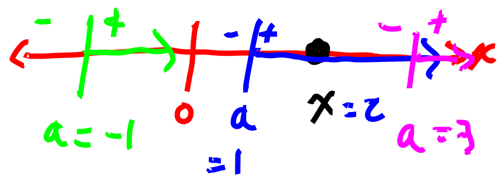
# Chapter 2: Training vs Testing (II)

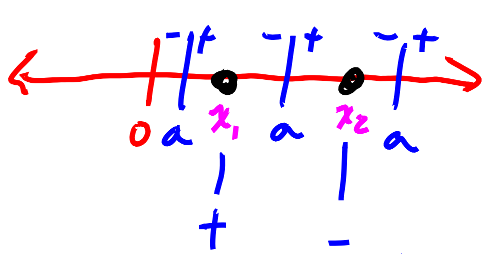
**Problem 1.** What is the VC dimension of the following hypothesis classes? (These hypothesis classes all come from either Example 2.2 on page 43-45 or the problems in the back of the chapter.)

1. The set of positive rays in 1 dimension:

$$\mathcal{H} = \left\{ x \mapsto \operatorname{sign}(x - a) : a \in \mathbb{R} \right\} \tag{1}$$




$d_{vc} \geq 1$




$\mathcal{H}(x_1, x_2) = 3 \neq 2^2 \Rightarrow$ not shattering $\Rightarrow d_{vc} < 2$

$\leq 1$

no value of a can achieve this labelling

2. The set of positive and negative rays in 1 dimension:

$$\mathcal{H} = \left\{ x \mapsto \sigma \operatorname{sign}(x - a) : a \in \mathbb{R}, \sigma \in \{+1, -1\} \right\} \tag{2}$$

will not shatter $N = 3$ dataset $\Rightarrow d_{vc} < 3$

$\boxed{d_{vc} = 2}$

3. The set of positive intervals in 1 dimension:

$$\mathcal{H} = \left\{ x \mapsto [\![a \leq x \leq b]\!] : a \in \mathbb{R}, b \in \mathbb{R} \right\} \tag{3}$$

4. The set of positive and negative intervals in 1 dimension:

$$\mathcal{H} = \left\{ x \mapsto \sigma [\![a \leq x \leq b]\!] : a \in \mathbb{R}, b \in \mathbb{R}, \sigma \in \{+1, -1\} \right\} \tag{4}$$

5. The set of centered spheres in $\mathbb{R}^d$:

$$\mathcal{H} = \left\{ \mathbf{x} \mapsto [\![ \|\mathbf{x}\|_2 \leq \alpha ]\!] : \alpha \in \mathbb{R}^+ \right\} \tag{5}$$

6. Convex sets in 2 dimensions: $\mathcal{H}$ is the set of all functions $h : \mathbb{R}^2 \to \{+1, -1\}$ that are positive inside some convex set and negative elsewhere. (Recall that a set is convex if the line segment connecting any two points in the set lies entirely within the set.)

**Problem 2.** For each statement below, indicate whether it is true or false and explain why. (The best possible explanation for true answers is a proof, and the best possible explanation for false answers is a counterexample.)

1. Let $f$ be the true labeling function. Then for all data distributions, $E_{\text{out}}(f) = 0$.

False

Any problem with noise $\Rightarrow E_{out}(f) > 0$

2. Let $\mathcal{H}$ be the perceptron hypothesis class, and $g$ be the result of the PLA. Also let $f$ be the true labeling function. Then $E_{\text{out}}(g) \geq E_{\text{out}}(f)$.

$\boxed{\text{True}}$

$$f = \underset{h \in \mathcal{H}^*}{\operatorname{argmin}} E_{out}(h)$$

union of all hyp. classes

3. Let $\mathcal{H}$ be the perceptron hypothesis class, and $g$ be the result of the PLA. Also let $f$ be the true labeling function. Then $E_{\text{in}}(g) \geq E_{\text{in}}(f)$.

4. Let $f$ be the true labeling function. There exists some hypothesis class $\mathcal{H}$ with hypothesis $g \in \mathcal{H}$ satisfying $E_{\text{out}}(g) < E_{\text{out}}(f)$.

5. If $\mathcal{H}$ is a finite hypothesis class, and $g$ is trained using the TEA algorithm, the it is always true that $|E_{\text{in}}(g) - E_{\text{out}}(g)| \leq E_{\text{in}}(g)$.

6. Let $g$ be a hypothesis selected from some hypothesis class $\mathcal{H}$ with finite dimension. Then it is possible for $E_{\text{in}}(g)$ to be less than $E_{\text{out}}(g)$.

7. The VC dimension of every finite hypothesis class is finite.

$\boxed{\text{True}}$

$d_{vc}$ = largest $N$ for which $m_{\mathcal{H}} = 2^N$

$$m_{\mathcal{H}}(N) = \max_{\text{all datasets}} |\mathcal{H}(x_1, x_2, \dots x_N)| \leq M$$

$$|\mathcal{H}(x_1, x_2, \dots x_N)| \leq |\mathcal{H}| = M$$

$$2^N \leq M$$
$$N \leq \log_2 M$$
$$\Rightarrow \boxed{d_{vc} \leq \log_2 M}$$

$$|E_{in} - E_{out}| = \tilde{O}\left(\sqrt{\frac{d_{vc}}{N}}\right)$$

$$|E_{in} - E_{out}| = O\left(\sqrt{\frac{\log M}{N}}\right)$$

8. The VC dimension of every infinite hypothesis class is infinite.

$\boxed{\text{False}}$

$\mathcal{H}$ perceptron is infinite hyp. class

$$d_{vc} = d+1 = \Theta(d)$$

9. If there exists a hypothesis $h \in \mathcal{H}$ such that $\min_{h \in \mathcal{H}} E_{\text{in}}(h) = 0$, then the VC dimension of $\mathcal{H}$ must be finite.

10. If $\mathcal{H}$ can shatter a set of size $N$, then $d_{\text{VC}}(\mathcal{H}) \geq N$.

11. For any hypothesis class $\mathcal{H}$, the value $d_{\text{VC}}(\mathcal{H}) + 1$ is a break point for $\mathcal{H}$.

12. If the hypothesis class $\mathcal{H}$ shatters some dataset of size $N$, then $m_{\mathcal{H}}(N) = 2^N$.

13. There exists some hypothesis class $\mathcal{H}$ with growth function $m_{\mathcal{H}}(N) = \Theta(2^{\sqrt{N}})$.

14. Let $\mathcal{H}$ be an arbitrary hypothesis class. Then $m_{\mathcal{H}}(N) = O(2^N)$.

15. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be hypothesis classes satisfying $\mathcal{H}_1 \subset \mathcal{H}_2$. Then $d_{\mathrm{VC}}(\mathcal{H}_1) \leq d_{\mathrm{VC}}(\mathcal{H}_2)$.

16. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be hypothesis classes satisfying $d_{\mathrm{VC}}(\mathcal{H}_1) \leq d_{\mathrm{VC}}(\mathcal{H}_2)$. Then $\mathcal{H}_1 \subset \mathcal{H}_2$.

17. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be hypothesis classes satisfying $d_{\mathrm{VC}}(\mathcal{H}_1) = d_{\mathrm{VC}}(\mathcal{H}_2)$. Then $\mathcal{H}_1 = \mathcal{H}_2$.