Teaching Dimension versus VC Dimension

Sina Dehghani, Neal Gupta, Karthik Abinav Sankararaman

Introduction

In this report, we give a brief survey of various results relating the Teaching Dimension and VC-Dimension. The concept of Teaching Dimension was first introduced by Goldman and Kearns, 1995 and Sinohara and Miyano, 1991. In this model, an algorithm tries to learn the hidden concept c from examples, called the teaching set, which uniquely identifies c from the rest of the concepts in the concept class $\mathcal{C}$. In a seemingly unrelated direction, the notion of sample compression was introduced by Warmuth and Littlestone, where they study if the number of samples needed to learn a concept can somehow be *compressed* so that the number of training examples needed is minimized. In this survey, we give results from recent literature which makes interesting connections between Teaching Sets and VC Dimension.

Recursive Teaching Dimension

Formally, given a set of examples $\mathcal{E} = \{E_1, E_2, \ldots, E_m\}$ and a concept class(possibly infinite and uncountable) $\mathcal{C}$, a *teaching set* for a concept $c \in \mathcal{C}$ is a subset $S \subseteq \mathcal{E}$ such that, $\forall c' \neq c \in \mathcal{C} \;\; \exists x \in S \;\; s.t. \;\; c(x) \neq c'(x)$. In other words, a teaching set is a subset of examples that can uniquely identify c in the class $\mathcal{C}$. A *minimum teaching set* for a concept $c \in \mathcal{C}$ is the smallest set $S \subseteq \mathcal{E}$ such that S is a teaching set for c. *Teaching dimension*($\mathcal{TD}$) for a concept class $\mathcal{C}$ is the maximum size of the minimum teaching set for any concept in $\mathcal{C}$. In other words, $\mathcal{TD} = \max_{c \in \mathcal{C}} \min_{S \subseteq \mathcal{E}}$ (S is a teaching set for c).

Often, the notion of teaching complexity is too restrictive. The following example, which is a folklore, illustrates this. Consider the set of concept class to be the set of all singletons on $[m]$ and the empty set. Note that for each of the singletons, the minimum teaching set is a single element, while for the empty set, the minimum teaching set is the entire set of examples.

Even for an extremely simple concept class, the teaching dimension is very large. Hence, this worst case description can be slightly refined to get was is known as the *recursive teaching dimension*.

Consider the subset of concepts which are the *simplest* to learn i.e. has the minimum sized teaching set among all concepts. Intuitively, one can first learn those concepts and remove them from the concept class. Now, for the remaining set of concepts we can recursively learn the simplest concepts and so on. With this goal in mind, recursive teaching dimension is formally defined as follows. Let k_1 denote the size of the minimum teaching set for the current concept class $\mathcal{C}$. Remove all concepts from $\mathcal{C}$ such that the size of their teaching sets is k_1. Let k_2 be the size of the minimum teaching set of the remaining concepts. Recursively, do the following and let k_i be the size of the minimum teaching set at iteration i. Recursive teaching dimension is $\max_i k_i$.

Dual Class

In this section, we will briefly describe the notion of dual class. We will first represent the set of examples and hypothesis in the form of a matrix. Define a matrix M such that the rows correspond to the hypothesis and the columns correspond to the set of points in the domain. The entry (h, x) in the matrix M is either 0 or 1 based on whether the hypothesis h labels the point x with 0 or 1. With this representation of the matrix, the dual class is simple to describe. Dual class is the set of hypothesis represented by the transpose of the matrix M. In other words, with every point x in the domain we associate a concept c_x. And the domain of this new concepts is the set of all concepts from the original class. $c_x(x) = 1$ if and only if $M(c, x) = 1$.

We will now prove the following simple claim regarding the Dual Class.

Claim 0.1. *For a given concept class $\mathcal{C}$ with VC-Dimension d, the dual class has a VC-Dimension at most $2^{d+1} - 1$.*

Proof. With the matrix representation of the concept class, it is straight forward to prove this claim. We will prove the contrapositive of this as follows. Suppose the dual class has a VC-Dimension of 2^{d+1}. Then in the transpose of the matrix M, we have a sub-matrix of size $2^{2^{d+1}} * 2^{d+1}$ such that for every binary vector b of length 2^{d+1}, there exists a row such that the entries there correspond to the vector b. Now, consider the transpose of this sub-matrix. It is clear that it is a sub-matrix of M. We will now see that, we can find $d+1$ columns which are shattered. Define $k = 2^{d+1}$. Consider the transpose of the sub-matrix. Note that, all columns in this untransposed sub-matrix are unique. Hence, in the transposed version, all the rows are unique. To have k unique rows, we need to have at least $\log(k)$ columns such that in those columns, all combinations of the binary vector is seen. Hence, we have $d+1$ columns where all combinations of the binary vector is seen. By definition, this is the VC-Dimension. □

