Vapnik-Chervonenkis Dimension

PAC Learnability of Function Classes

- Finite function classes are learnable
  - Sufficient but not necessary condition
  - Infinite classes can be PAC learnable too

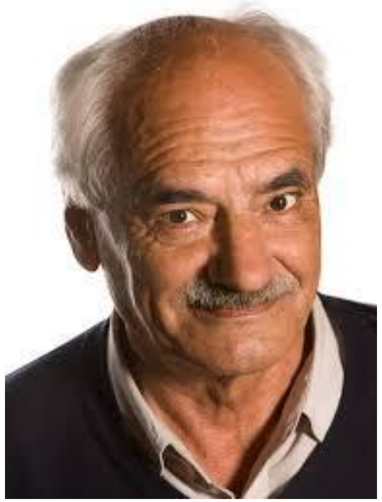
- Sample complexity bounded by a function of $\log|H|$
  - Measure of richness of a function class H

$$m_{\mathcal{H}}(\epsilon, \delta) = O\left(\frac{\log(|\mathcal{H}|/\delta)}{\epsilon}\right)$$

Measures of Capacity of Function Classes

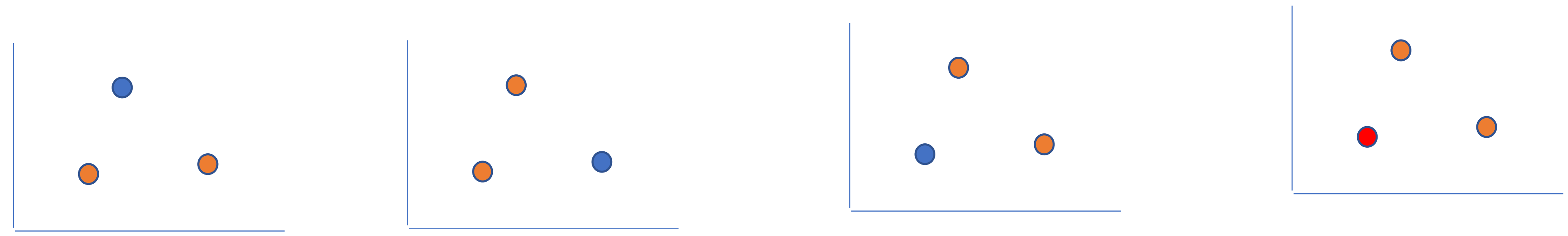
- Vapnik-Chervonenkis dimension

- Finite VC-dimension implies PAC learnability
- Infinite function classes can have finite VC-dimension



