Core Idea of the Sauer Lemma

1. The Main Statement to Be Proved

The entire proof is ultimately aimed at establishing the following key inequality:

$$\boxed{\forall C, \quad |\mathcal{H}_C| \;\le\; \big|\{\, B \subseteq C : \mathcal{H} \text{ shatters } B \,\}\big|} \tag{$\star$}$$

Here,

- the left-hand side,
$$|\mathcal{H}_C|,$$
denotes the number of distinct labelings that the hypothesis class $\mathcal{H}$ can realize on the set C;
- the right-hand side denotes the number of subsets of C that are shattered by $\mathcal{H}$.

Equivalently,

$$\text{number of labelings} \;\le\; \text{number of shattered subsets.}$$

This inequality constitutes the core mechanism of the Sauer Lemma.

2. Why This Implies the Sauer Lemma

Assume that the VC dimension of $\mathcal{H}$ satisfies $\mathrm{VCdim}(\mathcal{H}) \le d$. Then, for any subset $B \subseteq C$,

$$|B| > d \;\Rightarrow\; B \text{ cannot be shattered by } \mathcal{H}.$$

Consequently, only subsets of size at most d can be shattered, and we obtain

$$\big|\{\, B \subseteq C : \mathcal{H} \text{ shatters } B \,\}\big| \;\le\; \sum_{i=0}^{d} \binom{m}{i},$$

where $m = |C|$.

Therefore, once inequality $(\star)$ is established, it immediately follows that

$$|\mathcal{H}_C| \;\le\; \sum_{i=0}^{d} \binom{m}{i},$$

which is precisely the conclusion of the Sauer Lemma.

3. What the Proof Actually Does (Structural Overview)

The proof relies on a single decomposition argument.

(1) Decomposition of Labelings

We decompose the set of labelings as

$$|\mathcal{H}_C| = |Y_0| + |Y_1|,$$

where

- Y_0 consists of labelings for which the label of c_1 is not fixed;
- Y_1 consists of labelings for which both label values of c_1 are attainable.

Example (Why $|\mathcal{H}_C| = |Y_0| + |Y_1|$)

Let

$$C = \{c_1, c_2, c_3\}.$$

Consider the hypothesis class whose induced labelings on C are

$$\mathcal{H}_C = \{000,\ 100,\ 010,\ 011\}.$$

Thus,

$$|\mathcal{H}_C| = 4.$$

We now explain how this number can be recovered by the decomposition

$$|\mathcal{H}_C| = |Y_0| + |Y_1|.$$

Step 1: Decomposition into front and tail

Each labeling is written as

$$(y_1 \mid y_2, y_3),$$

where (y_2, y_3) is referred to as the *tail pattern.* The labelings decompose as follows:

Labeling	y_1	(y_2, y_3)
000	0	$(0,0)$
100	1	$(0,0)$
010	0	$(1,0)$
011	0	$(1,1)$

Step 2: Definition of Y_0

Let Y_0 be the set of tail patterns that appear at least once. From the table above,

$$Y_0 = \{(0,0), (1,0), (1,1)\}, \qquad |Y_0| = 3.$$

Each element of Y_0 contributes at least one labeling.

Step 3: Definition of Y_1

Let Y_1 be the set of tail patterns for which both values $y_1 = 0$ and $y_1 = 1$ occur.

- For $(0,0)$, both $(0,0,0)$ and $(1,0,0)$ appear $\Rightarrow$ included in Y_1.
- For $(1,0)$, only $(0,1,0)$ appears $\Rightarrow$ not included.
- For $(1,1)$, only $(0,1,1)$ appears $\Rightarrow$ not included.

Hence,

$$Y_1 = \{(0,0)\}, \qquad |Y_1| = 1.$$

Step 4: Counting

Each tail pattern contributes one labeling by default, and each element of Y_1 contributes one additional labeling. Therefore,

$$|\mathcal{H}_C| = |Y_0| + |Y_1| = 3 + 1 = 4.$$

Interpretation

The set Y_0 accounts for the minimal number of labelings, while Y_1 accounts for the additional labelings arising from the ability to flip the label of c_1. Thus, Y_0 provides the base count and Y_1 provides the surplus.

