Absolute versus probabilistic classification in a logical setting

Sanjay Jain[1], Eric Martin[2,3] and Frank Stephan[4]

[1] School of Computing, National University of Singapore, Singapore 117543, Republic of Singapore
sanjay@comp.nus.edu.sg *
[2] School of Computer Science and Engineering, The University of New South Wales, UNSW Sydney NSW 2052, Australia
[3] National ICT Australia, UNSW Sydney NSW 2052, Australia
emartin@cse.unsw.edu.au **
[4] School of Computing and Department of Mathematics, National University of Singapore, Singapore 117543, Republic of Singapore
fstephan@comp.nus.edu.sg *

Abstract. Given a set $\mathcal{W}$ of logical structures, or *possible worlds*, a set of logical formulas called *possible data*, and a logical formula φ, we consider the classification problem of determining in the limit and *almost always correctly* whether a possible world $\mathfrak{M}$ satisfies φ, from a complete enumeration of the possible data that are true in $\mathfrak{M}$. One interpretation of *almost always correctly* is that the classification might be wrong on a set of possible worlds of measure 0, with respect to some natural probability distribution over the set of possible worlds. Another interpretation is that the classifier is only required to classify a set $\mathcal{W}'$ of possible worlds of measure 1, without having to produce any claim in the limit on the truth of φ in the members of the complement of $\mathcal{W}'$ in $\mathcal{W}$. We compare these notions with absolute classification of $\mathcal{W}$ with respect to a formula that is almost always equivalent to φ in $\mathcal{W}$, hence investigate whether the set of possible worlds on which the classification is correct is definable. We mainly work with the probability distribution that corresponds to the standard measure on the Cantor space, but we also consider an alternative probability distribution proposed by Solomonoff and contrast it with the former. Finally, in the spirit of the kind of computations considered in Logic programming, we address the issue of computing *almost correctly* in the limit witnesses to leading existentially quantified variables in existential formulas.

1 Introduction

Paradigms of inductive inference are often highly idealized, even for those that impose very tight restrictions on the learning scenario. There might be constraints on how data are presented to a learner, or on the resources that are made available to a learner, or on the criterion that formalizes what is meant by 'learning.' Still learning paradigms are often 'merciless' when it comes

* Sanjay Jain is supported by the NUS research grant R252–000–127–112; Frank Stephan is supported by the NUS research grant R252–000–212–112.

** National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence Program.

to qualifying a learner as successful. They expect a successful learner to be correct with respect to all possible realities (languages in the numerical setting, structures in the logical setting) of the paradigm. This is certainly an extreme demand, that can be the object of theoretical investigation, but that would not be imposed in most practical contexts. Agents or processes are nor expected to be infallible. Allowing the learning process to succeed with respect to *almost* all realities—intuitively, successfully learning with probability 1—appears as a reasonable requirement that deserves to be investigated.

Probabilistic elements have already been considered in inductive inference, but they relate more to the learning process than to the class of languages learnt by a machine (sample references are [6, 8, 12, 13, 16]). For example, learning functions in the limit with probability $1/n$ turns out to be equivalent to having n nonprobabilistic learners such that at least one of them succeeds [13, 16]. Furthermore, most concepts have a break-even point at some probability $c < 1$ in the sense that whenever such concepts are learnable with probability c, they are already learnable by a deterministic machine [1]. Meyer [11] showed that exact monotonic and exact conservative learning with any probability $c < 1$ is more powerful than deterministic learning; still in case $c = 1$, the probabilistic and deterministic variants are again the same. In [7], the notions of effective measure and category are used to discuss the relative sizes of inferable sets and their complements.

An example for a setting in inductive inference where learning with probability 1 is more powerful than deterministic learning is the following. Assign to each set A to be learnt the distribution p_A with $p_A(x) = 2^{-1-x}$ for $x \in A$, $p_A(x) = 0$ for $x \notin A$ and $p_A(\#) = \sum_{x \notin A} 2^{-1-x}$. Then any class of sets that is learnable from informant is also learnable from text with probability 1, provided that for every member A of the class, the elements of a text for A are drawn with probability p_A. As some classes are learnable from informant but not from text, these classes witness that learning from almost all texts is more powerful than learning from all texts.

The main reason why probabilistic elements are restricted to the learning process and not to the class of realities being considered is that in a countable domain, 'almost all objects' would normally mean 'cofinitely many objects' and finite exceptions can often be handled by suitably patching the machine. Therefore it is much more appropriate to consider classification where one deals with a continuum of possible realities which can be identified with the Cantor-Space. Then 'almost all' can be interpreted in two major ways: 'of second category' as defined in topology, or 'of measure 1' as defined in measure theory.

Classification was already implicitly considered by recursion theorists when they investigated computation in the limit relative to an oracle, which subsumes classification [14]. Ben-David [3] characterized classification in the limit in topological terms. Subsequent work then established a connection between classification on one side and logic on the other side [2, 9]. In [10] the relationship between classification and topology was brought one step further, by casting the classification in a logical setting that considered arbitrary sets of data, each set determining a particular topology, and arbitrary sets of structures. Of particular importance in [10] are (usually axiomatized) classes of structures consisting of Henkin structures only, where every individual of the domain is denoted by a term in the underlying language. The set of atomic sentences (that is, closed atomic formulas) true in such structures uniquely determines the structure, and can be identified with a point in the Cantor space. A probability distribution can then be defined on the set of structures which represent the possible realities, that generalizes the classical probability distribution on the Cantor space. Still it is legitimate to consider alternative probability distributions on the set of possible realities, and this paper will explore one such alternative. It is also natural to examine what can be derived from the assumption that the set of possible realities is equipped with an arbitrary probability distribution, and we will provide one such general result.

The setting chosen for this paper is a particular instance of the logical framework investigated in [10]. It conceives of a logical paradigm as a vocabulary $\mathcal{V}$, a set $\mathcal{W}$ of structures over $\mathcal{V}$, or *possible worlds*, and a set $\mathcal{D}$ of closed formulas over $\mathcal{V}$, or *possible data*. Important cases of possible data are obtained by taking for $\mathcal{D}$ the set of atomic sentences or the set of basic sentences (atomic sentences or their negations). Both choices determine counterparts to the numerical notions of text and informant, in the form of enumerations of all possible data that are true in an underlying possible world $\mathfrak{M}$, yielding an *environment* for $\mathfrak{M}$. Given a formula φ, we consider the task of determining in the limit, from an environment for a possible world $\mathfrak{M}$, whether $\mathfrak{M}$ satisfies φ. In other words, the task is to classify a possible world as a member of one of two classes: the class of structures that satisfy φ, and the class of structures that don't. But we allow the classification to fail on a set of environments for a small set of possible worlds—either of first category or of measure 0.

It is important to distinguish between failing to converge to some answer and misclassifying. In the case of misclassification with respect to φ, an interesting question is whether perfect classification of $\mathcal{W}$ is achieved on the basis of another formula ψ, whose set of models in $\mathcal{W}$ is of course equal to the set of models of φ up to a set of measure 0. Whether the failure to classify correctly is due to nonconvergence or to genuine misclassification, we only measure on which class of possible worlds $\mathfrak{M}$ a correct classification is achieved from all environments for $\mathfrak{M}$. We do not assume that the possible data are generated following some underlying probability distribution, nor do we impose any condition on the speed of convergence. In other words, we remain in the realm of inductive inference and our use of probabilities is essentially different to its use in the PAC framework.

We now proceed as follows. In Section 2 we introduce the basic notions, that we apply to the classification task in Sections 3 to 7. More precisely, Section 3 is based on arbitrary measures whereas from Section 4 to Section 6, we focus on the measure that corresponds to the usual measure on the Cantor space. In Section 7, we show that we get a different picture if the measure proposed by Solomonoff [15] is used instead. In Section 8 we get back to the measure on the Cantor space and particularize the framework to φ being of the form $\exists x \psi(x)$, with the aim of not only classifying φ, but of computing in the limit a witness to the existentially quantified variable x. Provided that $\mathcal{W}$ is axiomatizable by a logic program, this corresponds to 'error tolerant' computations in Logic programming, where ψ is assumed to be quantifier free or to only contain bounded quantifiers [19]. When ψ is universal, and also with some assumptions on $\mathcal{W}$, this corresponds to 'error tolerant' computations in Limiting resolution [5].

2 Absolute and probabilistic classification

Let a class X be given. The class of finite sequences of members of X, including the empty sequence $()$, is represented by $X^\star$. The length of $\sigma \in X^\star$ is denoted $\mathrm{lt}(\sigma)$. The class of sequences of members of X of length ω is represented by X^ω.

Given a nonempty vocabulary V, that is, a set of (possibly nullary) predicate and function symbols, we shall consider both *first-order formulas over* V and *monadic second-order formulas over* V, built from the symbols in V, equality, the usual Boolean operators, first-order variables and quantifiers over those, and in the case of monadic second-order formulas, unary predicate variables and quantifiers over those. A (first-order or monadic second-order) *sentence over* V refers to a closed (first-order or monadic second-order) formula over V. We adopt the following conventions.

