Machine learning II

Learning theory

Souhaib Ben Taieb

February 23, 2022

University of Mons

Table of contents

Proof that $m_{\mathcal{H}}(N)$ is polynomial (if there is any break point)

Table of contents

Main result

Main result

No break point $\implies$ $m_{\mathcal{H}}(N) = 2^N$

Any break point $\implies$ $m_{\mathcal{H}}(N)$ is **polynomial** in N

Break point

k is a break point if $m_{\mathcal{H}}(k) < 2^k$

	N					
	1	2	3	4	5	$\cdots$
2-D perceptron	2	4	8	14	$\cdots$	
1-D pos. ray	2	3	4	5	$\cdots$	
2-D pos. rectangles	2	4	8	16	$< 2^5$	$\cdots$

Quiz I

For every set of k^* points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k^*}$, $\mathcal{H}$ implements $< 2^{k^*}$ dichotomies. Which statement(s) is (are) true?

1. k^* is a break point
2. k^* is not a break point
3. all $k \geq k^*$ are break points
4. all $k < k^*$ are break points
5. this has nothing to do with break points!

Quiz I

For every set of k^* points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k^*}$, $\mathcal{H}$ implements $< 2^{k^*}$ dichotomies:

1. k^* is a break point.
2. k^* is not a break point.
3. all $k \geq k^*$ are break points.
4. all $k < k^*$ are break points.
5. this has nothing to do with break points!

Quiz II

To show that k is **not** a break point for $\mathcal{H}$:

1. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ can shatter.
2. Show $\mathcal{H}$ can shatter any set of k points.
3. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ cannot shatter.
4. Show $\mathcal{H}$ cannot shatter any set of k points.
5. Show $m_{\mathcal{H}}(k) = 2^k$.

Quiz II

To show that k **is not** a break point for $\mathcal{H}$:

1. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ can shatter.
2. Show $\mathcal{H}$ can shatter any set of k points. (Overkill)
3. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ cannot shatter.
4. Show $\mathcal{H}$ cannot shatter any set of k points.
5. Show $m_{\mathcal{H}}(k) = 2^k$.

Quiz III

To show that k **is** a break point for $\mathcal{H}$:

1. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ can shatter.
2. Show $\mathcal{H}$ can shatter any set of k points.
3. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ cannot shatter.
4. Show $\mathcal{H}$ cannot shatter any set of k points.
5. Show $m_{\mathcal{H}}(k) = 2^k$.

Quiz III

To show that k **is** a break point for $\mathcal{H}$:

1. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ can shatter.
2. Show $\mathcal{H}$ can shatter any set of k points.
3. Show a set of k points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ which $\mathcal{H}$ cannot shatter.
4. Show $\mathcal{H}$ cannot shatter any set of k points.
5. Show $m_{\mathcal{H}}(k) = 2^k$.

Back to our combinatorial puzzle

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○

Try to add a 6th dichotomy.

We cannot add a 6th dichotomy

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	●	●	○

If $k = 2$ is a break point, the maximum number of dichotomies on $N = 4$ points is 5.

Intuition: any break point implies a huge combinatiorial restriction, i.e. an enormous constraint on the number of dichotomies. The number of dichotomies, which is equal to 2^N (without a break point) will reduce to a polynomial (with a break point).

Table of contents

Bounding $m_{\mathcal{H}}(N)$

- To show $m_{\mathcal{H}}(N)$ is polynomial, we will show that

$$m_{\mathcal{H}}(N) \leq \cdots \leq \cdots \leq \text{ a polynomial}$$

- The key quantity is $B(N, k)$ which gives the **maximum number of dichotomies on N points with break point k**.
- How many dichotomies can you list on 4 points with break point 2?