(2) Bounding Each Term by Shattered Sets

Each term is bounded as follows:

$$|Y_0| \;\le\; \big|\{\, B \subseteq C : c_1 \notin B,\ \mathcal{H} \text{ shatters } B \,\}\big|,$$

$$|Y_1| \;\le\; \big|\{\, B \subseteq C : c_1 \in B,\ \mathcal{H} \text{ shatters } B \,\}\big|.$$

(3) Combining the Bounds

Summing the two inequalities yields

$$\begin{aligned}
|\mathcal{H}_C| &= |Y_0| + |Y_1| \\
&\le \big|\{\, B \subseteq C : c_1 \notin B,\ \mathcal{H} \text{ shatters } B \,\}\big| \\
&\quad + \big|\{\, B \subseteq C : c_1 \in B,\ \mathcal{H} \text{ shatters } B \,\}\big| \\
&= \big|\{\, B \subseteq C : \mathcal{H} \text{ shatters } B \,\}\big|.
\end{aligned}$$

This is exactly inequality $(\star)$.

Example illustrating the Y_0/Y_1 bounds

Hypothesis class.

$$\mathcal{H} = \left\{ \mathbf{1}_{\{|x| \le r\}} : r \ge 0 \right\}.$$

Point set.

$$C = \{c, e_2\} = \{e_1 = (1, 0),\ e_2 = (0, 1)\}.$$

Note that

$$|e_1| = |e_2| = 1, \qquad \mathrm{VCdim}(\mathcal{H}) = 1.$$

From the previous computation,

$$\mathcal{H}_C = \{(0, 0), (1, 1)\}.$$

1. Definition of Y_0 and Y_1 (OR / AND)

Since

$$C \setminus \{c\} = \{e_2\},$$

the possible labels for e_2 are $y_2 \in \{0, 1\}$.

Y_0 (OR)

$$Y_0 = \{y_2 : (0, y_2) \in \mathcal{H}_C \ \vee\ (1, y_2) \in \mathcal{H}_C\}.$$

- $y_2 = 0$: $(0, 0) \in \mathcal{H}_C$ ✓
- $y_2 = 1$: $(1, 1) \in \mathcal{H}_C$ ✓

Hence,

$$Y_0 = \{0, 1\}, \qquad |Y_0| = 2.$$

Y_1 (AND)

$$Y_1 = \{y_2 : (0, y_2) \in \mathcal{H}_C \ \wedge\ (1, y_2) \in \mathcal{H}_C\}.$$

- $y_2 = 0$: $(1, 0) \notin \mathcal{H}_C$ ✗
- $y_2 = 1$: $(0, 1) \notin \mathcal{H}_C$ ✗

Therefore,

$$Y_1 = \varnothing, \qquad |Y_1| = 0.$$

2. Counting shattered subsets on the right-hand side

(A) $c \notin B$ and B is shattered

$$\{B \subseteq C : c \notin B\} = \{\emptyset, \{e_2\}\}.$$

- $\emptyset$: always shattered.
- $\{e_2\}$: $r < 1 \Rightarrow 0,\ r \geq 1 \Rightarrow 1$, hence shattered.

Thus,

$$\bigl|\{B \subseteq C : c \notin B,\ \mathcal{H} \text{ shatters } B\}\bigr| = 2.$$

(B) $c \in B$ and B is shattered

$$\{B \subseteq C : c \in B\} = \{\{e_1\}, \{e_1, e_2\}\}.$$

- $\{e_1\}$: singleton, hence shattered.
- $\{e_1, e_2\}$: labelings $(0, 1)$ and $(1, 0)$ are impossible, so it is not shattered.

Hence,

$$\bigl|\{B \subseteq C : c \in B,\ \mathcal{H} \text{ shatters } B\}\bigr| = 1.$$

3. Verifying the inequalities numerically

First inequality

$$|Y_0| = 2 \ \leq\ 2 = \bigl|\{B \subseteq C : c \notin B,\ \mathcal{H} \text{ shatters } B\}\bigr|.$$

Equality holds.

