Uniform Computability of PAC Learning

Vasco Brattka[1]
Guillaume Chirache[2]

[1] Theoretical Computer Science and Mathematical Logic
Universität der Bundeswehr München, Germany

[2] École polytechnique, Palaiseau, France

Computability and Complexity in Analysis
Trier, Germany
30 July 2026

PAC Learning

Hypotheses and learners

We consider samples of input-label pairs

$$S = ((x_1, y_1), ..., (x_n, y_n)) \in \mathcal{S} := (\mathbb{N} \times \{0, 1\})^*.$$

The job of a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ is to map such a sample to a hypothesis $h : \mathbb{N} \to \{0, 1\}$ (*i.e.*, an element of the Cantor space $2^{\mathbb{N}}$).

$\mathcal{L}$ is the set of learners.

Empirical risk minimizer

Intuitively, the best solution is to choose a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ that minimizes the empirical risk

$$L_S(h) = \frac{|\{i \in \{1, ..., n\} : h(x_i) \neq y_i\}|}{n}.$$

Hypotheses and learners

We consider samples of input-label pairs

$$S = ((x_1, y_1), ..., (x_n, y_n)) \in \mathcal{S} := (\mathbb{N} \times \{0, 1\})^*.$$

The job of a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ is to map such a sample to a hypothesis $h : \mathbb{N} \to \{0, 1\}$ (*i.e.*, an element of the Cantor space $2^{\mathbb{N}}$).

$\mathcal{L}$ is the set of learners.

Empirical risk minimizer

Intuitively, the best solution is to choose a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ that minimizes the empirical risk

$$L_S(h) = \frac{|\{i \in \{1, ..., n\} : h(x_i) \neq y_i\}|}{n}.$$

Hypotheses and learners

We consider samples of input-label pairs

$$S = ((x_1, y_1), ..., (x_n, y_n)) \in \mathcal{S} := (\mathbb{N} \times \{0, 1\})^*.$$

The job of a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ is to map such a sample to a hypothesis $h : \mathbb{N} \to \{0, 1\}$ (*i.e.*, an element of the Cantor space $2^{\mathbb{N}}$).

$\mathcal{L}$ is the set of learners.

Empirical risk minimizer

Intuitively, the best solution is to choose a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ that minimizes the empirical risk

$$L_S(h) = \frac{|\{i \in \{1, ..., n\} : h(x_i) \neq y_i\}|}{n}.$$

The need for hypothesis classes

An ERM can be a **very bad** learner: there is no constraint on inputs that have not been seen.

We need to make some assumptions about what reasonable hypotheses look like.

- A concept class or hypothesis class is a set $\mathcal{C} \subseteq 2^{\mathbb{N}}$. We sometimes deal with two such classes $\varnothing \neq \mathcal{C} \subseteq \mathcal{H} \subseteq 2^{\mathbb{N}}$.
- We consider probability distributions $\mathcal{D} : \mathbb{N} \times \{0, 1\} \to \mathbb{R}$.
- The true error (risk) of h is $L_{\mathcal{D}}(h) := \mathbb{P}_{(x,y)\sim\mathcal{D}}[h(x) \neq y]$.

PAC learning (informally)

$\mathcal{C}$ is PAC learnable relative to $\mathcal{H}$ if there exists a learner $A : \mathcal{S} \to \mathcal{H}$ which makes the risk $L_{\mathcal{D}}(A(S))$ arbitrarily close to the optimal risk $\inf_{h\in\mathcal{C}} L_{\mathcal{D}}(h)$, with arbitrarily high probability, provided the sample size is sufficiently large and regardless of the probability distribution $\mathcal{D}$ it is drawn from.

The need for hypothesis classes

An ERM can be a **very bad** learner: there is no constraint on inputs that have not been seen.

We need to make some assumptions about what reasonable hypotheses look like.

- A concept class or hypothesis class is a set $\mathcal{C} \subseteq 2^{\mathbb{N}}$. We sometimes deal with two such classes $\varnothing \neq \mathcal{C} \subseteq \mathcal{H} \subseteq 2^{\mathbb{N}}$.
- We consider probability distributions $\mathcal{D} : \mathbb{N} \times \{0, 1\} \to \mathbb{R}$.
- The true error (risk) of h is $L_{\mathcal{D}}(h) := \mathbb{P}_{(x,y)\sim\mathcal{D}}[h(x) \neq y]$.

PAC learning (informally)

$\mathcal{C}$ is PAC learnable relative to $\mathcal{H}$ if there exists a learner $A : \mathcal{S} \to \mathcal{H}$ which makes the risk $L_{\mathcal{D}}(A(S))$ arbitrarily close to the optimal risk $\inf_{h\in\mathcal{C}} L_{\mathcal{D}}(h)$, with arbitrarily high probability, provided the sample size is sufficiently large and regardless of the probability distribution $\mathcal{D}$ it is drawn from.

