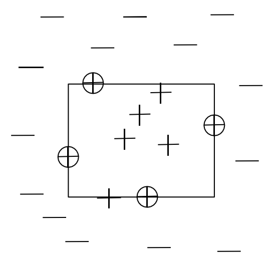
8 Compressing to VC Dimension Many Points

Manfred K. Warmuth[15]

Any set of labeled examples consistent with some hidden orthogonal rectangle can be "compressed" to at most four examples: An upmost, a leftmost, a rightmost and a bottommost positive example. These four examples represent an orthogonal rectangle (the smallest such rectangle that contains them) that is consistent with all examples.



Note that the VC dimension of orthogonal rectangles is four and this is exactly the number of examples needed to represent the consistent orthogonal rectangle.

A compression scheme of size k for a concept class C picks from any set of examples consistent with some concept in C a subset of up to k examples and this subset represents (via a mapping that that is specific to the class C) a hypothesis consistent with the whole original set of examples.

Conjecture: Any concept class of VC dimension d has a compression scheme of size d.

What evidence do we have that this conjecture might be true?

Call a concept class of VC dimension d maximum if for every subset of m instances, the concept class induces exactly $\sum_{i=1}^{d} \binom{m}{i}$ concepts on the subset. That is, for every subset of m instances, we have exactly the maximum number of induced concepts so that Sauer's lemma is tight.

In [1] it was shown that every maximum concept class of VC dimension d has a compression scheme that compresses example sets of size larger than d to a subset of exactly d examples.

Compression schemes were introduced in [2]. In that note a PAC bound is given (See also appendix of [1]). The size of the compression scheme takes the role of the VC dimension in the PAC bound and the constants are small. The proof is very short and interesting in its own right.

Various variants of the above definition of compression scheme are discussed in [2] and [1], Section 9.

Monetary rewards: \$600 for resolving the conjecture by either proving that for any concept class of VC dimension d there always is a compression scheme of size $O(d)$, or by providing a counter example that shows that such compressions schemes are not always possible.

[15] UC Santa Cruz

References

[1] S. Floyd and M. K. Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, 21(3):269–304, 1995.

[2] N. Littlestone and M. K. Warmuth. Relating data compression and learnability. Unpublished manuscript, obtainable at http://www.cse.ucsc.edu/m̃anfred, June 10 1986.