20.1 VC Dimension

Lemma 1. (Sauer-Shelah-Vapnik-Chervonenkis)
S is a set of cardinality n. Assume that $\Delta^{\mathcal{C}}(S) > \binom{n}{\leq k-1}$. Then $\exists\ F \subset S$ with $\text{card}(F) = k$ such that F is shattered by $\mathcal{C}$.
Proof: $S = \{x_1, \dots, x_n\}$, $T = \{I_C(x_1), \dots, I_C(x_n), C \in \mathcal{C}\}$, $\text{card}\, T > \binom{n}{\leq k-1}$. $J \subseteq \{1, \dots, n\}$, $t = (t_1, \dots, t_n)$, $\pi^J t = (t_{j1}, \dots, t_{jk})$, where $J = \{j1, \dots, jk\}$.
Idea: construct a sequence of sets

$$T = T_0 \to T_1 \to \cdots T_n$$

such that
(a) $\text{card}(T_i) = \text{card}(T_{i+1})$ for $i = 0, \dots, n-1$ and
(b) J is shattered by T_i implies it is shattered by T_{i+1}.

J is shattered $\Leftrightarrow \pi_J T = \{0,1\}^{\text{card}(J)}$:

$$T_0 \to^{\tau_1} T_1$$

is constructed as follows:

$$t = (t_1, \dots, t_n) \mapsto t' = (t'_1, \dots, t'_n)$$

where

1. $t'_2 = t_2, \dots, t'_n = t_n$
2. If $t_1 = 0$ then $t'_1 = 0$
3. If $t_1 = 1$ and $(0, t_2, \dots, t_n) \in T$ then $t'_1 = 1$
4. If $t_1 = 1$ and $(0, t_2, \dots, t_n) \notin T$ then $t'_1 = 0$

(a) and (b) hold. Similarly, for $i = 2, \dots, n-1$, we construct τ_i

1. $t'_{i+1} = t_{i+1}, \dots, t'_n = t_n$
2. If $t_i = 0$ then $t'_i = 0$
3. If $t_i = 1$ and $(t_1, \dots, 0, \dots, t_n) \in T$ then $t'_i = 1$
4. If $t_1 = 1$ and $(t_1, \dots, 0, \dots, t_n) \notin T$ then $t'_i = 0$

Then assume $t \in T_n$ and $s \le t$. Then $s \in T_n$, because $t \in T_n$ and $t_i = 1$ for some i, which implies $(t_1, \ldots, 0, \ldots, t_n) \in T_n$.

If $\text{card}(T_n) > \binom{n}{\le k-1}$, then there exists a set J of cardinality k such that J is shattered by T_n. If $t_* \in T_n$ such that $\text{card}\{j : t_{*j} = 1\} \ge k$, then that set is shattered. The set of different vectors t such that $\text{card}\{j : t_j = 1\} = i$ is

$$S_{n,i} = \{t : \text{card}\{j : t_j = 1\} = i\}.$$

Then $\text{card}(S_{n,i}) = \binom{n}{i}$ and

$$\text{card}(S_{n,0} \cup \cdots \cup S_{n,k-1}) = \binom{n}{\le k-1} \Rightarrow \exists t \text{ s.t. } t \in S_{n,i}$$

for $i \ge k$.

We will prove two facts about VC classes.
Let $A_1, \ldots, A_k$ be a collection of subsets of S. Then $\mathcal{A}(A_1, \ldots, A_k)$ is the Borel algebra generated by the A_i.

Theorem 1. Let $\mathcal{C}$ have finite VC dimension. Define

$$\mathcal{C}^{(k)} = \bigcup \{\mathcal{A}(C_1, \ldots, C_k) : C_i \in \mathcal{C}\}.$$

Then $\mathcal{C}^{(k)}$ has finite VC dimension.
Proof: $\text{card}(\mathcal{A}(C_1, \ldots, C_k)) \le 2^{2^k}$. There are 2^k possible disjoint constituent sets from taking complements and intersections; therefore there are 2^{2^k} ways to combine them.
Let F be a finite set of cardinality n.

$$\mathcal{C}^{(k)} \cap F = (\mathcal{C} \cap F)^{(k)} = \bigcup \{\mathcal{A}(C_i, \ldots, C_k) : C_i \in \mathcal{C}\}.$$

Then $\Delta^{\mathcal{C}^{(k)}}(F) = \text{card}((\mathcal{C} \cap F)^{(k)})$. Also $\text{card}(\mathcal{C} \cap F) \le m^{\mathcal{C}}(n)$. There are at most $m^{\mathcal{C}}(n)^k$ choices of $C_1, \ldots, C_k \in \mathcal{C} \cap F$. For each of these there are at most 2^{2^k} subsets in $\mathcal{A}(C_1, \ldots, C_k)$. Then $\text{card}((\mathcal{C} \cap F)^{(k)}) \le (m^{\mathcal{C}}(n))^k 2^{2^k}$. Since $\mathcal{C}$ has finite VC dimension

$$m^{\mathcal{C}}(n) \le \left(\frac{ne}{V}\right)^V,$$

and

$$(m^{\mathcal{C}}(n))^k 2^{2^k} \le \left(\frac{ne}{V}\right)^{Vk} 2^{2^k},$$

which is still polynomial in N. Thus the conclusion follows from last week's theorem.