Weihrauch complexity of PAC learning problems

École polytechnique third-year research internship

Guillaume Chirache
Supervised by Vasco Brattka

Universität der Bundeswehr München
Fakultät für Informatik

11 July 2025

Outline

Computability of functions $\mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$

We consider Turing machines with an *input tape* containing an infinite sequence and a **one-way** *output type*.

Computable functions

$F :\subseteq \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ is *computable* if there exists such a machine that upon input $p \in \mathrm{dom}(F)$ produces $F(p)$ in the long run.

Examples

- The function squarring each symbol of the input is computable.
- The equality test with $\hat{0}$ is not computable.

Continuity over the Baire space

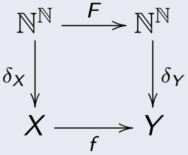
$$\forall F :\subseteq \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}},\; F \text{ computable} \implies F \text{ continuous}$$

Representations and problems

A *representation* of a set X is a surjective partial map $\delta :\subseteq \mathbb{N}^{\mathbb{N}} \to X$.
If $\delta(p) = x$, then $p \in \mathbb{N}^{\mathbb{N}}$ is a δ-*name* of $x \in X$.

Definition

A *problem* is a partial multifunction $f :\subseteq (X, \delta_X) \rightrightarrows (Y, \delta_Y)$.
$F :\subseteq \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ *realizes* f if $\delta_Y F(p) \in f\delta_X(p)$ for any $p \in \mathrm{dom}(f\delta_X)$.



This way, we can consider problems involving **many mathematical objects**.
Some representations (*admissible representations*) are appropriate for a given topology over X.

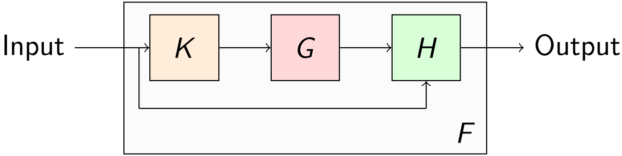
Weihrauch complexity

We have a many-one reduction for problems.

Weihrauch reducibility

$f \leqslant_W g$ if there are computable $H, K :\subseteq \mathbb{N}^\mathbb{N} \to \mathbb{N}^\mathbb{N}$ such that

$$G \vdash g \implies H\langle \mathrm{Id}_{\mathbb{N}^\mathbb{N}}, GK \rangle \vdash f.$$



This reduction induces a lattice of *Weihrauch degrees* full of **natural problems**.
It has a **strong counterpart** $\leqslant_{sW}$ in which H cannot use the initial input.
We have **unary operators** (jump, parallelization) on problems which behave well with (strong) Weihrauch degrees.

Outline

Samples and hypotheses

We have a *sample* of input-label pairs

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

The job of a learner is to find the best hypothesis $h : \mathbb{N} \to \{0, 1\}$ (i.e., an element of the Cantor space $2^{\mathbb{N}}$).

Empirical risk minimizer

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

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

The need for hypothesis classes

An ERM can be a **very bad** learner: there is no constraint on inputs that have not be seen.
We need to make some assumptions about what good hypotheses look like.

Hypothesis class

A *hypothesis class* $\mathcal{H}$ is a non-empty subset of $2^{\mathbb{N}}$.
A learner $A : \mathcal{S} \to \mathcal{H}$ is an $\mathrm{ERM}_{\mathcal{H}}$ if

$$\forall S \in \mathcal{S},\ A(S) \in \operatorname*{argmin}_{h \in \mathcal{H}} L_S(h).$$

PAC learning

Some hypothesis classes are *PAC learnable*, meaning that the error can be made as small as needed, with a probability as high as needed, for a sample big enough drawn i.i.d. from any probability distribution over $\mathbb{N} \times \{0, 1\}$.

VC dimension and the fundamental theorem of PAC learning

We need to characterize PAC learnable classes.

VC dimension

A finite subset $X \subseteq \mathbb{N}$ is *shattered* by $\mathcal{H}$ if

$$\forall f : X \to \{0,1\},\ \exists h \in \mathcal{H},\ \ h|_X = f.$$

The VC dimension of $\mathcal{H}$ is $\mathrm{VCdim}(\mathcal{H}) = \sup\{|X| \mid \mathcal{H} \text{ shatters } X\} \in \mathbb{N}_\infty$.

For instance, $\mathrm{VCdim}\{\mathbb{1}_{[\![a,b]\!]} \mid a < b \in \mathbb{N}\} = 2$.

Fundamental theorem of PAC learning

The following are equivalent:

- $\mathcal{H}$ is PAC learnable.
- Any $\mathrm{ERM}_{\mathcal{H}}$ is a PAC learner for $\mathcal{H}$.
- $\mathrm{VCdim}(\mathcal{H}) < \infty$.

Outline

Representing closed hypotheses classes

For cardinality reasons, we choose to focus on **closed** hypothesis classes. There are several interesting ways to represent a closed set $\mathcal{H} \subseteq 2^{\mathbb{N}}$.

Different representations of $\mathcal{H}$

- With *negative information*: we give all prefixes $w \in 2^*$ such that $w2^{\mathbb{N}} \cap \mathcal{H} = \varnothing$.
- With *positive information*: we give all prefixes $w \in 2^*$ such that $w2^{\mathbb{N}} \cap \mathcal{H} \neq \varnothing$.
- With *full information*: we give both.