The need for hypothesis classes

An ERM can be a **very bad** learner: there is no constraint on inputs that have not been seen.

We need to make some assumptions about what reasonable hypotheses look like.

- A concept class or hypothesis class is a set $\mathcal{C} \subseteq 2^{\mathbb{N}}$. We sometimes deal with two such classes $\varnothing \neq \mathcal{C} \subseteq \mathcal{H} \subseteq 2^{\mathbb{N}}$.
- We consider probability distributions $\mathcal{D} : \mathbb{N} \times \{0, 1\} \to \mathbb{R}$.
- The true error (risk) of h is $L_{\mathcal{D}}(h) := \mathbb{P}_{(x,y)\sim\mathcal{D}}[h(x) \neq y]$.

PAC learning (informally)

$\mathcal{C}$ is PAC learnable relative to $\mathcal{H}$ if there exists a learner $A : \mathcal{S} \to \mathcal{H}$ which makes the risk $L_{\mathcal{D}}(A(S))$ arbitrarily close to the optimal risk $\inf_{h\in\mathcal{C}} L_{\mathcal{D}}(h)$, with arbitrarily high probability, provided the sample size is sufficiently large and regardless of the probability distribution $\mathcal{D}$ it is drawn from.

Probably approximately correct learning

A sample complexity is a function $m : \mathbb{N} \to \mathbb{N}$.

PAC learning (formally)

We say that $A : \mathcal{S} \to \mathcal{H}$ PAC learns $\mathcal{C}$ *relative to $\mathcal{H}$ with* m if for all $i, j \in \mathbb{N}$ and $n \geq m\langle i, j\rangle$ and for every probability distribution $\mathcal{D}$:

$$\mathbb{P}_{S \sim \mathcal{D}^n}\left[L_{\mathcal{D}}(A(S)) \leq \inf_{h \in \mathcal{C}} L_{\mathcal{D}}(h) + 2^{-i}\right] > 1 - 2^{-j}.$$

If $\mathcal{H} = \mathcal{C}$, then we say that A (properly) PAC learns $\mathcal{C}$ with m.
If $\mathcal{H} = 2^{\mathbb{N}}$, then we say that A improperly PAC learns $\mathcal{C}$ with m.

- $\varepsilon = 2^{-i}$ is called accuracy and $\delta = 2^{-j}$ confidence.
- **Question:** do we have a good characterization of PAC learnable classes?

Probably approximately correct learning

A sample complexity is a function $m : \mathbb{N} \to \mathbb{N}$.

PAC learning (formally)

We say that $A : \mathcal{S} \to \mathcal{H}$ PAC learns $\mathcal{C}$ *relative to* $\mathcal{H}$ *with* m if for all $i, j \in \mathbb{N}$ and $n \geq m\langle i, j \rangle$ and for every probability distribution $\mathcal{D}$:

$$\mathbb{P}_{S \sim \mathcal{D}^n} \left[L_{\mathcal{D}}(A(S)) \leq \inf_{h \in \mathcal{C}} L_{\mathcal{D}}(h) + 2^{-i} \right] > 1 - 2^{-j}.$$

If $\mathcal{H} = \mathcal{C}$, then we say that A (properly) PAC learns $\mathcal{C}$ with m.
If $\mathcal{H} = 2^{\mathbb{N}}$, then we say that A improperly PAC learns $\mathcal{C}$ with m.

- $\varepsilon = 2^{-i}$ is called accuracy and $\delta = 2^{-j}$ confidence.
- **Question:** do we have a good characterization of PAC learnable classes?

The Fundamental Theorem of Statistical Learning

- A finite subset $A \subseteq \mathbb{N}$ is shattered by $\mathcal{C} \subseteq 2^{\mathbb{N}}$ if for any $f : A \to \{0, 1\}$, there exists $h \in \mathcal{C}$ such that $h|_A = f$.
- The VC dimension of $\mathcal{C} \subseteq 2^{\mathbb{N}}$ is defined by
  $\mathrm{VCdim}(\mathcal{C}) := \sup\{|A| \in \mathbb{N} : A \subseteq \mathbb{N} \text{ finite \& shattered by } \mathcal{C}\} \in \mathbb{N}_\infty$.

Theorem (Blumer, Ehrenfeucht, Haussler, and Warmuth 1989)

$\mathcal{C}$ is PAC learnable relative to $\mathcal{H}$ if and only if $\mathrm{VCdim}(\mathcal{C}) < \infty$. In this case, any ERM is a PAC learner.

