Growth function: For fixed $x_1, \cdots, x_n \in \mathcal{X}$, and a collection of subsets of $\mathcal{X}$: $\mathcal{D}$, we define

$$\Delta^{\mathcal{D}}(x_1, \cdots, x_n) = |\mathcal{D} \cap \{x_1, \cdots, x_n\}|$$

Example If $\mathcal{X} = \mathbb{R}$, $\mathcal{D} = \{(-\infty, t], t \in \mathbb{R}\}$, then

$$\Delta^{\mathcal{D}}(x_1, \cdots, x_n) \leq n+1, \quad \forall \{x_1, \cdots, x_n\} \subset \mathbb{R}$$

Connection between $\Delta^{\mathcal{D}}(X_1, \cdots, X_n)$ and covering numbers.

For a collection $\mathcal{D}$ and $X_1, \cdots, X_n \overset{iid}{\sim} (\mathcal{X}, \Sigma, P)$, define $\mathcal{F} = \{\mathbb{1}_D ; D \in \mathcal{D}\}$ and $N := N(\varepsilon, \mathcal{F}, d_n^{(\infty)})$, where $d_n^{(\infty)}(f, g) = \max_i |f(X_i) - g(X_i)|$, then

$$\Delta^{\mathcal{D}}(X_1, \cdots, X_n) = N, \quad \forall 0 < \varepsilon < 1$$

Theorem $\frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, \cdots, X_n) = o_p(1) \Rightarrow \mathcal{F}$ is P-GC

Pf: It's sufficient to prove

$$N(\varepsilon, \mathcal{F}, d_n^{(\infty)}) \geq N(\varepsilon, \mathcal{F}, d_n^{(1)})$$

Since $d_n^{(1)}(f, g) = \frac{1}{n} \sum_{i=1}^{n} |f(X_i) - g(X_i)| \leq d_n^{(\infty)}(f, g)$

$\forall f \in \bar{B}^{(\infty)}(g, \varepsilon)$, $f \in \bar{B}^{(1)}(g, \varepsilon)$

$$\Rightarrow \bar{B}^{(\infty)}(g, \varepsilon) \subset \bar{B}^{(1)}(g, \varepsilon) \quad \square$$

Def $\Delta^{\mathcal{D}}(n) = \sup_{x_1, x_2, \cdots, x_n \in \mathcal{X}} \Delta^{\mathcal{D}}(x_1, \cdots, x_n)$, if $\exists$ $V > 0$ s.t. $\Delta^{\mathcal{D}}(n) \lesssim n^V$, we call $\mathcal{D}$ is Vapnik-Chervonenkis Class.

Remark: If $\mathcal{D}$ is VC-class, $\mathcal{F} = \{\mathbb{1}_D, D \in \mathcal{D}\}$ is P-GC class, $\forall$ law P. We call that $\mathcal{F}$ is a uniform GC class.

Example If $\mathcal{D} = \{(-\infty, t], t \in \mathbb{R}\}$, $\mathcal{D}$ is VC class. And the instant result is that

$$\sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| = o_p(1)$$

Example $\mathcal{D} = \{$ all convex set of $\mathbb{R}^2\}$ is not.

Def VC dimension

$V(\mathcal{D}) = \max\{n, \Delta^{\mathcal{D}}(n) = 2^n\}$, which is a break point s.t. $\forall D \in \mathcal{D}$, if $|D| \leq V(\mathcal{D})$, D is shattered.

Remark: If $V(\mathcal{D}) < \infty$, $\sum_{k=0}^{V(\mathcal{D})} \binom{n}{k}$ is the number of the subset $\Lambda \subset \{x_1, \cdots, x_n\}$ which is shattered.

Theorem Sauer's Lemma

$\mathcal{D}$ is VC-class iff $V(\mathcal{D}) < \infty$

Pf: "$\Rightarrow$" If $V(\mathcal{D}) = \infty$, which means that $\Delta^{\mathcal{D}}(n) = 2^n$, $\forall n$, contradiction.

