Continuous Logic and Learning Bounds

Aaron Anderson

UPenn

May 8, 2025

Probably Approximately Correct Learning

A class $\mathcal{H}$ of functions $X \to [0,1]$ is PAC learnable when for every $\varepsilon, \delta > 0$, there is n such that when...

- $(x_1, y_1), \ldots, (x_n, y_n) \in X \times [0,1]$ are i.i.d. random,
- we can choose $h \in \mathcal{H}$ (hoping that $h(x_{n+1}) \approx y_{n+1}$) such that
- with probability at least $1 - \delta$,
- $\mathbb{E}\left[|y_{n+1} - h(x_{n+1})|\right]$ is within ε of the best case for all $h \in \mathcal{H}$.

We call $n = n(\varepsilon, \delta)$ the *sample complexity*.

Fat-Shattering Dimension

Theorem (Almost Folklore)

$\mathcal{H}$ is PAC-learnable if and only if the γ-fat-shattering dimension is finite for all $\gamma > 0$.

Definition

Let $\mathcal{H}$ be a class of functions $X \to [0,1]$ and let $\gamma > 0$. We say $\mathcal{H}$ has *γ-fat-shattering dimension* at least n when there are

- $x_1, \ldots, x_n \in X$
- $s_1, \ldots, s_n \in [0,1]$
- For every $E \subseteq \{1, \ldots, n\}$, a function $h_E \in \mathcal{H}$ satisfying
  - if $i \in E$, $h_E(x_i) \geq s_i + \gamma$
  - if $i \notin E$, $h_E(x_i) \leq s_i - \gamma$.

PAC Learning Bound

Theorem (Bartlett, Long)

The sample complexity $n(\varepsilon, \delta)$ of PAC-learning $\mathcal{H}$ is bounded by

$$O\left(\frac{1}{\epsilon^2} \cdot \left(\mathrm{FatSHDim}_{\frac{\epsilon}{9}}(\mathcal{H}) \cdot \log^2\left(\frac{1}{\epsilon}\right) + \log\left(\frac{1}{\delta}\right)\right)\right)$$

Hu et al. extended this to learning a class of measures on $\mathcal{H}$, at the cost of a much worse bound.

We'll use logic to find examples of learnable classes and improve the Hu et al. bound.

The Boolean-Valued Case

- Recall $\mathcal{H}$ is a class of functions $X \to [0,1]$.
- Consider the case where the functions $h \in \mathcal{H}$ are $\{0,1\}$-valued.
- These are the characteristic functions of subsets of X
- Where can we get interesting classes $\mathcal{C}$ of subsets of X?

Definable Classes

Let M be a first-order $\mathcal{L}$-structure, and let $\phi(\bar{x}, \bar{y})$ be an $\mathcal{L}$-formula, with $|\bar{x}| = m, |\bar{y}| = n$.

Definition

Let $\mathcal{C}_\phi$ be the class of subsets of M^m, indexed by M^n, given by

$$\mathcal{C}_\phi = (C_{\bar{b}} : \bar{b} \in M^n)$$
$$C_{\bar{b}} = \{\bar{a} : M \vDash \phi(\bar{a}, \bar{b})\}.$$

Any class of sets that arises this way we call *definable* in M by ϕ.

NIP

Definition

A formula $\phi(\bar{x}, \bar{y})$ is called *NIP* in a structure M when every class of sets definable by ϕ has finite VC-dimension.

We call M *NIP* when every formula is NIP in M.

NIP structures include

- The real field $(\mathbb{R}; 0, 1, +, \times, <)$
- Any other *o*-minimal structure
- The complex field $(\mathbb{C}; 0, 1, +, \times)$
- Any other stable structure

Model Theory to Learnability

For a definable class of sets $\mathcal{C}$, the properties in each row are equivalent:

Model Theory	Combinatorics	Learning Theory
NIP	finite VC dimension	PAC learnable
stable	finite Littlestone dimension	online learnable

These definitions have been generalized to the real-valued case - using *continuous logic.*

Continuous Logic

- There is a framework for *continuous logic*, where formulas take values in $[0, 1]$.
- Ask James Hanson for the details.
- A formula $\phi(\bar{x}, \bar{y})$ of continuous logic defines a class $\mathcal{H}$ of functions $M^m \to [0, 1]$.
- If all such classes have finite γ-fat-shattering dimension for all $\gamma > 0$, the formula is NIP.
- The connection from stability to real-valued online learning was understudied.

Stable Structures in Continuous Logic

Basic examples of stable (and thus NIP) metric structures:

Example

Let M be a boolean algebra with a probability measure μ. Can add

- metric $\mu(x \setminus y \cup y \setminus x)$
- functions $0, 1, ^c, \cap, \cup$
- relation $\mu(x)$.