- In recent years, there has been an increasing interest in non-uniform computability aspects of PAC learning. Among others:
  - Agarwal, Ananthakrishnan, Ben-David, Lechner, and Urner (2020) introduced computable PAC learning.
  - Sterkenburg (2022) studied this further.

Theorem (Delle Rose, Kozachinskiy, Rojas, and Steifer 2023)

$\mathcal{C}$ admits a computable learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ (with a computable sample complexity m) if and only if the effective VC dimension of $\mathcal{C}$ is finite.

- **Question**: Can we **algorithmically find** the learner A **given** $\mathcal{C}$?

The Fundamental Theorem of Statistical Learning

- A finite subset $A \subseteq \mathbb{N}$ is shattered by $\mathcal{C} \subseteq 2^{\mathbb{N}}$ if for any $f : A \to \{0, 1\}$, there exists $h \in \mathcal{C}$ such that $h|_A = f$.
- The VC dimension of $\mathcal{C} \subseteq 2^{\mathbb{N}}$ is defined by
  $\mathrm{VCdim}(\mathcal{C}) := \sup\{|A| \in \mathbb{N} : A \subseteq \mathbb{N} \text{ finite \& shattered by } \mathcal{C}\} \in \mathbb{N}_\infty$.

Theorem (Blumer, Ehrenfeucht, Haussler, and Warmuth 1989)

$\mathcal{C}$ is PAC learnable relative to $\mathcal{H}$ if and only if $\mathrm{VCdim}(\mathcal{C}) < \infty$. In this case, any ERM is a PAC learner.

- In recent years, there has been an increasing interest in non-uniform computability aspects of PAC learning. Among others:
  - Agarwal, Ananthakrishnan, Ben-David, Lechner, and Urner (2020) introduced computable PAC learning.
  - Sterkenburg (2022) studied this further.

Theorem (Delle Rose, Kozachinskiy, Rojas, and Steifer 2023)

$\mathcal{C}$ admits a computable learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ (with a computable sample complexity m) if and only if the effective VC dimension of $\mathcal{C}$ is finite.

- **Question**: Can we **algorithmically find** the learner A **given** $\mathcal{C}$?

The Fundamental Theorem of Statistical Learning

- A finite subset $A \subseteq \mathbb{N}$ is shattered by $\mathcal{C} \subseteq 2^{\mathbb{N}}$ if for any $f : A \to \{0,1\}$, there exists $h \in \mathcal{C}$ such that $h|_A = f$.
- The VC dimension of $\mathcal{C} \subseteq 2^{\mathbb{N}}$ is defined by
  $$\mathrm{VCdim}(\mathcal{C}) := \sup\{|A| \in \mathbb{N} : A \subseteq \mathbb{N} \text{ finite \& shattered by } \mathcal{C}\} \in \mathbb{N}_\infty.$$

Theorem (Blumer, Ehrenfeucht, Haussler, and Warmuth 1989)

$\mathcal{C}$ is PAC learnable relative to $\mathcal{H}$ if and only if $\mathrm{VCdim}(\mathcal{C}) < \infty$. In this case, any ERM is a PAC learner.

- In recent years, there has been an increasing interest in non-uniform computability aspects of PAC learning. Among others:
  - Agarwal, Ananthakrishnan, Ben-David, Lechner, and Urner (2020) introduced computable PAC learning.
  - Sterkenburg (2022) studied this further.

Theorem (Delle Rose, Kozachinskiy, Rojas, and Steifer 2023)

$\mathcal{C}$ admits a computable learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ (with a computable sample complexity m) if and only if the effective VC dimension of $\mathcal{C}$ is finite.

- **Question**: Can we **algorithmically find** the learner A **given** $\mathcal{C}$?

The Fundamental Theorem of Statistical Learning

- A finite subset $A \subseteq \mathbb{N}$ is shattered by $\mathcal{C} \subseteq 2^{\mathbb{N}}$ if for any $f : A \to \{0,1\}$, there exists $h \in \mathcal{C}$ such that $h|_A = f$.
- The VC dimension of $\mathcal{C} \subseteq 2^{\mathbb{N}}$ is defined by
  $$\mathrm{VCdim}(\mathcal{C}) := \sup\{|A| \in \mathbb{N} : A \subseteq \mathbb{N} \text{ finite \& shattered by } \mathcal{C}\} \in \mathbb{N}_\infty.$$

Theorem (Blumer, Ehrenfeucht, Haussler, and Warmuth 1989)

$\mathcal{C}$ is PAC learnable relative to $\mathcal{H}$ if and only if $\mathrm{VCdim}(\mathcal{C}) < \infty$. In this case, any ERM is a PAC learner.