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
●	○	○

$B(3, 2) = 4 \ (< 8)$

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○

$B(4, 2) = 5 \ (< 16)$

$m_{\mathcal{H}}(N)$ is bounded by $B(N,k)$

Suppose that $\mathcal{H}$ has a break point at k. Then

$$m_{\mathcal{H}}(N) \leq B(N,k).$$

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\ldots$	$\mathbf{x}_N$
○	○	○	○	$\ldots$	●
○	○	○	●	$\ldots$	○
○	○	●	○	$\ldots$	○
○	●	○	○	$\ldots$	○
●	○	○	○	$\ldots$	●
○	○	●	●	$\ldots$	●
○	●	○	●	$\ldots$	○
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$

- Consider any k points. They cannot be shattered (otherwise k woud not be a break point)
- $B(N,k)$ is largest such list

$m_{\mathcal{H}}(N)$ is bounded by $B(N, k)$

Suppose that $\mathcal{H}$ has a break point at k. Then

$$m_{\mathcal{H}}(N) \leq B(N, k).$$

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	...	$\mathbf{x}_N$
○	○	○	○	...	●
○	○	○	●	...	○
○	○	●	○	...	○
○	●	○	○	...	○
●	○	○	○	...	●
○	○	●	●	...	●
○	●	○	●	...	○
⋮	⋮	⋮	⋮	...	⋮

- Consider any k points. They cannot be shattered (otherwise k woud not be a break point)
- $B(N, k)$ is largest such list

How can we bound $B(N, k)$?

$B(N, k)$ for boundary cases

- $B(N, 1) = ?$
- $B(1, k) = ?$ for $k > 1$
- $B(N, N) = ?$

$B(N, k)$ for boundary cases

- $B(N, 1) = ?$
- $B(1, k) = ?$ for $k > 1$
- $B(N, N) = ?$

- $B(N, 1) = 1$
- $B(1, k) = 2$ for $k > 1$
- $B(N, N) = 2^N - 1$

We then assume $N \geq 2$ and $k \geq 2$ and try to develop a **recursion**.

Let us try to bound $B(4,3)$

How many dichotomies can you list on 4 points with break point 3 (i.e. no subset of 3 is shattered).

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	○	●	●
○	●	○	●
●	○	○	●
○	●	●	○
●	○	●	○
●	●	○	○

We need to find a recursion, i.e. bound $B(4,3)$ using $B(3,\cdot)$.

Let us try to bound $B(4,3)$

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	○	●	●
○	●	○	●
●	○	○	●
○	●	●	○
●	○	●	○
●	●	○	○

- Some dichotomies (out of the 2^4) are missing because of the break point constraint ($k = 3$). The remaining dichotomies are unique.
- What can we say about the prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ of these remaining dichotomies?

Two types of dichotomies

Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears once or twice.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
○	○	●	○
○	●	○	○
●	○	○	○
○	○	●	●
○	●	○	●
●	○	○	●
○	●	●	○
●	○	●	○
●	●	○	○

$\rightarrow$ Let us reorder the dichotomies to simplify the couting.

Reordering the dichotomies

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

- α: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears once
- β: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears twice

Reordering the dichotomies

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

- α: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears once
- β: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears twice
- $B(4, 3) = \alpha + 2\beta$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

- α: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears once
- β: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears twice
- $B(4, 3) = \alpha + 2\beta$
- Strategy for bounding $B(4, 3) = \alpha + 2\beta$:

$$\begin{aligned} \alpha + \beta &\leq Q \\ \beta &\leq R \\ \implies \alpha + 2\beta &\leq Q + R \end{aligned}$$

Reordering the dichotomies

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
	○	●	●	○
α	●	○	●	○
	●	●	○	○
	○	○	○	○
β	○	○	●	○
	○	●	○	○
	●	○	○	○
	○	○	○	●
β	○	○	●	●
	○	●	○	●
	●	○	○	●

- α: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears once
- β: Prefix $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ appears twice
- $B(4,3) = \alpha + 2\beta$
- Strategy for bounding $B(4,3) = \alpha + 2\beta$:

$$\alpha + \beta \leq Q$$
$$\beta \leq R$$
$$\implies \alpha + 2\beta \leq Q + R$$

- $Q = ?$, $R = ?$ $\implies$ exploit the fact that $k = 3$ is a break point.

First, bound $\alpha + \beta$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
	○	●	●	○
α	●	○	●	○
	●	●	○	○
	○	○	○	○
β	○	○	●	○
	○	●	○	○
	●	○	○	○
	○	○	○	●
β	○	○	●	●
	○	●	○	●
	●	○	○	●

- $\alpha + \beta \leq$?