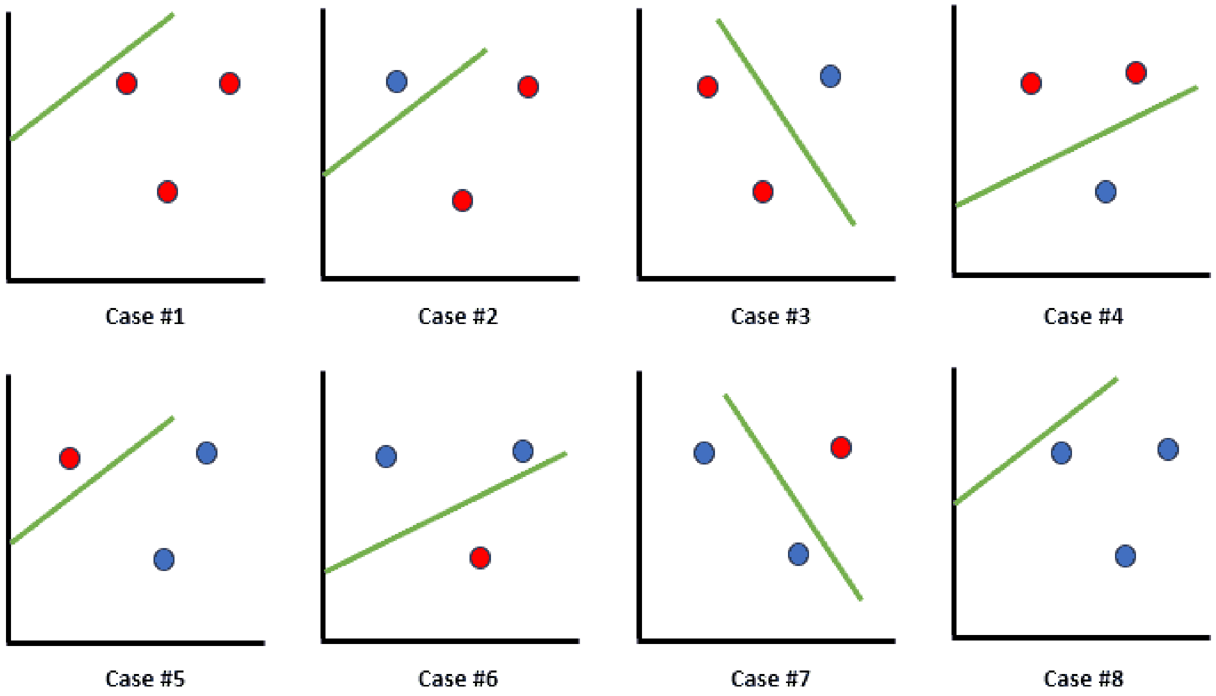
Dichotomies of a Set of Points

- All $2^{|C|}$ combinations of binary labelling of a set of points C
  - Each dichotomy corresponds to a indicator function for subsets of C



Realization of the Dichotomies

- Can all of these dichotomies be realized by some $h \in H$?
    - Need not use the same h for all the dichotomies



Restrictions of *H* to *C*

DEFINITION 6.2 (Restriction of $\mathcal{H}$ to C) Let $\mathcal{H}$ be a class of functions from $\mathcal{X}$ to $\{0,1\}$ and let $C = \{c_1, \ldots, c_m\} \subset \mathcal{X}$. The restriction of $\mathcal{H}$ to C is the set of functions from C to $\{0,1\}$ that can be derived from $\mathcal{H}$. That is,

$$\mathcal{H}_C = \{(h(c_1), \ldots, h(c_m)) : h \in \mathcal{H}\},$$

where we represent each function from C to $\{0,1\}$ as a vector in $\{0,1\}^{|C|}$.

Shattering

- If all the $2^{|C|}$ dichotomies of C are realizable by a member of H
  - H is said to shatter C

If the restriction of $\mathcal{H}$ to C is the set of all functions from C to $\{0, 1\}$, then we say that $\mathcal{H}$ *shatters* the set C. Formally:

DEFINITION 6.3 (Shattering) A hypothesis class $\mathcal{H}$ shatters a finite set $C \subset \mathcal{X}$ if the restriction of $\mathcal{H}$ to C is the set of all functions from C to $\{0, 1\}$. That is, $|\mathcal{H}_C| = 2^{|C|}$.

VC Dimension

DEFINITION 6.5 (VC-dimension) The VC-dimension of a hypothesis class $\mathcal{H}$, denoted $\text{VCdim}(\mathcal{H})$, is the maximal size of a set $C \subset \mathcal{X}$ that can be shattered by $\mathcal{H}$. If $\mathcal{H}$ can shatter sets of arbitrarily large size we say that $\mathcal{H}$ has infinite VC-dimension.

VC Dimension

- To show that VC-dimension of *H* is *d*, we need to show –

1. There exists a set C of size d that is shattered by $\mathcal{H}$.
2. Every set C of size $d+1$ is not shattered by $\mathcal{H}$.

Example:

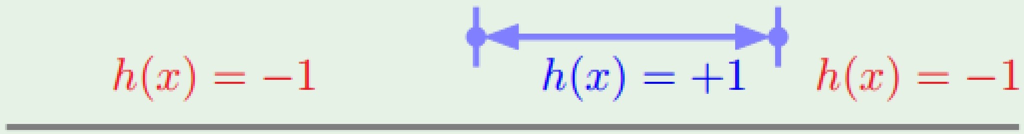
- Two function classes with Real line as domain

- Threshold Function on Real Line
- Intervals on Real Line

Example:

Positive rays ($d_{\text{VC}} = 1$):

Positive intervals ($d_{\text{VC}} = 2$):