"$\Leftarrow$" Supposed that $\Delta^{\mathcal{D}}(n) \leq \sum_{k=0}^{V(\mathcal{D})} \binom{n}{k}$, $n > V(\mathcal{D})$

$\forall \{x_1, \cdots, x_n, x_{n+1}\} \subset \mathcal{X}$, define

$\mathcal{D}_1 = \{D \in \mathcal{D};\ x_{n+1} \in D\}$, $\mathcal{D}_2 = \mathcal{D} / \mathcal{D}_1$, then

$\Delta^{\mathcal{D}}(x_1, \cdots, x_{n+1}) = |\{x_1, \cdots, x_n, x_{n+1}\} \cap \mathcal{D}|$

$\leq |\{x_1, \cdots, x_{n+1}\} \cap \mathcal{D}_1| + |\{x_1, \cdots, x_{n+1}\} \cap \mathcal{D}_2|$

$$= |\{x_1,\dots,x_n\}\cap (D_1/x_{n+1})| + |\{x_1,\dots,x_n\}\cap D_2|$$

$$\le \underbrace{\sum_{k=0}^{VC(D_1)} \binom{n}{k} + \sum_{k=0}^{VC(D_2)} \binom{n}{k}}_{\text{We claim that: } \le \sum_{k=0}^{VC(D)} \binom{n+1}{k}} \quad ①$$

Notice that ① $= |A_1| + |A_2|$, $A_i = \{\Lambda \subset \{x_1,\dots,x_n\};\ \Lambda$ is shattered by $D_i\}$, $A = \{\Lambda \subset \{x_1,\dots,x_{n+1}\};\ \Lambda$ is shattered by $D\}$

If $\Lambda \in A_1 \Delta A_2$, $\Lambda \in A$

If $\Lambda \in A_1 \cap A_2$, $\Lambda \in A$, but Λ is double counted.

Let $\Lambda_1 = \Lambda \cup \{x_{n+1}\}$, then $\Lambda_1 \notin A_1 \cup A_2$, but $\Lambda_1 \in A$ since that: Λ is shattered by both D_1 and D_2

$$\Rightarrow \quad \forall \Sigma \subset \Lambda,\ \exists B_i \in D_i \text{ s.t. } \Sigma = \Lambda \cap B_i$$

then $\forall \Sigma_1 \subset \Lambda_1$, $\exists \Sigma \subset \Lambda$ s.t. $\Sigma_1 = \{x_{n+1}\} \cup \Sigma$ or Σ

$$\Rightarrow \Sigma_1 = \Sigma = \Lambda_1 \cap B_2 \text{ or } = \{x_{n+1}\} \cup \Sigma = \Lambda_1 \cap B_1 \quad \square$$

Remark: Let $v = VC(D)$

$$\sum_{k=0}^{v} \binom{n}{k} \le \sum_{k=0}^{v} \binom{n}{k} \left(\frac{n}{v}\right)^k \Big/ \left(\frac{n}{v}\right)^v$$

$$\le \left(\frac{n}{v}+1\right)^v \Big/ \left(\frac{n}{v}\right)^v$$

$$= \left(1+\frac{n}{v}\right)^v$$

$$\le \left(\frac{en}{v}\right)^v$$

VC stability

$\mathcal{D}_i$ is a collection of $\mathcal{X}_i$, if $\phi: \mathcal{D}_1 \mapsto \mathcal{D}_2$ is a surjection, then $\Delta^{\mathcal{D}_1}(n) \geq \Delta^{\mathcal{D}_2}(n)$. $V(\mathcal{D}_1) \geq V(\mathcal{D}_2)$.

Property

① $\mathcal{D}$ is VC class $\Longleftrightarrow$ $\mathcal{D}^c$ is VC class.

② $\Delta^{\mathcal{D}_1 \cup \mathcal{D}_2}(n) \leq \Delta^{\mathcal{D}_1}(n)\Delta^{\mathcal{D}_2}(n)$, if $\mathcal{X}_1 = \mathcal{X}_2$

③ $\Delta^{\mathcal{D}_1 \times \mathcal{D}_2}(n) \leq \Delta^{\mathcal{D}_1}(n)\Delta^{\mathcal{D}_2}(n)$

Example $\mathcal{D} = \{(-\infty, t],\ t \in \mathbb{R}^d\}$, then $\mathcal{D}$ is VC class and $\Delta^{\mathcal{D}}(n) \leq (n+1)^d$.

Uniform bound for covering number

We focus on a VC class: $\mathcal{D}$ and $\mathcal{F} = \{\mathbb{1}_D ;\ D \in \mathcal{D}\}$, $d^{(2)}(f,g) = (P|f-g|^2)^{\frac{1}{2}}$, $d_n^{(2)}(f,g) = (P_n|f-g|^2)^{\frac{1}{2}}$. Note that:

$$N(\varepsilon, \mathcal{F}, d^{(1)}) \leq N(\varepsilon, \mathcal{F}, d^{(2)})$$

Let $v = V(\mathcal{D}) < \infty$, $0 \leq \varepsilon \leq 1$.

