Teaching Dimension

COMS 6998-4 Learning Theory

Benjamin Kuykendall

brk2117@columbia.edu

1 November 2017

Outline

Outline

Consistent learners and Helpful Directors

[Goldman, Rivest, & Shapire 1993]

Definition (Consistent learner)

A learner is *consistent* when for all t the is some $f \in \mathcal{C}$ such that

$$\forall i < t, \quad f(x_i) = f^*(x_i) \qquad \text{and} \qquad f(x_t) = y_t$$

Consistent learners and Helpful Directors

[Goldman, Rivest, & Shapire 1993]

Definition (Consistent learner)

A learner is *consistent* when for all t the is some $f \in \mathcal{C}$ such that

$$\forall i < t, \quad f(x_i) = f^*(x_i) \qquad \text{and} \qquad f(x_t) = y_t$$

In the online model, after inputs $x_1, x_2, \ldots, x_i$:

No consistent learner will make a mistake at $t > i$

$\Leftrightarrow$

Exactly one consistent hypothesis is consistent with the $x_{<t}$

Teaching dimension

[Goldman & Kearns 1995]

Definition (Teaching Sequence)

Inputs $x_1, \ldots, x_m$ are a *teaching sequence* for f when there is no other function $g \in \mathcal{C}$ such that $g(x_i) = f(x_i)$ for all $i \leq m$.

Teaching dimension

[Goldman & Kearns 1995]

Definition (Teaching Sequence)

Inputs $x_1, \ldots, x_m$ are a *teaching sequence* for f when there is no other function $g \in \mathcal{C}$ such that $g(x_i) = f(x_i)$ for all $i \leq m$.

Definition (Teaching Dimension)

The class $\mathcal{C}$ has *teaching dimension* of t when t is the smallest integer such that each $f \in \mathcal{C}$ has a teaching sequence of length at most t.

Outline

Trivial Teaching Sequence

Theorem (Teaching Upper Bound)

Any finite class has a teaching dimension at most

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

Enumerate $\mathcal{C} = f,\ f_1, \ldots,\ f_{|\mathcal{C}|-1}$.

To teach f, choose x_i such that $f(x_i) \neq f_i(x_i)$.

Counting Teaching Sequences

Theorem (Teaching Lower Bound)

Any finite class $\mathcal{C}$ over X has a teaching dimension at least

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

Each f uniquely identified by some $x_1, \ldots, x_t$ with $f(x_1), \ldots, f(x_t)$.

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

Summary of Generic Bounds

Theorem (Teaching Bounds)

Any finite class $\mathcal{C}$ over X has a teaching dimension t such that

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

Outline

Least Teachable Class

Example (Least Teachable Class)

Consider the following concept class over $\{1, 2, \ldots, n\}$:

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

To teach $X \setminus \{i\}$ use teaching sequence i.

To teach X need sequence $1, 2, \ldots, n$.

So teaching dimension is $n = |\mathcal{C}| - 1$.

Outline

Rectangles in the Plane

Example (Rectangles in $\mathbb{Z}^2$)

Two points $\boldsymbol{x},\ \boldsymbol{y} \in \mathbb{Z}^2$ define a rectangle

$$R_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{z}) = 1 \Leftrightarrow z_1 \in [x_1, y_1] \text{ and } z_2 \in [x_2, y_2].$$

