

# Download the FULL Version with Token NOW!

CSCI4230 Computational Learning Theory  
Lecturer: *Siu On Chan*

Spring 2019  
Based on *Rocco Servedio's* notes

## Notes 5: VC dimension

### 1. VAPNIK-CHERVONENKIS DIMENSION

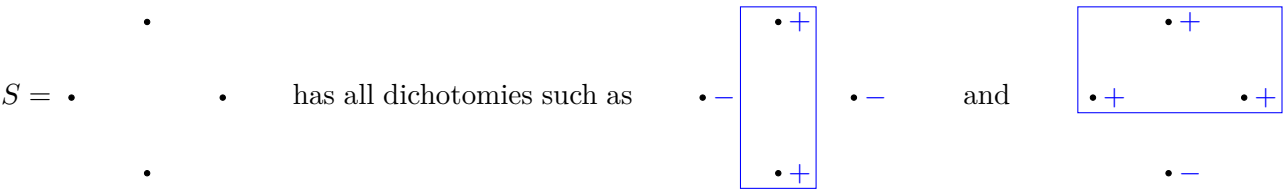
Related to mistake lower bounds in Online Learning

Usually an integer, telling us how expressive a concept class  $\mathcal{C}$  is

Given concept class  $\mathcal{C}$  over instance space  $X$ , subset  $S \subseteq X$  is **shattered** by  $\mathcal{C}$  if all “dichotomies” of  $S$  can be induced by  $\mathcal{C}$ , i.e.:

$$\forall T \subseteq S, \exists c \in \mathcal{C} \text{ s.t. } c \cap S = T$$

$X = \text{the plane} = \mathbb{R}^2$       $\mathcal{C} = \text{axis-aligned rectangles}$



$\text{VCDim}(\mathcal{C})$  is the size of the largest subset  $S \subseteq X$  shattered by  $\mathcal{C}$

$\text{VCDim}(\mathcal{C}) = d$  if and only if

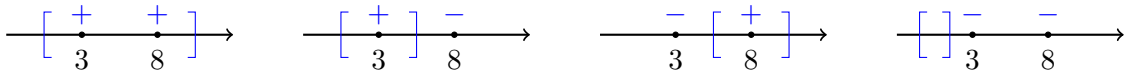
- (1) some subset  $S \subseteq X$  with  $|S| = d$  is shattered by  $\mathcal{C}$ ; and
- (2) all subsets of size  $d + 1$  is not shattered by  $\mathcal{C}$

$\text{VCDim}(\mathcal{C})$  can be  $\infty$

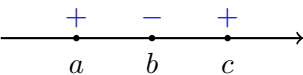
Example: Closed intervals of the real line

$X = \mathbb{R}$       $\mathcal{C} = \text{closed intervals} = \{[a, b] \mid a, b \in \mathbb{R}\}$  where  $[a, b] = \{x \in \mathbb{R} \mid a \leq x \text{ and } x \leq b\}$

Every two points (e.g. 3 and 8) can be shattered  $\implies \text{VCDim}(\mathcal{C}) \geq 2$



No three points ( $a < b < c$ ) can be shattered  $\implies \text{VCDim}(\mathcal{C}) \leq 2$



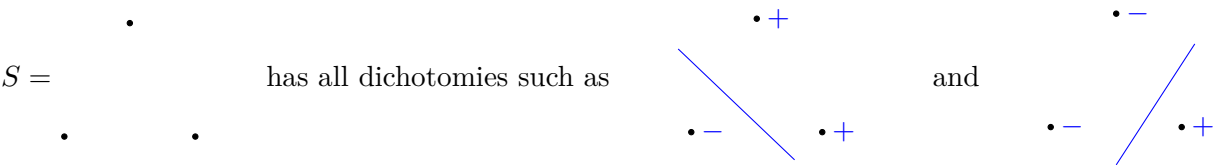
Example: Halfspaces in the plane

$X = \mathbb{R}^2$

$\mathcal{C} = \text{LTF}$

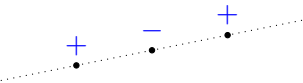
Any three non-collinear points can be shattered