We assume that $\mathcal{F}$ is totally bound, $\{f_i\}$ is the maximal ε-packing, i.e.

$$d(f_i, f_j) > \varepsilon,$$

and $N =: |\{f_i\}| \geq N(\varepsilon ; \mathcal{F}, d^{(2)})$

Lemma $\exists\ \varepsilon_0 < \varepsilon$, if n is large enough,

$$\mathbb{P}\left(\cap_{i\neq j} \{ d_n^{(2)}(f_i, f_j) > \varepsilon_0 \} \right) > 0$$

Pf: Define $h(x) = (f_i(x) - f_j(x))^2$, for fixed i, j

$$\Rightarrow \mathbb{P}\left(|(d_n^{(2)}(f_i, f_j))^2 - (d^{(2)}(f_i, f_j))^2| > \varepsilon_1 \right) \quad \forall \varepsilon_1 > 0$$

$$= \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^{n} h(X_i) - \mathbb{E}\, h(X_i)\right| > \varepsilon_1 \right)$$

$$= \mathbb{P}\left(|\Sigma_i (h(X_i) - \mathbb{E}\, h(X_i))| > n\varepsilon_1 \right) \quad ①$$

Note that $|h(X_i) - \mathbb{E}\, h(X_i)|$ is bounded.

$$\Rightarrow ① \leq 2e^{-c\frac{n^2\varepsilon_1^2}{n}} = 2e^{-cn\varepsilon_1^2}$$

$$\Rightarrow \mathbb{P}\left(|(d_n^{(2)}(f_i, f_j))^2 - (d^{(2)}(f_i, f_j))^2| \leq \varepsilon_1 \right) \geq 1 - 2e^{-cn\varepsilon_1^2}$$

And $d^{(2)}(f_i, f_j) > \varepsilon \Rightarrow$

$$\mathbb{P}\left(d_n^{(2)}(f_i, f_j) \geq \sqrt{\varepsilon^2 - \varepsilon_1} \right) \geq 1 - 2e^{-cn\varepsilon_1^2}$$

Let $\varepsilon_0 = \sqrt{\varepsilon^2 - \varepsilon_1}$, $\varepsilon_1 \leq \varepsilon^2$, then

$$\mathbb{P}\left(\cap_{i\neq j} \{ d_n^{(2)}(f_i, f_j) > \varepsilon_0 \} \right) \geq 1 - 2N^2 e^{-cn\varepsilon^4} \quad \square$$

Remark: $A = \cap_{i\neq j} \{ \omega ; d_n^{(2)}(f_i, f_j) > \varepsilon_0 \}$ is not empty, take $\omega \in A$, $x_i = X_i(\omega)$,

we define a uniform law on $\{x_i\}$: $P^{(n)}$

Now we mark $d_P : d$, to emphasize the influence of P on metric $d^{(2)}$.

Theorem 1 $\sup_p N(\delta; \mathcal{F}, d_p) \lesssim (\frac{1}{\delta})^{cv}$, $0 \leq \delta \leq 1$

Pf: Note that $\{f_i\}$ is ε_0-packing of $\mathcal{F}$ with metric $d_{p^{(n)}}^{(2)}$

then

$N = P(\delta; \mathcal{F}, d_p) \leq P(\delta_0; \mathcal{F}, d_{p^{(n)}})$
$\leq N(\frac{\delta_0}{2}; \mathcal{F}, d_{p^{(n)}})$

Define $d_{p^{(n)}}^{\infty}$:

$$d_{p^{(n)}}^{\infty}(f,g) = \max_i |f(x_i) - g(x_i)|, f, g \in \mathcal{F}$$

Then $d_{p^{(n)}}^{\infty}(f,g) \geq d_{p^{(n)}}(f,g)$

$\leq N(\frac{\delta_0}{2}; \mathcal{F}, d_{p^{(n)}}^{\infty})$ $(\delta_0 < 2)$

the growth function of $\mathcal{F}$ given $x_1, \cdots, x_n$

$\leq (\frac{en}{v})^{v}$ (1), v is VC-dim of $\mathcal{F}$.

Let $1 - 2N^2 e^{-cn\delta^4} = 0 \Rightarrow n = \frac{2\log N - \log\frac{1}{2}}{c\delta^4}$

Take $n = \lceil \frac{2\log N - \log\frac{1}{2}}{c\delta^4} \rceil \lesssim \delta^{-4} \log N$.

