Notes 13: Sauer–Shelah lemma

1. Sauer–Shelah Lemma

Claim 1. \( |\Pi_C(S)| \leq |\{T \subseteq S \mid \mathcal{C} \text{ shatters } T\}| \)

Proof. Apply following Proposition with \( \mathcal{F} = \Pi_C(S) \)

\( \text{ Note that } T \text{ is shattered by } \mathcal{C} \text{ if and only if } T \text{ is shattered by } \mathcal{F} = \Pi_C(S) \)

Proposition 2 (Pajor). A finite family \( \mathcal{F} \) of subsets over \( S \) shatters at least \( |\mathcal{F}| \) subsets, i.e.

\[ |\mathcal{F}| \leq \# \text{subsets } \mathcal{F} \text{ shatters} = |\{T \subseteq S \mid \mathcal{F} \text{ shatters } T\}| \]

\[ \mathcal{F} = \left\{ \{1, 2, 3\}, \{2, 3, 4\}, \{1, 2, 3, 4\} \right\}, \quad \{1\}, \{4\}, \emptyset \]

Proof of Proposition. Base case \( |\mathcal{F}| = 0 \): trivial

Base case \( |\mathcal{F}| = 1 \): \( \mathcal{F} \) shatters \( \emptyset \)

Induction step for \( |\mathcal{F}| > 1 \):

Fix \( x \in S \) belonging to some but not all of the sets in \( \mathcal{F} \)

Split \( \mathcal{F} \) into \( \mathcal{F}_{\geq x} \) and \( \mathcal{F}_{\leq x} \) (those containing \( x \) and those do not)

Induction hypothesis implies \( \mathcal{F}_{\geq x} \) shatters \( > |\mathcal{F}_{\leq x}| \) subsets, \( \mathcal{F}_{\leq x} \) shatters \( > |\mathcal{F}_{\geq x}| \) subsets

\[ |\mathcal{F}| = |\mathcal{F}_{\geq x}| + |\mathcal{F}_{\leq x}| \leq \# \text{subsets } \mathcal{F}_{\geq x} \text{ shatters} + \# \text{subsets } \mathcal{F}_{\leq x} \text{ shatters} \]

Remains to show right-hand-side \( \leq \# \text{subsets } \mathcal{F} \) shatters

Any set shattered by \( \mathcal{F}_{\geq x} \) cannot contain \( x \), since all sets in \( \mathcal{F}_{\geq x} \) contain \( x \)

Any set shattered by \( \mathcal{F}_{\leq x} \) cannot contain \( x \), since all sets in \( \mathcal{F}_{\leq x} \) do not contain \( x \)

Thus any set of the form \( T \cup \{x\} \) cannot be shattered by \( \mathcal{F}_{\geq x} \) or \( \mathcal{F}_{\leq x} \)

If \( T \) is shattered by only one of \( \mathcal{F}_{\geq x} \) or \( \mathcal{F}_{\leq x} \), \( T \) contributes 1 to \#subsets \( \mathcal{F} \) shatters

If \( T \) is shattered by both \( \mathcal{F}_{\geq x} \) and \( \mathcal{F}_{\leq x} \), then \( T \) and \( T \cup \{x\} \) are both shattered by \( \mathcal{F} \)

\( T \) and \( T \cup \{x\} \) together contribute 2 to \#subsets \( \mathcal{F} \) shatters

Lemma 3 (Perles–Sauer–Shelah). When \( \text{VCDim}(\mathcal{C}) = d \), \( \Pi_C(m) \leq \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{d} \)

Proof. By above Claim, at most \( \sum_{0 \leq k \leq d} \binom{m}{k} \) choices for shattered subset \( T \)

No subset larger than \( d = \text{VCDim}(\mathcal{C}) \) is shattered

Corollary 4. When \( \text{VCDim}(\mathcal{C}) = d \) and \( m \geq d \), \( \Pi_C(m) \leq \left( \frac{em}{d} \right)^d \)

