

# **SDS 384 11: Theoretical Statistics**

## **Lecture 12: Uniform Law of Large Numbers- VC dimension**

---

Purnamrita Sarkar  
Department of Statistics and Data Science  
The University of Texas at Austin

# Rademacher Complexity for general function classes

Recall that for  $|f(x)| \leq 1$ ,

$$\begin{aligned}\|\hat{P}_n - P\|_{\mathcal{F}} &\leq 2\mathcal{R}_{\mathcal{F}} + \epsilon = 2E[E[\sup_{f \in \mathcal{F}} \sum_i \epsilon_i f(X_i)/n] | X] + \epsilon \\ &\leq 2E\sqrt{\frac{2 \log(|\mathcal{F}(X_1^n) \cup -\mathcal{F}(X)|)}{n}} + \epsilon \\ &\leq \sqrt{\frac{8 \log 2 |\mathcal{F}(X_1^n)|}{n}} + \epsilon\end{aligned}$$

- How do I control  $|\mathcal{F}(X_1^n)|$ ?
- How big is  $\max_X |\mathcal{F}(X_1^n)|$ ?
- Let us focus on binary functions, i.e.  $f(X_i) \in \{0, 1\}$

# Growth function

## Definition

For a binary valued function class  $\mathcal{F}$ , the growth function is:

$$\Pi_{\mathcal{F}}(n) = \max\{|\mathcal{F}(x_1^n)| \mid x_1, \dots, x_n \in \mathcal{X}\}$$

- $\mathcal{X}$  could be  $\mathbb{R}^d$ .
- $\mathcal{R}_{\mathcal{F}} \leq \sqrt{\frac{2 \log(2\Pi_{\mathcal{F}}(n))}{n}}$
- $\Pi_{\mathcal{F}}(n) \leq 2^n$  (which is not really useful)
- We are looking for  $\Pi_{\mathcal{F}}(n)$  growing polynomially with  $n$ .
  - Because then  $\|\hat{P}_n - P\|_{\mathcal{F}} \xrightarrow{P} 0$

# Vapnik-Chervonenkis Dimension

## Definition

A dichotomy of a set  $S$  is a partition of  $S$  into two disjoint subsets.

## Definition (In words)

A set of instances  $S$  is shattered by a binary function class  $\mathcal{F}$  iff for every dichotomy of  $S$ , there is some function in  $\mathcal{F}$  consistent with this dichotomy.

## Definition (In math)

A binary function class  $\mathcal{F}$  shatters  $(x_1, \dots, x_n) \subseteq \mathcal{X}$ , implies that  $|\mathcal{F}(x_1^d)| = 2^d$ .

# Vapnik-Chervonenkis Dimension

## Definition

The VC dimension of a binary function class  $\mathcal{F}$  is given by

$$\begin{aligned} d_{VC}(\mathcal{F}) &= \max\{d : \text{some } x_1, \dots, x_d \in \mathcal{X} \text{ is shattered by } \mathcal{F}\} \\ &= \max\{d : \Pi_{\mathcal{F}}(d) = 2^d\} \end{aligned}$$

- If the VC dimension of a function class is small, then  $\Pi_{\mathcal{F}}(n)$  is small.

## Theorem

If  $d_{VC}(F) \leq d$ , then

$$\Pi_F(n) \leq \sum_{i=0}^d \binom{n}{i}.$$

If  $n \geq d$ , the latter sum is no more than  $(en/d)^d$ .

- So we have the growth function is either polynomially growing with  $d$ , or  $2^n$ .

$$\Pi_F(n) = \begin{cases} = 2^n & \text{If } n \leq d \\ \leq \left(\frac{en}{d}\right)^d & \text{If } n > d \end{cases}$$

## Example

Let  $\mathcal{F} = \{1_{(-\infty, t]} : t \in \mathbb{R}\}$  and  $\mathcal{X} = \mathbb{R}$ . Then  $d_{VC}(\mathcal{F}) = 1$ .

- First show that there exists some configuration of one point, which can be shattered by  $\mathcal{F}$ .
  - For any point  $x$ , if  $x$  has label 1, use  $t > x$
  - If  $x$  has label 0, use  $t < x$ .
- Now show that there exists no two points which can be shattered by  $\mathcal{F}$ . (this takes a bit of an argument in more complex cases.)
  - For any two points  $(x, y)$  the labeling  $(0, 1)$  cannot be achieved by any function in  $\mathcal{F}$ .