- We fix a vocabulary $\mathcal{V}$ containing at least a constant $\overline{0}$, a unary function symbol s and a unary predicate symbol P, possibly enriched with either the binary function symbol $+$, or with the binary function symbols $+$ and $*$. Given $n \in \mathbb{N}$, we denote by $\overline{n}$ the term obtained from $\overline{0}$ by n applications of s (hence $\overline{n+1} = s(\overline{n})$ for all $n \in \mathbb{N}$). We say that $\mathcal{V}$ is *standard* if $\mathcal{V}$ consists of $\overline{0}$, s and P only.
- We fix a language $\mathcal{L}$ equal either to the set of first-order sentences over $\mathcal{V}$ or to the set of monadic second-order sentences over $\mathcal{V}$.
- We fix an infinite set $\mathcal{D}$ of first-order sentences over $\mathcal{V}$, referred to as *possible data*.
- We fix a set $\mathcal{W}$ of structures over $\mathcal{V}$, referred to as *possible worlds*, all of whose individuals interpret a unique closed term of the form $\overline{n}$, $n \in \mathbb{N}$. When $\mathcal{V}$ contains $+$, we assume that the interpretation of $+$ in all members of $\mathcal{W}$ is given by the standard interpretation of $+$ in $\mathbb{N}$. When $\mathcal{V}$ contains $*$, we assume that the interpretation of $*$ in all members of $\mathcal{W}$ is given by the standard interpretation of $*$ in $\mathbb{N}$. We assume that for every subset X of $\mathbb{N}$, there exists a unique $\mathfrak{M} \in \mathcal{W}$ with $\{n \in \mathbb{N} : \mathfrak{M} \models P(\overline{n})\} = X$.

We say *term* for term over $\mathcal{V}$ and *sentence* for member of $\mathcal{L}$. Given $T \subseteq \mathcal{L}$, $\mathrm{Mod}_{\mathcal{W}}(T)$ represents the set of models of T in $\mathcal{W}$. Given $\varphi \in \mathcal{L}$, we write $\mathrm{Mod}_{\mathcal{W}}(\varphi)$ for $\mathrm{Mod}_{\mathcal{W}}(\{\varphi\})$. We will also use the following terminology.

Definition 1. Given $X \subseteq \mathcal{L}$ and a possible world $\mathfrak{M}$, we define the *X-diagram of* $\mathfrak{M}$ as the set of all members of X that are true in $\mathfrak{M}$.

Note that by convention, every subset of $X = \{P(\overline{n}) : n \in \mathbb{N}\}$ is the X-diagram of some possible world. We use *environment* to refer to an enumeration of the $\mathcal{D}$-diagram of a member of $\mathcal{W}$:

Definition 2. Given a possible world $\mathfrak{M}$, an *environment for* $\mathfrak{M}$ is any member e of $(\mathcal{D} \cup \{\sharp\})^\omega$ such that for all $\varphi \in \mathcal{D}$, φ occurs in e iff φ is true in $\mathfrak{M}$.

We denote by μ a measure on $\mathcal{W}$ such that for all $\varphi \in \mathcal{L}$, $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi))$ is measurable. Of particular importance is the (unique) measure on $\mathcal{W}$ that is directly derived from the standard measure on the Cantor space. More precisely, put $X = \{P(\overline{n}) : n \in \mathbb{N}\}$. The X-diagram of a possible world can then be identified with a point in the Cantor space. The next definition makes the relationship explicit, and uses the relationship to define the standard measure on $\mathcal{W}$, as well as the topological notions of sets of first and second category.

Definition 3. Given a possible world $\mathfrak{M}$, the *standard informant for* $\mathfrak{M}$ is the (unique) member e of $\{0,1\}^\omega$ such that for all $n \in \mathbb{N}$, $e(n) = 1$ iff $\mathfrak{M} \models P(\overline{n})$.

– We say that μ is *standard* iff for all subsets W of $\mathcal{W}$, the following holds. Let X be the set of standard informants of the members of W. Then $\mu(W)$ is defined iff X is Lebesgue measurable in the Cantor space. Moreover, if $\mu(W)$ is defined then it is equal to the measure of X in the Cantor space.
– Let a set W of possible worlds be given, and let X be the set of standard informants of the members of W. If X is of first, respectively, second, category in the Cantor space then we say that W is of *first*, respectively, *second*, *category*.

Note that in case $\mathcal{D} = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$, the standard informant for $\mathfrak{M}$ can be identified with an environment for $\mathfrak{M}$.

Definition 4. Two sentences φ and ψ are said to be *almost equivalent* iff $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi \leftrightarrow \neg\psi))$ is defined and null.

Given $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$, $\mathrm{cnt}(\sigma)$ denotes the set of members of $\mathcal{D}$ that occur in σ. The proofs of many propositions will make use of the next technical definition.

Definition 5. We say that a member σ of $(\mathcal{L} \cup \{\sharp\})^\star$ is *consistent in* $\mathcal{W}$ just in case there exists a member $\mathfrak{M}$ of $\mathcal{W}$ such that $\mathfrak{M} \models \mathrm{cnt}(\sigma)$.

The classification task will be performed by a classifier, defined next.

Definition 6. Given a set X of sentences, an X*-classifier* is a partial function from $(X \cup \{\sharp\})^\star$ into $\{0,1\}$. We say *classifier* for $\mathcal{D}$-classifier.

The following pair of definitions capture the absolute notion of classification.

Definition 7. Let a classifier f and a subset $\mathcal{W}'$ of $\mathcal{W}$ be given.

Given a subset X of $\mathcal{W}$, we say that f *classifies* $\mathcal{W}'$ *in the limit following* X just in case for all $\mathfrak{M} \in \mathcal{W}'$ and environments e for $\mathfrak{M}$:

– $\mathfrak{M} \in X$ iff $\{\sigma \in (\mathcal{D} \cup \{\sharp\})^\star : \sigma \subset e$ and $f(\sigma) = 1\}$ is cofinite;
– $\mathfrak{M} \notin X$ iff $\{\sigma \in (\mathcal{D} \cup \{\sharp\})^\star : \sigma \subset e$ and $f(\sigma) = 0\}$ is cofinite.

Given a sentence φ, we say that f *classifies* $\mathcal{W}'$ *in the limit following* φ iff f classifies $\mathcal{W}'$ in the limit following $\mathrm{Mod}_{\mathcal{W}}(\varphi)$.

Definition 8. Given $\varphi \in \mathcal{L}$ and $\mathcal{W}' \subseteq \mathcal{W}$, we say that $\mathcal{W}'$ *is classifiable*, respectively, *computably classifiable*, *in the limit following* φ iff some classifier, respectively, computable classifier, classifies $\mathcal{W}'$ in the limit following φ.

We are interested in classifiers that classify all possible worlds, but misclassify a subset of $\mathcal{W}$ of measure 0, as captured in the next pair of definitions.

Definition 9. Let a classifier f and a sentence φ be given. We say that f *classifies* $\mathcal{W}$ *in the limit following* φ *almost everywhere* iff there exists a subset X of $\mathcal{W}$ such that:

- $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi)\triangle X)$ is defined and null;
- f classifies $\mathcal{W}$ in the limit following X.

Definition 10. Given $\varphi \in \mathcal{L}$, we say that $\mathcal{W}$ is *classifiable*, respectively, *computably classifiable*, *in the limit following* φ *almost everywhere* iff some classifier, respectively, computable classifier, classifies $\mathcal{W}$ in the limit following φ almost everywhere.

3 General measures

We start with the simple observation that in measure-theoretic terms, misclassification of a small set of possible worlds implies absolute classification of almost all possible worlds:

Property 11. *Let a sentence φ be such that $\mathcal{W}$ is classifiable in the limit following φ almost everywhere. Then there exists a subset $\mathcal{W}'$ of $\mathcal{W}$ with $\mu(\mathcal{W}') = 1$ that is classifiable in the limit following φ.*

The well-known relationships between classification in the limit and Σ_2 sentences has a counterpart in the probabilistic setting being investigated:

Proposition 12. *Suppose that $\mathcal{D}$ is closed under negation. Let a sentence φ be given. Then there exists a subset $\mathcal{W}'$ of $\mathcal{W}$ with $\mu(\mathcal{W}') = 1$ such that $\mathcal{W}'$ is computably classifiable in the limit following φ.*

Proof. Assume that $\mathcal{D}$ is closed under negation. Let a sentence φ be of the form

$$Q_1x_1Q_2x_2\dots Q_nx_n\psi(x_1,\dots,x_n)$$

where $n \in \mathbb{N}$, $Q_1, Q_2, \dots, Q_n$ are either existential or universal quantifiers, and $\psi(t_1,\dots,t_n)$ is a finite Boolean combination of $\mathcal{D}$ for all closed terms $t_1, t_2, \dots, t_n$. Let $\tau_0, \tau_1, \dots$ be an enumeration of all closed terms. Remember that μ is chosen in such a way that $\mathrm{Mod}_{\mathcal{W}}(\varphi)$ is measurable. Moreover, by the choice of $\mathcal{W}$, the set of models in $\mathcal{W}$ of a closed formula of the form $\exists x\chi(x)$ is equal to the set of models in $\mathcal{W}$ of $\bigvee\{\chi(\overline{n}) : n \in \mathbb{N}\}$, and the set of models in $\mathcal{W}$ of a closed formula of the form $\forall x\chi(x)$ is equal to the set of models in $\mathcal{W}$ of $\bigwedge\{\chi(\overline{n}) : n \in \mathbb{N}\}$. We infer that for all $d \in \mathbb{N}$, there exists $n^d, i^d_1, \dots, i^d_n \in \mathbb{N}$ with the following property. For all $d \in \mathbb{N}$ and $m \leq n$, let φ^d_m denote

$$Q_1x_1 < \overline{i^d_1}\, Q_2x_2 < \overline{i^d_2} \dots Q_mx_m < \overline{i^d_m}\, Q_{m+1}x_{m+1} \dots Q_nx_n\psi(\tau_{i^d_1}, \dots, \tau_{i^d_m}, x_{m+1}, \dots, x_n).$$

Then for all $d \in \mathbb{N}$ and $m < n$, $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi^d_m)\Delta\mathrm{Mod}_{\mathcal{W}}(\varphi^d_{m+1})) < 2^{-d}/n$. Note that for all $d \in \mathbb{N}$, $\varphi^d_0 = \varphi$ and $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi^d_0)\Delta\mathrm{Mod}_{\mathcal{W}}(\varphi^d_n)) < 2^{-d}$.

