VCdim and Sauer Lemma

Definitions. A *range space* $\mathcal{S}$ is a pair (X, C), where X is finite set of *points* and $C \subseteq 2^X$ is a set of *ranges*. For a set $Y \subseteq X$ the *projection* of $\mathcal{S}$ on Y is

$$P_C(Y) = \{f \cap Y | f \in C\}.$$

We say that Y is *shattered* if $P_C(Y) = 2^Y$. The Vapnik-Chervonenkis dimension, VCdim, of $\mathcal{S}$, denoted by $VCdim(\mathcal{S})$, is the maximal cardinality of a shattered subset of X.

[**Sauer Lemma**] If $\mathcal{S} = (X, C)$ is a range space of VCdim d with $|X| = n$ points, then

$$|C| \leq \sum_{i=0}^{d} \binom{n}{i} \leq n^d$$

Proof. For a range space $\mathcal{S}$ we will write $d(\mathcal{S})$ and $n(\mathcal{S})$ for $VCdim(\mathcal{S})$ and $|X|$, respectively. The proof will be by induction on $d(\mathcal{S})+n(\mathcal{S})$. When $d(\mathcal{S})+n(\mathcal{S}) = 0$ we have $|C| \leq 1$. This is because if C contains two elements f_1 and f_2 then any element $x \in (f_1 \backslash f_2) \cup (f_2 \backslash f_1)$ is shattered and then $VCdim(C) \geq 1$.

Assume the result holds for all $n(\mathcal{S}) + d(\mathcal{S}) \leq r$. Let $x \in X$. Consider the following two range spaces

$$\mathcal{S}_1 = (X \backslash \{x\}, C_1) \qquad C_1 = \{f \backslash \{x\} | f \in C\}$$

and

$$\mathcal{S}_2 = (X \backslash \{x\}, C_2) \qquad C_2 = \{f \in C | x \notin f, f \cup \{x\} \in C\}. \square$$

We define

$$g(d, n) = \sum_{i=0}^{d} \binom{n}{i}.$$

Exercise. Show that $|C| = |C_1| + |C_2|$.

Now, $VCdim(\mathcal{S}_1) \leq d$ because if Y is shattered in $\mathcal{S}_1$ then it is also shattered in $\mathcal{S}$. Also since $n(\mathcal{S}_1) + d(\mathcal{S}_1) \leq n - 1 + d < r$ we have

$$|C_1| \leq \sum_{i=0}^{d} \binom{n-1}{i}.$$

Suppose Y is shattered in $\mathcal{S}_2$. We will show that $Y \cup \{x\}$ is shattered in S. For every $Z \subseteq Y$ there is $f_Z \in C_2 \subseteq C$ such that $Y \cap f_Z = Z$. Since $x \notin Y$ we also have $(Y \cup \{x\}) \cap C_Z = Z$. Since $f_Z \in C_2$ we also have $\{x\} \cup f_Z \in C$ and therefore

$$(Y \cup \{x\}) \cap (\{x\} \cup f_Z) = Z \cup \{x\}.$$

Therefore $Y \cup \{x\}$ is shattered in $\mathcal{S}$. Since $Vcdim(\mathcal{S}) = d$ we have $d(\mathcal{S}_2) = VCdim(\mathcal{S}_2) \leq d-1$. Therefore $d(\mathcal{S}_2) + n(\mathcal{S}_2) \leq d+n-2$ and by the induction hypothesis we have

$$|C_2| \leq \sum_{i=0}^{d} \binom{n-1}{i-1}.$$

Therefore,

$$|C| = |C_1| + |C_2| \leq \sum_{i=0}^{d} \binom{n-1}{i} + \sum_{i=0}^{d} \binom{n-1}{i-1} = \sum_{i=0}^{d} \binom{n}{i}. \square$$

Example. Let $[n] = \{1, \ldots, n\}$ and for two real numbers a, b and r let $O_{a,b,r} = \{(x,y) \in [n]^2 | (x-a)^2 + (y-b)^2 \leq r\}$. That is, $O_{a,b,r}$ is the set of all integer points that are bounded in a circle with the center at the point (a,b) and radius r. Let

$$C = \{O_{a,b,r} | a, b, r \in \mathcal{R}\}.$$

Find an upper bound for $|C|$.
Solution. Let $X = [n]^2$ and consider the range space (X, C). It is easy to prove that the VCdim of C is 3. This can easily be proved by showing that for any 4 points in X there is set of points $Y \subset X$ such that any circle that contain Y will also contain another point from $X \backslash Y$. Therefore by the lemma we have

$$|C| \leq \sum_{i=0}^{3} \binom{n^2}{i} = \frac{n^6}{6} + O(n^4).$$

We now show

Let $\mathcal{S} = (X, C)$ be a range space of VC-dimension $d \geq 2$ and let $g : \{0,1\}^h \to \{0,1\}$ be any boolean function. Let $S(g) = (X, g(C))$ be a range space on X in which

$$g(C) = \{g(f_1, \ldots, f_h) \mid f_1, \ldots, f_h \in C\}.$$

Then

$$VCdim(S(g)) \leq 2dh \log(dh).$$

Proof. Let $A \subset X$, $|A| = n = VCdim(S(g))$ shattered in $S(g)$. By Sauer Lemma

$$|P_C(A)| \leq g(d,n) \leq n^d.$$

Since each member of $P_{g(C)}(A)$ is g of h members of $P_C(A)$ it follows that

$$|P_{g(C)}| \leq \binom{g(d,n)}{h} \leq n^{dh}.$$

Since A is shattered in $S(g)$ then we must have $n^{dh} \geq 2^n$. This implies the result when $dh \geq 4$.$\square$

[1] N. Sauer. On the density of families of sets, *J. Combinatorial Theory*, Ser. A 13, pp. 145-147, (1972)