Abstract

Machine Learning has become a widely used tool, yet its theoretical theoretical counterpart (and in particular computational learning theory) has been around since the XX's century. In particular Leslie Valiant proposed in 1984 the framework of *probably approximately correct* (shortened PAC) *learning*. The framework of Valiant is rooted in the concept of *learner*: an algorithm that generates with high probability an answer that has low probability to be wrong. If a learner exists for a given problem, the problem is said to be *learnable*.

The definition of learner relies heavily on notions from probability theory, but surprising results have been obtained relating the notion of PAC–learnability, to notions of *dimension* and *compressions*. Namely for a class of concept C the following statements were shown to be equivalent:

1. C is PAC–learnable.
2. The *VC–dimension* of C is finite [Blu+89].
3. C is compressible in the sense of Littlestone and Warmuth [LW03] , [MY15].

From such remarkable results one can wonder whether they can be generalized on to other settings, such as *Estimating the Maximum* (denoted EMX) learning. This is the work of Ben-David and al. in the article 'Learnability can be undecidable' [Ben+19]. The thesis is an analysis and decomposition of that work.

In [Ben+19] Ben-David and al. successfully show that there exists a notion of *compression schemes* that captures the notion of EMX learnability. On the other hand the authors have no luck in finding a suiting notion of dimension. This is in fact related to the fact that EMX learnability is not always a decidable problem.

To obtain the undecidability result, a method is proposed, still found in [Ben+19]. First, it is observed that the notion of compressibility and that of cardinality of a given class are related. To be precise, The existence of a compression scheme for a class C implies that the class is bounded by an infinite cardinal $\aleph_k$.

From this observation comes the undecidability of EMX learning; carefully choosing the class C such that the statement 'C is EMX–learnable' (and so is compressible) implies the continuum hypothesis, i.e. $\aleph_0 = \aleph_1$. The renowned set–theoretic result of Cohen [Coh64] ensures that the continuum hypothesis is undecidable, hence, it must be that the EMX-learnability also falls under the same predicament.

Bibliography

[Coh64] P. J. Cohen. "THE INDEPENDENCE OF THE CONTINUUM HYPOTHESIS, II". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 51.1 (Jan. 1964). PMC300611[pmcid], pp. 105–110. ISSN: 0027-8424. DOI: 10.1073/pnas.51.1.105. URL: https://doi.org/10.1073/pnas.51.1.105.

[Blu+89] Anselm Blumer et al. "Learnability and the Vapnik-Chervonenkis Dimension". In: *J. ACM* 36.4 (Oct. 1989), pp. 929–965. ISSN: 0004-5411. DOI: 10.1145/76359.76371. URL: https://doi.org/10.1145/76359.76371.

[LW03] Nick Littlestone and Manfred K. Warmuth. "Relating Data Compression and Learnability". In: 2003.

[MY15] Shay Moran and Amir Yehudayoff. *Sample compression schemes for VC classes*. 2015. DOI: 10.48550/ARXIV.1503.06960. URL: https://arxiv.org/abs/1503.06960.

[Ben+19] Shai Ben-David et al. "Learnability can be undecidable". In: *Nature Machine Intelligence* 1.1 (Jan. 2019), pp. 44–48. ISSN: 2522-5839. DOI: 10.1038/s42256-018-0002-3. URL: https://doi.org/10.1038/s42256-018-0002-3.