COMS 6998-4 Fall 2017 Presenter: Benhamin Kuykendall (brk2117)
November 1, 2017 Scribe: Yiyang Li (yl3789)

Teaching Dimensions

1 Introduction

1.1 Learning Model

Thinks through the setting of the online learning with different directors trying to give a query sequence for a *consistent* learner in their own purpose. To be clear, let's define the *consistency.*

Definition 1 (Consistency)**.** *A learner is consistent if for any t, there exist some $c \in \mathcal{C}$ such that*

$$\forall i < t, c(x_i) = c * (x_i), \text{ and } f(x_t) = y_t.$$

As you can see, the basic idea is that the learner won't provide an unreasonable answer on purpose. Its predictions are all come from history. In the online model, the problem of no consistent learner will make a mistake at $t > i$ is completely equivalent with that in finding one exact hypothesis that is consistent with $x_{<t}$.

Now let us go further and think about this teacher setting. How can we define the capability of this model in a practical problem? We can start from the definition of the query sequence, or from teachers point, teaching sequence.

Definition 2 (Teaching Sequence)**.** *Input $x_1, \dots, x_m$ are a teaching sequence for $c \in \mathcal{C}$ if there's no other $g \in \mathcal{C}$ that*

$$\forall i \leq m, g(x_i) = f(x_i).$$

So a teaching sequence is a set of instances from sample space which can uniquely specify the target concept. Apparently, the length of a teaching sequence can highly depend on what kind of concept the teacher is trying to teach. Still, we will need a standard criterion to measure the complexity of a problem, i.e. the concept class $\mathcal{C}$. Thus we have the teaching dimension.

Definition 3 (Teaching dimension)**.** *The teaching dimension of a concept class $\mathcal{C}$ is the smallest integer t that for all $c \in \mathcal{C}$ will have the teaching sequence of length at most t.*

$$t = TD(\mathcal{C}) = \max_{c \in \mathcal{C}} \left(\min_{\tau \in \mathcal{T}(c)} |\tau| \right).$$

1.2 Generic bounds

Here we are trying to introduce some bound of teaching dimension in various settings.

Theorem 1 (Teaching Upper Bound on finite set)**.** *Any finite concept class will have a teaching dimension at most*

$$t \leq |\mathcal{C}| - 1$$

Proof. This is trivial. For the arbitrary $c \in \mathcal{C}$, the teacher just enumerate its difference between every other hypothesis, i.e. for $c_i \neq c$, teacher can choose x_i s.t. $c_i(x_i) \neq c(x_i)$. ☐

Theorem 2 (Teaching Lower Bound on finite set)**.** *Any finite concept class will have a teaching dimension at least*

$$t \geq \frac{\log |\mathcal{C}| - 1}{\log |X|}$$

Proof. For each c, it will uniquely be identified by some $x_1, \dots, x_t$, then

$$|\mathcal{C}| \leq 2^t \binom{|X|}{t} \leq 2|X|^t.$$

Combine two-part and do the log we can get the result. ☐

2 Motivation example

Here we are trying to give some examples to give motivation for specific problems. We will start with the most difficult concept class for teaching in the finite setting.

2.1 Least Teachable Class

We are now considering the concept class over $\{1, 2, \dots, n\}$:

$$\mathcal{C} = \{X \setminus \{1\}, X \setminus \{2\}, \dots, X \setminus \{n\}\} \cup X.$$

You will have to list all $i = 1, \dots, n$ to teach the X. Thus we have the teaching dimenson $t = |\mathcal{C}| - 1$.

2.2 Axis Aligned Boxes

In this problem, we are trying to teach the learner of the joint of some box-areas that aligned to the axis. First, we consider a simple situation. In $\mathbb{Z}^2$, the boxes shrink to the rectangles in the Plane. The teacher is trying to teach the rectangle area has integer vertices.

Teaching sequence:

- Positive examples: x and y;

- Negative examples: $x-(1,0)$, $x-(0,1)$, $y+(1,0)$, $y+(0,1)$.

