results

- Now, we discuss our new results.
- First, we established the first ever upper bound on $\mathrm{opt}_p(\mathcal{F}_q)$ when $p, q \in (1, 2)$, confirming its finiteness.
- Second, we proved a conjecture from Geneson and Zhou (2023) that intuitively means that polynomials in $\mathcal{F}_q$ are not any easier to learn than any general function in $\mathcal{F}_q$, for any $q \geq 1$.

An Upper Bound on $\text{opt}_p(\mathcal{F}_q)$ for $p, q \in (1, 2)$

We establish the first upper bound for the case where $p, q \in (1, 2)$.

Theorem

For all $\delta, \epsilon \in (0, 1)$, we have $\text{opt}_{1+\delta}(\mathcal{F}_{1+\epsilon}) = O(\min(\delta, \epsilon)^{-1})$.

As such, we achieve a complete characterization of all $p, q \geq 1$ that result in $\text{opt}_p(\mathcal{F}_q)$ being finite, a problem open since Kimber and Long (1995) defined the model. Our result confirms a conjecture by Geneson and Zhou (2023).

Corollary

The worst-case learning error $\text{opt}_p(\mathcal{F}_q)$ is finite if and only if $p, q > 1$.

Current bounds for all p, q

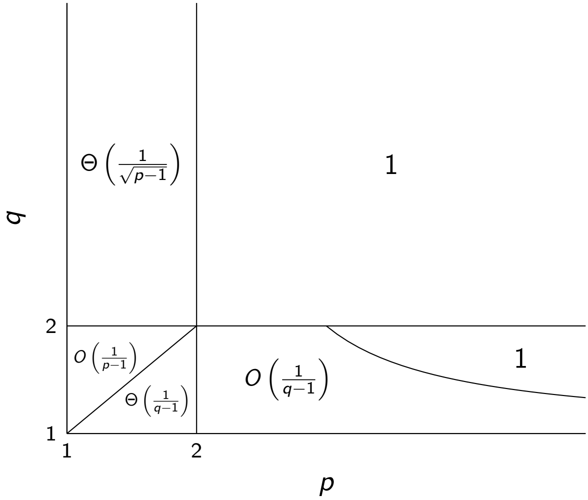


Figure 2: All bounds on $\text{opt}_p(\mathcal{F}_q)$ for $p, q > 1$ including new results

A Result About Learning Polynomials

- We confirmed a conjecture by Geneson and Zhou (2023) regarding the online learning of polynomials
- Intuitively, the result states that it is not any easier to learn a polynomial in $\mathcal{F}_q$ than any general function in $\mathcal{F}_q$.
- For any $q \geq 1$, let $\mathcal{P}_q$ denote the family of polynomials P such that $P \in \mathcal{F}_q$. Then, we have:

Theorem

For all $p > 0$ and $q \geq 1$, we have $\mathrm{opt}_p(\mathcal{P}_q) = \mathrm{opt}_p(\mathcal{F}_q)$.

Acknowledgements

I am immensely grateful for:

- My mentor, Dr. Geneson, for his time and patience
- My parents, for their support of my passion for math
- My school teachers, Mr. Almeida and Dr. Eso, for their support of my research through granting an Independent Study in school
- PRIMES-USA, for this wonderful research opportunity

References

J. Geneson and E. Zhou
Online learning of smooth functions
Theoretical Computer Science 979C: 114203 (2023).

D. Kimber and P. M. Long
On-line learning of smooth functions of a single variable
*Theoretical Computer Science **148** (1995) 141–156.*

P. M. Long
Improved bounds about on-line learning of smooth functions of a single variable
*Theoretical Computer Science **241** (2000) 25–35.*