Recall the definition of VC-dimension. Consider some examples:

- $\mathcal{C} = \{(-\infty, a) \text{ and } (a, \infty) : a \in \mathbb{R}\}$. $VC(\mathcal{C}) = 2$.
- $\mathcal{C} = \{(a, b) \cup (c, d)\}$. $VC(\mathcal{C}) = 4$.
- $f_1, \ldots, f_d : \mathcal{X} \to \mathbb{R}$, $\mathcal{C} = \{\{x : \sum_{k=1}^{d} \alpha_k f_k(x) > 0\} : \alpha_1, \ldots, \alpha_d \in \mathbb{R}\}$

Theorem 9.1. *$VC(\mathcal{C})$ in the last example above is at most d.*

Proof. Observation: For any $\{x_1, \ldots, x_{d+1}\}$ if we cannot shatter $\{x_1, \ldots, x_{d+1}\} \longleftrightarrow \exists I \subseteq \{1 \ldots d+1\}$ s.t. we cannot pick out $\{x_i, i \in I\}$. If we can pick out $\{x_i, i \in I\}$, then for some $C \in \mathcal{C}$ there are $\alpha_1, \ldots, \alpha_d$ s.t. $\sum_{k=1}^{d} \alpha_k f_k(x) > 0$ for $i \in I$ and $\sum_{k=1}^{d} \alpha_k f_k(x) \le 0$ for $i \notin I$.
Denote

$$\left(\sum_{k=1}^{d} \alpha_k f_k(x_1), \ldots, \sum_{k=1}^{d} \alpha_k f_k(x_{d+1}) \right) = F(\alpha) \in \mathbb{R}^{d+1}.$$

By linearity,

$$F(\alpha) = \sum_{k=1}^{d} \alpha_k \left(f_k(x_1), \ldots, f(x_{d+1}) \right) = \sum_{k=1}^{d} \alpha_k F_k \subseteq H \subset \mathbb{R}^{d+1}$$

and H is a d-dim subspace. Hence, $\exists \phi \neq 0$, $\phi \cdot h = 0, \forall h \in H$ (ϕ orthogonal to H). Let $I = \{i : \phi_i > 0\}$, where $\phi = (\phi_1, \ldots, \phi_{d+1})$. If $I = \emptyset$ then take $-\phi$ instead of ϕ so that ϕ has positive coordinates.
Claim: We cannot pick out $\{x_i, i \in I\}$. Suppose we can: then $\exists \alpha_1, \ldots, \alpha_d$ s.t. $\sum_{k=1}^{d} \alpha_k f_k(x_i) > 0$ for $i \in I$ and $\sum_{k=1}^{d} \alpha_k f_k(x_i) \le 0$ for $i \notin I$. But $\phi \cdot F(\alpha) = 0$ and so

$$\phi_1 \sum_{k=1}^{d} \alpha_k f_k(x_1) + \ldots + \phi_{d+1} \sum_{k=1}^{d} \alpha_k f_k(x_{d+1}) = 0.$$

Hence,

$$\sum_{i \in I} \underbrace{\phi_i \left(\sum_{k=1}^{d} \alpha_k f_k(x_i) \right)}_{>0} = \sum_{i \notin I} \underbrace{(-\phi_i)}_{\ge 0} \underbrace{\left(\sum_{k=1}^{d} \alpha_k f_k(x_i) \right)}_{\le 0}.$$

Contradiction. □

- Half-spaces in $\mathbb{R}^d$: $\{\{\alpha_1 x_1 + \ldots + \alpha_d x_d + \alpha_{d+1} > 0\} : \alpha_1, \ldots, \alpha_{d+1} \in \mathbb{R}\}$.

By setting $f_1 = x_1, \ldots, f_d = x_d, f_{d+1} = 1$, we can use the previous result and therefore $VC(\mathcal{C}) \le d+1$ for half-spaces.
Reminder: $\triangle_n(\mathcal{C}, x_1, \ldots, x_n) = \text{card}\{\{x_1, \ldots, x_n\} \cap C : C \in \mathcal{C}\}$.

Lemma 9.1. *If $\mathcal{C}$ and $\mathcal{D}$ are VC classes of sets,*

1. *$\mathcal{C} = \{C^c : C \in \mathcal{C}\}$ is VC*
2. *$\mathcal{C} \cap \mathcal{D} = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC*
3. *$\mathcal{C} \cup \mathcal{D} = \{C \cup D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC*

(1) obvious - we can shatter $x_1, \ldots, x_n$ by $\mathcal{C}$ iff we can do the same by $\mathcal{C}^c$.

(a) By Sauer's Lemma,

$$\triangle_n(\mathcal{C} \cap \mathcal{D}, x_1, \ldots, x_n) \le \triangle_n(\mathcal{C}, x_1, \ldots, x_n)\triangle_n(\mathcal{C} \cap \mathcal{D}, x_1, \ldots, x_n)$$
$$\le \left(\frac{en}{V_{\mathcal{C}}}\right)^V_{\mathcal{C}} \left(\frac{en}{V_{\mathcal{D}}}\right)^V_{\mathcal{D}} \le 2^n$$

for large enough n.

(b) $(C \cup D) = (C^c \cap D^c)^c$, and the result follows from (1) and (2).

Example 9.1. Decision trees on $\mathbb{R}^d$ with linear decision rules: $\{C_1 \cap \ldots C_\ell\}$ is VC and $\bigcup_{\text{leaves}}\{C_1 \cap \ldots C_\ell\}$ is VC.

Neural networks with depth ℓ and binary leaves.