Second inequality

$$|Y_1| = 0 \ \leq\ 1 = \bigl|\{B \subseteq C : c \in B,\ \mathcal{H} \text{ shatters } B\}\bigr|.$$

This inequality holds trivially.

4. Key takeaway

- Y_0 can attain its maximal possible size (equality case).
- Y_1 may be completely empty.
- Nevertheless, both are always controlled by the number of shattered subsets.

This demonstrates that the corresponding step in the proof of Sauer's lemma reflects an actual combinatorial phenomenon rather than a purely formal manipulation.

Examples Illustrating the Tightness of the Sauer Inequality

Let $\mathcal{X} = \mathbb{R}^d$. We will demonstrate all the 4 combinations using hypothesis classes defined over $\mathcal{X} \times \{0,1\}$. Remember that the empty set is always considered to be shattered.

- $(<,=)$: Let $d \geq 2$ and consider the class

$$\mathcal{H} = \left\{ \mathbf{1}_{\{\|x\|_2 \leq r\}} : r \geq 0 \right\}$$

of concentric balls. The VC-dimension of this class is 1. To see this, we first observe that if $x \neq (0, \ldots, 0)$, then $\{x\}$ is shattered. Second, if $\|x_1\|_2 \leq \|x_2\|_2$, then the labeling $y_1 = 0, y_2 = 1$ is not obtained by any hypothesis in $\mathcal{H}$. Let $A = \{e_1, e_2\}$, where e_1, e_2 are the first two elements of the standard basis of $\mathbb{R}^d$. Then,

$$\mathcal{H}_A = \{(0,0),(1,1)\}, \quad \{B \subseteq A : \mathcal{H} \text{ shatters } B\} = \{\emptyset, \{e_1\}, \{e_2\}\},$$

and

$$\sum_{i=0}^{d} \binom{|A|}{i} = 3.$$

example (dimension $d \geq 2$ is necessary). Let

$$\mathcal{X} = \mathbb{R}^2, \qquad \mathcal{H} = \left\{ \mathbf{1}_{\{\|x\| \leq r\}} : r \geq 0 \right\}.$$

Choice of points. Let

$$A = \{e_1, e_2\}, \qquad e_1 = (1,0),\ e_2 = (0,1).$$

Both points satisfy

$$\|e_1\| = \|e_2\| = 1.$$

Possible labelings.

- If $r < 1$, then $(0,0)$.
- If $r \geq 1$, then $(1,1)$.

The labelings $(1,0)$ and $(0,1)$ are impossible.

Hence,

$$|\mathcal{H}_A| = 2.$$

Shattered subsets. The subsets of A shattered by $\mathcal{H}$ are

$$\emptyset, \quad \{e_1\}, \quad \{e_2\}.$$

Therefore,

$$\big|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}\big| = 3.$$

Sauer bound (VC-dimension $= 1$).

$$\sum_{i=0}^{1} \binom{2}{i} = 3.$$

Final comparison.

$$|\mathcal{H}_A| = 2 \ < \ 3 = \sum_{i=0}^{1} \binom{2}{i}.$$

Thus this is indeed a $(<, =)$ example.

- $(=, <)$: Let $\mathcal{H}$ be the class of axis-aligned rectangles in $\mathbb{R}^2$. We have seen that the VC-dimension of $\mathcal{H}$ is 4. Let $A = \{x_1, x_2, x_3\}$, where

$$x_1 = (0,0), \quad x_2 = (1,0), \quad x_3 = (2,0).$$

All the labelings except $(1,0,1)$ are obtained. Thus,

$$|\mathcal{H}_A| = 7, \quad |\{B \subseteq A : \mathcal{H} \text{ shatters } B\}| = 7,$$

and

$$\sum_{i=0}^{d} \binom{|A|}{i} = 8.$$

example

Domain.

$$\mathcal{X} = \mathbb{R}^2.$$

Hypothesis class.

$$\mathcal{H} = \left\{ \mathbf{1}_{\{x \in R\}} : R \text{ is an axis-aligned rectangle} \right\},$$

where

$$R = [a,b] \times [c,d].$$

The VC-dimension of this class is a known fact and equals 4 (we do not prove it here).