(1) $\Rightarrow N \lesssim (\frac{\log N}{v\delta^4})^{v} \Rightarrow N^{\frac{1}{v}} \lesssim \frac{2}{\delta^4} \log N^{\frac{1}{2v}}$

$\Rightarrow N^{\frac{1}{v}} \lesssim \frac{1}{\delta^4} \cdot N^{\frac{1}{2v}} \Rightarrow N \lesssim (\frac{1}{\delta^4})^{2v} = (\frac{1}{\delta})^{cv}$ $\square$

VC Class of functions

Lemma

Let $\mathcal{F}$ is a vector space of function: $\mathcal{X} \mapsto \mathbb{R}$
$d = \dim(\mathcal{F}) < \infty$, then $\mathcal{D} = \{ \{x ; f(x) \le 0\}, f \in \mathcal{F} \}$ is a VC class and $V(\mathcal{D}) \le \dim(\mathcal{F})$

Pf: If $V(\mathcal{D}) > \dim(\mathcal{F})$, which means that $\forall \{x_1, \dots, x_d, x_{d+1}\} \subset \mathcal{X}$, $\forall I \subset \{1, \dots, d+1\}$
$\exists f \in \mathcal{F}$ s.t.

$$f(x_i) \le 0, \ i \in I, \quad f(x_i) > 0, \ i \notin I \qquad ①$$

Define $T : \mathcal{F} \mapsto \mathbb{R}^{d+1}$, $Tf = (f(x_1), \dots, f(x_{d+1}))$

Since $\dim(T\mathcal{F}) \le d$, then $\exists H$ s.t. $\dim(H) \ge 1$

$$\mathbb{R}^{d+1} = T\mathcal{F} \oplus H$$

$\Rightarrow \exists (y_1, \dots, y_{d+1}) \ne 0$, $\sum_{i=1}^{d+1} y_i f(x_i) = 0$, $\forall f \in \mathcal{F}$

$\Rightarrow \quad \sum_{\{i, y_i \le 0\}} y_i f(x_i) = -\sum_{\{i, y_i > 0\}} y_i f(x_i)$

Let $I = \{i, y_i \ge 0\}$, from ①, contradiction.

Example Half space $\{ \{x \in \mathbb{R}^d ; \theta^T x \ge t\}, \binom{\theta}{t} \in \mathbb{R}^{d+1} \}$
$V(\text{Half Space}) \le d+1$

Example (Sphere in $\mathbb{R}^d$)

$$\{ \underbrace{\bar{B}(a ; \rho)}, \binom{a}{\rho} \in \mathbb{R}^{d+1} \}$$

$$\{ x \in \mathbb{R}^d ; \underbrace{\|x-a\|^2 - \rho^2} \le 0 \}$$

$\|x\|^2 - 2\langle x, a\rangle + \|a\|^2 - \rho^2 \Rightarrow (a, \|a\|^2 - \rho^2)$

$\Rightarrow \|x\|^2 - 2\langle x, d\rangle + d \overset{\text{linear}}{\Longleftrightarrow} (c, d) \in \mathbb{R}^{d+1}$

$\Rightarrow$ embedding in $\mathbb{R}^{d+1} \rightarrow VC(\text{Sphere}) \leq d+1$

Def

$\mathcal{F}$ is a function space : $\mathcal{X} \mapsto \mathbb{R}$. $\mathcal{F}$ is called VC class if the subgraph

$D_{\mathcal{F}} = \{ \{(x,t) ; f(x) \geq t\}, f \in \mathcal{F}\}$ is VC class.

And $VC(\mathcal{F}) \triangleq VC(D_{\mathcal{F}})$

Example $\mathcal{D}$ is VC class $\Leftrightarrow \mathcal{F} = \{\mathbb{1}_D ; D \in \mathcal{D}\}$ is VC

Pf: $\forall D \in \mathcal{D}$,

$\{(x,t) ; \mathbb{1}_D(x) \geq t\} = (\mathcal{X} \times (-\infty, 0]) \cup (D \times (0,1]) \cup (\phi \times (1, +\infty))$

$\overset{\phi}{\Longleftrightarrow} D$

Since ϕ is bijection, claim holds.