First, bound $\alpha + \beta$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
	○	●	●	○
α	●	○	●	○
	●	●	○	○
	○	○	○	○
β	○	○	●	○
	○	●	○	○
	●	○	○	○
	○	○	○	●
β	○	○	●	●
	○	●	○	●
	●	○	○	●

- $\alpha + \beta \leq$?
- $\alpha + \beta$: the total number of <u>different</u> dichotomies on the first 3 points
- No subset of <u>$k = 3$</u> of these first 3 points can be shattered (since no 3-subset of all 4 points can be shattered).

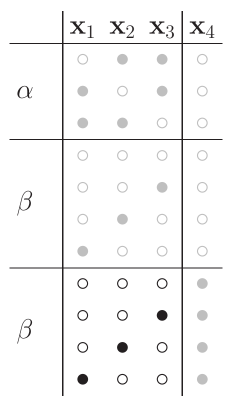
First, bound $\alpha + \beta$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
	○	●	●	○
α	●	○	●	○
	●	●	○	○
	○	○	○	○
β	○	○	●	○
	○	●	○	○
	●	○	○	○
	○	○	○	●
β	○	○	●	●
	○	●	○	●
	●	○	○	●

- $\alpha + \beta \leq$?
- $\alpha + \beta$: the total number of <u>different</u> dichotomies on the first 3 points
- No subset of $\underline{k = 3}$ of these first 3 points can be shattered (since no 3-subset of all 4 points can be shattered).

$\Longrightarrow \quad \alpha + \beta \leq B(3, 3)$

Second, bound β



- $\beta \leq$?

Second, bound β

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

- $\beta \leq$?
- If 2 points are shattered, then using the mirror dichotomies you shatter 3 points

Second, bound β

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

- $\beta \leq$?
- If 2 points are shattered, then using the mirror dichotomies you shatter 3 points
- $\beta \leq B(3, 2)$

Combining, to bound $\alpha + 2\beta$

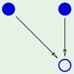
	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
α	○	●	●	○
	●	○	●	○
	●	●	○	○
β	○	○	○	○
	○	○	●	○
	○	●	○	○
	●	○	○	○
β	○	○	○	●
	○	○	●	●
	○	●	○	●
	●	○	○	●

- $B(4, 3) = \alpha + \beta + \beta \leq B(3, 3) + B(3, 2)$
- The argument generalizes to (N, k):

$$B(N, k) \leq B(N-1, k) + B(N-1, k-1)$$

Numerical computation of $B(N,k)$ bound

$$B(N,k) \leq B(N-1,k) + B(N-1,k-1)$$



N	k: 1	2	3	4	5	6	..
1	1	2	2	2	2	2	..
2	1	3	4	4	4	4	..
3	1	4	7	8	8	8	..
4	1	5	11	..	..	..	..
5	1	6	:	.			
6	1	7	:		.		
:	:	:	:			.	

Analytic solution for $B(N, k)$ bound

Sauer's Lemma:

$$B(N, k) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$

Proof by induction on N.

Proof of Sauer's Lemma

1. Verify it for the boundary conditions
    - The statement is true whenever $k = 1$ or $N = 1$, by inspection.
    - Since the statement is already true when $k = 1$ (for all values of N) by the initial condition, we only need to worry about $k \geq 2$.
2. Suppose $B(N, k) \leq \sum_{i=0}^{k-1} \binom{N}{i}$ for $k \geq 2$
3. Assuming (2), show that $B(N+1, k) \leq \sum_{i=0}^{k-1} \binom{N+1}{i}$ for $k \geq 2$.
    - You can use the following Lemma: $\binom{N+1}{i} = \binom{N}{i} + \binom{N}{i-1}$. In other words, when choosing i objects from N + 1 objects, either the first object is not included, in $\binom{N}{i}$ ways, or the first object is included in $\binom{N}{i-1}$ ways.

Proof of Sauer's Lemma

$$
\begin{aligned}
B(N+1,k) &\leq B(N,k) + B(N,k-1) \\
&\leq \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=0}^{k-2} \binom{N}{i} \\
&= \sum_{i=0}^{k-1} \binom{N}{i} + \sum_{i=1}^{k-1} \binom{N}{i-1} \\
&= 1 + \sum_{i=1}^{k-1} \left[\binom{N}{i} + \binom{N}{i-1} \right] \\
&= 1 + \sum_{i=1}^{k-1} \binom{N+1}{i} \\
&= \sum_{i=0}^{k-1} \binom{N+1}{i}
\end{aligned}
$$

It is polynomial!