- In recent years, there has been an increasing interest in non-uniform computability aspects of PAC learning. Among others:
  - Agarwal, Ananthakrishnan, Ben-David, Lechner, and Urner (2020) introduced computable PAC learning.
  - Sterkenburg (2022) studied this further.

Theorem (Delle Rose, Kozachinskiy, Rojas, and Steifer 2023)

$\mathcal{C}$ admits a computable learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ (with a computable sample complexity m) if and only if the effective VC dimension of $\mathcal{C}$ is finite.

- **Question**: Can we **algorithmically find** the learner A **given** $\mathcal{C}$?

The Fundamental Theorem of Statistical Learning

Theorem (Blumer, Ehrenfeucht, Haussler, and Warmuth 1989)

$\mathcal{C}$ admits a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ with some sample complexity $m : \mathbb{N} \to \mathbb{N}$ if and only if the VC dimension of $\mathcal{C}$ is finite.

We want to answer three questions from different perspectives:

- **Logic:** How non-constructive is the Fundamental Theorem?
- **Computability theory:** How are computability properties of the concept class $\mathcal{C}$ related to computability properties of the learner?
- **Theoretical computer science:** What kind of algorithms can exist to compute the learner A, given the concept class $\mathcal{C}$?

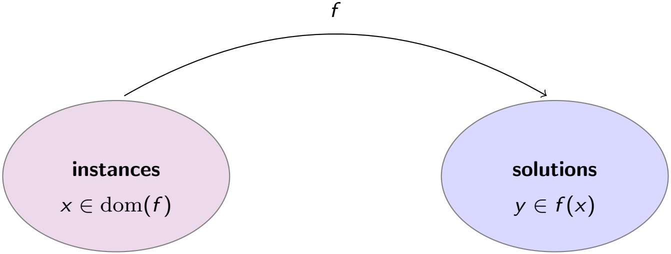
Weihrauch complexity allows us to answer all three questions at once!

The Fundamental Theorem of Statistical Learning

Theorem (Blumer, Ehrenfeucht, Haussler, and Warmuth 1989)

$\mathcal{C}$ admits a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ with some sample complexity $m : \mathbb{N} \to \mathbb{N}$ if and only if the VC dimension of $\mathcal{C}$ is finite.

- In Weihrauch complexity mathematical problems are formalized as partial multivalued functions $f :\subseteq X \rightrightarrows Y$.



The Fundamental Theorem of Statistical Learning

Theorem (Blumer, Ehrenfeucht, Haussler, and Warmuth 1989)

$\mathcal{C}$ admits a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ with some sample complexity $m : \mathbb{N} \to \mathbb{N}$ if and only if the VC dimension of $\mathcal{C}$ is finite.

- In Weihrauch complexity mathematical problems are formalized as partial multivalued functions $f :\subseteq X \rightrightarrows Y$.
- It turns out that we consider PAC learning problems most appropriately as problems of type

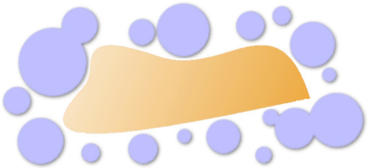
$$(\mathcal{C}, d) \mapsto \{(A, m) : A \text{ PAC learns } \mathcal{C} \text{ with } m\}$$

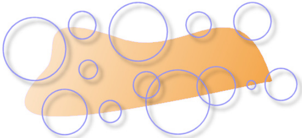
  where we map $\mathcal{C}$ (possibly also $\mathcal{H} \supseteq \mathcal{C}$) and an upper bound $d > \mathrm{VCdim}(\mathcal{C})$ to the learner with complexity (A, m).
- One can also omit the sample complexity m on the output side, but learners without complexity are practically useless!
- We restrict everything to topologically closed concept classes, as the VC dimension is invariant under taking the closure.

Closed sets in computable analysis

There are three ways of representing the hyperspace $\mathcal{A}(2^{\mathbb{N}})$ of closed subsets $A \subseteq 2^{\mathbb{N}}$ of the Cantor space in computable analysis:

- $\mathcal{A}_-$ (negative information): we give all prefixes $w \in \{0,1\}^*$ such that $w2^{\mathbb{N}} \cap \mathcal{H} = \varnothing$.
- $\mathcal{A}_+$ (positive information): we give all prefixes $w \in \{0,1\}^*$ such that $w2^{\mathbb{N}} \cap \mathcal{H} \neq \varnothing$.
- $\mathcal{A}$ (full information): we give both.