We write $\mathscr{A}_+(2^{\mathbb{N}})$, $\mathscr{A}_-(2^{\mathbb{N}})$, or $\mathscr{A}_\pm(2^{\mathbb{N}})$.
For instance, we can define the problem:

$$\begin{array}{rcccc} \mathrm{VCdim}_{\mathscr{A}_+(2^{\mathbb{N}})} & : & \mathscr{A}_+(2^{\mathbb{N}}) & \to & \mathbb{N}_\infty \\ & & \mathcal{H} & \mapsto & \mathsf{VCdim}(\mathcal{H}) \end{array}$$

Positive information and the SORT problem

By exploring the positive information, we can give at any point a lower bound for the VC dimension.
Any shattered finite set $X \subseteq \mathbb{N}$ is eventually discovered, so this minimal estimate tends towards the VC dimension.
Finding the VC dimension is therefore equivalent to finding the (finite or infinite) limit of a non-decreasing sequence of natural numbers.

$$\lim_{\mathbb{N}_\infty}$$

$$\downarrow$$

$$\mathrm{SORT} \equiv_{\mathsf{sW}} \lim^{\nearrow}_{\mathbb{N}_\infty} \equiv_{\mathsf{sW}} \mathrm{VCdim}_{\mathscr{A}_+(2^{\mathbb{N}})}$$

$$\downarrow$$

$$\lim_{\mathbb{N}} \equiv_{\mathsf{sW}} \lim^{\nearrow}_{\mathbb{N}}$$

The "irrelevance" of the negative information

The negative information is "useless" as such: even if you know that a set of given cardinality is not shattered, another set of the same cardinality with bigger indices could very well be shattered.
It cannot be used in a better way than being translated to the positive information, which is equivalent to a **Turing jump**.

VC dimension with negative information

Having the full information is no better than having the positive information:

$$\mathrm{VCdim}_{\mathscr{A}_{\pm}(2^{\mathbb{N}})} \equiv_{\mathsf{sW}} \mathrm{VCdim}_{\mathscr{A}_{+}(2^{\mathbb{N}})} \equiv_{\mathsf{sW}} \mathrm{SORT}.$$

The negative information alone needs to be translated via a jump:

$$\mathrm{VCdim}_{\mathscr{A}_{-}(2^{\mathbb{N}})} \equiv_{\mathsf{sW}} \mathrm{SORT}'.$$

Outline

Computation with full information

The ERM problem is defined as such (with $*$ being $+$, $-$, or $\pm$):

$$\begin{array}{rcccl} \mathrm{ERM}_{\mathscr{A}_*(2^{\mathbb{N}})} & : & \mathscr{A}_*(2^{\mathbb{N}}) \setminus \{\varnothing\} & \rightrightarrows & (2^{\mathbb{N}})^{\mathcal{S}} \\ & & \mathcal{H} & \mapsto & \mathsf{ERM}_{\mathcal{H}} \end{array}$$

Theorem

Computing an ERM is possible when the hypothesis class is described by its full information, i.e., $\mathrm{ERM}_{\mathscr{A}_\pm(2^{\mathbb{N}})}$ is computable.

Proof sketch.

- Given a sample, the empirical risk does not depend on hypothesis values beyond a certain index.
- Use positive and negative information to find the best prefix up to that index.
- Extend this prefix with positive information.

Weihrauch degree with positive information

With only the positive information, it is **not possible** to know whether a missing prefix will eventually appear or whether it actually belongs to the negative information.
We need to take the whole sequence into account and choose the best prefix it contains.
This is equivalent to finding the minimum of a sequence.
This must be done for all samples, and this parallelization is equivalent to finding the limit of a sequence of sequences of natural numbers.

Theorem

$$\mathrm{ERM}_{\mathscr{A}_+(2^{\mathbb{N}})} \equiv_{\mathrm{W}} \lim_{\mathbb{N}^{\mathbb{N}}}$$

This is also true under the assumption that the VC dimension is finite.

Conclusion

- The Weihrauch lattice: a structure in which natural problems are well embedded.
- Is the problem of finding a PAC learner easier?

$$\mathrm{SORT}' \equiv_{\mathrm{sW}} \mathrm{VCdim}_{\mathscr{A}_-(2^{\mathbb{N}})}$$

$$\downarrow$$

$$\lim_{\mathbb{N}^{\mathbb{N}}} \equiv_{\mathrm{W}} \mathrm{ERM}_{\mathscr{A}_+(2^{\mathbb{N}})}$$

$$\downarrow$$

$$\lim_{\mathbb{N}_\infty}$$

$$\downarrow$$

$$\mathrm{SORT} \equiv_{\mathrm{sW}} \lim^{\nearrow}_{\mathbb{N}_\infty} \equiv_{\mathrm{sW}} \mathrm{VCdim}_{\mathscr{A}_\pm(2^{\mathbb{N}})} \equiv_{\mathrm{sW}} \mathrm{VCdim}_{\mathscr{A}_+(2^{\mathbb{N}})}$$

$$\downarrow$$

$$\lim_{\mathbb{N}} \equiv_{\mathrm{sW}} \lim^{\nearrow}_{\mathbb{N}}$$

$$\downarrow$$

$$\min_{\mathbb{N}} \equiv_{\mathrm{W}} \mathrm{ERM}^{\mathrm{loc}}_{\mathscr{A}_+(2^{\mathbb{N}})}$$

$$\downarrow$$

$$\mathrm{Id}_{\mathbb{N}^{\mathbb{N}}} \equiv_{\mathrm{W}} \mathrm{ERM}^{\mathrm{loc}}_{\mathscr{A}_\pm(2^{\mathbb{N}})} \equiv_{\mathrm{W}} \mathrm{ERM}_{\mathscr{A}_\pm(2^{\mathbb{N}})} \text{ (computable)}$$