Online Learning of Smooth Functions

Ethan Zhou
Mentor: Jesse Geneson

October 15, 2022
MIT PRIMES Conference

Introduction

- Online learning is a model of machine learning where a learner trains on data revealed sequentially.
  - For example, predicting stock prices
- We investigate the online learning of real-valued functions, where a learning algorithm predicts outputs of a real-valued function f based on previously revealed points $(x, f(x))$.

The setup

Fix a domain X and a class $\mathcal{F}$ of functions from X to $\mathbb{R}$. (These are known to the learner.)

A learning algorithm, A, learns $\mathcal{F}$ as follows:

- An adversary selects $f \in \mathcal{F}$.
- Learning then proceeds in trials. On trial $i \geq 0$:
  - The adversary gives A an input $x_i \in X$
  - A produces $\hat{y}_i \in \mathbb{R}$, its prediction for $f(x_i)$ based on all previously revealed points $(x_j, f(x_j))$
  - The adversary reveals the true value of $f(x_i)$

Measuring performance

Question

How can we measure the difficulty of learning some class $\mathcal{F}$ of real-valued functions?

- Fix $p \geq 1$. On each trial $i \geq 1$, if the learner A makes a mistake, it gains a penalty term $|\hat{y}_i - f(x_i)|^p$.
  - For this to work, $\mathcal{F}$ should contain "nice" functions
  - As p increases, these penalties decay faster
- We are interested in worst-case performance, where the adversary selects $f \in \mathcal{F}$ and inputs x_i to maximize total penalty.

opt

For $p \geq 1$ and a class $\mathcal{F}$ of real-valued functions, $\text{opt}_p(\mathcal{F})$ is the best upper bound on the sum of penalties, $\sum_{i \geq 1} |\hat{y}_i - f(x_i)|^p$, a learner can guarantee while learning $\mathcal{F}$, against any adversary.

"Smooth" single-variable functions

Consider the following classes of "nice" functions:

$\mathcal{F}_q$

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

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

- $\mathcal{F}_\infty$ can be thought of as the limit of $\mathcal{F}_q$ as $q \to \infty$.
- As q increases, $\mathcal{F}_q$ shrinks.
  - Thus $\mathrm{opt}_p(\mathcal{F}_q)$ decreases as well, for any $p \geq 1$.

Example

We show that for $p, q \geq 1$, $\text{opt}_p(\mathcal{F}_q) \geq 1$:

- The adversary sets $x_0 = 0$ and reveals $f(0) = 0$.
- The function $f(x) = x$ is in $\mathcal{F}_q$, since

$$\int_0^1 |f'(x)|^q \text{d}x = \int_0^1 1\text{d}x = 1 \leq 1.$$

  Likewise the function $f(x) = -x$ is in $\mathcal{F}_q$.
- Thus the adversary can set $x_1 = 1$ and reveal $f(1) = \pm 1$, whichever is farther from the learner's prediction.
- This forces a penalty of at least $|\hat{y}_1 - f(x_1)|^p \geq 1$.

Main questions

- For which $p, q \geq 1$ is $\text{opt}_p(\mathcal{F}_q)$ infinite?
- For which $p, q \geq 1$ is $\text{opt}_p(\mathcal{F}_q)$ equal to 1?
- How does $\text{opt}_p(\mathcal{F}_q)$ vary with p, q?
  - Especially as it tends to infinity

Previous results

Theorem [Kimber and Long, 1995]

For $p, q \geq 1$, $\text{opt}_1(\mathcal{F}_q) = \text{opt}_1(\mathcal{F}_\infty) = \text{opt}_p(\mathcal{F}_1) = \infty$.

Theorem [Kimber and Long, 1995]

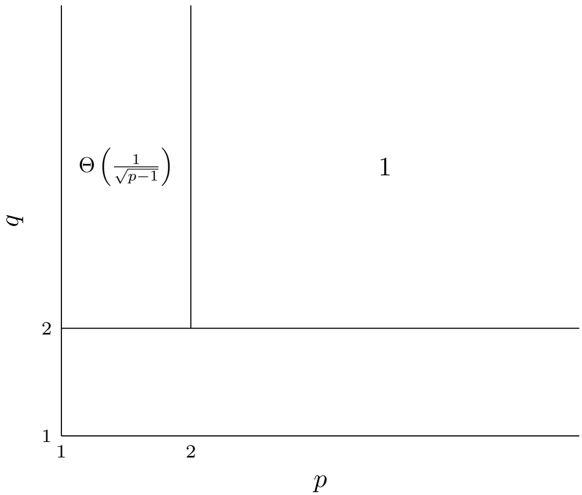
For $p, q \geq 2$, $\text{opt}_p(\mathcal{F}_q) = 1$.

