Review of *Probably Approximate Correct,* by Leslie Valiant. (Basic Books, 2013, x+195 pps.)

In 2010, Leslie Valiant won the Turing Award, the "Nobel prize of computer science". The citation mentioned a number of contributions; the best known of these is the theory of probably approximately correct (PAC) learning, which Valiant first developed in the early 1980's. In his new book *Probably Approximately Correct,* Valiant discusses the theory of PAC learning and its applications to artificial intelligence; and he introduces a more specialized concept "evolvable learning", which, he claims, characterizes Darwinian evolution.

PAC learning is an elegant framework which characterizes the objective and the computational difficulty of carrying out certain kinds of learnings. Imagine a person or a computer program that is trying to determine the characteristics of mammals from viewing a sample of individual animals, each correctly labelled either as a mammal or as not a mammal. Suppose, as is likely, that the sample does not include any platypuses or echidnas. Then the learner might well come up with a rule like "hair, warm-blooded, suckles their young, live-bearing." This rule is not quite correct, because platypuses and echidnas are mammals that lay eggs, but it is *approximately correct,* because it works correctly for every animal except platypuses and echidnas. Can we develop a learning algorithm that always generates an approximately correct answer given a large enough random sample? No, because one could be very unlucky in how the sample is constituted. For instance there is a small but non-zero chance, if animals are sampled randomly, that all the mammals in the sample are platypuses and echidnas. If so, the learner might reasonably posit the additional condition that mammals are egg-laying; and that rule would not be approximately correct. So what one can hope for is an algorithm that is *probably* approximately correct; that is, an algorithm that, given a random sample, will probably come up with a rule that is approximately correct.

One particularly elegant feature of PAC learning is that, because the same distribution is used in the "probably" as in the "approximately", the algorithm does not have to make any prior assumptions about the distribution used to select the sample. Consider, for example, an anomalous distribution over mammals that chooses platypuses and echidnas 99.9% of the time. If samples are being generated according to that distribution, then the learner is likely to come up with a rule that requires mammals to be egg-laying. But now that rule *is* approximately correct, relative to this anomalous distribution.

Another attractive feature of the definition, from the standpoint of mathematical analysis, is that it provides a number of numerical parameters to play with. There is $1-\delta$, the accuracy of the rule; $1-\epsilon$, the probability of attaining an approximate correct rule; n, the number of samples; s, the size of an individual instance; and t, the running time.

PAC-learnability is a property of a collection $\mathcal{C}$ of categories in some space of instances $\mathcal{X}$; for example, the collection of all categories definable by a simple conjunction of unary properties like "warm-blooded and hairy and suckles-young". The collection $\mathcal{C}$ is *PAC-learnable* if there is an algorithm that, given large enough values of n and t, will probably find an approximately correct definition of each category in $\mathcal{C}$ for any distribution over $\mathcal{X}$ and for arbitrarily small δ and ϵ.

Finally, Valiant, following standard practice in complexity theory, characterizes a problem as "tractable" if it is solvable by an algorithm whose running time is bounded by a polynomial function of the parameters. Thus, the collection $\mathcal{C}$ is *efficiently PAC-learnable* if the running time t and the number of samples n are bounded by a polynomial in $1/\epsilon, 1/\delta$, and s. A rather small number of learning problems have been proven to be efficiently PAC-learnable, by Valiant and other researchers.

§

However, despite the title, the focus of the book is not PAC learning, but a problem motivated by Darwinian evolution, specifically by the question whether it is plausible that life could have evolved to its current state within the time that the earth has existed, if the mechanism for evolution is Darwinian. Valiant approaches the analysis of this problem by identifying Darwinian evolution with a narrow class of algorithms called "statistical query" algorithms. These are not allowed to use information about individual instances; they can only use statistical information about the sample as a whole. A class of learning problems is considered to be evolvable if it satisfies the conditions of PAC learning when the class of algorithms under consideration is limited to statistical query algorithms. The theory of evolvable learning is much more recent and less developed than the theory of PAC learning, but some theoretical results have been obtained.

Valiant applies this theory to the question of the plausibility of Darwinism as follows. Each species is confronted by the terrestrial environment with the challenge of maximizing its "performance" (essentially fitness). The adaptation of a species through evolution is conceptualized as the execution of a learning algorithm (Valiant calls these "ecorithms") to solve that problem. The terrestrial environment, indeed, has a collection of different problems that it could set to the species, and the ecorithm must be such that the species succeeds whatever problem in this collection it faces. Valiant is not very explicit about the exact relation of the complexity classes to the theories of evolution; but presumably if it is found that the class of problems set by the environment is evolvable, then Darwinism is plausible; if it is PAC-learnable but not evolvable, then some other evolutionary mechanism, such as Lamarckianism, should be sought; if it is not even PAC-learnable, then we must fall back on intelligent design, or on the anthropic principle. This kind of analysis presumably could not distinguish an evolutionary theory that requires 40 million years from one that requires 4 billion years; but it might distinguish these from theories that would require $10^{20,000}$ years.

I am not at all an expert, but it does not seem to me that Valiant makes a very cogent case that this is a useful abstract model for this question. To what extent is it reasonable to view adaptation as the execution of an algorithm, and specifically a statistical query algorithm? In particular, to what extent is this a reasonable view of the early stages of the emergence of life, the formation of self-replicating biological chemicals? To what extent is it reasonable to view the environment as potentially posing any of a collection of problems, and to require that the ecorithm be able to solve all of them? How is one to find a characterization of the class of problems that the environment could potentially present? Moreover, the devil is very much in the details in this kind of analysis; small changes to the formulation of a problem can make a large different to its learnability. Unless the abstraction reflects reality with high fidelity, an analysis based on that abstraction may well be useless.

§

Valiant also includes an extended discussion of machine learning and artificial intelligence. However, his view of machine learning is disappointingly narrow. PAC learning applies in a natural way to only a rather narrow, though very important, class of learning problems: learning categories from labelled data (so-called "supervised" learning), in cases where there is an exact definition (or at least arbitrarily good definitions). Therefore, if PAC learning is to be viewed as a universal framework for learning, other forms of learning must be either shoehorned into this form, declared unimportant, or ignored. Valiant does some of each of these. Most strikingly, he specifically denies that unsupervised learning (learning from unlabelled examples, such as grouping examples into clusters) is a useful concept in studying natural learning, though he concedes it may have some value in machines.

In one case, an errant form of learning sneaks into the book behind his back. Valiant remarks in passing, in a discussion of learning from positive examples, "We may have figured out that each animal belongs to one species." On the whole, one suspects that the way we figured this out is by learning it; and the problem of learning propositions like "each animal belongs to one species" does not fit well into the framework of supervised learning of categories.[1]

[1]An observer in a state of nature does not directly perceive species; he perceives animals, their features, and

PAC learning is only one of many general frameworks that have been proposed for learning; and in many ways, one of the most limited. Other frameworks include Bayesian learning, minimum description length learning, the Vapnik-Chervonenkis (VC) theory of learning, and classical frequentist statistics. VC theory is mentioned in an endnote; the others not at all.

PAC learning is an elegant and useful approach to a specific class of learning problems. I am doubtful that this kind of approach is useful in characterizing evolution; and it is certainly not a universal framework applicable to all forms of either natural or machine learning.

their behaviors. Species are a higher-order construct. Learning the proposition "each animal belongs to one species" involves determining that the class of animals is partitioned into categories where features and behavior of two animals in the same category is generally much more similar than features and behavior of two animals of different categories, and animals breed only within the same category.