Teoria de Vapnik-Chervanenkis (VC)

Motivação: Obter majorações uniformes para probabilidade de forte discrepância entre risco e risco empírico, quando a clase de funções classificadors NÃO é contínua.

(X, Y) , $X \in \mathcal{X}$, $Y \in \{0,1\}$

$\mathcal{F} \doteq$ clase de funções clasificadors de $\mathcal{X}$ en $\{0,1\}$

risco de clasificadors $f \in \mathcal{F}$

$$R(f) = \mathbb{P}(f(X) \neq Y)$$

Risco empírico: Dada amostra $(X_1, Y_1), \ldots, (X_n, Y_n)$ iid

$$R_n(f) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{f(X_i) \neq Y_i\}}$$

Des. Hoeffding

$$\mathbb{P}\left(|R_n(f) - R(f)| > \varepsilon \right) \leq 2 e^{-2n\varepsilon^2}$$

depende de f !!!

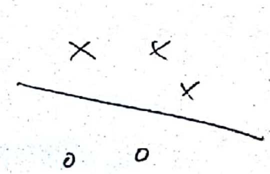
$$\mathbb{P}\left(\bigcup_{f \in \mathcal{F}} \{ |R_n(f) - R(f)| > \varepsilon \} \right) \leq \sum_{f \in \mathcal{F}} \mathbb{P}\left(|R_n(f) - R(f)| > \varepsilon \right) = 2 |\mathcal{F}| e^{-2n\varepsilon^2}$$

Se $\mathcal{F}$ for finito ou numerable $\Rightarrow$ Se $\mathcal{F}$ é finito

Problemão: Em geral $\mathcal{F}$ é contínuo.

Exemplo:

1) $\mathcal{X} = [0,1]$, $\mathcal{F} = \{ f(x) = \mathbb{1}_{\{x < t\}} , t \in [0,1] \}$

2)  $\mathbb{R}^2$ $\mathcal{F} =$ clase de rectas em $\mathbb{R}^2$

$f : \mathcal{X} \to \{0,1\}$

$x_1, x_2, \ldots, x_n$ amostra dada.

$\downarrow$ clasifico

$\{f(x_1), f(x_2), \ldots, f(x_n) : (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n\}$.

De quantas maneiras podemos classificar de forma binaria uma sequência de n pontos? R/ $2^n \leftarrow \boxed{VC}$.

Outra forma de atacar o problema: (mais ingenua).

No exemplo: $\mathcal{X} = [0,1]$

$\mathcal{F} = \{f(x) = \mathbb{1}\{x<t\},\ t \in [0,1]\}$

$0 < r < 1$

$0 \quad \frac{1}{r} \quad \frac{2}{r} \quad 1$.

$\mathcal{F}_r = \{f(x) = \mathbb{1}\{x < kr\} : k = 1, \ldots, r\}$.

Observe que $\forall f \in \mathcal{F}$, $\forall \varepsilon > 0$, $\exists\ \bar{r} = \bar{r}(\varepsilon)$ tal que $\forall r > \bar{r}$ $\exists f' \in \mathcal{F}_r$ satisfazendo $|R(f) - R(f')| \leq \varepsilon$

Dado $\mathcal{X}$ e $\mathcal{F}$ classes de funções de $\mathcal{X}$ em $\{0,1\}$

Amostra de treinamento:

$(X_1, Y_1), \ldots, (X_n, Y_n)$

Definição: <u>Coeficiente de fragmentação</u> (shatter coefficient) $S(\mathcal{F}, n)$

$$S(\mathcal{F}, n) = \max_{(x_1, x_2, \ldots x_n) \in \mathcal{X}^n} |N_{\mathcal{F}}(x_1, \ldots, x_n)| \quad \text{onde}$$

$$N_{\mathcal{F}}(x_1, \ldots, x_n) = \{f(x_1), \ldots, f(x_n) : f \in \mathcal{F}\}.$$

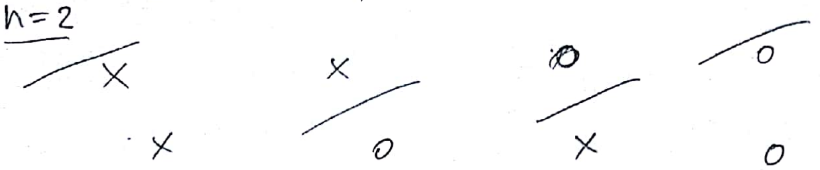
Observe que $S(\mathcal{F}, n) \leq 2^n$.

