汎用的学習理論
General Learning Theory

宮部賢志 *

明治大学理工学部数学科

Kenshi Miyabe

Department of Mathematics, Meiji University

概要

Solomonoff に始まる万能推論の理論は，学習に関していくつかの示唆を与える．しかし，予測関数が計算不可能であるため，実装可能な予測アルゴリズムについて語ることは簡単ではない．ここに，汎用性という概念を導入することで，計算可能な予測関数について語れるようになる．本稿では，この枠組の可能性と展望についてまとめる．

Abstract

The theory of universal induction initiated by Solomonoff provides some insight into learning. However, because of the noncomputability of the prediction function, it is not easy to talk about implementable prediction algorithms. We introduce the notion of generality to talk about computable prediction functions. This note surveys the possibilities and prospects of this framework.

1　はじめに

データが与えられたときに，そこに含まれる規則を発見・学習し，次に来るデータの予測をしたい．どういう予測をしたら良いだろうか？予測の良さはどのように測るべきか？

これまで，様々な学習に関する理論が提案され議論されてきた．最初に，いくつかの理論の特徴を紹介し，比較する．特に独立同分布性や空間の構造を頼りに学習していることを説明する．その後，Solomonoff に始まる万能推論の理論を紹介する．万能推論の理論では，記述の複雑さを頼りに学習していることを説明する．最後に汎用性の概念を紹介する．この概念を使うことにより，万能推論の理論と他の学習理論を比較できるようになる．

* research@kenshi.miyabe.name

2 学習理論の比較

本節ではいくつかの学習理論を紹介し，その特徴を比較する．

2.1 統計的決定理論

最初に統計学を見てみよう．詳細は例えば竹村 [18, Chapter 5] などを参照せよ．統計的推測を統一的に論じる枠組みとして，Wald (1950) が導入した統計的決定理論がある．n 次元の確率変数 $X = (X_1, X_2, \ldots, X_n)$ が独立同分布である確率分布 P_θ に従うとする．その確率分布は θ でパラメータ付けられたある確率分布族に含まれていると仮定する．真の値 θ_0 は未知である．観測できるデータ X に基づき何らかの決定 $D = \delta(X)$ を行う．その決定に基づく損失 $L(\theta, D)$ を定め，その期待値をリスク関数 $R(\theta, \delta) = E_\theta[L(\theta, \delta(X))]$ と呼ぶ．この設定のもとで，リスクの小さい決定関数を求める問題を統計的決定問題という．

統計的決定理論において学習する対象はもとの確率分布 P_θ であり，それを表現するパラメータ θ である．ここでの最も大きな仮定は，想定する確率分布族を仮定することであり，この仮定により確率論の言葉で議論できるようになる．このような想定が常にできるわけではないので，想定する分布族を仮定しないノンパラメトリックな手法の研究も盛んに行われてきた．

最近発展してきた分野として高次元の統計学がある．詳細は例えば青嶋・矢田 [3] などを参照せよ．通常の統計学では，標本の次元 d よりも標本のサイズ n の方が大きいことを仮定することが多い．最近になって d の方が n よりもずっと大きい場合を考えたいことがでてきた．高次元の統計学では，次元が互いに独立である場合の振る舞いを調べることで，通常の統計学と異なる振る舞いを明らかにし，新たな統計学を構築している．

2.2 PAC 学習

次に機械学習の数学的枠組みを見てみよう．詳細は例えば Shalev-Shwartz and Ben-David [15] などを参照せよ．ここでは PAC 学習を紹介する．X で入力空間を表し，Y でラベル空間を表す．単純のため $Y = \{0, 1\}$ としよう．ある概念 $c : X \to Y$ を学習したい．c は概念空間と呼ばれる関数族 $\mathcal{C}$ に含まれているとしよう．X 上の未知の確率分布 D に基づいて入力 $S = (x_1, x_2, \ldots, x_m)$ が与えられ，そのラベル $(c(x_1), c(x_2), \ldots, c(x_m))$ も与えられる．ここでの課題は以下の汎化誤差 R を小さくする仮説 $h : X \to Y$ を選択することで