Let a classifier f be defined as follows. First modify φ^0_n such that φ^0_n is not satisfiable in $\mathcal{W}$, in order to avoid that f be undefined on some input. Let $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ be given. Let $d \in$

$\{0, 1, \ldots, |\sigma|\}$ be greatest such that it can be decided from the data seen so far whether these data are consistent with φ_n^d in $\mathcal{W}$: this can be done for at least those $d \in \mathbb{N}$ such that for all $j_1 < i_1, j_2 < i_2, \ldots, j_n < i_n$ and for all $\theta \in \mathcal{D}$ that occur in $\psi(\tau_{j_1}, \tau_{j_2}, \ldots, \tau_{j_n})$ either θ or $\neg\theta$ is one of the data seen so far. Thus d is monotonically increasing and unbounded in the number of received data. Let $\mathcal{W}'$ be the set of all possible worlds in which the interpretation of φ_0^d is equal to the interpretation of φ_n^d for almost all d. As $\mathrm{Mod}_{\mathcal{W}}(\varphi_0^d)$ and $\mathrm{Mod}_{\mathcal{W}}(\varphi_n^d)$ differ by a set of measure 2^{-d} at most, $\mathcal{W}'$ has measure 1. Moreover, f classifies $\mathcal{W}'$ following φ, which completes the proof of the proposition. □

4 Failing to classify versus misclassifying

In this section, as well as in Sections 5, 6 and 8, we assume that μ is standard, as defined in Definition 3.

The first use of the standard measure is to show that the converse of Property 11 does not necessarily hold. Indeed, a classifier that correctly classifies a subset $\mathcal{W}'$ of $\mathcal{W}$ of measure 1 might be forced to diverge on some environments for some members of $\mathcal{W} \setminus \mathcal{W}'$. The next proposition shows that this might indeed happen.

Proposition 13. *Suppose that $\mathcal{V}$ is enriched with $+$ only, $\mathcal{L}$ is the set of first-order sentences and $\mathcal{D} = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$. Then there exists a sentence φ with the following properties.*

- *There exists a subset $\mathcal{W}'$ of $\mathcal{W}$ with $\mu(\mathcal{W}') = 1$ such that $\mathcal{W}'$ is computably classifiable in the limit following φ.*
- *$\mathcal{W}$ is not classifiable in the limit following φ almost everywhere.*

Proof. Write $x \leq y$ for $\exists z(x + z = y)$ and $x < y$ for $x \leq y \wedge x \neq y$, and define

- a formula $\psi(x)$ whose meaning is that $P(y)$ holds for all y between $\frac{x}{2}$ and x;
- a sentence φ whose meaning is that $\psi(x)$ holds for finitely many x's only, and the maximum x such that $\psi(x)$ holds is even.

Formally,

$$\begin{array}{rcl}\psi(x) & \equiv & \forall y((x < y + y \wedge y < x) \rightarrow P(y)) \quad \text{and} \\ \varphi & \equiv & \exists x(\psi(x + x) \wedge \forall y(\psi(y) \rightarrow y \leq x + x)).\end{array}$$

Note that for all possible worlds $\mathfrak{M}$, the reduct of $\mathfrak{M}$ to $\{\overline{0}, s, +, <\}$ is isomorphic to $\mathbb{N}$ with the standard interpretation of $\overline{0}$, s, $+$ and $<$. Also note that for all $\mathfrak{M}, \mathfrak{N} \in \mathcal{W}$ that agree on all members of $\mathcal{D}$, $\mathfrak{N} \models \varphi$ iff $\mathfrak{M} \models \varphi$. Let a computable classifier f be defined as follows. Let a member σ of $(\mathcal{D} \cup \{\sharp\})^\star$ be given. Let $n \in \mathbb{N}$ be maximal such that for all $m < n$, either $P(\overline{m})$ or $\neg P(\overline{m})$ occurs in σ and all models of $\mathrm{cnt}(\sigma) \cap \{P(\overline{m}), \neg P(\overline{m}) : m < n\}$ in $\mathcal{W}$ are models of $\psi(\overline{n})$. Put $f(\sigma) = 1$ if n is even, and put $f(\sigma) = 0$ otherwise. Let X be the set of all $\mathfrak{M} \in \mathcal{W}$ for which there exist infinitely many $n \in \mathbb{N}$ with $\mathfrak{M} \models \psi(\overline{n})$. It is immediately verified that

- $\mu(X) = 0$;

– for all $\mathfrak{M} \in \mathrm{Mod}_{\mathcal{W}}(\varphi)$ and environments e for $\mathfrak{M}$, f outputs 1 in response to cofinitely many finite initial segments of e;
– for all $\mathfrak{M} \in \mathrm{Mod}_{\mathcal{W}}(\neg\varphi) \setminus X$ and environments e for $\mathfrak{M}$, f outputs 0 in response to cofinitely many finite initial segments of e.

This shows that $\mathcal{W} \setminus X$ is computably classifiable in the limit following φ.

Let a classifier g be given. Suppose for a contradiction that g classifies $\mathcal{W}$ in the limit following some subset Y of $\mathcal{W}$ with $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi) \triangle Y) = 0$. Note that for all $\sigma \in \mathcal{D}^\star$ that are consistent in $\mathcal{W}$, neither $\mu(\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma) \cup \{\varphi\}))$ nor $\mu(\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma) \cup \{\neg\varphi\}))$ is equal to 0, hence that g has to output 1 in response to some extension σ_1 of σ in $\mathcal{D}^\star$, and 0 in response to another extension σ_2 of σ in $\mathcal{D}^\star$, with both $\mathrm{cnt}(\sigma_1)$ and $\mathrm{cnt}(\sigma_2)$ being consistent in $\mathcal{W}$. Using this observation, it is easy to construct an environment e for a member of X such that $g(\sigma)$ is equal to 1 for infinitely many finite initial segments σ of e and equal to 0 for infinitely many finite initial segments σ of e. Contradiction. ▮

Considering only computable classification, as opposed to noncomputable classification, a similar result to Proposition 13 can be established using Peano arithmetics instead of Presburger arithmetics:

Proposition 14. *Suppose that $\mathcal{V}$ is enriched with both $+$ and $*$, $\mathcal{L}$ is the set of first-order sentences and $\mathcal{D} = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$. Then there exists a sentence φ with the following properties.*

– *There exists a subset $\mathcal{W}'$ of $\mathcal{W}$ with $\mu(\mathcal{W}') = 1$ that is computably classifiable in the limit following φ.*
– *$\mathcal{W}$ is not computably classifiable in the limit following φ almost everywhere.*
– *$\mathcal{W}$ is classifiable in the limit following φ.*

Proof. Let $(\phi_i)_{i \in \mathbb{N}}$ denote an acceptable indexing of the unary partial recursive functions from $\mathbb{N}$ into $\{0,1\}$. Given $i, x \in \mathbb{N}$, let $\chi(\overline{i}, \overline{x})$ be a formula which expresses that $\phi_i(x)$ is undefined, and let $\xi(\overline{i}, \overline{x})$ be a formula which expresses that $\phi_i(x)$ is defined and equal to 0. Consider the sentence φ defined as

$$\exists i \exists x [i < x \quad \wedge \quad P(x) \quad \wedge \quad P(i) \quad \wedge \quad \forall y (y < i \rightarrow \neg P(y)) \quad \wedge \\ (\chi(i,x) \vee (\xi(i,x) \leftrightarrow P(x))) \quad \wedge \quad \forall y (y < x \rightarrow \neg(\chi(i,y) \vee (\xi(i,y) \leftrightarrow P(y))))].$$

So φ states the existence of $i, x \in \mathbb{N}$ with $i < x$ such that i is the least $n \in \mathbb{N}$ such that $P(\overline{n})$ holds, $P(\overline{x})$ holds, and x is the least $n \in \mathbb{N}$ such that either $\phi_i(n)$ is undefined or $\phi_i(n)$ is defined but disagrees with the interpretation of $P(\overline{n})$. Let a computable classifier f have the following properties. If $\mathrm{cnt}(\sigma)$ contains no formula of the form $P(\overline{n})$ then $f(\sigma) = 0$. Otherwise let i be the least number such that $P(\overline{i}) \in \mathrm{cnt}(\sigma)$. Let x be the least number such that either $x \geq |\sigma|$ or $\phi_i(x)$ does not converge in $|\sigma|$ steps or $\phi_i(x)$ converges in $|\sigma|$ steps to a value different from $P(\overline{x})$. Output 1 if $P(\overline{x}) \in \mathrm{cnt}(\sigma)$ and output 0 otherwise. Let $\mathfrak{M} \in \mathcal{W}$ be such that there exists a least $i \in \mathbb{N}$ such that $\mathfrak{M} \models P(\overline{i})$. Let e be an environment for $\mathfrak{M}$. If ϕ_i is not total, or if ϕ_i is total and the interpretation of P in $\mathfrak{M}$ disagrees with ϕ_i, then f converges on e to 1 if $\mathfrak{M} \models \varphi$

and to 0 otherwise. This proves that a subset of $\mathcal{W}'$ of measure 1 is computably classifiable in the limit following φ.