$\implies \text{VCDim}(\mathcal{C}) \geq 3$



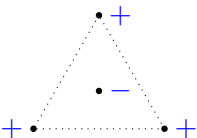
No four points can be shattered  $\implies \text{VCDim}(\mathcal{C}) \leq 3$

Case 1: contains three collinear points



Case 2: No three points collinear

Case 2a: Some point inside the triangle formed by three other points

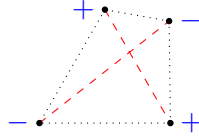


Case 2b: Four points form a convex quadrilateral

$\iff$

the **two diagonals** cross

# Download the FULL Version with Token NOW!



endpoints of two diagonals get different labels

Example:  $X = \{0, 1\}^n$   $\mathcal{C} = \{\text{monotone conjunction}\}$  e.g.  $c(x) = x_2 \wedge x_5 \wedge x_7 \in \mathcal{C}$   
 $\text{VCDim}(\mathcal{C}) \geq n$ :  $S = \{a_j = \text{vector with 0 at position } j \text{ and 1 everywhere else} \mid 1 \leq j \leq n\}$

$$\text{e.g. } n = 4 \quad S = \begin{Bmatrix} 0111, \\ 1011, \\ 1101, \\ 1110 \end{Bmatrix}, \quad T = \begin{Bmatrix} 0111, \\ 1101, \\ 1110 \end{Bmatrix} \text{ induced by } c(x) = x_2$$

Every subset  $T \subseteq S$  is induced by  $c \in \mathcal{C}$  containing precisely variables  $x_j$  s.t.  $a_j \notin T$   
 $\text{VCDim}(\mathcal{C}) \leq n$ : because  $|\mathcal{C}| = 2^n$  and  
**Observation:**  $\text{VCDim}(\mathcal{C}) \geq d$  implies  $|\mathcal{C}| \geq 2^d$

## 2. ONLINE MISTAKE LOWERBOUNDS FROM VC DIMENSION

**Claim 2.1.** Any deterministic algorithm for learning  $\mathcal{C}$  makes  $\geq \text{VCDim}(\mathcal{C})$  mistakes on some sample sequence

*Proof.*  $S = \{x^1, \dots, x^d\}$  be shattered set of size  $d = \text{VCDim}(\mathcal{C})$

Instance sequence is  $x^1, \dots, x^d$

On sample  $x^i$ , algorithm predicts  $b_i \in \{0, 1\}$

Can find  $c \in \mathcal{C}$  s.t.  $c(x^i) = \bar{b}_i$  for all  $1 \leq i \leq n$  (opposite of all predictions)  $\square$

**Claim 2.2.** Some fixed sample sequence causes every randomized algorithm for learning  $\mathcal{C}$  to make  $\geq \text{VCDim}(\mathcal{C})/2$  mistakes in expectation

Previous claim follows from the next claim (via Yao's minimax principle, not covered in this course)

**Claim 2.3.** Some distribution of random sample sequences causes every deterministic algorithm for learning  $\mathcal{C}$  to make  $\geq \text{VCDim}(\mathcal{C})/2$  mistakes in expectation

*Proof.*  $S = \{x^1, \dots, x^d\}$  be shattered set of size  $d = \text{VCDim}(\mathcal{C})$

Sample sequence is  $(x^1, y^1), \dots, (x^d, y^d)$ , where  $y^1, \dots, y^d$  are uniformly random bits

Any algorithm predicting  $d$  uniformly random bits makes  $d/2$  mistakes in expectation

For every choice of random bits  $y^1, \dots, y^d$ , some  $c \in \mathcal{C}$  correctly labels all instances  $x^1, \dots, x^d$   $\square$

**This is the bottom of preview version.  
Please download the full version with token.**