ある：

$$L_{D,c}(h) = P_{x\sim D}[h(x) \neq c(x)].$$

h は仮説空間と呼ばれる関数族 $\mathcal{H}$ から選ぶことにする．ある仮説空間 $\mathcal{H}$ が PAC 学習可能とは，次を満たす関数 $m_{\mathcal{H}} : (0,1)^2 \to \mathbb{N}$ と入力 S から予測 $h_S \in \mathcal{H}$ への関数とが存在することをいう：任意の $\epsilon, \delta \in (0,1)$ と，X 上の分布 D，概念 $c \in \mathcal{C}$ に対して，$L_{D,c}(h) = 0$ となる $h \in \mathcal{H}$ が存在するなら，$m_{\mathcal{H}}(\epsilon, \delta)$ 以上のサンプルサイズ m に対し，

$$P_{S\sim D^m}[L_{D,c}(h_S) \leq \epsilon] \geq 1 - \delta.$$

つまり，高い確率で (high probability) 良い近似 (approximately correct) が得られるのに必要な標本サイズが一様に抑えられることをいう．

PAC 学習には様々な変種があり，h_S に多項式時間計算可能性を要求したり，$m_{\mathcal{H}}$ が多項式で表現できることを要求したりすることもある．

PAC 学習において学習する対象は仮説空間 $h \in \mathcal{H}$ である．データのサンプリングとして確率分布 D が使われるが，D が学習の対象ではない．agnostic な PAC 学習ではラベルへの関数 c にもランダム性を入れて，選択する h も完全なラベル付けではなく $\mathcal{H}$ の中で最良に近いものを返せば良いとする．この場合も選択する h は決定的な関数である．

PAC 学習可能かどうかは想定する仮説空間 $\mathcal{H}$ の複雑さに依存する．X の何個の有限部分集合に対するラベル付け関数の取りうる値が完全に散らばるか (shatter) により，VC 次元を定義する．このとき適当な可測性の条件のもとでは，PAC 学習可能性と VC 次元の有限性が同値になる．

概念空間 $\mathcal{C}$ と仮説空間 $\mathcal{H}$ としてできるだけ広いものをとりたい．Soloveichik (2008) [16] は，概念空間 $\mathcal{C}$ として $\{0,1\}^*$ から $\{0,1\}$ へのすべての全域計算可能関数，仮説空間 $\mathcal{H}$ として $\{0,1\}^*$ から $\{0,1\}$ へのすべての部分計算可能関数をとり，標本サイズ m が概念 $c \in \mathcal{C}$ に依存して良いことにすれば，学習が可能であることを示した．すべての部分計算可能関数という仮説空間は非常に広く，特に VC 次元は無限大である．それにも関わらず学習が可能なのは，概念空間 $\mathcal{C}$ および仮説空間 $\mathcal{H}$ が可算集合に制限されていることが主要因である．また標本サイズ m に一様の上界が存在しないのは，計算時間が非常に長い仮説が存在するからという，計算の側面からの帰結であり，統計的な理由ではない．

また，Ben-David ら [4, 6, 5] が ZFC から学習可能性を決定できない問題があることを示した．これを契機に Agarwal ら [2] は PAC 学習の計算可能性版としての CPAC 学習可能性の研究を始めた．CPAC 学習可能性の文脈では，仮説空間 $\mathcal{H}$ として全域計算可能関数全体の族 (Decidable Representation, DR) または部分計算可能関数全体の族 (Recursively

Enumerable Representation, RER) をとる．関連研究として，CPAC 学習可能性の特徴付け [17] や，計算可能解析学の言葉を使った入力空間 X の連続空間への拡張 [1] などがある．

可能な限り様々な概念を学習する汎用的学習の性質を調べたい場合，計算可能性を導入して，計算論的側面を考察する必要があるように思われる．

3　万能推論

3.1　設定

本節では Solomonoff に始まる万能推論の理論 (the theory of universal induction) を紹介する．詳細は Hutter のモノグラフ [8] を参照せよ．

考察する空間はアルファベット Σ の無限列である．ここでは単純のため $\Sigma = \{0,1\}$ とし Cantor 空間 $\{0,1\}^{\mathbb{N}}$ としよう．$\{0,1\}^{\mathbb{N}}$ 上の未知の計算可能なモデル測度 μ を考える．課題は，$\{0,1\}^{\mathbb{N}}$ 上の元で μ からランダムな列 $X = X_1X_2\cdots$ の接頭辞が順番に与えられたときに，次の文字の条件付き確率

$$\mu(\cdot|X_{<n}) = \frac{\mu(X_{\leq n})}{\mu(X_{<n})}$$

を予測することである．

ランダムな標本から未知の測度を予測するという点では統計的決定理論に似ている．しかし，独立同分布性は仮定していない．測度が計算可能であるという点が，この予測を可能にする (もしくは予測に関する議論を可能にする) 要因になっている．

