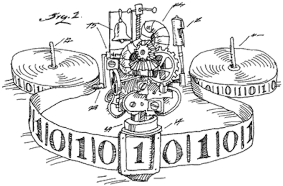
Algorithmic learning of probability distributions from random data in the limit



by George Barmpalias

Joint work with Nan Fang and Frank Stephan

January 2018, Oberwolfach

Chinese Academy of Sciences

One of the central problems in statistics is:

given a set of random data, find a distribution with respect to which the given data are random.

Vinanyi and Chater recently proposed to study this question from the point of view of Algorithmic Learning Theory.

In Algorithmic Learning Theory the basic problem is:

given samples of a formal language, find a grammar that generates the given language.

Algorithmic Learning Theory (Gold 1967)

Given increasingly long initial segments of a computable $X \in 2^\omega$, eventually find a machine that computes X.

A learner is a function $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$.

We say that learner $\mathcal{L}$:

EX-succeeds on X if $\lim_n \mathcal{L}(X \upharpoonright_n)$ is an index of X

BC-succeeds on X if for almost all n, $\mathcal{L}(X \upharpoonright_n)$ is an index of X

Partially succeeds on X if $\mathcal{L}(X \upharpoonright_n)$ is a fixed index of X for infinitely many n, and any other guess appears finitely often.

(Osherson, Stob and Weinstein)

Classic facts in Algorithmic Learning Theory

- The computable reals are not EX or BC learnable. (Gold)
- Learnability is not closed under union. (Blum and Blum)
- The computable reals are partially learnable.
  (Osherson, Stob and Weinstein)
- An oracle can EX-learn all computable reals iff it is high.
  (Adleman and Blum).
- Low for EX is exactly 1-generics below $0'$
  (Slaman, Solovay, Pleszkoch, Gasarch, Jain)

Learning probability distributions

Given increasingly long initial segments of $X \in 2^\omega$ which is μ-random for a computable measure μ, eventually find a description of μ' such that X is μ'-random.

A learner is a function $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$. We say that $\mathcal{L}$:

EX-succeeds on X if $\lim_n \mathcal{L}(X \restriction_n)$ is an index of some μ such that X is μ-random.

BC-succeeds on X if there exists a computable μ such that X is μ-random and for almost all n, $\mathcal{L}(X \restriction_n)$ is an index of μ

Partially succeeds on X if there exists μ such that X is μ-random, $\mathcal{L}(X \restriction_n)$ is an index of X for infinitely many n, and any other approximation appears finitely often.

Learning probability distributions

Given any uniformly computable family of measures $\mathcal{C}$, there exists a computable learner that EX-succeeds on every μ-random real for any $\mu \in \mathcal{C}$.

(Vitanyi/Chater)

There is no computable learner which EX or BC succeeds on all reals that are μ-random for some computable (continuous) μ.

(Bienvenu, Monin, Shen)

An oracle EX-succeeds on all μ-random reals for any computable (continuous) μ iff it is high.

(Barmpalias/Stephan).

There exists a learner which partially succeeds on every μ-random real for any computable μ.

(Barmpalias/Stephan).

Learning classes of measures

A class $\mathcal{C}$ of computable measures is weakly EX-learnable if there exists a computable learner $\mathcal{L}$ such that for every $\mu \in \mathcal{C}$ and every μ-random real X the limit $\lim_n \mathcal{L}(X \upharpoonright_n)$ exists and equals an index some μ' such that X is μ'-random.

A class $\mathcal{C}$ of computable measures is EX-learnable if there exists a computable learner $\mathcal{L}$ such that for every $\mu \in \mathcal{C}$ and every μ-random real X, $\lim_n \mathcal{L}(X \upharpoonright_n)$ exists and equals an index of some $\mu' \in \mathcal{C}$ such that X is μ'-random.

Similarly for weak BC-learnability.

Related to the layerwise learnability of Bienvenu/Monin.

Question: Closure of EX and BC learnability under subsets?

Equivalence of learnability of reals and measures

If a class of measures $\mathcal{B}$ is nicely parametrized by a class of reals $\mathcal{C}$ then $\mathcal{B}$ is EX/BC learnable iff $\mathcal{C}$ is EX/BC learnable.

Formally, let $\mathcal{M}$ be the space of Borel measures on 2^ω:

Let $\mathrm{f} : 2^\omega \to \mathcal{M}$ be computable and $\mathcal{D} \subseteq 2^\omega$ effectively closed such that for any $\mathrm{X} \neq \mathrm{Y}$ in $\mathcal{D}$ the measures $\mathrm{f}(\mathrm{X}), \mathrm{f}(\mathrm{Y})$ are effectively orthogonal.

If $\mathcal{D}^* \subseteq \mathcal{D}$ is a class of computable reals, $\mathcal{D}^*$ is EX-learnable if and only if $\mathrm{f}(\mathcal{D}^*)$ is EX-learnable.

The same is true of the BC learnability of $\mathcal{D}^*$.

(Barmpalias/Fang).

Applications of the equivalence theorem

There exist two EX-learnable classes of (Bernoulli) measures whose union is not EX-learnable.

The class of computable Bernoulli measures is EX-learnable with oracle A iff A is high.

Every low for EX (for measures) is 1-generic and below $0'$.

Oracles $\not\leq_T 0'$ are not low for EX for measures.

The computable measures are EX-learnable with finitely many queries on an oracle A if and only if $\emptyset'' \leq_T A \oplus \emptyset'$.

Thanks! – and main references

- Barmpalias/Stephan. Algorithmic learning of probability distributions from random data in the limit. Arxiv:1710.11303 (2017)
- Barmpalias/Fang. An equivalence between learning of data and probability distributions Arxiv:1801.02566 (2018)
- Bienvenu/Figueira/Monin/Shen. Algorithmic identification of probabilities is hard. ArXiv:1405.5139 (2017, also ALT 2014)
- Bienvenu/Monin. Von Neumann's Biased Coin Revisited LICS'12
- Vitanyi/Chater. Identification of probabilities JMP (2017)