Now, suppose for a contradiction that a (partial) computable function f classifies $\mathcal{W}$ following φ almost everywhere. Given a member σ of $\{0,1\}^\star$, let $\widehat{\sigma}$ be the sequence obtained from σ by replacing $\sigma(n)$ by $P(\overline{n})$ if $\sigma(n) = 1$, and by $\neg P(\overline{n})$ otherwise, for all natural numbers n smaller than the length of σ. Then, by the recursion theorem there is an i such that ϕ_i is the function whose graph is constructed by the following finite extension method: $\sigma_0 = 0^i 1$ and σ_{j+1} is the first extension of σ_j found such that $f(\widehat{\sigma_{j+1}}) \neq f(\widehat{\sigma_j})$. As f converges on all environments, there exists $j \in \mathbb{N}$ such that σ_{j+1} is undefined, implying that $\phi_i = \bigcup_{k \leq j} \sigma_k = \sigma_j$. Let n be the first number where $\sigma_j(n)$ and thus $\phi_i(n)$ is undefined. Then f has to incorrectly converge on all environments for any possible world $\mathfrak{M}$ in which the interpretation of $P(\overline{m})$ is fixed for all $m < n$ and $\mathfrak{M} \models P(\overline{n})$, or on all environments for any possible world $\mathfrak{M}$ in which the interpretation of $P(\overline{m})$ is fixed for all $m < n$ and $\mathfrak{M} \models \neg P(\overline{n})$. This shows that f does not classify $\mathcal{W}$ in the limit following φ almost everywhere.

Finally let a noncomputable classifier f have the following properties. If $\mathrm{cnt}(\sigma)$ contains no formula of the form $P(\overline{n})$ then $f(\sigma) = 0$. Otherwise let i be the least number such that $P(\overline{i})$ belongs to $\mathrm{cnt}(\sigma)$. If there exists $x > i$ such that

- $P(\overline{x}) \in \mathrm{cnt}(\sigma)$,
- for all $n < x$, either $P(\overline{n})$ or $\neg P(\overline{n})$ occurs in σ,
- x is the least $n \in \mathbb{N}$ such that either $\phi_i(n)$ is undefined or $\phi_i(n)$ is defined but disagrees with which of $P(\overline{n})$ or $\neg P(\overline{n})$ occurs in $\mathrm{cnt}(\sigma)$,

then $f(\sigma) = 1$; otherwise $f(\sigma) = 0$. Obviously, f classifies $\mathcal{W}$ in the limit following φ. ∎

5 Definability versus nondefinability of misclassified sets

The next fundamental result shows that a classifier which uses φ to partition the set of possible worlds might have to misclassify a subset of $\mathcal{W}$ that is not only of measure 0, but also necessarily not definable. Hence almost correct classification is not equivalent to absolute classification with respect to a partition of the possible worlds given by another formula than φ (that formula would of course be almost equivalent to φ). It is worth noting that the next proposition uses a set of possible data that is neither $\{P(\overline{n}) : n \in \mathbb{N}\}$ nor $\{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$, henceforth exploiting the generality and flexibility allowed by the parameter $\mathcal{D}$.

Proposition 15. *Assume that $\mathcal{V}$ is standard and $\mathcal{L}$ is the set of first-order sentences. For some choice of $\mathcal{D}$, there exists $\varphi \in \mathcal{L}$ such that:*

- *$\mathcal{W}$ is computably classifiable in the limit following φ almost everywhere;*
- *$\mathcal{W}$ is not classifiable in the limit following any sentence that is almost equivalent to φ.*

Proof. Let φ be a sentence which expresses that the characteristic function of P is lexicographically greater than the characteristic function of $2\mathbb{N} = \{0, 2, 4, \ldots\}$. Formally,

$$\varphi \;\equiv\; P(0) \wedge \exists x (P(x) \wedge P(s(x)) \wedge \forall y < x\, (P(y) \leftrightarrow \neg P(s(y)))).$$

Let B be the subset of $\mathbb{N}$ whose characteristic function can be represented as the concatenation of all strings of even length, in lexicographic order and in increasing length. Hence the ω-sequence

$$00\,01\,10\,11\,0000\,0001\,0010\,\ldots\,1111\,000000\,000001\,000010\ldots$$

represents the characteristic function of B. Put $\mathcal{D} = \{P(\overline{n}) : n \in B\} \cup \{\neg P(\overline{n}) : n \notin B\}$. Note that B is not definable by a member of $\mathcal{L}$ (this property is key to the proof of the proposition). Another essential property of B used in the proof is that

(†) for every $\tau \in \{0,1\}^\star$ there are infinitely many even numbers x such that for all $y < \mathrm{lt}(\tau)$, $\tau(y) = B(x+y)$.

We first define a computable classifier f which classifies $\mathcal{W}$ following φ almost everywhere; in other words, f converges on all environments for all members of $\mathcal{W}$, but f's conjectures can be false in the limit on some environments for a set of possible worlds of measure 0. The classifier f outputs 0 until it is presented with a datum of the form $P(\overline{n})$ for an odd $n \in \mathbb{N}$, or of the form $\neg P(\overline{n})$ for an even $n \in \mathbb{N}$. Then f takes this n as a parameter and computes from now on for any stage s the set R_s consisting of all $m \in \{0,1,\ldots,n\} \cap B$ such that $P(\overline{m})$ has appeared in the first s elements of the input, and all members m of $\{0,1,\ldots,n\} \setminus B$ such that $\neg P(\overline{m})$ has not appeared in the first s elements of the input. Note that when s is large enough, the restriction of the characteristic function of R_s to $\{0,1,\ldots,n\}$ is equal to the restriction of the characteristic function of the interpretation of P in the possible world one of whose environments is fed to f. Let f conjecture in the limit the value 0 if the characteristic function of R_s is lexicographically smaller than the characteristic function of $2\mathbb{N}$, and 1 if the characteristic function of $2\mathbb{N}$ is lexicographically smaller than the characteristic function of R_s. Clearly, f has the desired properties.

Now let ψ be a member of $\mathcal{L}$ (that is, a first-order sentence over $\mathcal{V}$) and assume for a contradiction that a classifier g classifies $\mathcal{W}$ in the limit following ψ. Let $\mathfrak{M}$ be the (unique) possible world such that the characteristic function of the interpretation of P in $\mathfrak{M}$ is isomorphic to the characteristic function of $2\mathbb{N}$. Then there is a finite initial segment σ of $\mathcal{D}^\star$ such that $\mathfrak{M}$ is a model of $\mathrm{cnt}(\sigma)$ and for all $\tau \in \mathcal{D}^\star$, if τ extends σ and $\mathfrak{M} \models \mathrm{cnt}(\tau)$ then $g(\tau)$ is defined and equal to $g(\sigma)$. Since ψ contains only finitely many occurrences of $\overline{0}$, s and variables, there exists $k \in \mathbb{N}$ such that:

- for all $n \in \mathbb{N}$, if either $P(\overline{n})$ or $\neg P(\overline{n})$ occurs in σ then $n < 2k$;
- for all members $\tau_0, \tau_1, \ldots$ of $\{0,1\}^\star$ of even length and for all $\mathfrak{N}, \mathfrak{N}' \in \mathcal{W}$, if the characteristic functions of P in $\mathfrak{N}$ and $\mathfrak{N}'$ are both of the form $(01)^k(01)^*\tau_0(01)^k(01)^*\tau_1(01)^k(01)^*\tau_2\ldots$ then $\mathfrak{N} \models \psi$ iff $\mathfrak{N}' \models \psi$.

