Warm Up

1. \( \binom{k}{i} + \binom{k}{i-1} = ? \)
   - \( \binom{k}{i} \): don’t choose the element
   - \( \binom{k}{i-1} \): do choose the element
   
   We can recover the choice based on whether or not "k+1" in set. So the term count disjunctions.
   
   Thus: \( \binom{k}{i} + \binom{k}{i-1} = \binom{k+1}{i} \)

2. \( \sum_{i=0}^{k} \binom{k}{i} = ? \)
   - Solution 1:
     choose subsets of all, size = 2^k
     \( \binom{k}{0} + \binom{k}{1} + \ldots + \binom{k}{k-1} + \binom{k}{k} = 2^k \)
   - Solution 2:
     We know: \( (1 + x)^k = \sum_{i=0}^{k} \binom{k}{i} x^i \)
     Here x=1 and we get:
     \( \sum_{i=0}^{k} \binom{k}{i} = 2^k \)

\( \epsilon \)-nets and Inductive Generalization

Objective: Show an strengthening of Occam’s razor using VC-dimension of \( H \)

For any target concept \( c \), the class of error regions of \( H \) with respect to \( c \) is:
\[ \Delta_H(c) = \{ [h \neq c](x) : h \in H \} \]

The set of error regions of probability \( \epsilon \) under a fixed distribution \( D \) is:
\[ \Delta_{H,\epsilon} = \{ r \in \Delta_H(c) : Pr(x \in D)[x \in r] \geq \epsilon \} \]

For any \( \epsilon > 0 \) we say that \( S \subseteq X \) is an \( \epsilon \)-net for:
\[ \Delta_{H,\epsilon} - r \] if for every \( r \in \Delta_{H,\epsilon} \) there is some \( x \in r \cap S \).

Example: Target concept \( C(n) = \Phi(0) \)

\( H \): intervals
\( D \): uniform distribution on \([0, 1]\]
\( \Delta_{H,\epsilon} = \text{intervals of size } > \epsilon \)
\( S = [k\epsilon, k=1, \ldots, \frac{1}{\epsilon}] \)
\( S \subseteq [0, 1] \)
\( \Gamma \subseteq \Delta_{H,\epsilon} \)
\( |S| = m \rightarrow Pr(S \text{ doesn’t hit gamma}) \leq (1 - \epsilon)^m \)

Definition of Sauer’s Lemma

For a set of points \( S = \{x^1, \ldots, x^m\} \), the set of dichotomies realized by a class \( H \) on \( S \), \( \prod_H(S) \), is the set of all distinct labeling \( H \) can give \( S \)

The VC-dimension of \( H \) is the size of the largest set \( S \), shattered by \( H \).

The growth function of a representation class \( H \), \( \prod_H : N \rightarrow N \) is given by:
\[ \prod_H(m) : \text{max}(S \subseteq X : |S| = m) | \prod_H(S) | \]

Sauer’s Lemma: If VCD(\( H \)) = d, then for every \( m \),
\[ \prod_H(m) \leq \sum_{i=0}^{d} \binom{m}{i} = O(m^d) \]
Proof of Sauer’s Lemmas

Lemma 1
For any \( m, d \in \mathbb{N} \), define \( \Phi_d(m) \) inductively by: \( \Phi_0(m) = \Phi_d(0) = 1 \), otherwise \( \Phi_d(m) = \Phi_{d-1}(m-1) = \Phi_d(m-1) \). If \( \text{VCD}(H) = d \), then: \( \Pi_H(m) \leq \Phi_d(m) \).

Proof: By definition on \( m, d \)

Base: \( d = 0 \) or \( m = 0 \). If \( d = 0 \), \( H \) must only contain constant function \( \Pi_H(0) = 1 \, \leq \Phi_d(0) \); if \( m = 0 \), the only subset is \( \emptyset \) so \( \Pi_H(\emptyset) = 1 \leq \Phi_d(0) \).

Induction hypothesis: for any \( m', d' \), such that \( m' \geq m \), or \( d' \geq d \), \( \Pi_H(m') = 1 \, \leq \Phi_{d'}(m') \).

