Assume $f \in \mathcal{F} = \{f : \mathcal{X} \mapsto \mathbb{R}\}$ and $x_1, \ldots, x_n$ are i.i.d. Denote $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(x_i)$ and $\mathbb{P}f = \int f dP = \mathbb{E}f$. We are interested in bounding $\frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbb{E}f$.

Worst-case scenario is the value

$$\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f| .$$

The Glivenko-Cantelli property $GC(\mathcal{F}, P)$ says that

$$\mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f| \to 0$$

as $n \to \infty$.

- Algorithm can output any $f \in \mathcal{F}$
- Objective is determined by $\mathbb{P}_n f$ (on the data)
- Goal is $\mathbb{P}f$
- Distribution P is unknown

The most pessimistic requirement is

$$\sup_P \mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f| \to 0$$

which we denote

$$\text{uniform}GC(\mathcal{F}).$$

VC classes of sets

Let $\mathcal{C} = \{C \subseteq X\}$, $f_C(x) = I(x \in C)$. The most pessimistic value is

$$\sup_P \mathbb{E}\sup_{C\in\mathcal{C}} |\mathbb{P}_n(C) - \mathbb{P}(C)| \to 0.$$

For any sample $\{x_1, \ldots, x_n\}$, we can look at the ways that $\mathcal{C}$ intersects with the sample:

$$\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\}.$$

Let

$$\triangle_n(\mathcal{C}, x_1, \ldots, x_n) = \text{card}\,\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\},$$

the number of different subsets picked out by $C \in \mathcal{C}$. Note that this number is at most 2^n.

Denote

$$\triangle_n(\mathcal{C}) = \sup_{\{x_1,\ldots,x_n\}} \triangle_n(\mathcal{C}, x_1, \ldots, x_n) \le 2^n.$$

We will see that for some classes, $\triangle_n(\mathcal{C}) = 2^n$ for $n \le V$ and $\triangle_n(\mathcal{C}) < 2^n$ for $n > V$ for some constant V.

What if $\triangle_n(\mathcal{C}) = 2^n$ for all $n \ge 1$? That means we can always find $\{x_1, \ldots, x_n\}$ such that $C \in \mathcal{C}$ can pick out any subset of it: "$\mathcal{C}$ shatters $\{x_1, \ldots, x_n\}$". In some sense, we do not learn anything.

Definition 8.1. *If $V < \infty$, then $\mathcal{C}$ is called a* VC class. *V is called* VC dimension *of $\mathcal{C}$.*

Sauer's lemma states the following:

Lemma 8.1.

$$\forall \{x_1, \ldots, x_n\}, \qquad \triangle_n(\mathcal{C}, x_1, \ldots, x_n) \le \left(\frac{en}{V}\right)^V \text{ for } n \ge V.$$

Hence, $\mathcal{C}$ will pick out only very few subsets out of 2^n (because $\left(\frac{en}{V}\right)^V \sim n^V$).

Lemma 8.2. *The number $\triangle_n(\mathcal{C}, x_1, \ldots, x_n)$ of subsets picked out by $\mathcal{C}$ is bounded by the number of subsets shattered by $\mathcal{C}$.*

Identify

$$\mathcal{C} := \{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\}$$

i.e. restrict $\mathcal{C}$ on $\{x_1, \ldots, x_n\}$.

We will say that $\mathcal{C}$ is hereditary if and only if whenever $C \in \mathcal{C}$, then any $B \subseteq C$ is in $\mathcal{C}$.

If $\mathcal{C}$ is hereditary, Lemma is obvious. Otherwise, we will transform $\mathcal{C} \to \mathcal{C}'$, hereditary, in such a way that card $\mathcal{C}$ = card C', i.e. the number of shattered subsets can only decrease.

$$\text{card } C = \text{card } C' = \#(\text{shattered by } \mathcal{C}') \le \#(\text{shattered by } \mathcal{C})$$

Define

$$T_i(C) = \begin{cases} C - \{x_i\} & \text{if } C - \{x_i\} \text{ is not in } \mathcal{C} \\ C & \text{otherwise} \end{cases}$$

Define

$$T_i(\mathcal{C}) = \{T_i(C) : C \in \mathcal{C}\}.$$

Note that card $T_i(\mathcal{C}) =$ card $\mathcal{C}$. Moreover, if C is shattered by $T_i(\mathcal{C})$, it is shattered by $\mathcal{C}$. Indeed, if $x_i \notin C$, then obvious. Otherwise, let $B \in T_i(\mathcal{C})$, but $B \in \mathcal{C}$. Since $x_i \in B$, x_i was not removed from B. This means that $B - \{x_i\} \in \mathcal{C}$. This proves that $\mathcal{C}$ shatters C.

Let

$$T = T_1 \circ \ldots \circ T_n$$

and consider $T^k(\mathcal{C})$ until $T^{k+1}(\mathcal{C}) = T^k(\mathcal{C})$. This will happen because if $T^{k+1}(\mathcal{C}) \neq T^k(\mathcal{C})$, it means that for some C and some i, point x_i was removed from C, $T_i(C) = C - \{x_i\}$, $k \leq 2^n \cdot n$.

$T(T^k(\mathcal{C})) = T^k(\mathcal{C})$ implies that $T^k(\mathcal{C})$ is hereditary because for any $C \in T^k(\mathcal{C})$ and any $x_i \in C$, $C - \{x_i\}$ is also in $T^k(\mathcal{C})$. This is our $\mathcal{C}' = T^k(\mathcal{C})$.

Corollary 8.1. *If $V < \infty$, then*

$$\triangle_n(\mathcal{C}) \leq \sum_{i=0}^{V} \binom{n}{i} \leq \left(\frac{en}{V}\right)^V$$

Indeed, for arbitrary $\{x_1, \ldots, x_n\}$,

$$\begin{aligned} \triangle_n(\mathcal{C}, x_1, \ldots, x_n) &\leq \text{card (shattered subsets of } \{x_1, \ldots, x_n\}) \\ &\leq \text{card (subsets of size } \leq V) \\ &= \sum_{i=0}^{V} \binom{n}{i}. \end{aligned}$$