It is known from the theory of randomness that the characteristic function of any random set coincides with the characteristic function of $2\mathbb{N}$ on infinitely many even places for at least $2k$ consecutive bits. Furthermore, the measure of all random sets starting with $(01)^k$ equals 2^{-2k}. Fix a random set R whose characteristic function extends $(01)^k$. Thus we can choose some members $\tau_0, \tau_1, \ldots$ of $\{0,1\}^\star$ of even length such that the characteristic function of R is of the form

$$(01)^k\tau_0(01)^k\tau_1(01)^k\tau_2\ldots$$

Using (†) above, there exist $a_0, a_1, a_2, \ldots \in \mathbb{N}$ such that the set R' whose characteristic function is $(01)^{k+a_0}\tau_0(01)^{k+a_1}\tau_1(01)^{k+a_2}\tau_2\ldots$ satisfies the following property. Given $n \in \mathbb{N}$, put

$$x_n = \sum_{i \leq n}(2k + a_i) + \sum_{i<n} \mathrm{lt}(\tau_i).$$

For all $n \in \mathbb{N}$ and $y < \mathrm{lt}(\tau_n)$, if y is odd and $\tau_n(y) = 1$ then $B(x_n + y) = 0$, whereas if y is even and $\tau_n(y) = 0$ then $B(x_n + y) = 1$. Let $\mathfrak{N}$, respectively $\mathfrak{N}'$, be the (unique) possible world such that the characteristic function of the interpretation of P in $\mathfrak{N}$, respectively $\mathfrak{N}'$, is isomorphic to the characteristic function of R, respectively R'. Note that all members of $\mathcal{D}$ that are true in $\mathfrak{N}'$ are also true in $\mathfrak{M}$. Moreover, σ is a finite initial segment of some environment for $\mathfrak{N}'$. As a consequence, g converges in the limit to $g(\sigma)$ on all environments for $\mathfrak{N}'$ that extend σ. But since both $\mathfrak{N}$ and $\mathfrak{N}'$ agree on ψ and g is assumed to classify $\mathcal{W}$ in the limit following ψ, g converges in the limit to $g(\sigma)$ on all environments for $\mathfrak{N}$ that extend σ. It follows that all random sequences extending $(01)^k$ are classified as $g(\sigma)$, and the measure of all extensions of $(01)^k$ which are classified as $g(\sigma)$ is 2^{-2k}. On the other hand, those extensions of $(01)^k$ which are identified with models of φ in $\mathcal{W}$ have measure $2^{-2k}/3$. Therefore φ and ψ differ on a class of possible worlds of positive measure. Contradiction. ■

Proposition 15 cannot be generalized to arbitrary paradigms. This can be proved by using results from automata theory. Recall that a subset of $\{0,1\}^\omega$ (or *ω-language*) is *Büchi recognizable* if there exists a finite nondeterministic automaton A with a set F of accepting states such that for all $w \in \{0,1\}^\omega$, e belongs to S iff there exists a run of A on w that goes infinitely often through an accepting state. The next result plays a crucial role in the proof of Proposition 19 below.

Lemma 16. [4, Theorem 3.1] *$S \subseteq \{0,1\}^\omega$ is Büchi recognizable iff it is the disjoint union of:*

- *a sparse set (of measure 0);*
- *finitely many sets of the form $R \star Y$ where R is a prefix free regular language and Y is an ω-language of measure 0;*
- *finitely many sets $R_0 \star Y_0, \ldots, R_n \star Y_n$ where for all $i \leq n$, R_i is a regular language and Y_i is an ω-language of measure 1.*

Corollary 17. *Let a subset S of $\{0,1\}^\omega$ be Büchi recognizable. Then there are regular and prefix free subsets A and B of $\{0,1\}^\star$ such that:*

- $\mu(A \star \{0,1\}^\omega \cup B \star \{0,1\}^\omega) = 1$*;*
- $A \star \{0,1\}^\omega$ *and* $B \star \{0,1\}^\omega$ *are disjoint;*
- $\mu(A \star \{0,1\}^\omega \bigtriangleup S) = 0$.

Definition 18. Let a subset S of $\{0,1\}^\omega$ be Büchi recognizable. Choose regular and prefix free $A, B \subseteq \{0,1\}^\star$ that satisfy the three conditions expressed in Corollary 17, and let R_S^+ denote $A \star \{0,1\}^\omega$ and R_S^- denote $B \star \{0,1\}^\omega$.

Proposition 19. *Suppose that $\mathcal{V}$ is standard, $\mathcal{L}$ is the set of monadic second-order sentences and $\mathcal{D}$ equals $\{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$. For all $\varphi \in \mathcal{L}$, $\mathcal{W}$ is computably classifiable in the limit following some sentence that is almost equivalent to φ.*

Proof. Let a sentence φ be given. Let S be the set of standard informants for the models of φ in $\mathcal{W}$. By the choice of $\mathcal{L}$, S is Büchi recognizable. Since R_S^+ is of the form $C \star \{0,1\}^\omega$ for a prefix-free subset C of $\{0,1\}^\star$, it is immediately verified that there exists a computable classifier f such that for all standard informants (identified with environments) for some possible world, if $e \in R_S^+$ then f outputs 1 in response to cofinitely many finite initial segments of e, and if $e \notin R_S^+$ then f outputs 0 in response to cofinitely many finite initial segments of e. Let W be the set of all possible worlds whose standard informants belong to R_S^+. Using [4, Theorem 3.1] again, we infer that W is the set of models of some member ψ of $\mathcal{L}$. Moreover, it follows immediately from the definition of R_S^+ that $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi) \triangle W) = 0$, hence ψ is almost equivalent to φ. Since $\mathcal{W}$ is classifiable in the limit following ψ, we are done. ∎

Corollary 17 has other applications. For instance, considering classification from positive data only, and using a different language and set of possible worlds than in Proposition 13, one can strengthen the first statement of Proposition 13:

Proposition 20. *Suppose that $\mathcal{V}$ is standard, $\mathcal{L}$ is the set of monadic second-order sentences, and $\mathcal{D} = \{P(\overline{n}) : n \in \mathbb{N}\}$.*

1. *There exists a subset $\mathcal{W}'$ of $\mathcal{W}$ with $\mu(\mathcal{W}') = 1$ such that for all sentences ψ, $\mathcal{W}'$ is computably classifiable in the limit following ψ.*
2. *There is a sentence φ such that $\mathcal{W}$ is not classifiable almost everywhere in the limit following φ.*

Proof. Let a sentence ψ be given. Let S be the set of standard informants for the models of ψ in $\mathcal{W}$. Recall the definition of R_S^+ and R_S^- from Definition 18. Since $R_S^+ \cup R_S^-$ is of the form $C \star \{0,1\}^\omega$ for a prefix-free subset C of $\{0,1\}^\star$, there exists a computable $\{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$-classifier such that for all standard informants e for some possible world, if $e \in R_S^+$ then f outputs 1 in response to cofinitely many finite initial segments of e, and if $e \in R_S^-$ then f outputs 0 in response to cofinitely many finite initial segments of e. Let W^+, respectively, W^-, be the set of all possible worlds $\mathfrak{M}$ whose standard informants belong to R_S^+, respectively, R_S^-. It follows immediately from the definitions of R_S^+ and R_S^- that both $\mu(\mathrm{Mod}_{\mathcal{W}}(\psi) \triangle W^+)$ and $\mu(\mathrm{Mod}_{\mathcal{W}}(\neg\psi) \triangle W^-)$ are null. Hence $\mu(W^+ \cup W^-) = 1$ and $W^+ \cup W^-$ is classifiable in the limit following ψ almost everywhere. Since the number of sentences is countable, part *1.* of the proposition follows immediately.

For part *2.*, define φ as $\exists x \forall y (y < s(s(x)) \rightarrow (P(y) \leftrightarrow x \neq y))$. Suppose for a contradiction that a classifier f classifies $\mathcal{W}$ almost everywhere in the limit following φ. Since f converges (possibly to 1) on all environments for the possible world whose $\mathcal{D}$-diagram is $\{P(\overline{n}) : n \in \mathbb{N}\}$, we can choose a member σ of $\mathcal{D}^\star$ such that for all $\tau \in \mathcal{D}^\star$ that extend σ, $f(\tau) = f(\sigma)$. Let $a \in \mathbb{N}$ be greater than all members of $\mathrm{cnt}(\sigma)$. Put

$$\begin{array}{rcl} W_1 & = & \mathrm{Mod}_{\mathcal{W}}(P(\overline{0}) \wedge \ldots \wedge P(\overline{a}) \wedge \neg P(\overline{a+1}) \wedge P(\overline{a+2})) \quad \text{and} \\ W_2 & = & \mathrm{Mod}_{\mathcal{W}}(P(\overline{0}) \wedge \ldots \wedge P(\overline{a}) \wedge \neg P(\overline{a+1}) \wedge \neg P(\overline{a+2})). \end{array}$$

Note that $W_1 \subseteq \mathrm{Mod}_{\mathcal{W}}(\varphi)$ and $W_2 \subseteq \mathrm{Mod}_{\mathcal{W}}(\neg\varphi)$, and that both $\mu(W_1)$ and $\mu(W_2)$ are nonnull. But by the choice of σ, for all $\mathfrak{M} \in W_1 \cup W_2$ and for all environments e for $\mathfrak{M}$ that extend σ, f outputs $f(\sigma)$ in response to every finite initial segment of e that extends σ. Contradiction. ∎

6 Positive only versus positive and negative data

The next result shows that being allowed to misclassify a set of possible worlds of measure 0 when classifying from positive data does not always make up for non-access to negative data.

Proposition 21. *Suppose that $\mathcal{V}$ is enriched with $+$ only and $\mathcal{L}$ is the set of first-order sentences. Then there exists $\varphi \in \mathcal{L}$ with the following properties.*

- *If $\mathcal{D} = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$ then $\mathcal{W}$ is computably classifiable in the limit following φ.*
- *If $\mathcal{D} = \{P(\overline{n}) : n \in \mathbb{N}\}$ then $\mathcal{W}$ is not classifiable in the limit following φ almost everywhere.*

Proof. Define φ as $\exists x \forall y((x < y \wedge y < x + x + 3) \rightarrow \neg P(y))$. It is immediately verified that if $\mathcal{D} = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$ then $\mathcal{W}$ is computably classifiable in the limit following φ (with at most one mind change).