予測測度 $\xi : \{0,1\}^* \to [0,1]$ は条件付き確率として

$$\xi(\cdot|X_{<n}) = \frac{\xi(X_{\leq n})}{\xi(X_{<n})}$$

を与える．損失関数として Kullback-Leibler 情報量や Hellinger 距離が使われることが多いが，$\{0,1\}$ を予測値として出力して正解するかどうかを問うこともできる．また有限接頭辞での正解率というよりは，極限で収束するかどうかを議論することが多い．

考察する空間として一般の距離空間に拡張することもできる [11] が，より自然な拡張があるように思われる．強化学習の設定に拡張した AIXI というモデルの研究もある．

未知の測度の族は，計算可能な測度全体として固定することが多い．具体的な問題を学習する方法を研究するというより，最も汎用的な学習はどのような性質を持つかを考察している．

最初に確認すべき事実として，計算可能な予測測度では計算可能なモデル測度全体の族を

学習することはできない．実際，任意の計算可能な予測測度 ξ に対して，$\xi(1|\sigma)$ が 1 に近ければ $\mu(1|\sigma)$ を 0 に近く，$\xi(1|\sigma)$ が 1 に近くなければ $\mu(1|\sigma)$ を 1 に近く，となるような計算可能測度 μ を作ることができる．すなわち，ξ の予測はこの μ の条件付き確率に収束せず，ξ は μ を学習できない．

ただし，モデル測度として独立同分布の測度に限定すれば，多項式時間計算可能な万能な学習アルゴリズムが存在して，収束速度も速いことが分かっている [12]．

3.2 最適な測度

独立同分布に制限しない任意の計算可能なモデル測度を学習するため，予測測度として c.e. 半測度を許すように空間を広げる．

実数 $x \in \mathbb{R}$ が計算可能とは，計算可能な有理数列 $(a_n)_n$ で，$|a_n - x| < 2^{-n}$ かつ $\lim_n a_n = x$ となるものが存在することを言う．すなわち実数 x を任意の精度で有理数近似するアルゴリズムが存在することを意味する．実数 $x \in \mathbb{R}$ が左 c.e.(left-c.e.) であるとは，計算可能な単調増加の有理数列 $(a_n)_n$ で，$\lim_n a_n = x$ となるものが存在することを言う．これは実数 x よりも小さい有理数の集合 $\{q \in \mathbb{Q} \mid q < x\}$ が c.e. (computably enumerable) であることを意味する．左 c.e. 実数は下側から計算可能に近似可能な列だが，その近似列の有限項目が x にどれだけ近づいているか分からない．計算可能な実数は左 c.e. 実数だが，逆は成り立たない．

関数 $f : \{0,1\}^* \to \mathbb{R}$ が計算可能であるとは，$f(\sigma)$ が $\sigma \in \{0,1\}^*$ について一様に計算可能であることをいう．$f : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ が下側半計算可能 (lower semicomputable) とは，$f(\sigma)$ が $\sigma \in \{0,1\}^*$ について一様に左 c.e. であることをいう．$\{0,1\}^{\mathbb{N}}$ 上の測度 μ が計算可能であるとは，関数 $\sigma \mapsto \mu(\sigma) = \mu([\sigma])$ が計算可能であることをいう．ここで，$[\sigma] = \{X \in \{0,1\}^{\mathbb{N}} \mid \sigma \preceq X\}$ で，σ を接頭辞に持つ Cantor 空間の元の集合のことで，シリンダー集合と呼ばれるものである．Borel 測度はその開基での測度から一意に定まることに注意しておこう．μ が $\{0,1\}^{\mathbb{N}}$ 上の半測度 (semi-measure) であるとは，

$$\mu(\epsilon) \leq 1,\ \mu(\sigma) \geq \mu(\sigma 0) + \mu(\sigma 1) \tag{1}$$

となるものをいう．ここで ϵ は空文字 (長さ 0 の文字列) である．半測度 μ が c.e. であるとは，関数 $\sigma \mapsto \mu(\sigma)$ が下側半計算可能であることをいう．

定義 1. c.e. 半測度 μ が**最適** (optimal) であるとは，任意の c.e. 半測度 ν に対して，ある定数 $c \in \mathbb{N}$ が存在し，すべての $\sigma \in \{0,1\}^*$ に対し，

$$\nu(\sigma) \leq c \cdot \mu(\sigma)$$

が成り立つことを言う.