Thus we have teaching dimension of 6. In the higher dimension space as $\mathbb{Z}^d$, we just use d negative examples for each positive example, add them together we will get

$$t = 2(d+1).$$

3 Teaching versus Learning

3.1 Disparities

Definition 4 (Shattered set)**.** *The class $\mathcal{C}$ shatters a set $\mathcal{S} \subset \mathcal{X}$ when*

$$\{\mathcal{S} \cap c : c \in \mathcal{C}\} = \mathbb{P}(\mathcal{S}).$$

Definition 5 (VC Dimension)**.** *The integer d is the Vapnik-Chervonenkis dimension of a class $\mathcal{C}$ if it is the minimum d such that $\mathcal{C}$ shatters no sets of $d+1$ points.*

As the teaching dimension represents the difficulty of teaching the class, the VC dimension denotes the difficulty of learning the concept. So what is the connection between this two value, since in real world we take teaching and learning as two sides of a coin. Sadly, as for these two dimensions, we can't see the relation between them. We are now present some examples to show the disparities between them.

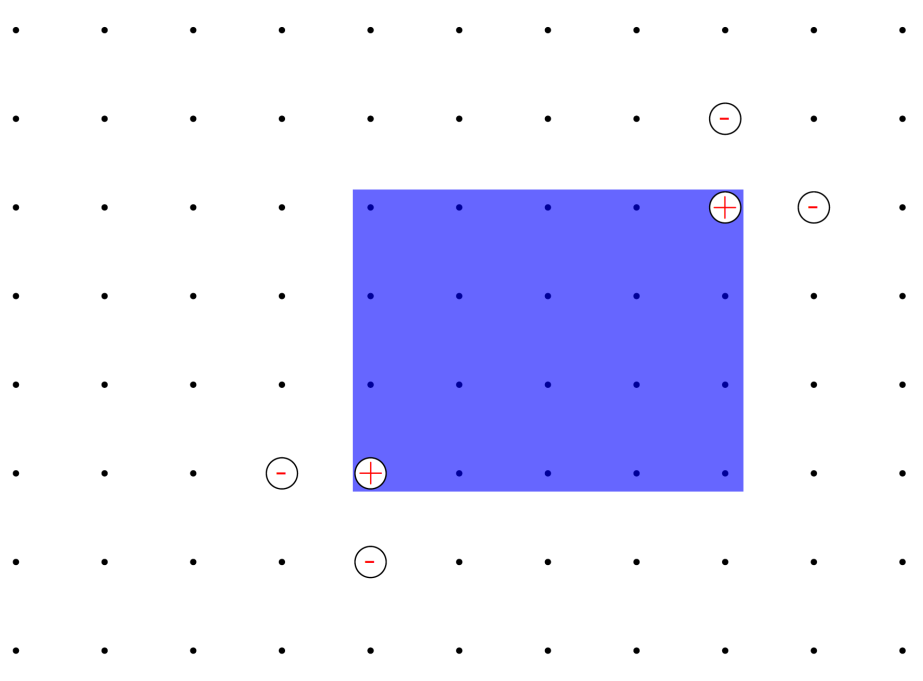

Figure 1: Illustration of learning rectangle in $\mathbb{Z}^2$

3.1.1 Least Teachable Class

Recall the example setting we introduced above, each concept will have all instance but one of the instance space. In this setting, teaching dimension will be n where n denote the number of the instance in the sample set and VC dimension 2 because there will be no hypothesis that induces $(1, 0, 0)$ to an arbitrary triplet. So this is an example of which is hard to teach but easy to learn.

3.1.2 Hybrid Concept

We define two set of functions. a set of n easy to teach function:

$$F = \{\{x\} \ : \ x \in [n]\}$$

.

And also a set of 2^m hard to learn functions:

$$G = 2^{[m]}$$

.

Choose $2^m = n$ and construct class over $[n] \cup [m]$ like illustrated in table 1.

	x_1	x_2	x_3	…	x_{n-1}	x_n	y_1	…	y_{m-1}	y_m
h_1	+	−	−	…	−	−	−	…	−	−
h_2	−	+	−	…	−	−	−	…	−	+
h_3	−	−	+	…	−	−	−	…	+	−
$\vdots$										
h_{n-1}	−	−	−	…	+	−	+	…	+	−
h_n	−	−	−	…	−	+	+	…	+	$\pm$

Table 1: Illustration for hybrid concept

We say it is still easy to teach because we h_i can be identified by positive example x_i. Also, it will be still hard to learn due to $y_1, \ldots, y_m$ is shattered Thus we have teaching Dimension of 1 but VC Dimension of $\log n$.