Theorem [Geneson, 2021]

For $\varepsilon \in (0, 1)$ and $q \geq 2$, $\text{opt}_{1+\varepsilon}(\mathcal{F}_q) = \Theta(\varepsilon^{-\frac{1}{2}})$, where the constant factors do not depend on q.

Previous results

Bounds on $\mathrm{opt}_p(\mathcal{F}_q)$ for $p, q > 1$



Our results

For $p \geq 1$ and $q \geq 2$, $\text{opt}_p(\mathcal{F}_q)$ is known up to a constant factor, so most of our work centers on understanding the $q \in (1, 2)$ case.

Theorem

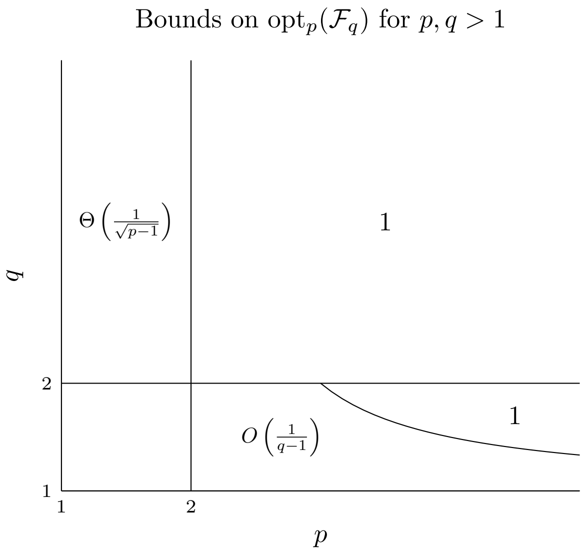
For $\varepsilon \in (0, 1)$, $\text{opt}_2(\mathcal{F}_{1+\varepsilon}) = \Theta(\varepsilon^{-1})$.

Theorem

For $\varepsilon \in (0, 1)$ and $p \geq 2 + \varepsilon^{-1}$, $\text{opt}_p(\mathcal{F}_{1+\varepsilon}) = 1$.

- For any $q > 1$, there exists p such that the adversary cannot do better than forcing the learner to guess between $f(x) = x$ and $f(x) = -x$.
  - Compare to $\text{opt}_p(\mathcal{F}_1) = \infty$ for all $p \geq 1$

Our results



Bounds on $\mathrm{opt}_p(\mathcal{F}_q)$ for $p, q > 1$
$\Theta\left(\frac{1}{\sqrt{p-1}}\right)$
1
2
$O\left(\frac{1}{q-1}\right)$
1
q
2
1
1
2
p

A multivariable generalization

We can also generalize to functions from $[0,1]^d$ to $\mathbb{R}$:

$\mathcal{F}_{q,d}$

For $q \geq 1$ and $d \in \mathbb{Z}_{>0}$, $\mathcal{F}_{q,d}$ is the class of functions $f : [0,1]^d \to \mathbb{R}$ such that any function $g : [0,1] \to \mathbb{R}$ formed by fixing $d-1$ arguments of f is in $\mathcal{F}_q$. $\mathcal{F}_{\infty,d}$ is defined similarly.

Our results

Proposition

For $p, q \geq 1$ and $d \in \mathbb{Z}_{>0}$:

- $\text{opt}_p(\mathcal{F}_{q,d}) \geq d \cdot \text{opt}_p(\mathcal{F}_q)$;
- $\text{opt}_p(\mathcal{F}_{\infty,d}) \geq d^p \cdot \text{opt}_p(\mathcal{F}_\infty)$.

Proposition

For $p \geq 1$ and $d \in \mathbb{Z}_{>0}$:

- If $p < d$ then $\text{opt}_p(\mathcal{F}_{\infty,d}) = \infty$;
- If $p > d$ then $\text{opt}_p(\mathcal{F}_{\infty,d}) \leq \frac{(2^d - 1)d^p}{1 - \frac{2^d}{2^p}}$.

Future work

Single-variable setup:

- Is $\text{opt}_p(\mathcal{F}_q)$ finite for all $p, q > 1$?
  - If so, how does it grow as $p, q \to 1$?
- What does the region of (p, q) for which $\text{opt}_p(\mathcal{F}_q) = 1$ look like?

Multivariable setup:

- Is $\text{opt}_d(\mathcal{F}_{\infty,d})$ infinite for all $d \in \mathbb{Z}_{>0}$?
- Algorithms for learning functions in $\mathcal{F}_{q,d}$

Acknowledgments

- My mentor, Dr. Jesse Geneson
- Prof. Etingof, Dr. Gerovitch, Dr. Khovanova, and the PRIMES program
- My family