DMML

Practical Lesson 7 (November 6, 2025)[1]

VC Dimension

1. (a) $\mathcal{X} = \mathbb{R}^2$, $\mathcal{F} = \{\text{linear classifiers}\}$, i.e., $\mathcal{H} = \{\text{half-planes}\}$.
   (b) $\mathcal{X} = \mathbb{R}^d$, $\mathcal{F} = \{\text{linear classifiers}\}$.
2. (a) Prove that if $\mathcal{H}_1 \subseteq \mathcal{H}_2$, then $\dim_{\mathrm{VC}}(\mathcal{H}_1) \leq \dim_{\mathrm{VC}}(\mathcal{H}_2)$.
   (b) Prove that if $\mathcal{H}$ is finite, then $\dim_{\mathrm{VC}}(\mathcal{H}) \leq \log_2 |\mathcal{H}|$.
3. $\mathcal{X} = \{1, 2, ..., 999\}$, $\mathcal{H} = \{H_0, H_1, ..., H_9\}$, where H_i is the set of numbers in $\mathcal{X}$ whose decimal representation contains the digit i. Determine $\dim_{\mathrm{VC}}(\mathcal{H})$.
4. (a) $\mathcal{X} = \mathbb{R}$, $\mathcal{F} = \left\{ f_{a,b}(x) = \mathbb{I}(a \leq x \leq b) : a, b \in \mathbb{R} \right\}$, i.e., $\mathcal{H} = \{\text{intervals}\}$.
   (b) $\mathcal{X} = \mathbb{R}$, $\mathcal{F} = \left\{ f_{a,b,s} : a, b \in \mathbb{R}, s \in \{-1, 1\} \right\}$ (“signed intervals”), where:

$$f_{a,b,s}(x) = \begin{cases} s, & \text{if } x \in [a, b] \\ -s, & \text{if } x \notin [a, b]. \end{cases}$$

   (c) $\mathcal{X} = \mathbb{R}$, $\mathcal{H} = \{\text{union of } k \text{ closed intervals}\}$.
5. (a) $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \{\text{axis-aligned rectangles}\}$.
   (b) $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \{\text{axis-aligned squares}\}$.
   (c) $\mathcal{X} = \mathbb{R}^d$, $\mathcal{H} = \{\text{axis-aligned rectangles}\}$.
6. (a) $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \{\text{triangles}\}$.
   (b) $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \{\text{convex } k\text{-gons}\}$ for some fixed k.
   (c) $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \{\text{convex polygons}\}$.
7. Let $k \leq |\mathcal{X}| < \infty$.
   (a) Determine $\dim_{\mathrm{VC}}(\mathcal{H}_{\leq k})$, where $\mathcal{H}_{\leq k} = \{H \subseteq \mathcal{X} : |H| \leq k\}$.
   (b) Determine $\dim_{\mathrm{VC}}(\mathcal{H}_{=k})$, where $\mathcal{H}_{=k} = \{H \subseteq \mathcal{X} : |H| = k\}$.
8. Let $\mathcal{X} = \{0, 1\}^n$. For $I \subseteq \{1, 2, ..., n\}$, the parity function f_I is defined as follows:

$$f_I(x_1, x_2, ..., x_n) = \left(\sum_{i \in I} x_i \right) \bmod 2.$$

   What is the VC dimension of the class of parity functions?

At the lecture:

- A class of sets $\mathcal{C}$ *shatters* a set A if

$$\forall S \subseteq A : \exists C \in \mathcal{C} : S = A \cap C.$$

- A *class of classifiers* $\mathcal{F}$ shatters a set $A \subseteq \mathcal{X}$, if the class generated by $\mathcal{F}$, i.e.,

$$\mathcal{C}_{\mathcal{F}} = \left\{ C \subseteq \mathcal{X} : \exists f \in \mathcal{F} : C = f^{-1}(\{1\}) \right\},$$

  shatters A.
- The *Vapnik–Chervonenkis* (VC) dimension of $\mathcal{F}$ is the largest h, such that there is a set A with $|A| = h$ that is shattered by $\mathcal{F}$.

---

[1] `https://apagyidavid.web.elte.hu/2025-2026-1/dmml`

Alternatively:

- Let $\mathcal{H}$ be a set system over $\mathcal{X}$. For $A \subseteq \mathcal{X}$, the *trace* of $\mathcal{H}$ on A is defined as as

$$\mathrm{tr}_A(\mathcal{H}) = \{H \cap A : H \in \mathcal{H}\}.$$

- A subset $A \subseteq \mathcal{X}$ is *shattered* by $\mathcal{H}$ if

$$\mathrm{tr}_A(\mathcal{H}) = 2^A.$$

- The *Vapnik–Chervonenkis* dimension of $\mathcal{H}$ is

$$\dim_{\mathrm{VC}}(\mathcal{H}) = \max\{|A| : A \subseteq \mathcal{X} \text{ is shattered by } \mathcal{H}\}.$$

Solutions:

1.(a) 3 1.(b) $d+1$ 3. 😶 4.(a) 2 4.(b) 3 4.(c) $2k$ 5.(a) 4 5.(b) 3 5.(c) $2d$ 6.(a) 7 6.(b) $2k+1$
6.(c) ∞ 7.(a) k 7.(b) $\min(k, |\mathcal{X}| - k)$ 8. n