PAC learning problems

We consider the following variants of PAC learning problems:

- The proper PAC learning problem

$$\begin{aligned} &\mathsf{PPAC} :\subseteq \mathcal{A} \times \mathbb{N} \rightrightarrows \mathcal{L} \times \mathbb{N}^{\mathbb{N}} \\ &(\mathcal{C}, d) \mapsto \{(A, m) : A \text{ properly PAC learns } \mathcal{C} \text{ with } m\} \end{aligned}$$

  with $\mathrm{dom}(\mathsf{PPAC}) = \{(\mathcal{C}, d) : \mathrm{VCdim}(\mathcal{C}) \leq d \text{ and } \mathcal{C} \neq \varnothing\}$.

- The improper PAC learning problem

$$\begin{aligned} &\mathsf{IPAC} :\subseteq \mathcal{A} \times \mathbb{N} \rightrightarrows \mathcal{L} \times \mathbb{N}^{\mathbb{N}} \\ &(\mathcal{C}, d) \mapsto \{(A, m) : A \text{ improperly PAC learns } \mathcal{C} \text{ with } m\} \end{aligned}$$

  with $\mathrm{dom}(\mathsf{IPAC}) = \{(\mathcal{C}, d) : \mathrm{VCdim}(\mathcal{C}) \leq d \text{ and } \mathcal{C} \neq \varnothing\}$.

- The relative PAC learning problem

$$\begin{aligned} &\mathsf{RPAC} :\subseteq \mathcal{A}_+ \times \mathcal{A}_- \times \mathbb{N} \rightrightarrows \mathcal{L} \times \mathbb{N}^{\mathbb{N}} \\ &(\mathcal{C}, \mathcal{H}, d) \mapsto \{(A, m) : A \text{ PAC learns } \mathcal{C} \text{ relative to } \mathcal{H} \text{ with } m\} \end{aligned}$$

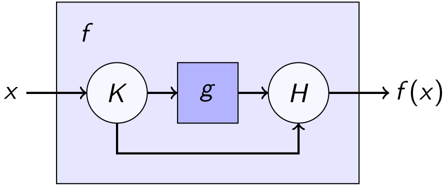
  with $\mathrm{dom}(\mathsf{RPAC}) = \{(\mathcal{C}, \mathcal{H}, d) : \mathrm{VCdim}(\mathcal{C}) \leq d \text{ and } \varnothing \neq \mathcal{C} \subseteq \mathcal{H}\}$.

For PPAC and IPAC, we add an upper index $+$ or $-$ if we replace the space $\mathcal{A}$ by either $\mathcal{A}_+$ or $\mathcal{A}_-$, respectively. For instance, we write PPAC^+.

Weihrauch Complexity

Weihrauch reducibility

Let $f :\subseteq X \rightrightarrows Y$ and $g :\subseteq Z \rightrightarrows W$ be two problems.

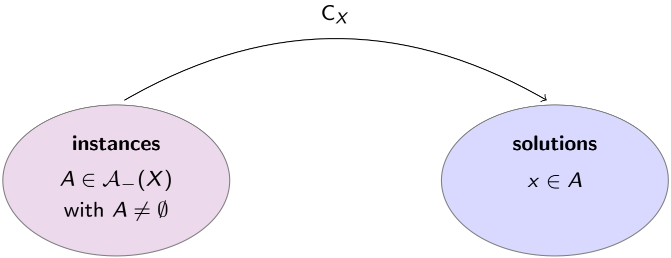


- f is Weihrauch reducible to g, $f \leq_{\mathrm{W}} g$, if there are computable $K :\subseteq X \rightrightarrows V \times Z$ and $H :\subseteq V \times W \rightrightarrows Y$ such that
$$\emptyset \neq H \circ (\mathrm{id}_V \times g) \circ K(x) \subseteq f(x)$$
for all $x \in \mathrm{dom}(f)$.
- f is strongly Weihrauch reducible to g, $f \leq_{\mathrm{sW}} g$, if the above holds with g instead of $\mathrm{id}_V \times g$.
- By $\equiv_{\mathrm{W}}$ and $\equiv_{\mathrm{sW}}$ we denote the corresponding equivalences.
- The distributive lattice induced by $\leq_{\mathrm{W}}$ is usually referred to as Weihrauch lattice.

Choice and limit problems

- A prototypical problem is the following: given a negatively represented closed subset A of a computable metric space X, find a point $x \in A$.

- Formally, this is the choice problem
$$\mathsf{C}_X :\subseteq \mathcal{A}_-(X) \rightrightarrows X, A \mapsto A$$
with $\mathrm{dom}(\mathsf{C}_X) = \{A : A \neq \emptyset\}$.