Relation between recursive teaching dimension and VC-Dimension

The key question regarding the concept of teaching sets is, the relation between Recursive Teaching Dimension and VC-Dimension. It is clear that the recursive teaching dimension of a concept class is at least the VC-Dimension. The question that remains unsolved is the following. Is it possible to upper-bound the recursive teaching dimension of a concept class with a function of its VC-Dimension.

It is a folklore that for concept class with infinite concepts, the answer is negative. The following concept class has a VC-Dimension of 2, while its recursive teaching dimension is unbounded. Let $\mathbb{Q}$ denote the set of all rational numbers. For every $q \in \mathbb{Q}$, associate a concept c_q which labels all points greater than equal to q as 1 and the remaining points as 0. The VC-Dimension of this concept class is 2, because this is same as the concept class of intervals. However, notice that for any given concept, the teaching set is an infinite set. This is because, one has to provide all the rationals greater or equal q to uniquely identify the concept c_q.

Hence, the question can be restricted to finite concept classes. In case of concept classes with VC-Dimension of 1, Kuhlmann showed that the recursive teaching dimension is at most 1. Hence, the quest for constant factor with respect to the VC-Dimension was solved. For concepts VC-Dimension of $d >= 2$ was open for a while, until recently Moran, Shpilka, Wigderson and Yehudayoff showed that for a special restriction of concept classes with VC-Dimension 2, the recursive teaching dimension can be bounded by 3. The restriction they defined is as follows. For concept classes with VC-Dimension of 2, it implies that any subset of 3 points there can be at most 7 out of the 8 patterns of the binary string. They call this the $(3,7)$ hypothesis class. Their result applies to the $(3,6)$ case of VC-Dimension 2. Formally, the following claim shows their result.

Claim 0.2 (Moran, Shpilka, Wigderson, Yehudayoff)**.** *For concept classes with VC-Dimension* 2 *and* $(3,6)$*-pairing the recursive teaching dimension can be upper bounded by* 3*.*

Before we prove the above claim, we make a note that this theorem is tight. In other words, we show the following example where the concept class has a $(3,7)$ pairing and the recursive dimension is 4.

Concepts	x_1	x_2	x_3	x_4	x_5
c_1	0	0	1	1	0
c_2	0	1	1	1	0
c_3	1	0	0	1	0
c_4	1	1	1	0	0
c_5	0	0	1	1	1
c_6	0	1	1	1	1
c_7	1	0	1	0	0
c_8	1	1	0	1	0

Note that in the above example, the VC Dimension is 2. And for the columns x_1, x_2, x_5, other than the pattern $1,1,1$ all other patterns are present. But in the first step, we need x_1, x_2, x_3, x_5 to teach any concept.

Proof. (Moran, Shpilka, Wigderson, Yehudayoff)
Let C denote a $(3,6)$ concept class. If VC-Dimension of C is 1, by Kuhlmann C has RTD of at most 1. Thus we assume C has VC-Dimension 2. Hence for a shattered pair $\{x, x'\} \subseteq X$ and $b, b' \in \{0,1\}$, there exists a non empty concept class $C_{x,x'}^{b,b'} := \{c \in C | c(x) = b, c(x') = b'\}$. It is not hard to see that the VC-Dimension of $C_{x,x'}^{b,b'}$ is 1. Thus the set of $\{x, x'\}$ in addition to a teaching set for $C_{x,x'}^{b,b'}$ is a teaching set for a concept in C. $\square$

Attempts

In this section, we briefly mention some of the attempts we made to make progress. We wanted to understand how the RTD grows with respect to adding more concepts into the class. In particular, we have the following conjecture.

Conjecture 0.3. *Given a class of concepts $\mathcal{C}$ whose teaching dimension is t, it is impossible to add a new concept $c' \notin \mathcal{C}$ such that the teaching dimension of the new concept class is $t+2$ or greater.*

The above conjecture is interesting, because if true, this says that adding concept by concept one cannot increase the teaching dimension by a lot. On the other hand, adding in more concepts increases the VC Dimension at least logarithmically. This will potentially give us a relation between the two, independent of the size of concept class.

