VC dimension : exercises

October 12, 2017

Notation :

We will consider functions $f : \chi \mapsto \{0,1\}$.

If F is a class of such functions and $x_1, \cdots x_n$ is a family of n points in χ, we define the set $N_F(x_1, \cdots x_n)$ as the set of all images of this family of points by the functions in F :

$$N_F(x_1, \cdots x_n) = \{(f(x_1), \cdots f(x_n), f \in F\}$$

We define then the shattering coefficient of F with respect to n points sets in χ, denoted $S(F, n)$, as :

$$S(F, n) = \max |N_F(x_1, \cdots x_n)|$$

where the maximum is taken over all possible set $(x_1, \cdots x_n) \in \chi^n$.

Finally, we define the VC dimension of F as :

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

Exercises :

Determine the VC dimension of the next sets of functions where $\chi = [0,1]$:

- $F = \{f : \chi \mapsto \{0,1\},\ f(x) = 1_{x<t},\ \mathrm{t}\ \in [0,1]\}$
- $F' = \{f : \chi \mapsto \{0,1\},\ f(x) = 1_{x<t}$ or $f(x) = 1 - 1_{x<t},\ \mathrm{t}\ \in [0,1]\}$
- $F = \{f : \chi \mapsto \{0,1\},\ f(x) = 1_{t_1 \leq x < t_2},\ t_1 < t_2 \in [0,1]\}$
- $F' = \{f : \chi \mapsto \{0,1\},\ f(x) = 1_{t_1 \leq x < t_2}$ or $f(x) = 1 - 1_{t_1 \leq x < t_2},\ t_1 < t_2 \in [0,1]\}$
- $F_k = \{f : \chi \mapsto \{0,1\},\ f(x) = \sum_{i=0}^{k} 1_{t_{2i} \leq x < t_{2i+1}}$, for $0 \leq C_0 < \cdots < C_{2k+1} \leq 1\}$ for any $k \geq 1$

Note here that for any F, F' is essentially the same set of functions, the only difference being that it allows to label the points indifferently 1 against 0, or 0 against 1. This apparently harmless technical enhancement is actually not totally insignificant as the VC dimension of F and F' are differents.