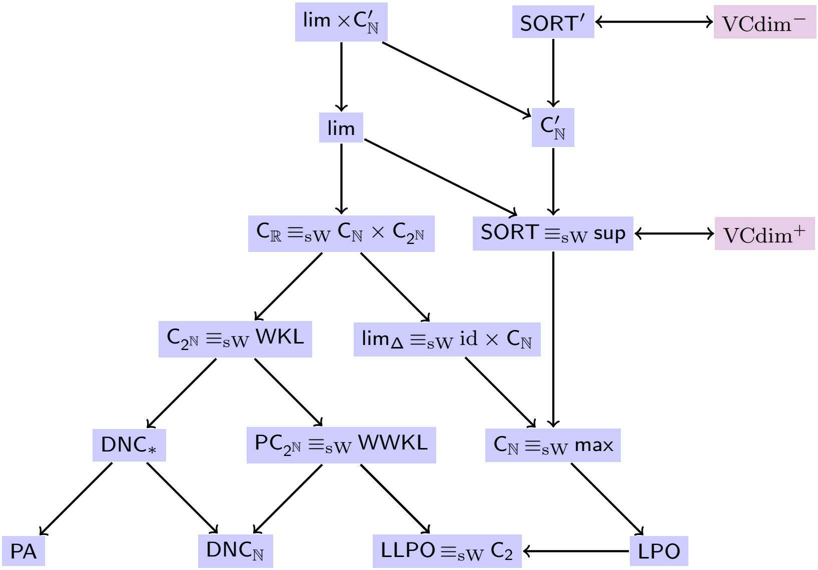
- By $\lim :\subseteq \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ we denote the limit map on Baire space.

Benchmark problems for PAC learning

$$
\begin{tikzcd}
 & & & \lim \times \mathsf{C}'_{\mathbb{N}} \arrow[d] \arrow[drr] & & \mathsf{SORT}' \arrow[d] & \\
 & & & \lim \arrow[d] \arrow[drr] & & \mathsf{C}'_{\mathbb{N}} \arrow[d] & \\
 & & & \mathsf{C}_{\mathbb{R}} \equiv_{\mathrm{sW}} \mathsf{C}_{\mathbb{N}} \times \mathsf{C}_{2^{\mathbb{N}}} \arrow[dl] \arrow[dr] & & \mathsf{SORT} \equiv_{\mathrm{sW}} \mathsf{sup} \arrow[dd] & \\
 & & \mathsf{C}_{2^{\mathbb{N}}} \equiv_{\mathrm{sW}} \mathsf{WKL} \arrow[dl] \arrow[dr] & & \lim_{\Delta} \equiv_{\mathrm{sW}} \mathrm{id} \times \mathsf{C}_{\mathbb{N}} \arrow[dr] & & \\
 & \mathsf{DNC}_{*} \arrow[dl] \arrow[dr] & & \mathsf{PC}_{2^{\mathbb{N}}} \equiv_{\mathrm{sW}} \mathsf{WWKL} \arrow[dl] \arrow[dr] & & \mathsf{C}_{\mathbb{N}} \equiv_{\mathrm{sW}} \mathsf{max} \arrow[dr] & \\
\mathsf{PA} & & \mathsf{DNC}_{\mathbb{N}} & & \mathsf{LLPO} \equiv_{\mathrm{sW}} \mathsf{C}_2 & & \mathsf{LPO} \arrow[ll]
\end{tikzcd}
$$

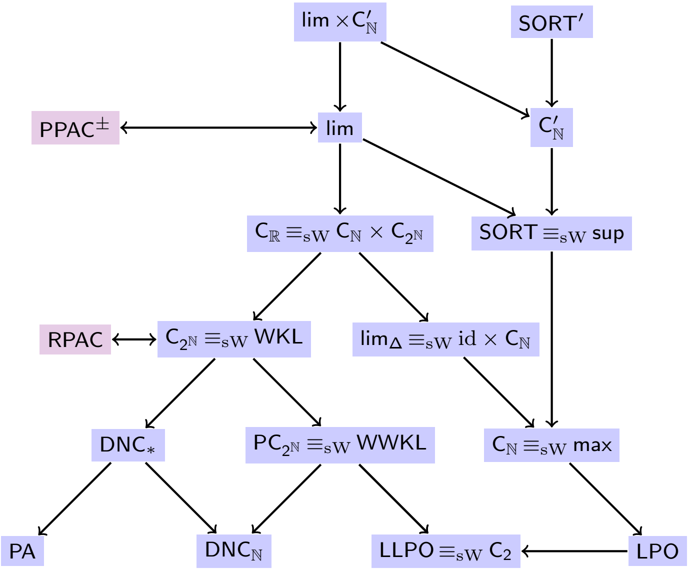
VC dimension

Benchmark problems for PAC learning



Proper and Relative PAC Learning

Benchmark problems for PAC learning



Proper PAC learning

We consider the proper PAC learning problem

$$\begin{aligned}&\mathsf{PPAC} :\subseteq \mathcal{A} \times \mathbb{N} \rightrightarrows \mathcal{L} \times \mathbb{N}^{\mathbb{N}}\\&(\mathcal{C}, d) \mapsto \{(A, m) : A \text{ properly PAC learns } \mathcal{C} \text{ with } m\}\end{aligned}$$

with $\mathrm{dom}(\mathsf{PPAC}) = \{(\mathcal{C}, d) : \mathrm{VCdim}(\mathcal{C}) \leq d \text{ and } \mathcal{C} \neq \varnothing\}$.