Now suppose that $\mathcal{D} = \{P(\overline{n}) : n \in \mathbb{N}\}$. Trivially, $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi)) > 0$. Moreover,

$$\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi)) \leq \Sigma_{n \in \mathbb{N}} 2^{-n-2} = 2^{-1}.$$

For a contradiction, let a classifier f be such that f classifies $\mathcal{W}$ in the limit following φ almost everywhere. For all $\sigma \in \mathcal{D}^\star$, define U_σ as the set of all $\mathfrak{M} \in \mathrm{Mod}_{\mathcal{W}}(\neg\varphi)$ such that for all $\tau \in \mathcal{D}^\star$, if τ extends σ and $\mathfrak{M} \models \mathrm{cnt}(\tau)$ then $f(\tau) = 0$. Choose $\sigma \in \mathcal{D}^\star$ with $\mu(U_\sigma) > 0$. Denote by a the maximal number in $\mathrm{cnt}(\sigma)$. Let U be the set of all $\mathfrak{M} \in \mathcal{W}$ with:

- $\mathfrak{M} \models \neg P(\overline{a+1}) \wedge \ldots \wedge \neg P(\overline{a+a+2})$;
- the $\mathcal{D}$-diagram of $\mathfrak{M}$ agrees with the $\mathcal{D}$-diagram of some member of U_σ, except perhaps on $\{P(\overline{a+1}), \ldots, P(\overline{a+a+2})\}$.

Note that all members of $\mathrm{Mod}_{\mathcal{W}}(U)$ are models of φ. Since $\mu(\mathrm{Mod}_{\mathcal{W}}(U))$ is at least equal to $2^{-a-2}\mu(U_\sigma)$, $\mu(\mathrm{Mod}_{\mathcal{W}}(U))$ is nonnull. Let $\mathfrak{M} \in U$ be given. Since the $\mathcal{D}$-diagram of $\mathfrak{M}$ is included in the $\mathcal{D}$-diagram of some member of U_σ, we infer that for all environments e for the $\mathcal{D}$-diagram of $\mathfrak{M}$ that extend σ, f outputs 0 in response to all finite initial segments of e that extend σ. Contradiction. ∎

Considering failing to classify a set of first category rather than misclassifying a set of measure 0, one can contrast Proposition 21 with the following one.

Proposition 22. *Let $\varphi \in \mathcal{L}$ be such that if $\mathcal{D} = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$ then $\mathcal{W}$ is classifiable, respectively, computably classifiable, in the limit following φ. Assume that $\mathcal{D} = \{P(\overline{n}) : n \in \mathbb{N}\}$. Then there exists a subset $\mathcal{W}'$ of $\mathcal{W}$ such that:*

- *$\mathcal{W}'$ is of second category;*
- *$\mathcal{W}'$ is classifiable, respectively, computably classifiable, in the limit following φ.*

Proof. Put $D = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$. Consider a D-classifier f such that if $\mathcal{D} = D$ then f classifies $\mathcal{W}$ in the limit following φ. Without loss of generality we can assume that f is total and f depends only on the content of the input and not on the order and number of repetitions of symbols. Let S be the set of all members σ of $D^\star$ such that $f(\tau) = f(\sigma)$ for all consistent $\tau \in D^\star$ that extend σ. Note that if f is computable then S is co-r.e. Let $\mathcal{W}'$ be the set of possible worlds that are models of $\mathrm{cnt}(\sigma)$ for some $\sigma \in S$. By the choice of f, for all $\sigma \in D^\star$, some member of S extends σ. This implies immediately that $\mathcal{W} \setminus \mathcal{W}'$ is of first category. Fix an enumeration $(\tau_i)_{i\in\mathbb{N}}$ of $D^\star \setminus S$ that in case f is computable, is itself computable. Assume that $\mathcal{D} = \{P(\overline{n}) : n \in \mathbb{N}\}$. Define a classifier g as follows. Let $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ be given, and let p denote the length of σ. Then $g(\sigma)$ is equal to $f(\tau)$ for the shortest $\tau \in D^\star \setminus \{\tau_0, \dots, \tau_p\}$ such that for all $n \leq p$, $P(\overline{n})$ occurs in τ iff $P(\overline{n})$ occurs in σ. Obviously, g is computable if f is computable. Moreover, for all $\mathfrak{M} \in \mathcal{W}'$ and environments e for $\mathfrak{M}$ (with respect to the choice of $\mathcal{D}$), some member τ of S satisfies $\{n \in \mathbb{N} : P(\overline{n}) \in \mathrm{cnt}(\tau)\} \subseteq \{n \in \mathbb{N} : P(\overline{n}) \in \mathrm{cnt}(e)\}$ and $\{n \in \mathbb{N} : \neg P(\overline{n}) \in \mathrm{cnt}(\tau)\} \cap \{n \in \mathbb{N} : P(\overline{n}) \in \mathrm{cnt}(e)\} = \emptyset$; hence g converges on e to $f(\tau)$. This shows that g classifies $\mathcal{W}'$ in the limit following φ. ∎

7 Alternative measures

The aim of this section is to illustrate that results obtained from the canonical measure on the Cantor space do not generalize to arbitrary measures. To this aim, we assume that μ is the measure defined in [15], used to define the complexity of a string as the length of the minimal program that generates it. It is an important notion (see [18]), hence properties of this particular μ are interesting in their own right, not only in contrast to the measure on the Cantor space. The key properties of μ are that:[5]

- μ is a K-recursive function from $\{0,1\}^\star$ into the set of rational numbers;
- for all recursive members x of $\{0,1\}^\omega$, $\mu(x) > 0$.

The decidability property of monadic second-order logic immediately yields the following property.

Property 23. *Assume that $\mathcal{V}$ is standard and $\mathcal{L}$ is the set of monadic second-order sentences. For all $\varphi \in \mathcal{L}$, if φ has a model in $\mathcal{W}$ then $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi)) > 0$.*

This allows to contrast Solomonoff's measure with the measure on the Cantor space:

Corollary 24. *For all sentences φ, if $\mathcal{W}$ is classifiable in the limit following a sentence that is almost everywhere equivalent to φ, then $\mathcal{W}$ is classifiable in the limit following φ.*

Proof. By Property 23, for all sentences φ and ψ, φ and ψ are either logically equivalent or $\mu(\mathrm{Mod}_{\mathcal{W}}(\varphi \leftrightarrow \neg\psi)) > 0$. The corollary follows immediately. ∎

Another difference between both measures is given by the next proposition, whose proof relies on another acceptance criterion by deterministic Rabin automata. The nondeterministic Büchi automata used in the previous sections and the deterministic Rabin automata of the current

[5] It can be shown that for every recursive measure μ', there is a recursive member x of $\{0,1\}^\omega$ such that $\mu'(x) = 0$. Hence the Solomonoff measure is not recursive.

section actually accept the same class of languages, but for the purpose of the next result, Rabin automata offer a better tool. Recall that a Rabin automaton A is a deterministic finite automaton that replaces the set of accepting states by a set X of pairs whose members are sets of states. A subset S of $\{0,1\}^\omega$ is then accepted by A iff for all $e \in \{0,1\}^\omega$, e belongs to S iff there exists $(I,F) \in X$ such that the unique run of A on e goes infinitely often through each member of I, and finitely often through the members of F.

Proposition 25. *Assume that $\mathcal{V}$ is standard and $\mathcal{L}$ is the set of monadic second-order sentences. Let $\varphi \in \mathcal{L}$ be given. Then one of the following statements holds.*

1. *There exists $n \in \mathbb{N}$ such that $\mathcal{W}$ is classifiable in the limit with at most n mind changes following ϕ or*
2. *there exists no $\mathcal{W}' \subseteq \mathcal{W}$ with $\mu(\mathcal{W}') = 1$ such that $\mathcal{W}'$ is computably classifiable in the limit following φ.*

Proof. By the choice of $\mathcal{L}$, there is a finite deterministic automaton A such that the set of models of φ in $\mathcal{W}$ is identified with a subset S of $\{0,1\}^\omega$ such that for all $e \in \{0,1\}^\omega$, $e \in S$ iff e is accepted by A. Let Q be the set of its reachable states. We distinguish between two cases, that correspond to the two alternatives in the statement of the proposition, and that we examine in turn.

Case *1.* For all $q \in Q$ and $x, y \in \{0,1\}^\omega$ such that A goes through q infinitely often on the run of A on x and on the run of A on y, A accepts x iff A accepts y. For all $q \in Q$, put $Acc(q) = 1$ if there exists $x \in \{0,1\}^\omega$ such that A accepts x and A goes through q infinitely often on the run of A on x; otherwise put $Acc(q) = 0$. Let f be the unique classifier such that for all $\sigma \in \mathcal{D}^\star$, $f(\sigma) = Acc(q)$ for the state q reached by A when run on (the member of $\{0,1\}^\star$ identified with) σ. For all members q, q' of the same strongly connected component of A, $Acc(q) = Acc(q')$. Since Q is finite, the number of strongly connected components of A has to be an upper bound on the number of mind changes that f can make when classifying $\mathcal{W}$ in the limit following any sentence. Furthermore, for all $x \in \{0,1\}^\omega$, there exists a finite initial segment σ of x such that A remains in the same strong connected component C after σ has been processed; obviously, there has to be a member q of C such that A goes through q infinitely often when run on x, and f accepts x iff $Acc(q) = 1$, which itself is equivalent to A accepting x. This shows that f correctly classifies $\mathcal{W}$ in the limit following φ.