Proof. Want to show \( \sum_{0 \leq k \leq d} \binom{m}{k} \leq \left( \frac{em}{d} \right)^d \) for \( m \geq d \)

\[ \left( \frac{d}{m} \right)^d \sum_{0 \leq k \leq d} \binom{d}{k} \binom{m}{k} \leq \sum_{0 \leq k \leq d} \binom{d}{k} \binom{m}{k} = \left( 1 + \frac{d}{m} \right)^m \leq \left( \frac{e}{d/m} \right)^m = e^d \]

First inequality due to \( d/m \leq 1 \)

Second inequality due to \( d \leq m \)

Next equality is binomial theorem

Last inequality is \( 1 + x \leq e^x \) for all real \( x \)
2. Consistent Hypothesis

**Theorem 5.** Given \( m \) independent labelled samples, with prob. \( \geq 1 - \delta \), any hypothesis consistent with all \( m \) samples has error at most \( \varepsilon \), provided

\[
m \geq \Omega \left( \frac{1}{\varepsilon} \log \frac{\Pi_C(2m)}{\delta} \right)
\]

Compared with notes09, now \( C \) may be infinite
notes09 was union bound over \( \mathcal{H} \); now over dichotomies on \( 2m \) samples

**Proof.** Imagine drawing \( 2m \) labelled samples \((x^i, c(x^i))\) from \( \text{EX}(c, \mathcal{D}) \)

Call first \( m \) samples \( S_1 \); last \( m \) samples \( S_2 \)

Event \( A \): Some bad \( h \in C \) is consistent with \( S_1 \)
Recall \( h \) is bad if \( \text{err}_D(h, c) \geq \varepsilon \); Goal: show \( P[A] \leq \delta \)

Event \( B \): Some \( h \in C \) is consistent with \( S_1 \) but wrong on \( \geq \varepsilon m/2 \) samples in \( S_2 \)

**Claim 6.** If \( m \geq 8/\varepsilon \), then \( P[A] \leq 2P[B] \)

**Proof of Claim.** \( P[B] \geq P[B \text{ and } A] = P[A]P[B \mid A] \)
Suffice to show \( P[B \mid A] \geq 1/2 \)

When \( A \) occurs, fix any bad \( h \), \( P[h \text{ makes at most } \varepsilon m/2 \text{ mistakes on } S_2] \leq e^{-\frac{1}{2}\varepsilon m} \leq 1/e \leq 1/2 \)

Using Claim, suffices to show \( P[B] \leq \delta/2 \)

Equivalent way to view \( B \):

(1) First draw \( 2m \) independent labelled samples \( S \)
(2) Randomly split \( S \) into two halves, \( S_1 \) and \( S_2 \) (first and second halves)
(3) Event \( B \): \( S_1 \) contains no mistakes, \( S_2 \) contains \( \geq \varepsilon m/2 \) mistakes

Now fix any \( 2m \) instances \( S \) and a labeling/dichotomy of \( S \) (from \( \Pi_C(S) \)) from step (1)
Event \( B \) is equivalent to \( \geq \varepsilon m/2 \) mistakes in \( S \) all falling in \( S_2 \)

Combinatorial experiment: \( 2m \) balls (\( S \)), each colored red (mistake) or blue (correct)

- Exactly \( \ell \) are red (\( \ell \geq \varepsilon m/2 \))
- Randomly put \( m \) balls into \( S_1 \) and the other \( m \) balls into \( S_2 \)
- Probability that all red balls fall into \( S_2 \) is \( \frac{\binom{m}{\ell}}{\binom{2m}{\ell}} \)

\[
\frac{\binom{m}{\ell}}{\binom{2m}{\ell}} = \frac{m}{2m} \frac{m-1}{2m-1} \cdots \frac{m-\ell+1}{2m-\ell+1} \leq \left( \frac{1}{2} \right)^\ell
\]

Union bound over at most \( \Pi_C(S) \) labelings of \( S \) with \( \ell \geq \varepsilon m/2 \):

\[
P[B] \leq \frac{\Pi_C(2m)}{2^{\varepsilon m/2}} \leq \frac{\delta}{2} \quad \text{when } m \geq \frac{2}{\varepsilon} \log \frac{2\Pi_C(2m)}{\delta}
\]

**Advantage of Event \( B \) over Event \( A \):**

union bound over finitely many (in fact \( \Pi_C(2m) \)) labelings; even when \( C \) is infinite