3.2 Bounds

Theorem 3 (Lower bound)**.** *With VC Dimension d, we have the lower bound of*

$$t \geq \frac{d-1}{\log |X|}.$$

Proof. Follows directly from previous:

$$t \geq \frac{\log |\mathcal{C}| - 1}{\log |X|} \qquad \text{and} \qquad \log |\mathcal{C}| \geq d.$$

□

Theorem 4 (Upper bound). *With VC Dimension d, we have the upper bound of*

$$t \leq |\mathcal{C}| - 2^d + d.$$

Proof. We can build the teaching sequence like this: we have shattered set of size d, then we use one example to exclude each remaining hypothesis. First step removes $2^d - 1$ hypotheses with d examples. Then we just removes $|\mathcal{C}| - (2^d - 1) - 1$ hypotheses, 1 example each. □

4 Recursive Teaching

4.1 Alomost maximal Classes

Theorem 5 (Concentration of Teaching Dimension). *If the teaching dimension of $\mathcal{C}$ is $t \geq |\mathcal{C}| - k$, then for some $f \in \mathcal{C}$ the class $\mathcal{C} \setminus \{f\}$ has teaching dimension at most k.*

Since fixing f requiring a teaching sequence like $x_1, x_2, \ldots, x_t$ of length t.In order to prove, we may fix some f_1 in the class $\mathcal{C} \setminus \{f\}$ and wlog take $f_1(x_1) \neq f(x)$.

The idea is to partition $\mathcal{C} \setminus \{f\}$ into two set:

- S will be a large set that disagrees with f_1 on x_1;
- T will be a small set.

To teach f_1, use sequence x_i plus one x to distinguish from each $g \in T$.

We will construct S and T inductively. First, let $C = \mathcal{C} \setminus (\{f\} \cup S \cup T)$ the remaining concepts. Define $D(x)$ the set of $g \in C$ such that $g(x) \neq f(x)$. Initially, we set $S = \{f_1\}$ and $T = D(x_1) \setminus \{f_1\}$. Then for $i = 2, \ldots, t$:

- Pick an arbitrary $f_i \in D(x_i)$.
- Add f_i to S.
- Add any remaining $D(x_i) \setminus \{f_i\}$ to T.

Proof. Now we try to validate the algorithm. We will begin by claiming that: $f_i \in S$ disagrees with f_1 on x_1.

- Assume $f_i(x_1) = f_1(x_1)$.
- $f_i(x_1) \neq f(x_1)$ by construction.
- But then in first step $f_i \in D(x_1)$ so $f_i \in T$
- T and S are disjoint, so $f_i \notin S$.

Then we will claim that $|T| = k - 1$:

- $D(x_i)$ non-empty at each step, otherwise $\{x_j\} \setminus x_i$ a learning sequence
- One f_i gets added to S each round, have $|S| = t$
- $\mathcal{C} \setminus \{f\} = S \cup T$ implies $|T| = |\mathcal{C}| - 1 - |S|$
- Assumed $t = |\mathcal{C}| - k$ so $|T| = k - 1$. □

□

4.2 Recursive Teaching Dimension

Let MinTD($\mathcal{C}$) be the set of $f \in \mathcal{C}$ with the shortest teaching sequences. Then we can construct levels of $\mathcal{C}$ like this:

$$\mathcal{C}_i = \text{MinTD}\left(\mathcal{C} \setminus \bigcup_{j<i} \mathcal{C}_j\right).$$

Then we can define a robust notion of teaching dimension.

Definition 6 (Recursive Teaching Dimension)**.** *The* recursive teaching dimension *of $\mathcal{C}$ is the maximum of the teaching dimensions of the levels $\mathcal{C}_i$ constructed above.*

Bibliographic notes

The concept of teaching sequence comes from [2], and teaching dimension from [1]. The part of recursive teaching dimension is from [3]

References

[1] Sally A Goldman and Michael J Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1):20–31, 1995.

[2] Sally A Goldman, Ronald L Rivest, and Robert E Schapire. Learning binary relations and total orders. *SIAM Journal on Computing*, 22(5):1006–1034, 1993.

[3] Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *Journal of Machine Learning Research*, 12(Feb):349–384, 2011.