Example

Let M be the unit ball of an infinite-dimensional Hilbert space, with the metric, $\langle \cdot, \cdot \rangle$, scalar multiplication, and partial addition.

The Expectation Class

Suppose (Ω, Σ, μ) is a probability space and $\mathcal{F} = (\mathcal{H}_\omega : \omega \in \Omega)$ is a family of function classes $\mathcal{H}_\omega = (h_{\omega,y} : y \in Y)$.

Definition

Assuming measurability, define $\mathbb{E}\mathcal{F}_y : X \to [0,1]$ by

$$\mathbb{E}\mathcal{F}_y(x) = \mathbb{E}\left[h_{\omega,y}(x)\right].$$

We call the class $\mathbb{E}\mathcal{F} = \{\mathbb{E}\mathcal{F}_y : y \in Y\}$ the *expectation class* of $\mathcal{F}$.

Think of $\mathbb{E}\mathcal{F}$ and every $\mathcal{H}_\omega$ as a class of functions $h : X \to [0,1]$ indexed by Y.

Theorem (Ben Yaacov, Keisler)

If $\mathcal{F}$ is uniformly NIP/stable, then $\mathbb{E}\mathcal{F}$ is NIP/stable.

The Randomization

We can apply this to randomize a single class $\mathcal{H}$ of functions $X \to [0,1]$, indexed by Y.

Definition

- Let $\mathbf{X}, \mathbf{Y}$ be appropriate spaces of random variables $\Omega \to X$ and $\Omega \to Y$.
- Define $\mathcal{H}_\omega = \{h_{\omega,\mathbf{y}} : \mathbf{y} \in \mathbf{Y}\}$ by

$$h_{\omega,\mathbf{y}}(\mathbf{x}) = h_{\mathbf{y}(\omega)}(\mathbf{x}(\omega)).$$

- Let $R\mathcal{H} : X \times Y \to [0,1]$ be the expectation class of this family.
- We call this new class the *expectation class* of $\mathcal{H}$.

PAC Learning The Randomization

Theorem (A., Benedikt)

If $\mathcal{H}$ has $\mathrm{FatSHDim}_{\frac{\epsilon}{50}}(\mathcal{H}) \leq d$, one can PAC learn the randomization class of $\mathcal{H}$ with sample complexity

$$O\left(\frac{d}{\epsilon^4} \cdot \log^2 \frac{d}{\epsilon} + \frac{1}{\epsilon^2} \cdot \log \frac{1}{\delta}\right).$$

- $\mathrm{FatSHDim}$ can be used to bound *Rademacher mean width*
- Rademacher mean width can be used to bound sample complexity
- Adapt Ben Yaacov's proof that *Gaussian* mean width is preserved under randomization

Online Learning

- At step i, an adversary chooses $(x_i, y_i) \in X \times [0, 1]$
- Given x_i, you guess $y_i' \approx y_i$ (you can use randomness)
- The adversary tells you y_i, penalizes you $|y_i - y_i'|$
- After n steps, compare to the best strategy $y_i' = h(x_i)$ for $h \in \mathcal{H}$.
- Call the difference in penalty the *regret*.
- $\mathcal{H}$ is online learnable if whatever the adversary does, regret is sublinear in n.

Online Learning Bounds

To bound regret in online learning, replace our existing notions with *sequential* versions, replacing subsets $E \subseteq \{1, \ldots, n\}$ with branches of a binary tree of depth n:

Theorem (Rakhlin, Sridharan, Tewari)

Finite γ-sequential-fat-shattering dimension is equivalent to online learnability, with bounds given.

Their proof goes through sequential Rademacher mean width.

Our Online Learning Results

Theorem (A., Benedikt)

- *Stability in continuous logic is equivalent to finite γ-sequential-fat-shattering dimension for all $\gamma > 0$.*
- *Sequential Rademacher mean width, and thus online learnability, is preserved under randomization.*

Theorem (A., Benedikt)

The minimax regret of online learning for the randomization class of $\mathcal{H}$ with γ-sequential-fat-shattering dimension at most d on a run of length n is at most

$$4 \cdot \gamma \cdot n + 12 \cdot (1 - \gamma) \cdot \sqrt{d \cdot n \cdot \log\left(\frac{2 \cdot e \cdot n}{\gamma}\right)}.$$

Realizable Learning

- In *realizable* PAC or online learning, assume there is some $h \in \mathcal{H}$ such that for all i, $y_i = h(x_i)$.
- Classically, this assumption does not change learnability.
- We show that proposed definitions of realizable PAC and online learnability for real-valued function classes are not closed under basic operations like dualization, continuous combinations, or randomization.
- We propose better definitions that are, inspired by model theory.

Thank you, ICICL!