In a different direction we wanted to find a counter example to our original problem. In other words, we wanted to find a concept class such that the recursive teaching dimension is much bigger than the VC Dimension. We wrote computer codes to enumerate the concept classes. For small values of $m <= 30$ and $n <= 8$, where m is the number of concepts and n is the number of examples, we generated thousands of random matrices and we were unable to find a single example where the RTD exceeded the VCD more than 2. This gives an evidence that, for most matrices the RTD is very close to the VCD exactly as conjectured. Currently, we are writing a code to systematically enumerate all examples to see if there is a special example where it might break. The challenge with all concept enumeration is that, the number of concept classes are really big that even for $n >= 4$, it quickly becomes infeasible to generate all possible concept classes.

Moran et al. provide a general lower bound for RTD of a concept class C of VC-Dimension d.

Theorem 0.4 (Moran, Shpilka, Wigderson, Yehudayoff)**.** *Let C be a concept class of VC-Dimension d. Then there exists $c \in C$ with a teaching set of size at most*

$$d2^{d+3}(\log(4e^2) + \log\log|C|).$$

Proof Idea. They show if $|C| > (4e^2)^{d2^{d+2}}$, there exist two distinct x and x' in X such that the set of concepts $c \in C$ such that $c(x) \neq c(x')$ is much smaller than $|C|$. More precisely they show $|\{c \in C : c(x) = 0 \text{ and } c(x') = 1\}| \leq |C|^{1-1/d2^{d+2}}$. Then they add x and x' to the teaching set. They recursively do this until there exists a teaching set of size 1. They show $d2^{d+2}\log\log|C|$

Theorem 0.4 implies the RTD is also upper bounded by $d2^{d+3}(\log(4e^2) + \log\log|C|)$. However they leave the following open question.

Problem 0.5. *Let C be a concept class of VC-Dimension d. Does there exist any upper bound for the RTD of C which is independent of $|C|$?*

Geometric examples

As indicated by the simple example before, there exist concept classes with infinite RTD and finite VC dimension. A direct translation of geometric concepts like the interiors of circles or rectangles do not have finite teaching dimension. For example, for the concept class of concentric circles centered at the origin with radius in $\mathbb{Q}$, there is no finite teaching set for the circle centered at 1, as given any finite teaching set, we can always find a slightly larger circle than the true concept consistent with the teaching examples, by taking a point in the annulus between the minimum radius of a "no" instance found in the teaching set and the true radius.

However we can construct a related (infinite) concept class that has RTD similar to the VC dimension: circles (containing their interior) centered around the origin with integer radius.

The teaching set for a circle with radius $r \geq 1$ is of size 2, as the yes instance $(r, 0)$ and the no instance $(r+1, 0)$ suffice. The only concept not in the minimal teaching set is the concept all 0's (circle with radius 0), so the RTD is 2. This is identical to the VC dimension.

Similar discretizations exist for slightly more complex shapes.

Consider rectangles on the two-dimensional lattice Z^2. These will have RTD 4, as the two corners (and points just beyond the corners) suffice. The VC dimension for this concept class is 4.

Both of these concept classes are closed under intersection.

A result proved in Doliwa, Simon, and Zilles is that for **finite** concept classes which are closed under intersection, the RTD is always less than the VC dimension. The examples above are not finite, but a review of the proof sketch confirms that the techniques used still apply to these examples.

Note that since RTD is monotonic in the sense that removing a concept cannot increase RTD, geometric concept classes whose closure have small VC dimension also must have small RTD.

References

1. Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank, VC dimension and spectral gaps. Electronic Colloquium on Computational Complexity (ECCC), 21:135, 2014.

2. T. Doliwa, H.-U. Simon, and S. Zilles. Recursive teaching dimension, learning complexity, and maximum classes. In ALT, pages 209-223, 2010.

3. S. A. Goldman and M. J. Kearns. On the complexity of teaching. J. Comput. Syst. Sci., 50(1):20-31, 1995.

4. C. Kuhlmann. On teaching and learning intersection-closed concept classes. In EuroCOLT, pages 168-182, 1999.

5. A. Shinohara and S. Miyano. Teachability in computational learning. In ALT, pages 247-255, 1990.

6. Shay Moran, Amir Shpilka, Avi Wigderson, Amir Yehudayoff, Teaching and compressing for low VC-dimension. In FOCS 2015