Case *2.* There exists $q \in Q$ and $x, y \in \{0,1\}^\omega$ such that A goes through q infinitely often on the run of A on x and on the run of A on y, A accepts x, and A rejects y. Notice that for all $x \in \{0,1\}^\omega$, whether A accepts or rejects x only depends on the set of states that are visited infinitely often when A is run on x. It follows that there exist $\eta, \tau \in \{0,1\}^\star$ such that for all strings $\sigma \in \{0,1\}^\star$:

- A is in state q after it has processed any of σ, $\sigma\eta$, and $\sigma\tau$;
- A accepts $\sigma\eta^\omega$ and rejects $\sigma\tau^\omega$.

Let a recursive classifier f be given. Then there exists an infinite sequence $x = \sigma_0\sigma_1\ldots$ such that A is in state q after σ_0 has been processed and

$$\sigma_{n+1} = \begin{cases} \eta & \text{if } f(\sigma_0\sigma_1\ldots\sigma_n) = 0; \\ \tau & \text{if } f(\sigma_0\sigma_1\ldots\sigma_n) = 1. \end{cases}$$

Since f is computable, x is recursive, so $\mu(\{x\}) > 0$. Moreover, it is easy to verify that:

- if f converges on x to 0 then $\sigma_n = \eta$ for cofinitely many n's;
- if f converges on x to 1 then $\sigma_n = \tau$ for cofinitely many n's.

This shows that there exists a subset $\mathcal{W}'$ of $\mathcal{W}$ with $\mu(\mathcal{W}') > 0$ such that f fails to classifies $\mathcal{W}'$ in the limit following φ. ∎

8 Learning witnesses

In this section, we go back to the standard measure on the Cantor space, and focus on classifiability of existential sentences. By the choice of $\mathcal{W}$, such a sentence, of the form $\exists x\psi(x)$, is true in a member $\mathfrak{M}$ of $\mathcal{W}$ iff $\psi(t)$ is true for some closed term t. It is then natural not only to determine that $\exists x\psi(x)$ is true, but also to learn in the limit such a witness t. This amounts to a computation in the limit that generalizes the kind of computations done by Prolog systems. We now formalize the concepts that have just been introduced.

Definition 26. A *learner* is a partial function from $(\mathcal{D} \cup \{\sharp\})^\star$ into the union of $\{0\}$ with the set of closed terms.

Definition 27. We say that a classifier g is *associated with* a learner f iff for all $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$, $g(\sigma)$ is defined iff $f(\sigma)$ is defined and $g(\sigma) = 0$ iff $f(\sigma) = 0$.

Definition 28. Let a learner f, a sentence of the form $\exists x\psi(x)$, and a subset $\mathcal{W}'$ of $\mathcal{W}$ be given.

We say that f *learns* $\exists x\psi(x)$ *in the limit in* $\mathcal{W}'$ just in case for all $\mathfrak{M} \in \mathcal{W}'$ and environments e for $\mathfrak{M}$, the following holds.

- If $\mathfrak{M} \models \exists x\psi(x)$ then there exists a closed term t such that $\mathfrak{M} \models \psi(t)$ and the set of all $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ such that $\sigma \subset e$ and $f(\sigma) = t$ is cofinite.
- If $\mathfrak{M} \not\models \exists x\psi(x)$ then $\{\sigma \in (\mathcal{D} \cup \{\sharp\})^\star : \sigma \subset e \text{ and } f(\sigma) = 0\}$ is cofinite.

Definition 29. Let a learner f and an existential sentence φ be given. We say that f *learns* φ *in the limit in* $\mathcal{W}$ *almost correctly* just in case:

- the classifier associated with f classifies $\mathcal{W}$ in the limit following φ almost everywhere;
- there exists $\mathcal{W}' \subseteq \mathcal{W}$ with $\mu(\mathcal{W}') = 1$ such that f learns φ in the limit in $\mathcal{W}'$.

Definition 30. We say that an existential sentence φ is *learnable*, respectively, *computably learnable*, *in the limit in* $\mathcal{W}$ *almost correctly* iff some learner, respectively, computable learner, learns φ in the limit in $\mathcal{W}$ almost correctly.

Adapting the proof of Proposition 19, we obtain an important particular case where limiting computation of witnesses for existentially quantified sentences is always possible:

Proposition 31. *Suppose that $\mathcal{V}$ is standard, $\mathcal{L}$ is the set of monadic second-order sentences, and $\mathcal{D}$ is equal to $\{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$. Then for all existential sentences φ, φ is computably learnable in the limit in $\mathcal{W}$ almost correctly.*

Proof. Let a sentence of the form $\exists x\psi(x)$ be given. Let Z be the set of all $e \in \{0,1\}^\omega$ that are identified with a model of $\exists x\psi(x)$ in $\mathcal{W}$. Let Z_n be the set of all $e \in Z$, for which n is the minimal i such that $\psi(\overline{i})$ holds. Recall the notation of R_Z^+ and R_Z^- given in Definition 18. So by Corollary 17, R_Z^+ and R_Z^- are of respective form $C_Z^+ \star \{0,1\}^\omega$ and $C_Z^- \star \{0,1\}^\omega$ for regular prefix-free subsets C_Z^+ and C_Z^- of $\{0,1\}^\star$.

Let an extra unary predicate symbol Q be given, and consider the following monadic second-order sentence ξ over $\mathcal{V} \cup \{Q\}$:

$$\exists x[\psi(x) \ \wedge \ Q(x) \ \wedge \ \forall y[y < x \ \rightarrow [\ \neg\psi(y) \ \wedge \ \neg Q(y)]]].$$

Recall that $<$ is definable in monadic second order logic. Intuitively, ξ expresses that both ψ and Q are true on some x, and the minimal value on which they are true are the same.

Identify every standard structure $\mathfrak{M}$ over $\mathcal{V} \cup \{Q\}$, that is, every structure over $\mathcal{V} \cup \{Q\}$ all of whose individuals interpret a closed term, with the unique point e in the Cantor space such that:

- for all $n \in \mathbb{N}$, $\mathfrak{M} \models P(\overline{n})$ iff $e(2n) = 1$;
- for all $n \in \mathbb{N}$, $\mathfrak{M} \models Q(\overline{n})$ iff $e(2n+1) = 1$.

Let S be the set of all $e \in \{0,1\}^\omega$ that are identified with a standard structure over $\mathcal{V} \cup \{Q\}$ that is a model of ξ. For all $n \in \mathbb{N}$, let S_n be the set of all $e \in S$ for which n is the least $i \in \mathbb{N}$ such that $\psi(\overline{i})$ holds in e. Recall the notation of R_S^+ and R_S^- given in Definition 18. So by Corollary 17, R_S^+ and R_S^- are of respective form $C_S^+ \star \{0,1\}^\omega$ and $C_S^- \star \{0,1\}^\omega$ for regular prefix-free subsets C_S^+ and C_S^- of $\{0,1\}^\star$. Without loss of generality, we can assume that for all $\sigma \in C_S^+$, there exists $i \in \mathbb{N}$ such that $2i+1$ is smaller than the length of σ and $\sigma(2i+1) = 1$. One can thus divide C_S^+ into regular prefix free subsets $C_{S_n}^+$, where $C_{S_n}^+$ consists of those sequences σ in C_S^+ for which the minimal i such that $\sigma(2i+1) = 1$ is n. Given $n \in \mathbb{N}$, note that Z_n is same as (the set of possible worlds identified with) the set of restrictions to $\mathcal{V}$ of the members of S_n. Now let $R_{S_n}^+ = C_{S_n}^+ \star \{0,1\}^\omega$. Let $C_{Z_n}^+$ be the set of sequences of the form $\sigma(0)\sigma(2)\sigma(4)\ldots$ where σ ranges over $C_{S_n}^+$. Let $R_{Z_n}^+ = C_{Z_n}^+ \star \{0,1\}^\omega$. Using Corollary 17, $R_{S_n}^+$ differs from S_n on a set of measure 0. As the interpretation of $Q(\overline{i})$ in all members of S_n is fixed for all $i \leq n$, but arbitrary for all $i > n$, $\mu(S_n) = 2^{-(n+1)}\mu(Z_n)$. Thus, $R_{Z_n}^+$ differs from Z_n on a set of measure 0. Using regularity of $C_{S_n}^+$, it folows that $C_{Z_n}^+$, $n \in \mathbb{N}$, are regular sets (which could be made prefix free), which intersect neither with each other nor with C_Z^-. Moreover, one can find elements of the sets C_Z^- and $C_{Z_n}^+$, effectively from each n.

Now let a computable $\{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$-classifier f be defined as follows. Let a standard informant e for some possible world (that is, standard structure over $\mathcal{V}$, not $\mathcal{V} \cup \{Q\}$!) be given.

- Suppose that e extends a member σ of $C_{Z_n}^+$. Then, f outputs $\overline{n}$ in response to cofinitely many finite initial segments of e.
- If e extends a member of C_Z^- then f is constant and equal to 0 on e.

It follows that

– for all $n \in \mathbb{N}$, f converges to $\overline{n}$ on almost every (standard environment identified with a) member of $R^+_{Z_n}$;
– f converges to 0 on almost every member of R^-_Z.