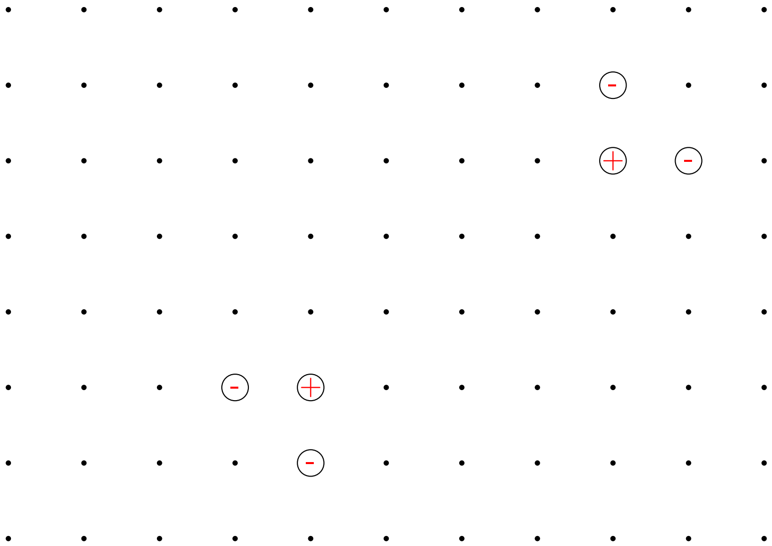
Teaching sequence

Positive examples: $\boldsymbol{x}$ and $\boldsymbol{y}$

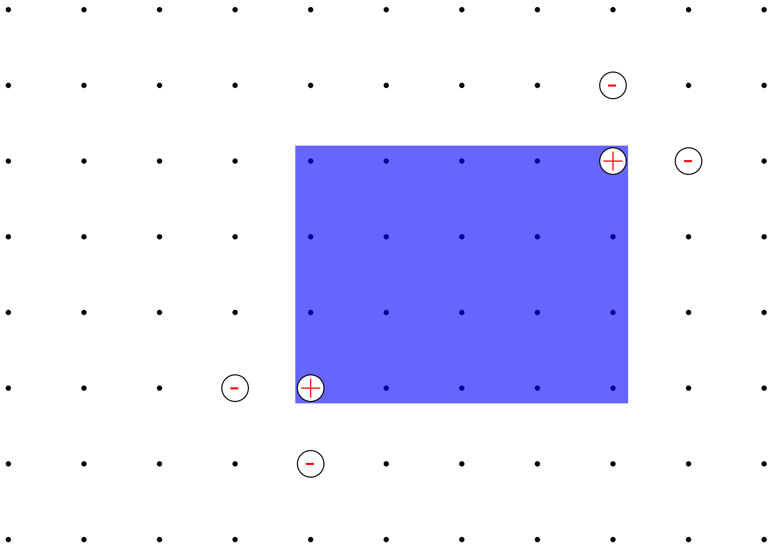
Negative examples: $\boldsymbol{x} - (1,0),\ \boldsymbol{x} - (0,1),\ \boldsymbol{y} + (1,0),\ \boldsymbol{y} + (0,1)$

Teaching dimension 6

Rectangles in the Plane



Rectangles in the Plane



Higher Dimensions

Example (Boxes in $\mathbb{Z}^d$)

Two points $\boldsymbol{x},\ \boldsymbol{y} \in \mathbb{Z}^d$ define a box

$$R_{\boldsymbol{x},\boldsymbol{y}}(z) = 1 \Leftrightarrow \forall i \in [d] \ \ z_i \in [x_i, y_i].$$

Teaching sequence

Positive examples: $\boldsymbol{x}$ and $\boldsymbol{y}$

Negative examples for each $i \in [d]$: $\boldsymbol{x} - \boldsymbol{e}_i$, $\boldsymbol{y} + \boldsymbol{e}_i$

Teaching dimension $2(1 + d)$

Unions of boxes

Example (Union of Boxes)

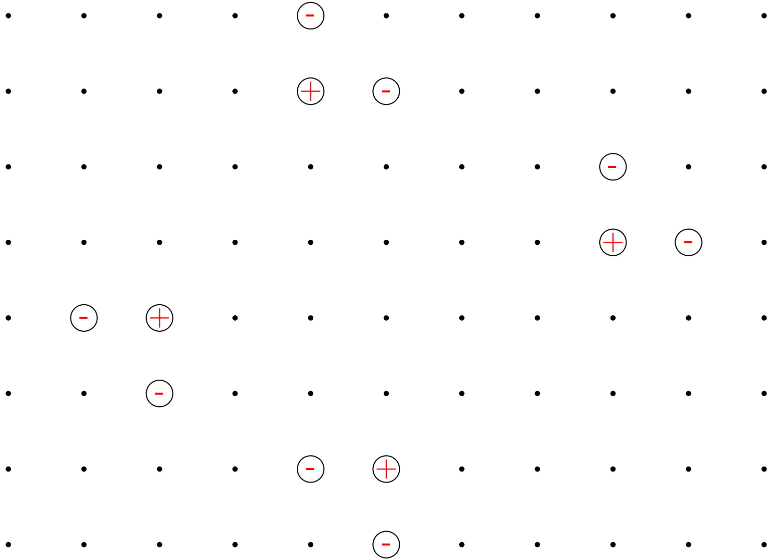
Fix k. For $R_{\mathbf{x}_1,\mathbf{y}_1}, \ldots, R_{\mathbf{x}_k,\mathbf{y}_k}$ disjoint each in $\mathbb{R}^d$ let

$$U_{\{\mathbf{x}_i,\mathbf{y}_i\}}(z) = \bigcup_{i=1}^{k} R_{\mathbf{x}_i,\mathbf{y}_i}.$$