Exemplo:

$\mathcal{X} = \mathbb{R}^2$

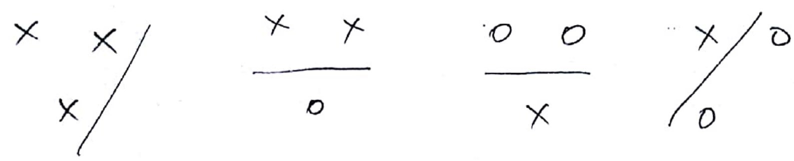
$\mathcal{F}$ = classificadores definidos por rectas

n=2



$S(\mathcal{F},2) = 2^2 = 4$

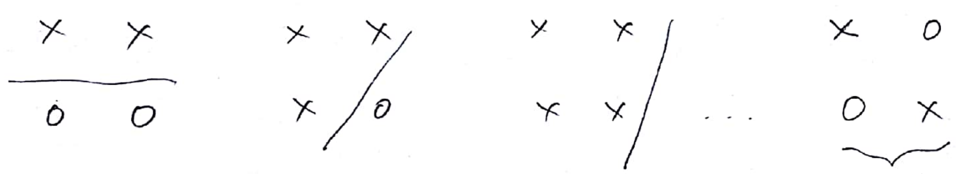
n=3



.... 2^3 casos possíveis

Sempre consigo classificar, então $S(\mathcal{F},3) = 2^3 = 8$.

n=4



Não temos como separar as duas classes com uma recta.

$S(\mathcal{F},4) < 2^4$

Definição A dimenção de VC da classe $\mathcal{F}$ é definida como

$$VC(\mathcal{F}) = \max\{n \geq 1 : S(\mathcal{F},n) = 2^n\}$$

"Tamanho efectivo" da classe $\mathcal{F}$.

Lembro que se $\mathcal{F}$ for finito ou enumeravel

$$\mathbb{P}\left(\bigcup_{f\in\mathcal{F}} |R_n(f) - R(f)| > \varepsilon\right) \leq \sum_{f\in\mathcal{F}} 2e^{-2n\varepsilon^2}$$

$$\leq \underbrace{\boxed{2Ne^{-2n\varepsilon^2}}}_{\delta} \qquad \text{Si } |\mathcal{F}| = N.$$

Teorema: Desigualdade VC

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |R_n(f) - R(f)| > \varepsilon \right) \leq 8\, S(\mathcal{F}, n)\, e^{-n\varepsilon^2/32}.$$

e

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} |R_n(f) - R(f)| \right] \leq 2 \sqrt{\frac{\log S(\mathcal{F}, n) + \log 2}{n}}$$

Lema de Saver

$$S(\mathcal{F}, n) \leq (n+1)^{VC(\mathcal{F})}$$

Usando o Lema de Saver e a desigualdade VC demostramos o seguinte colorano:

Corolánio: Seja $\hat{f}_n = \operatorname{argmin} \{ R_n(f) : f \in \mathcal{F} \}$.

Então $$\mathbb{E}\left[R(\hat{f}_n) \right] - \inf_{f \in \mathcal{F}} R(f) \leq 4 \sqrt{\frac{\log S(\mathcal{F}, n) + \log 2}{n}}$$

$$\leq 4 \sqrt{\frac{VC(\mathcal{F}) \log(n+1) + \log 2}{n}}$$

Demostração:

$$R(\hat{f}_n) - \inf_{f \in \mathcal{F}} R(f)$$

$$= R(\hat{f}_n) - R_n(\hat{f}_n) + R_n(\hat{f}_n) - \inf_{f \in \mathcal{F}} R(f)$$

$$= \left[R(\hat{f}_n) - R_n(\hat{f}_n) \right] + \sup_{f \in \mathcal{F}} \left[R_n(\hat{f}_n) - R(f) \right]$$

Note que: por definição: $R_n(\hat{f}_n) \leq R_n(f) \ \forall f \in \mathcal{F}$

$$\leq \left[R(\hat{f}_n) - R_n(\hat{f}_n) \right] + \sup_{f \in \mathcal{F}} \left[R_n(f) - R(f) \right]$$

$$\leq \underbrace{\left| R(\hat{f}_n) - R_n(\hat{f}_n) \right|}_{*} + \sup_{f \in \mathcal{F}} \left| R_n(f) - R(f) \right|$$

$$* \leq \sup_{f \in \mathcal{F}} \left| R(f) - R_n(f) \right|$$, por tanto

$$\leq 2 \sup_{f \in \mathcal{F}} \left| R(f) - R_n(f) \right|$$

Agora si calulamos esperanzca los dois lados, e usamos Teorema VC

$$\mathbb{E}\left[R(\hat{f}_n) \right] - \inf_{f \in \mathcal{F}} R(f) \leq 2\, \mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| R(f) - R_n(f) \right| \right]$$

$$\leq 2 \times 2 \sqrt{\frac{\log S(\mathcal{F}, n) + \log 2}{n}}$$