Theorem

Proper PAC learning PPAC *is strongly Weihrauch equivalent to*

- id *if* $\mathcal{C} \in \mathcal{A}$ *is given.*
- $\lim$ *if* $\mathcal{C} \in \mathcal{A}_+$ *or* $\mathcal{C} \in \mathcal{A}_-$ *is given.*

Relative PAC learning

We consider the relative PAC learning problem

$$\mathsf{RPAC} :\subseteq \mathcal{A}_+ \times \mathcal{A}_- \times \mathbb{N} \rightrightarrows \mathcal{L} \times \mathbb{N}^{\mathbb{N}}$$
$$(\mathcal{C}, \mathcal{H}, d) \mapsto \{(A, m) : A \text{ PAC learns } \mathcal{C} \text{ relative to } \mathcal{H} \text{ with } m\}$$

with $\mathrm{dom}(\mathsf{RPAC}) = \{(\mathcal{C}, \mathcal{H}, d) : \mathrm{VCdim}(\mathcal{C}) \leq d \text{ and } \varnothing \neq \mathcal{C} \subseteq \mathcal{H}\}$.

Theorem

$\mathsf{RPAC} \equiv_{\mathrm{sW}} \mathsf{WKL} \equiv_{\mathrm{sW}} \mathsf{C}_{2^{\mathbb{N}}}$.

Proof idea. "$\mathsf{RPAC} \leq_{\mathrm{sW}} \mathsf{WKL}$" Given $\mathcal{C} \in \mathcal{A}_+$, $\mathcal{H} \in \mathcal{A}_-$ and m, the condition that (A, m) fails to learn $\mathcal{C}$ relative to $\mathcal{H}$ is c.e. open.
Conversely, we show $\widehat{\mathsf{C}_2} \equiv_{\mathrm{sW}} \mathsf{WKL} \leq_{\mathrm{sW}} \mathsf{RPAC}$ by constructing concept classes $\mathcal{C}_B$ and $\mathcal{H}_B$ given a sequence $B = (B_k)_{k \in \mathbb{N}}$ of sets $B_k \in \mathcal{A}_-(\{0, 1\})$ with $B_k \neq \varnothing$. These classes are chosen such that a point $b \in B_k$ can be reconstructed for each $k \in \mathbb{N}$ by evaluating a learner on a suitable sample S of size $n \geq m\langle 2, 2\rangle$, due to the PAC condition for a Dirac distribution. $\square$

Relative PAC learning

We consider the relative PAC learning problem

$$\begin{aligned}&\mathsf{RPAC} :\subseteq \mathcal{A}_+ \times \mathcal{A}_- \times \mathbb{N} \rightrightarrows \mathcal{L} \times \mathbb{N}^{\mathbb{N}}\\ &(\mathcal{C}, \mathcal{H}, d) \mapsto \{(A, m) : A \text{ PAC learns } \mathcal{C} \text{ relative to } \mathcal{H} \text{ with } m\}\end{aligned}$$

with $\mathrm{dom}(\mathsf{RPAC}) = \{(\mathcal{C}, \mathcal{H}, d) : \mathrm{VCdim}(\mathcal{C}) \leq d \text{ and } \varnothing \neq \mathcal{C} \subseteq \mathcal{H}\}$.

Theorem

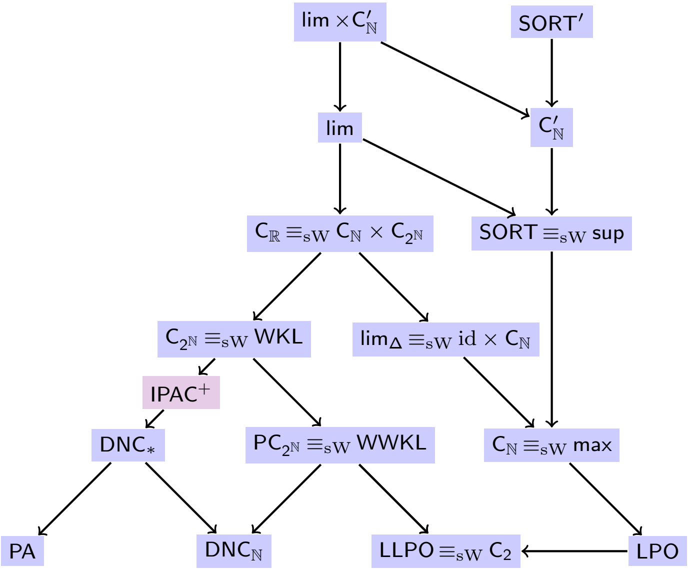
$\mathsf{RPAC} \equiv_{\mathrm{sW}} \mathsf{WKL} \equiv_{\mathrm{sW}} \mathsf{C}_{2^{\mathbb{N}}}$.