予測測度 (半測度) は確率の主観主義の立場に立てば，信念の度合いの分配の仕方を表している．すなわち，(半) 測度 μ は列が σ から始まる確率を $\mu(\sigma)$ と思っている．σ の次に来るのが 0 なのか 1 なのか，更なる計算の後に信念の度合いを増やす余地を残しているため，式 (1) において等号が成り立たないと解釈するのが良いだろう．予測測度が最適であるとは，どんな他の測度が与える信念の度合いの分配よりも，定数倍を除いては少なくないことを意味する.

これらの定義のもとで，計算論における万能 Turing 機械の存在の証明と同様の手法により，最適な c.e. 半測度が存在することが示せる．最適な計算可能測度は存在しない．このことは，計算可能測度は計算可能に枚挙できないが，c.e. 半測度は計算可能に枚挙できることに対応している.

最適な c.e. 半測度は予測測度として以下の意味で良い振る舞いをする.

定理 2 (Solomonoff). μ を $\{0,1\}^{\mathbb{N}}$ 上の計算可能なモデル測度，ξ を $\{0,1\}^{\mathbb{N}}$ 上の最適な半測度である予測測度とする．このとき，$k \in \{0,1\}$ に対して，確率 1 で，μ に従う X に対し，

$$|\xi(k|X_{<n}) - \mu(k|X_{<n})| \to 0,\ n \to \infty$$

ここで驚くべきなのは，μ に関して独立同分布などの仮定はされておらず，任意の計算可能なモデル測度に対して成り立つことである．最適な予測測度 ξ は，X の接頭辞から μ の情報を取り出して学習している.

最適な学習アルゴリズムが存在するという事実を聞くと，No Free Lunch 定理との関係が気になるだろう．機械学習の文脈でよく知られているように，あらゆる問題に対して良い予測を与えるような学習アルゴリズムは存在しない．うまい話にはわけがあり，タダより高いものはないとはよく言われる．学習アルゴリズムも，ある問題に対して良い予測をするものは，他のある問題に対しては悪い予測になる．この意味で最適な学習アルゴリズムは存在しない.

No Free Lunch 定理は通常有限情報での学習での話であるのに対し，上記の結果は極限での振る舞いの話である．また，定数倍という遊びもある．異なる種類の主張であり，数学的に矛盾するわけではない.

上記の定理において「確率 1 で」の部分を，「μ-Martin-Löf ランダムな列に対して」と置き換えることはできない [9, 10]．μ は下側計算可能であるが，μ から作られる予測はその条件付き確率でありその比になる．極限計算可能つまり Δ^0_2 ではあるので，2 ランダムな列であれば十分ではある．しかし，正確な特徴付けは知られていない.

また，収束速度は最適な測度のとり方に依存し，かなり遅くなりうる [9].

最適な予測測度の存在は喜ばしいことだが，どの最適な予測測度も計算不可能であり，実装できない．この計算不可能性が万能推論の理論の応用可能性をかなり狭めているように思われる．

3.3 圧縮可能性と予測可能性

最適な予測測度の性質を調べることで，良い予測に関する示唆が得られることはある．アルゴリズム的ランダムネスの理論 [7, 13] によれば，ランダムであることは，マルチンゲールによる予測不可能性で特徴づけられたり，コルモゴロフ複雑性による圧縮不可能性で特徴づけられたりする．それに対応するように，最適な c.e. 半測度の構成はすべての c.e. 半測度の線形和としてつくることもできるし，monotone complexity Km からつくることもできる．

通常の Turing 機械は入力に対して，出力が存在するなら，唯一つの文字列に定まる．これに対し monotone Turing 機械は，入力に対して順に延長して出力することがありうる．入力も延長した場合，出力も同様に出力することができる．monotone Turing 機械 M と文字列 σ に対して，

$$Km_M(\sigma) = \min\{p \in \{0,1\}^* \ : \ \sigma \preceq M(p)\}$$

を σ の monotone 複雑性という．通常の Turing 機械と同様に，万能 monotone Turing 機械が存在し，Km で表す．

A を計算可能な列とする．通常の接頭複雑性 (prefix-free complexity)K に対して，

$$K(A \upharpoonright n) \to \infty,\ n \to \infty$$

となる．$A \upharpoonright n$ を出力するには桁数 n の情報が必要なので，プログラムは長くなる必要がある．それに対して，

$$\sup_n Km(A \upharpoonright n) < \infty$$

となる．A は計算可能なので，1 つのプログラムで任意の n に対して $A \upharpoonright n$ を計算できる．

定義 3. 万能 monotone 機械 M に対し，

$$m(\sigma) = \sum_{p \ : \ M(p)=\sigma*} 2^{-|p|}$$

を Solomonoff の半測度 (Solomonoff semimeasure) と呼ぶ．

ここで $|p|$ は文字列 p の長さである．すなわち，m は monotone 複雑性に応じて測度を振り分けている．複雑な文字列には小さい測度，単純な文字列には大きい測度を与えている．特に計算可能な列に正の測度を与えていることに注意してほしい．