Induction step: Let any set \( S \) of size \( m \) be given, \( H \) of VC-dimension \( d \). Choose any \( x \in S \), since \( |S - \{x\}| \) has size \( m-1 \), by induction hypothesis \( |\Pi_H(S - \{x\})| \leq \Phi_d(m-1) \). \( \Pi_H(S - \{x\}) \) is the set of all labelings of \( S - \{x\} \) contains all labelings of \( S \), projected down to \( S - \{x\} \).

So

\[
\Pi_H(S) \leq |\Pi_H(S - \{x\})| + |\Pi_B(S - \{x\})|
\]

Where \( B = \{ r \in \Pi_H(S) : x \in r, r - \{x\} \in \Pi_H(S) \} \)

Consider \( \Pi_B(S - \{x\}) \), \( \text{VCD}(B) \leq d-1 \).

Suppose \( S' \subseteq S - \{x\} \) is shattered by \( B \). Then \( S' \cup \{x\} \) is shattered by \( H \), since both \( r \in S', r \cup \{x\} \in \Pi_H(S) \).

In other words, \( x \) can take two labels.

So \( \text{VCD}(B) \leq \text{VCD}(H) - 1 \).

Since for every arbitrary \( S \), \( 2|\Pi_B(S - \{x\})| \leq |\Pi_H(S)| \)

So \( d' = d - 1 \)

So \( |\Pi_B(S - \{x\})| \leq \Phi_{d-1}(m-1) \) by definition hypothesis, and hence

\[
\Pi_H(S) \leq \Phi_d(m-1) + \Phi_d - 1(m - 1) = \Phi_d(m) \quad \text{by definition}
\]

Example:

\[
\begin{array}{|c|c|c|}
\hline
x_1 & x_2 & x_3 \\
\hline
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
\hline
\end{array}
\]

\[\Pi_H(S)\]

\[
\begin{array}{|c|c|}
\hline
x_1 & x_2 \\
\hline
0 & 1 \\
1 & 1 \\
1 & 0 \\
\hline
\end{array}
\]

\[\Pi_H(S - \{x\}) \leq \Phi_d(m-1)\]

\[
\begin{array}{|c|c|}
\hline
x_1 & x_2 \\
\hline
0 & 1 \\
1 & 1 \\
1 & 0 \\
\hline
\end{array}
\]

\[B' = \{ b \in \Pi_H(S) : x \notin b, b \cup x \in \Pi_H(S) \} \]

\[B' = \Pi_{B'}(S - \{x\})\]
\[ |\Pi_H(S)| = |\Pi_H(S - x)| + |\Pi_H'(S - x)| \leq \Phi_d(m - 1) + \Phi_{d-1}(m - 1) = \Phi_d(m) \]

**Lemma 2**

\[ \Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i} \leq \left( \frac{me}{d} \right)^d = O(m^d) \quad \text{for} \quad (m > d) \]

**Proof:** By induction on \( m, d \)

Base: \( m = 0 \). \( \sum_{i=0}^{d} \binom{0}{i} = l = \Phi_d(0) \), since \( d=0 \sum_{i=0}^{0} \binom{m}{i} = 1 = \Phi_0(m) \)

Induction hypothesis: As in claim

Induction step:

\[ \Phi_d(m) = \Phi_d(m - 1) + \Phi_{d-1}(m - 1) \]
\[ = \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d-1} \binom{m-1}{i} \]
\[ = \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d} \binom{m-1}{i-1} \]
\[ = \sum_{i=0}^{d} \left( \binom{m-1}{i} + \binom{m-1}{i-1} \right) \]
\[ = \sum_{i=0}^{d} \binom{m}{i} \]

Complete inductor

For the second part, for \( m > d, \frac{d}{m} < 1 \)

\[ \frac{d}{m} \sum_{i=0}^{d} \binom{m}{i} \leq \sum_{i=0}^{d} \frac{d}{m} \binom{m}{i} \]
\[ \leq \sum_{i=0}^{m} \frac{d}{m} \binom{m}{i} \]
\[ = (1 + \frac{d}{m})^m \]
\[ \leq e^d \]

For \( d > 0 \)

\[ \sum_{i=0}^{d} \binom{m}{i} \leq \left( \frac{me}{d} \right)^d \]