Corollary

Relative PAC learning is non-deterministically computable and even complete for this class.

Improper PAC Learning

Benchmark problems for PAC learning



Witness functions

- A function $w : \mathbb{N}^{k+1} \to \{0,1\}^{k+1}$ is called a k–witness or just a witness function for $\mathcal{C} \subseteq 2^{\mathbb{N}}$, if

$$w(x_1, \ldots, x_{k+1}) \neq (h(x_1), \ldots, h(x_{k+1}))$$

  for all $x_1 < \ldots < x_{k+1} \in \mathbb{N}$ and $h \in \mathcal{C}$.
- $\mathrm{VCdim}(\mathcal{C}) = \min\{k \in \mathbb{N} : \mathcal{C} \text{ admits a } k\text{–witness}\}$.
- $\mathrm{eVCdim}(\mathcal{C}) = \min\{k \in \mathbb{N} : \mathcal{C} \text{ admits a computable } k\text{–witness}\}$ is called the effective VC dimension.

Theorem (Delle Rose, Kozachinskiy, Rojas, and Steifer 2023)

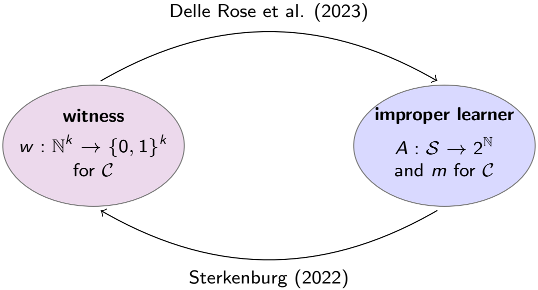
$\mathcal{C}$ admits a computable learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ (with a computable sample complexity m) if and only if the effective VC dimension of $\mathcal{C}$ is finite.

We have analyzed the proof of this theorem and we have extracted the uniform computational content of it.

Witness functions

Theorem (implicit in Delle Rose et al. (2023) and Sterkenburg (2022))

- *Given a witness w of some concept class $\mathcal{C}$ (which is not given), we can computably find a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ and a sample complexity $m \in \mathbb{N}^{\mathbb{N}}$ such that A improperly learns $\mathcal{C}$ with m.*
- *Given a learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ and a sample complexity $m \in \mathbb{N}^{\mathbb{N}}$ such that A improperly learns a concept class $\mathcal{C}$ (which is not given) with sample complexity m, we can computably find a witness w for $\mathcal{C}$.*



Witness functions

The witness problem WIT can be formalized as a Weihrauch problem.

Corollary

$\mathsf{WIT}^{\pm} \equiv_{\mathrm{sW}} \mathsf{IPAC}^{\pm}$.

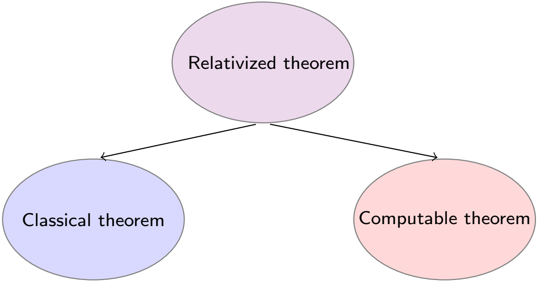
We have proved our main results on IPAC^{+} actually by showing them on WIT^{+}.

Relativized Fundamental Theorem

Corollary

For a non-empty concept class $\mathcal{C} \subseteq 2^{\mathbb{N}}$ and a Turing degree $\mathbf{a}$ the following are equivalent:

- *$\mathcal{C}$ admits an improper learner $A : \mathcal{S} \to 2^{\mathbb{N}}$ that is computable from $\mathbf{a}$ (with computable sample complexity m).*
- *$\mathcal{C}$ admits a witness function that is computable from $\mathbf{a}$.*



Thank you for your attention

Sunset from Les Espaces d'Abraxas in Noisy-le-Grand, near Paris (June 2026)