万能 monotone 機械のとり方により Solomonoff の半測度は複数存在して，いずれも最適な c.e. 半測度になるが，逆は成り立たない [19].

monotone 機械にはランダム性の要素はないのだが，その monotone 複雑性による半測度が任意の計算可能測度の予測ができるのが面白い.

ところで古代ギリシャの哲学者 Epicurus の多説明原理によれば，複数の理論が観察と合致するならすべての理論を保持せよという．一方 Occam の剃刀によれば，必要以上の仮定を置かず，観察と合致する現象の中で最も単純なものを選べという．Newton は自然は単純さを好むと言った．Solomonoff の半測度はその良いところどりをして，観察に合致するプログラムが単純なものに高い確率を，そのようなプログラムが複雑なものしかなければ低い確率を与えている．このような Solomonoff の万能推論の理論の歴史および哲学的解説については [14] に詳しい.

4　汎用的学習

4.1　十分汎用的な予測

万能推論において万能性を導くのは優越性 (domination) の概念である．$\{0,1\}^{\mathbb{N}}$ 上の測度 ξ, ν に関して，ξ が ν より優勢である (dominate) ことを，ある定数 $c \in \mathbb{N}$ が存在して，すべての $\sigma \in \{0,1\}^*$ について

$$\nu(\sigma) \leq c \cdot \xi(\sigma)$$

が成立することとして定義する．また，これを還元の関係と見て $\nu \leq_d \xi$ で表す．この定義は半測度にも自然に拡張される.

最適な c.e. 半測度とは，すべての c.e. 半測度より優勢な c.e. 半測度である．計算可能な測度に対して，この還元からなる次数 (degree) を考えると，最大元は存在しない．では，2 つの計算可能な測度について $\nu \leq_d \xi$ が成り立つとき，これらから作られる予測の良さにどんな関係があるだろうか．最近の筆者が得た結果によれば，計算可能なモデル測度 μ に対して，ν, ξ の条件付き予測測度と条件付きモデル測度の KL 情報量の和の期待値を考える．$\nu \leq_d \xi$ であることは，任意の μ に対して ξ でのこの値が ν でのこの値と定数の和で抑えられることが同値になる．つまり，誤差の和の期待値に関して，ν より ξ の方が常に定数を除いて小さい．この意味で ξ から作られる予測の方が良い予測である．このような議論ができるのは，想定するモデル測度が計算可能で，その族が可算であるからである.

次に最も良い予測が存在しなかったとしても，「良い」予測であれば必ず成り立つ性質を

知りたい．**十分汎用的な** (sufficiently general) 計算可能な予測測度 ξ に対して性質 $P(\xi)$ が成り立つということを，ある計算可能な予測測度 ν が存在して，ν より優勢なすべての計算可能な予測測度 ξ 対して $P(\xi)$ が成り立つこととして定義する．ちょうど，十分大きな $n \in \mathbb{N}$ に対して $P(n)$ が成り立つとは，ある $m \in \mathbb{N}$ が存在して，すべての $n \geq m$ に対して $P(n)$ が成立することを意味するのと同様である．汎用的な学習とは 1 つのアルゴリズムで様々な課題を解くことができるときにいう．汎用的な予測測度 ξ は多くのモデル測度 μ に対して学習ができるものである．

これらの定義のもとで，次のことが示せる．任意の計算可能なモデル測度 μ に対して，十分汎用的な計算可能予測測度 ξ は，以下を満たす．それぞれの条件付き測度の KL 情報量で測った誤差の和の期待値が有限でルベーグ測度に対する Martin-Löf ランダムな左 c.e. 実数になる．このことは条件付き測度の収束の速度が任意の計算可能な列よりも遅いことを意味する．「十分汎用的」という概念は学習の極限の振る舞いを考える上で有用な概念のように思える．