Theorem. If k is any break point for $\mathcal{H}$, i.e. $m_{\mathcal{H}}(k) < 2^k$, then

$$m_{\mathcal{H}}(k) \leq B(N,k) \leq \sum_{i=0}^{k-1} \binom{N}{i}.$$

But, we also have that

$$\sum_{i=0}^{k-1} \binom{N}{i} \leq \begin{cases} N^{k-1}+1 \\ (\frac{eN}{k-1})^{k-1} \end{cases} \qquad \text{(polynomial in } N\text{)}$$

$\rightarrow$ If we can replace M by $m_{\mathcal{H}}(N)$ in the bound, then learning is feasible.

Three examples

$$\sum_{i=0}^{k-1} \binom{N}{i}$$

- $\mathcal{H}$ is **positive rays:** (break point $k = 2$)

$$m_{\mathcal{H}}(N) = N + 1 \;\leq\; N + 1$$

- $\mathcal{H}$ is **positive intervals**: (break point $k = 3$)

$$m_{\mathcal{H}}(N) = \tfrac{1}{2}N^2 + \tfrac{1}{2}N + 1 \;\leq\; \tfrac{1}{2}N^2 + \tfrac{1}{2}N + 1$$

- $\mathcal{H}$ is **2D perceptrons**: (break point $k = 4$)

$$m_{\mathcal{H}}(N) = \;?\; \leq\; \tfrac{1}{6}N^3 + \tfrac{5}{6}N + 1$$

Proof that $m_{\mathcal{H}}(N)$ can replace M

Proof that $m_{\mathcal{H}}(N)$ can replace M

What we want

Instead of:

$$\mathbb{P}[\,|E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\,] \;\leq\; 2 \quad M \quad e^{-2\epsilon^2 N}$$

We want:

$$\mathbb{P}[\,|E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\,] \;\leq\; 2 \; m_{\mathcal{H}}(N) \; e^{-2\epsilon^2 N}$$

Pictorial proof

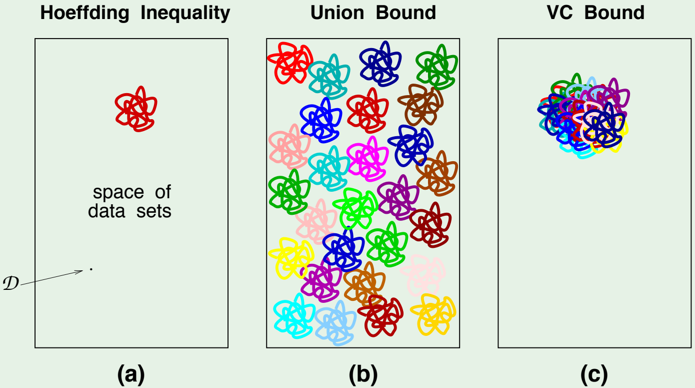
Pictorial proof ☺

- How does $m_{\mathcal{H}}(N)$ relate to overlaps?

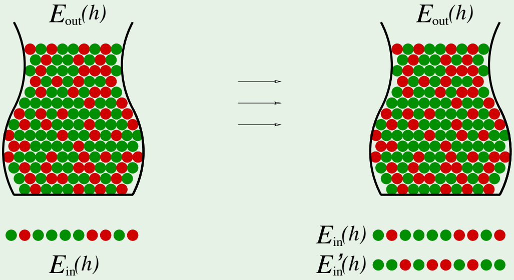
- What to do about E_{out}?

- Putting it together

Overlaps between events



What to do about E_{out}



The Vapnik-Chervonenkis (VC) Inequality

Putting it together

Not quite:

$$\mathbb{P}[\ |E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\ ] \ \leq\ 2\ \ m_{\mathcal{H}}(\ N\ )\ \ e^{-\,2\,\epsilon^2 N}$$

but rather:

$$\mathbb{P}[\ |E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\ ] \ \leq\ 4\ \ m_{\mathcal{H}}(2N)\ \ e^{-\,\frac{1}{8}\,\epsilon^2 N}$$

The Vapnik-Chervonenkis Inequality