Hence $\exists x\psi(x)$ is computably learnable in the limit in $\mathcal{W}$ almost correctly. ▮

When investigating the connections between classifiability and learnability of sentences of the form $\exists x\psi(x)$, it is reasonable to assume that $\mathcal{W}$ is classifiable in the limit following any of the sentences $\psi(\overline{n})$; otherwise one could obtain a separation by taking a sentence ξ such that $\mathcal{W}$ is not almost everywhere classifiable in the limit following ξ and then consider

$$\exists x\ ((x = \overline{0} \wedge \neg\xi) \vee (x = \overline{1} \wedge \xi)).$$

Furthermore, if $\mathcal{W}$ is classifiable in the limit following $\exists x\psi(x)$ as well as any of the sentences $\psi(\overline{n})$, then the parameter x is also learnable, though computability might be lost, for example with $\exists x\ (\phi_x \text{ is total } \wedge \phi_x(0) = \min\{z : P\overline{z}\} \wedge \forall y < x\ (\phi_y \neq \phi_x))$. This example can be formalized in full arithmetic. So it is natural to ask what happens if one only postulates that $\mathcal{W}$ is almost everywhere classifiable in the limit following $\exists x\psi(x)$ and the other conditions are kept as such. The next result shows that learnability is then lost.

Proposition 32. *Suppose that $\mathcal{V}$ is enriched with $+$ only, $\mathcal{L}$ is the set of first-order sentences, and $\mathcal{D} = \{P(\overline{n}), \neg P(\overline{n}) : n \in \mathbb{N}\}$. Let $\psi(x)$ be the formula*

$$\forall y \exists u \exists v \exists w\ (x + y + v = u \wedge u + w = x + x + y + y \wedge P(u)).$$

Then:

– *the set of models of $\exists x\psi(x)$ in $\mathcal{W}$ is of measure 1;*
– *for all $n \in \mathbb{N}$, $\mathcal{W}$ is classifiable in the limit with at most 1 mind change following $\psi(\overline{n})$;*
– *$\exists x\psi(x)$ is not almost correctly learnable in the limit in $\mathcal{W}$.*

Proof. The formula $\psi(x)$ expresses that for every $y \in \mathbb{N}$, there exists $u \in \mathbb{N}$ such that $x + y \leq u \leq 2x + 2y$ and $P(\overline{u})$ holds. By the law of large numbers, $\mu(\mathrm{Mod}_{\mathcal{W}}(\exists x\psi(x)))$ is equal to 1. Given $n \in \mathbb{N}$, a classifier that outputs 1, until it finds $y \in \mathbb{N}$ such that $\neg P(\overline{n + y}), \ldots, \neg P(\overline{2n + 2y})$ all appear in the data, at which point it outputs 0, classifies $\mathcal{W}$ in the limit with at most 1 mind change following $\psi(\overline{n})$. Suppose for a contradiction that a learner f learns $\exists x\psi(x)$ in the limit in $\mathcal{W}$ almost correctly. It is then easy to construct a $\subseteq$-increasing sequence $(\sigma_n)_{n\in\mathbb{N}}$ of members of $\mathcal{D}^\star$ such that:

– for all $i \in \mathbb{N}$, σ_{2i} is an initial segment of an environment for a model of $\psi(\overline{n})$, for some n, and f outputs $\overline{n}$ in response to σ_{2i};
– for all $i \in \mathbb{N}$, σ_{2i+1} is an initial segment of an environment for a model of $\neg\exists x\psi(x)$ and f outputs 0 in response to σ_{2i};
– $\bigcup_{i\in\mathbb{N}} \sigma_i$ is an environment for a possible world.

Since the classifier associated with f does not converge on $\bigcup_{i\in\mathbb{N}} \sigma_i$, it follows that f does not learn $\exists x\psi(x)$ almost correctly in the limit in $\mathcal{W}$. ▮

A further question would be whether classification can be parametrized. That is, given a sentence φ such that $\mathcal{W}$ can be classified in the limit following φ, is φ equivalent to a sentence of the form $\exists x \psi(x)$ such that $\mathcal{W}$ is classifiable with a constant number of mind changes, following any sentence $\psi(\overline{n})$? The answer is yes for computable classification in full arithmetic since one can transform every learner into a formula monitoring its behaviour and then quantify over the last time where the classifier makes a wrong conjecture.

9 Conclusion

In this paper the relationship between classifying all possible worlds and classifying almost all possible worlds has been investigated. The main results include the following. In the case of monadic second order logic with one successor, which is well known to be decidable thanks to its connections with Büchi automata, one can classify $\mathcal{W}$ following a formula φ almost everywhere from both positive and negative data. But with positive data only, convergence might fail on a set of measure 0. For other choices of possible data, $\mathcal{W}$ might even be classifiable in the limit following a given sentence φ almost everywhere, though no sentence differing from φ only on a null set of worlds allows for absolute classification. With Presburger arithmetic, there is a sentence φ such that some set of worlds of measure 1 can be classified in the limit following φ, though no classifier which is correct on a set of measure 1 converges on all environments for all possible worlds. The fact that these results depend crucially on the choice of measure as the standard measure on the Cantor space has been emphasized by considering the measure proposed by Solomonoff. Finally, some connections were made between classification and learning of an essential parameter, namely, the value of the first quantified variable in an existential sentence. It turns out that this can be achieved on a set of worlds of measure 1 in monadic second order logic with one successor, demonstrating that Prolog style inferences are almost everywhere possible in this case.

Acknowledgments. We would like to thank Wolfgang Merkle for very helpful discussions.

References

1. Andris Ambainis: *Probabilistic Inductive Inference: a Survey.* Theoretical Computer Science 264, pp. 155–167 (2001).
2. Janis Bārzdiņš, Rūsiņš Freivalds and Carl H. Smith: *Learning formulae from elementary facts.* Proceedings of the Third European Conference on Computational Learning Theory (EuroCOLT 1997), LNCS 1208, pp. 272-285 (1997).
3. Shai Ben-David: *Can Finite Samples Detect Singularities of Read-Valued Functions?* Proceedings of the 24th Annual ACM Symposium on the Theory of Computer Science, Victoria, B.C., pp. 390–399 (1992).
4. J. Richard Büchi: *On a decision method in restricted second order arithmetic.* In Proceedings of the International Congress on Logic, Methodology and Philosophy of Science, Stanford University Press, pp. 1–11 (1960)
5. Patrick Caldon and Eric Martin: *Limiting Resolution: from Foundations to Implementation.* In B. Demoen, V. Lifschitz: Proceedings of the 20th International Conference on Logic Programming, Springer-Verlag, LNCS 3132, pp. 149–164 (2004)

6. Robert P. Daley: *Inductive Inference Hierarchies: Probabilistic vs. Pluralistic.* Mathematical Methods of Specification and Synthesis of Software Systems, LNCS 215, pp. 73–82 (1985).
7. Lance Fortnow, Rūsiņš Freivalds, William Gasarch, Martin Kummer, Stuart A. Kurtz, Carl H. Smith and Frank Stephan: *On the relative sizes of learnable sets.* Theoretical Computer Science 197, pp. 139–156 (1998).
8. Rūsiņš Freivalds: *Finite Identification of General Recursive Functions by Probabilistic Strategies.* Proceedings of the Conference on Fundamentals of Computation Theory, Akademie Verlag, Berlin, pp. 138–145 (1979).
9. William Gasarch, Mark Pleszkoch, Frank Stephan and Mahendran Velauthapillai: *Classification Using Information.* Annals of Mathematics and Artificial Intelligence 23, pp. 147–168 (1998).
10. Eric Martin, Arun Sharma and Frank Stephan: *Logic, Learning, and Topology in a Common Framework.* In N. Cesa-Bianchi, M. Numao, R. Reischuk: Proceedings of the 13th International Conference on Algorithmic Learning Theory. Springer-Verlag, LNAI 2533 pp. 248-262 (2002).
11. Lea Meyer: *Probabilistic language learning under monotonicity constraints.* Theoretical Computer Science 185, pp. 81–128 (1997).
12. Leonard Pitt: *Probabilistic Inductive Inference.* Journal of the Association for Computing Machinery 36, pp. 383–333 (1989).
13. Leonard Pitt and Carl H. Smith: *Probability and plurality for aggregations of learning machines.* Information and Computation 77, pp. 77–92 (1988).
14. Hartley Rogers, Jr.: *Theory of recursive functions and effective computability.* McGraw-Hill Book Company, New York (1967).
15. Ray J. Solomonoff: *A formal theory of inductive inference, part 1 and part 2.* Information and Control 7, pp. 1–22 and pp. 224–254 (1964).
16. Carl H. Smith: *The power of pluralism for automatic program synthesis.* Journal of the Association for Computing Machinery 29, pp. 1144–1165 (1982).
17. I. Takeuti: *The measure of an omega regular language is rational.* Available at `http://www.sato.kuis.kyoto-u.ac.jp/~takeuti/art/regular.ps`
18. Ming Li and Paul Vitányi: *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd edition. Springer-Verlag (1997).
19. Andrei Voronkov: *Logic programming with bounded quantifiers.* In Proceedings of the third International Conference on Logic Programming and Automated Reasoning, Springer-Verlag, LNCS 592, pp. 486–514 (1992).