## Example

Let  $\mathcal{F}$  be linear classifiers in  $\mathcal{X} = \mathbb{R}^2$ . Then  $d_{VC}(\mathcal{F}) = 3$ .

- First show that there exists some configuration of 3 points, which can be shattered by  $\mathcal{F}$ .
  - Purna draws picture, and if you miss class, you can easily draw a picture to see this.
- Now show that there exists no 4 points which can be shattered by  $\mathcal{F}$ . (this takes a bit of an argument.)



## Example

Let  $\mathcal{F}$  be linear classifiers in  $\mathcal{X} = \mathbb{R}^2$ . Then  $d_{VC}(\mathcal{F}) = 3$ .

- Now show that there exists no 4 points which can be shattered by  $\mathcal{F}$ . (this takes a bit of an argument.)
  - Take 4 non-collinear points. If they are collinear, it is easy to find label configurations which cannot be shattered by a linear classifier.
  - The convex hull of these points will either be a triangle, or a quadrilateral.
  - In case the convex hull is a triangle, and there is a third point inside the convex hull, give all the points on the hull label 1 and the one inside label 0.
  - If three points are collinear or the convex hull is a quadrilateral, then just label the consecutive points with alternative labels.

# VC dimension: rectangles

## Example

Let  $\mathcal{F}$  be classifiers which classify the interior (plus boundary) as one of axis aligned rectangles in  $\mathcal{X} = \mathbb{R}^2$ . Then  $d_{VC}(\mathcal{F}) = 4$ .

- This is on your homework.

## Sauer's lemma proof [Courtesy: P. Bartlett]

- For a fixed  $x_1, \dots, x_n$ , consider the following table.
- Let  $\mathcal{F} = \{f_1, \dots, f_5\}$  and let  $\mathcal{F}$  have VC dimension  $d$ .

---

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$f_1$	0	1	0	1	1
$f_2$	1	0	0	1	1
$f_3$	1	1	1	0	1
$f_4$	0	1	1	0	0
$f_5$	0	0	0	1	0

---

- $|\mathcal{F}|$  is the number of distinct rows of the above table.

## Sauer's lemma proof [Courtesy: P. Bartlett]

- Consider the following shifting operation of the table.
- You start shifting columns from left to right.
- For each column, change a 1 to a zero unless it leads to a row which is already in the table.

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$f_1$	0	1	0	1	1
$f_2$	1	0	0	1	1
$f_3$	1	1	1	0	1
$f_4$	0	1	1	0	0
$f_5$	0	0	0	1	0



	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$f_1$	0	1	0	0	0
$f_2$	0	0	0	0	1
$f_3$	0	0	1	0	1
$f_4$	0	0	1	0	0
$f_5$	0	0	0	0	0

## Sauer's lemma proof [Courtesy: P. Bartlett]

- This operation is done column after column until nothing can be shifted.
- The number of rows does not change.
- An all zero column implies that any subset containing that datapoint is not shattered.
- Consider a row with some 1's. Let  $S$  be the set of points with the 1's.
  - Every configuration with any of these 1's turned into zeros is a row in this table.
  - In other words  $S$  is shattered by  $\mathcal{F}$ .

## Sauer's lemma proof [Courtesy: P. Bartlett]

- The column shifting never shatters a set that was not shattered already, i.e. a set of points can go from shattered to un-shattered but not the other way around.
  - If a column is all zeros after shifting, then any subset containing that datapoint is not shattered.
  - Say we shift a one to a zero, and there are other ones in than column left. Then there is another row which is identical but a zero in that column. So the new zero does not shatter a set.

## Sauer's lemma proof [Courtesy: P. Bartlett]

- Each row has at most  $d$  ones.
- Say there was a row with  $d + 1$  ones.
  - This means there is another identical row except for zeros in place of some of these 1's.
  - But that is the definition of shattering. This means this set of  $d + 1$  points (where there are ones in the row) is shattered by the function class. Which is a contradiction since VC dimension is  $d$  and the shifting operation cannot increase the VC dimension, since it does not shatters a set that was not shattered already.