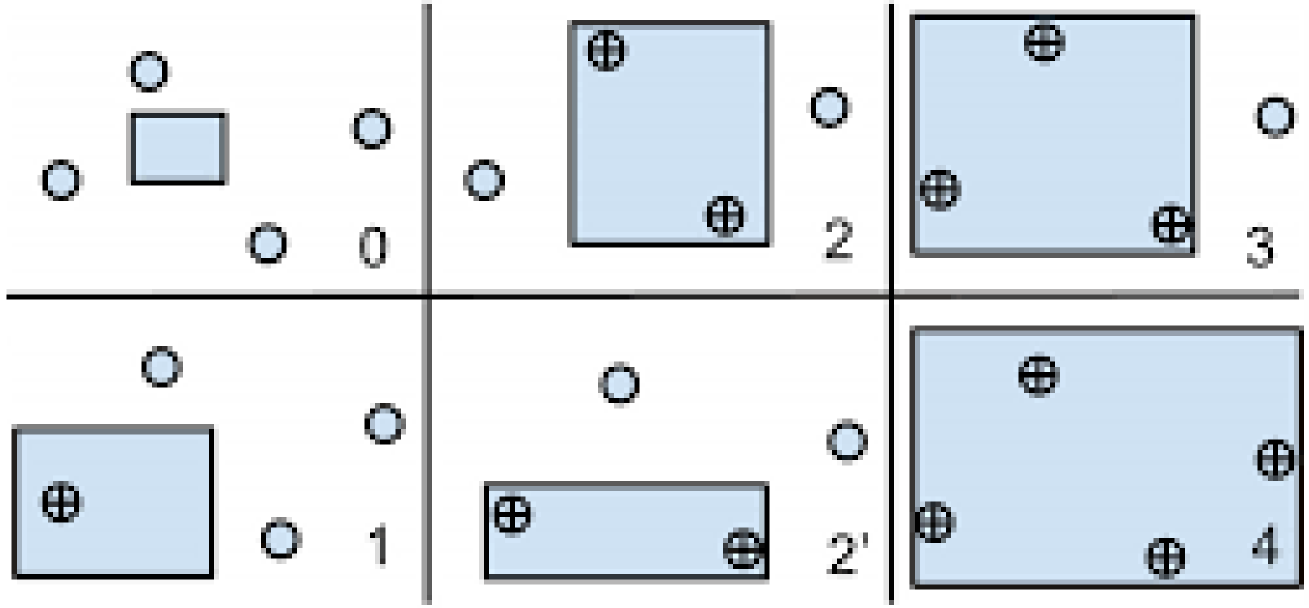
Example: Threshold Function on Real Line

Let $\mathcal{H}$ be the class of threshold functions over $\mathbb{R}$. Recall Example 6.2, where we have shown that for an arbitrary set $C = \{c_1\}$, $\mathcal{H}$ shatters C; therefore $\mathrm{VCdim}(\mathcal{H}) \geq 1$. We have also shown that for an arbitrary set $C = \{c_1, c_2\}$ where $c_1 \leq c_2$, $\mathcal{H}$ does not shatter C. We therefore conclude that $\mathrm{VCdim}(\mathcal{H}) = 1$.

Example: Intervals on Real Line

Let $\mathcal{H}$ be the class of intervals over $\mathbb{R}$, namely, $\mathcal{H} = \{h_{a,b} : a, b \in \mathbb{R}, a < b\}$, where $h_{a,b} : \mathbb{R} \to \{0, 1\}$ is a function such that $h_{a,b}(x) = \mathbb{1}_{[x \in (a,b)]}$. Take the set $C = \{1, 2\}$. Then, $\mathcal{H}$ shatters C (make sure you understand why) and therefore $\mathrm{VCdim}(\mathcal{H}) \geq 2$. Now take an arbitrary set $C = \{c_1, c_2, c_3\}$ and assume without loss of generality that $c_1 \leq c_2 \leq c_3$. Then, the labeling $(1, 0, 1)$ cannot be obtained by an interval and therefore $\mathcal{H}$ does not shatter C. We therefore conclude that $\mathrm{VCdim}(\mathcal{H}) = 2$.