4.2 PAC 学習との比較

PAC 学習に計算可能性に入れたものと，計算可能な測度に限定した汎用的学習を比較してみよう．

考察するモデル測度については，PAC 学習では独立同分布の任意の測度であるのに対し，汎用的学習では独立同分布とは限らない計算可能な測度に制限する．汎用的学習において独立同分布性を仮定することは可能だが，汎用的であることの意味も変化する．

データの空間は，agnostic な PAC 学習では入力空間 X とラベル空間 Y の直積で教師あり学習を考えているのに対し，汎用的学習では入力空間が $X = \{*\}$ という 1 点集合でラベル空間 Y が有限アルファベットを考えている．

このように PAC 学習と汎用的学習はそれぞれに一般化の道を進んだ結果，互いに近づいているように見える．私個人としては，具体的な問題を学習するアルゴリズムというより，より汎用的な学習が可能となる条件に興味がある．またそのアルゴリズムが多項式時間計算可能か，ニューラルネットワークで実装可能かなどを通して，機械学習の指標の一助になることを願っている．

参考文献

[1] Nathanael L. Ackerman, Julian Asilis, Jieqi Di, Cameron E. Freer, and Jean-Baptiste Tristan. On computable learning of continuous features. *CoRR*, Vol. abs/2111.14630, , 2021.

[2] Sushant Agarwal, Nivasini Ananthakrishnan, Shai Ben-David, Tosca Lechner, and Ruth Urner. On learnability wih computable learners. In Aryeh Kontorovich and Gergely Neu, editors, *Algorithmic Learning Theory, ALT 2020, 8-11 February 2020, San Diego, CA, USA*, Vol. 117 of *Proceedings of Machine Learning Research*, pp. 48–60. PMLR, 2020.

[3] 青嶋誠, 矢田和善. 高次元の統計学. 共立出版, 2019.

[4] Shai Ben-David, Pavel Hrubes, Shay Moran, Amir Shpilka, and Amir Yehudayoff. A learning problem that is independent of the set theory ZFC axioms. *CoRR*, Vol. abs/1711.05195, , 2017.

[5] Shai Ben-David, Pavel Hrubes, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Author correction: Learnability can be undecidable. *Nat. Mach. Intell.*, Vol. 1, No. 2, p. 121, 2019.

[6] Shai Ben-David, Pavel Hrubes, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Learnability can be undecidable. *Nat. Mach. Intell.*, Vol. 1, No. 1, pp. 44–48, 2019.

[7] Rodney G. Downey and Denis R. Hirschfeldt. *Algorithmic randomness and complexity.* Theory and Applications of Computability. Springer, New York, 2010.

[8] M. Hutter. *Universal artificial intelligence: Sequential decisions based on algorithmic probability.* Springer, 2005.

[9] Marcus Hutter and A. Muchnik. On semimeasures predicting Martin-Löf random sequences. *Theoretical Computer Science*, Vol. 382, pp. 247–261, 2007.

[10] Tor Lattimore and Marcus Hutter. On Martin-Löf (non-)convergence of Solomonoff's universal mixture. *Theoretical Computer Science*, Vol. 588, pp. 2–15, 2015.

[11] Kenshi Miyabe. An optimal superfarthingale and its convergence over a computable topological space. *Lecture Notes in Artificial Intelligence*, Vol. 7070, pp. 273–284, 11 2013.

[12] Preetum Nakkiran. Turing-universal learners with optimal scaling laws. *CoRR*, Vol. abs/2111.05321, , 2021.

[13] André Nies. *Computability and randomness*, Vol. 51 of *Oxford Logic Guides*. Oxford University Press, Oxford, 2009.

[14] S. Rathmanner and M. Hutter. A Philosophical Treatise of Universal Induction. *Entoropy*, Vol. 13, pp. 1076–1136, 2011.

[15] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

[16] David Soloveichik. Statistical learning of arbitrary computable classifiers. *CoRR*, Vol. abs/0806.3537, , 2008.

[17] Tom F. Sterkenburg. On characterizations of learnability with computable learners. *CoRR*, Vol. abs/2202.05041, , 2022.

[18] 竹村彰通. 新装改訂版 現代数理統計学. 学術図書出版社, 2020.

[19] Ian Wood, Peter Sunehag, and Marcus Hutter. (non-)equivalence of universal priors. In David L. Dowe, editor, *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence - Papers from the Ray Solomonoff 85th Memorial Conference, Melbourne, VIC, Australia, November 30 - December 2, 2011*, Vol. 7070 of *Lecture Notes in Computer Science*, pp. 417–425. Springer, 2011.