Example $\mathcal{F}$ is a vector space of function : $\mathcal{X} \mapsto \mathbb{R}$, $\dim(\mathcal{F}) < \infty$, then $\mathcal{F}$ is VC and $VC(\mathcal{F}) \leq \dim(\mathcal{F}) + 1$

Pf: $\{(x,t) ; f(x) \geq t\} = \{(x,t) ; t - f(x) \leq 0\}$

$t - f(x) \overset{\text{linear}}{\Longleftrightarrow} (f, t) \Rightarrow VC(\mathcal{F}) \leq \dim(\mathcal{F}) + 1$

Theorem

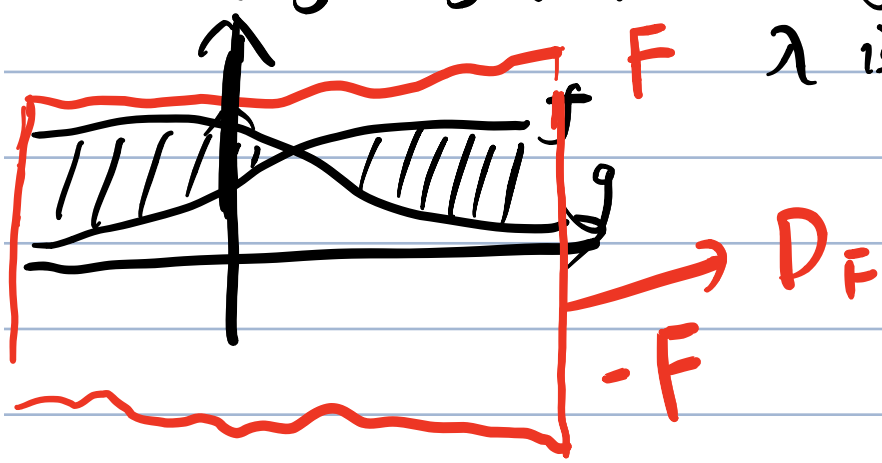
$\mathcal{F}$ is a VC-class of function with $v = V(\mathcal{F})$ and F is the envelope function of $\mathcal{F}$ ($|f(x)| \le F(x)$, $\forall f \in \mathcal{F}$) s.t. $R = (\mathbb{E} F(X)^2)^{\frac{1}{2}} < \infty$, $X \sim (\mathcal{X}, \Sigma, P)$

$$\sup_P N(\delta R; \mathcal{F}, d_P) \lesssim \left(\frac{1}{\delta}\right)^{cv}$$

Pf: $\forall f, g \in \mathcal{F}$, let $D_f = \{(x,t); f(x) \ge t\}$

then $\mathbb{E}|f(X) - g(X)| = P \times \lambda(D_f \triangle D_g)$,

λ is Lebesgue measure.



Then $\mathbb{E}(f(X)-g(X))^2 \le 2\,\mathbb{E}|f(X)-g(X)|F(X)$

$$= 2\int |f(x)-g(x)|\,F(x)\,dP(x)$$

$$= 2\,\mathbb{E}_Q|f(X)-g(X)| \quad \left(\frac{dQ}{dP} = F\right)$$

$$= 2\,Q\times\lambda(D_f \triangle D_g)$$

Note that $D_f \triangle D_g \subset \{(x,t); |t| \le F(x)\} \triangleq D_F$

Define new probability measure on D_F:

$$A \in \sigma(D_F),\ T(A) =: \frac{Q\times\lambda(A)}{Q\times\lambda(D_F)} = \frac{Q\times\lambda(A)}{2\mathbb{E}_Q F(X)} = \frac{Q\times\lambda(A)}{2R^2}$$

Then $d(f,g) \le \sqrt{2Q\times\lambda(D_f \triangle D_g)}$

$$= \sqrt{2T(D_f \triangle D_g)} \cdot \sqrt{2R^2}$$

$$= \sqrt{\mathbb{E}_T(\mathbb{1}_{D_f} - \mathbb{1}_{D_g})^2} \cdot 2R$$

$$\Rightarrow \quad \frac{d_P(f,g)}{2R} \leq d_T(\mathbb{1}_{D_f}, \mathbb{1}_{D_g}),$$

Let $\mathcal{G} := \{\mathbb{1}_{D_f} ; f \in \mathcal{F}\}$, then $\mathcal{G}$ is a class of functions: $\mathcal{X} \times \mathbb{R} \to \mathbb{R}$ and $V(\mathcal{G}) = V(\mathcal{F}) = v$, then

$$N(\delta; \mathcal{F}, d_P/2R) = N(R\delta; \mathcal{F}, d_P)$$
$$\leq N(\delta, \mathcal{G}, d_T) \lesssim \left(\frac{1}{\delta}\right)^{cv}. \qquad \square$$