Example: Axis Aligned Rectangles



Axis Aligned Rectangles

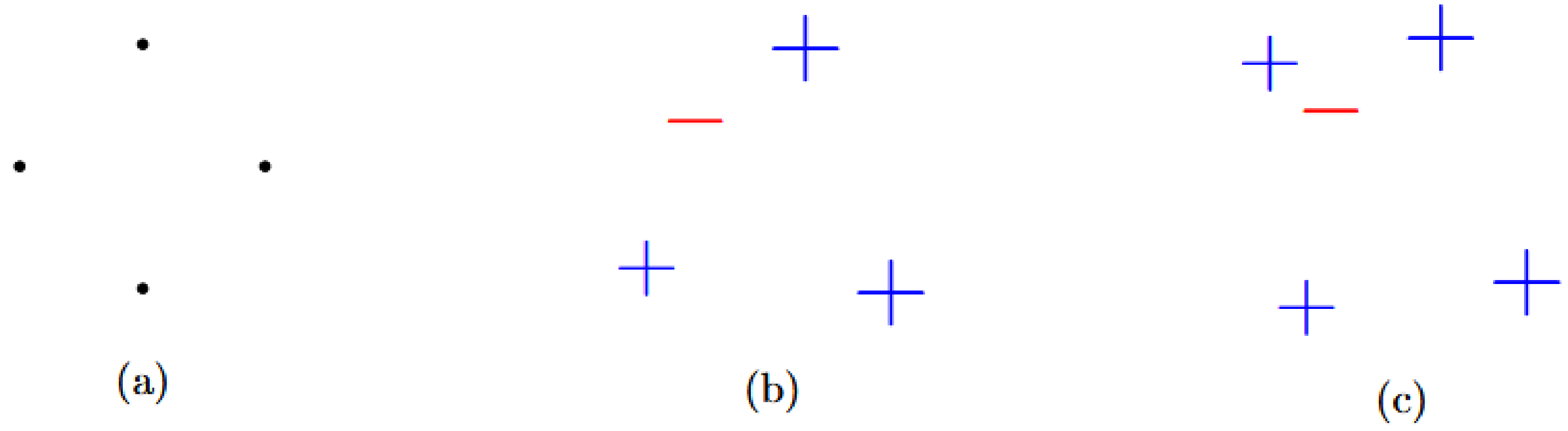


Figure 2: (a) A set of 4 points on which all dichotomies can be realised using rectangles. (b) A set of 4 points with a dichotomy that cannot be realised by rectangles. (c) Any set of 5 points always has a dichotomy that cannot be realised using rectangles.

Linear Threshold Functions in 2D

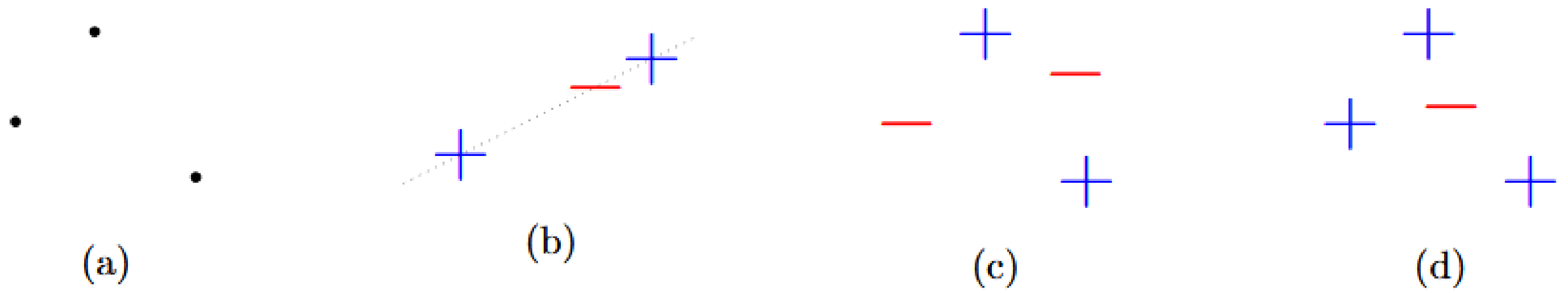


Figure 3: (a) A set of 3 points shattered by linear threshold functions. (b) A dichotomy on a set of 3 points that cannot be realised by linear threshold functions. (c) & (d) No set of 4 points can be shattered by linear threshold functions.