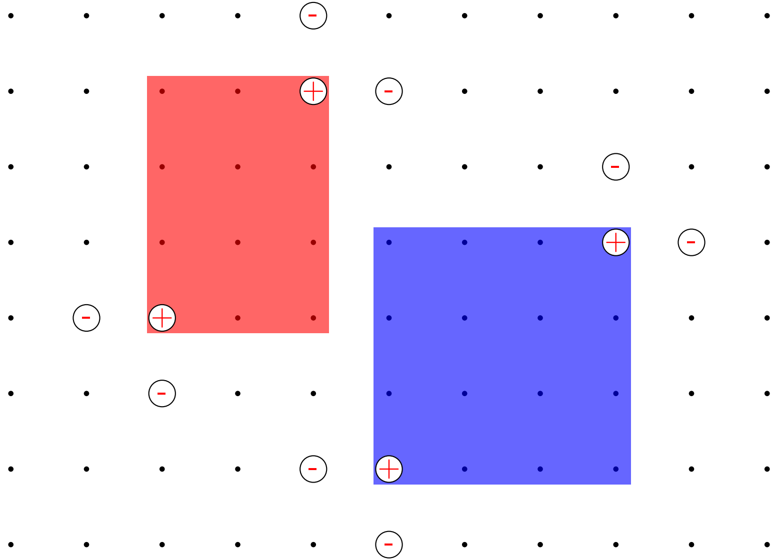
Use the union of the teaching sequences for each box
(with special case when boxes are adjacent)

Teaching dimension $2k(1+d)$.

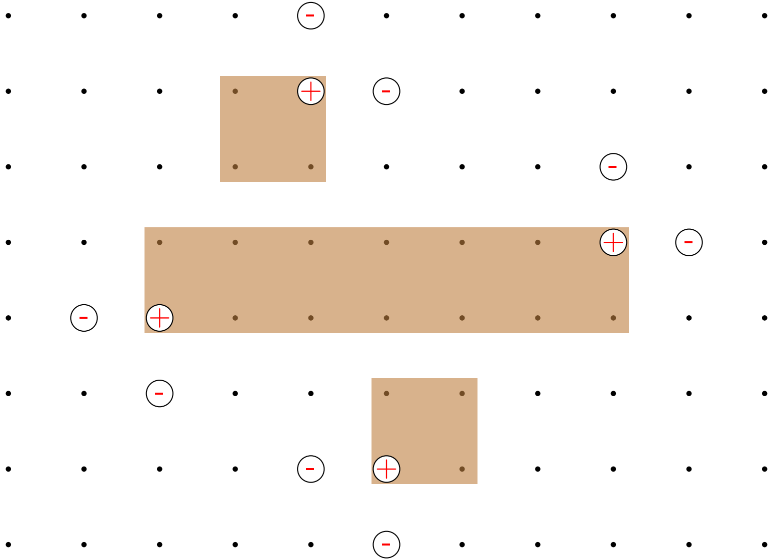
Union of boxes



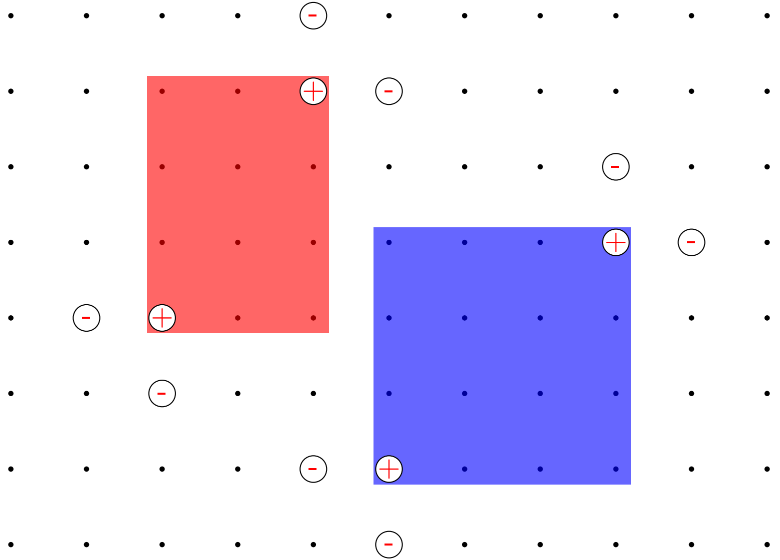
Union of boxes ($k = 2$)



Union of boxes (k = ?)



Union of boxes ($k = 2$)



Outline

VC Dimension

Definition (Shattered set)

The class $\mathcal{C}$ *shatters* a set $S \subset X$ when

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

Definition (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.

Hard to teach, easy to learn

Example (Least Teachable Class)

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

Teaching Dimension n

VC Dimension 2 as no hypothesis induces $(1, 0, 0)$ on three points

Infinite Teaching Dimension

[Moran, Shpilka, Wigderson, Yehudayoff 2015]

Example (Dedekind cuts)

Consider the class of sets of rational numbers less than some real

$$\mathcal{C} = \{(-\infty, r) \cap \mathbb{Q} \ : \ r \in \mathbb{R}\}.$$

VC Dimension 2 as for $q_1 < q_2 < q_3$ cannot induce $(1, 0, 1)$

Teaching Dimension ∞

Easy to teach, hard to learn

Set of n easy to teach functions:

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

Easy to teach, hard to learn

Set of n easy to teach functions:

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

Set of 2^m hard to learn functions:

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

Easy to teach, hard to learn

Set of n easy to teach functions:

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

Set of 2^m hard to learn functions:

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

Choose $2^m = n$ and construct class over $[n] \cup [m]$

Example (Hybrid Concept)

Enumerate $F = f_1, \ \ldots, \ f_n$ and $G = g_1, \ \ldots, \ g_m$ above. Define class

$$\mathcal{C} = \{h_i = f_i \cup g_i \ : \ i \in [n]\}.$$

Easy to teach, hard to learn

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

Easy to teach, hard to learn

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

Still easy to teach: h_i identified by positive example x_i

Still hard to learn: $y_1, \dots, y_m$ is shattered

Teaching Dimension 1 but VC Dimension $\log n$

Outline

Bounding Teaching by VC Dimension

Theorem (Lower bound)

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

Follows directly from previous:

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

Bounding Teaching by VC Dimension

Theorem (Upper bound)

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

Learning sequence:

Shattered set of size d

One example to exclude each remaining hypothesis

First step removes $2^d - 1$ hypotheses with d examples

Second step removes $|\mathcal{C}| - (2^d - 1) - 1$ hypotheses, 1 example each

Summary of VC Dimension Bounds

Theorem (Teaching versus Learning Bounds)

If $\mathcal{C}$ has teaching dimension t and VC dimension d then

$$|\mathcal{C}| - 2^d + d \geq t \geq \frac{d-1}{\log |X|}$$

Outline

Removing one function

Theorem (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.

Removing one function

Theorem (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.

Fix f requiring a teaching sequence $x_1, x_2, \ldots, x_t$ of length t.

To prove: fix some f_1 in the class $\mathcal{C} \setminus \{f\}$ and wlog take $f_1(x_1) \neq f(x)$.

Removing one function

Theorem (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.

Fix f requiring a teaching sequence $x_1, x_2, \ldots, x_t$ of length t.

To prove: fix some f_1 in the class $\mathcal{C} \setminus \{f\}$ and wlog take $f_1(x_1) \neq f(x)$.

Idea: partition $\mathcal{C} \setminus \{f\}$ into

- S a large set that disagrees with f_1 on x_1
- T a small set

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

Concentration Theorem (proof)

Construct S and T inductively.
Let $C = \mathcal{C} \setminus (\{f\} \cup S \cup T)$ the remaining concepts.
Define $D(x)$ the set of $g \in C$ such that $g(x) \neq f(x)$.

Concentration Theorem (proof)

Construct S and T inductively.
Let $C = \mathcal{C} \setminus (\{f\} \cup S \cup T)$ the remaining concepts.
Define $D(x)$ the set of $g \in C$ such that $g(x) \neq f(x)$.

First set $S = \{f_1\}$ and $T = D(x_1) \setminus \{f_1\}$.

Concentration Theorem (proof)

Construct S and T inductively.
Let $C = \mathcal{C} \setminus (\{f\} \cup S \cup T)$ the remaining concepts.
Define $D(x)$ the set of $g \in C$ such that $g(x) \neq f(x)$.

First 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.

Concentration Theorem (proof)

Construct S and T inductively.
Let $C = \mathcal{C} \setminus (\{f\} \cup S \cup T)$ the remaining concepts.
Define $D(x)$ the set of $g \in C$ such that $g(x) \neq f(x)$.

First 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.

Concentration Theorem (proof)

First 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.

Claim 1: $f_i \in S$ 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 $f_i \notin S$.

Concentration Theorem (proof)

First set $S = \{f_1\}$ and $T = D(x_1) \setminus \{f_1\}$.

Then for $i = 2, \dots, t$:

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.

Claim 2: $|T| = k - 1$:

$D(x_i)$ 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 $|T| = k - 1$. □

Outline

Recursive Teaching Dimension

[Zilles, Lange, Holte, Zinkevich 2011]

Let $\mathrm{MinTD}(\mathcal{C})$ be the set of $f \in \mathcal{C}$ with the shortest teaching sequences.

Construct levels of $\mathcal{C}$ as follows:

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

Recursive Teaching Dimension

[Zilles, Lange, Holte, Zinkevich 2011]

Let $\text{MinTD}(\mathcal{C})$ be the set of $f \in \mathcal{C}$ with the shortest teaching sequences.

Construct levels of $\mathcal{C}$ as follows:

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

Then we can define a robust notion of teaching dimension.

Definition (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.