Choice of the point set A (key step) Let

$$x_1 = (0,0), \quad x_2 = (1,0), \quad x_3 = (2,0).$$

All three points lie on the x-axis and are collinear.

Enumerating possible labelings We write labelings as (y_1, y_2, y_3).

Possible labelings

1. $(0,0,0)$ Choose a rectangle that contains none of the points.
2. $(1,1,1)$ For example, choose $[0,2] \times [-\varepsilon, \varepsilon]$.
3. $(1,1,0)$ For example, choose $[0,1] \times [-\varepsilon, \varepsilon]$.
4. $(0,1,1)$ For example, choose $[1,2] \times [-\varepsilon, \varepsilon]$.
5. $(1,0,0)$ For example, choose $[0,0] \times [-\varepsilon, \varepsilon]$.
6. $(0,1,0)$ For example, choose $[1,1] \times [-\varepsilon, \varepsilon]$.
7. $(0,0,1)$ For example, choose $[2,2] \times [-\varepsilon, \varepsilon]$.

Thus, 7 labelings are achievable.

Impossible labeling: $(1,0,1)$

Suppose we try to realize the labeling $(1,0,1)$.

- x_1 and x_3 must be inside the rectangle.
- x_2 must be outside the rectangle.

However, any axis-aligned rectangle has the form

$$[a,b] \times [c,d].$$

Since

$$x_1.x \le x_2.x \le x_3.x,$$

including both x_1 and x_3 necessarily includes x_2 as well.

Hence, it is impossible to exclude the middle point only.

Conclusion for $\mathcal{H}_A$

$$|\mathcal{H}_A| = 7.$$

Counting shattered subsets

- Every singleton subset is shattered.
- Every subset of size 2 is shattered.
- The full set A is *not* shattered.

Therefore,

$$\Big|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}\Big| = 7.$$

Sauer bound

Since $\text{VCdim}(\mathcal{H}) = 4$ and $|A| = 3$,

$$\sum_{i=0}^{4} \binom{3}{i} = \binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 8.$$

Final comparison

$$|\mathcal{H}_A| = 7 < 8 = \sum_{i=0}^{d} \binom{|A|}{i}.$$

Result. This example realizes the $(=,<)$ case exactly.

- $(<,<)$: Let $d \geq 3$ and consider the class

$$\mathcal{H} = \Big\{\text{sign}(\langle w, x\rangle) : w \in \mathbb{R}^d\Big\}$$

of homogeneous halfspaces. We will prove in Theorem 9.2 that the VC-dimension of this class is d. However, here we only rely on the fact that $\text{VCdim}(\mathcal{H}) \geq 3$. This follows by observing that the set $\{e_1, e_2, e_3\}$ is shattered. Let

$$A = \{x_1, x_2, x_3\}, \quad x_1 = e_1,\ x_2 = e_2,\ x_3 = (1, 1, 0, \ldots, 0).$$

Note that all the labelings except $(1, 1, -1)$ and $(-1, -1, 1)$ are obtained. It follows that

$$|\mathcal{H}_A| = 6, \quad |\{B \subseteq A : \mathcal{H} \text{ shatters } B\}| = 7,$$

and

$$\sum_{i=0}^{d} \binom{|A|}{i} = 8.$$

example

Domain.

$$\mathcal{X} = \mathbb{R}^3.$$

Hypothesis class (homogeneous halfspaces).

$$\mathcal{H} = \left\{ h_w(x) = \mathrm{sign}(\langle w, x \rangle) : w \in \mathbb{R}^3 \right\}.$$

All separating hyperplanes pass through the origin.

2. Choice of the set A (important)

$$A = \{x_1, x_2, x_3\} = \{e_1 = (1,0,0),\ e_2 = (0,1,0),\ x_3 = (1,1,0)\}.$$

We now enumerate all labelings realizable on this set.

3. Enumeration of possible labelings

Let $(y_1, y_2, y_3) \in \{\pm 1\}^3$ denote a labeling.

Achievable labelings (examples)

- $(1,1,1)$: choose $w = (1,1,0)$.
- $(1,-1,1)$: choose $w = (1,-0.2,0)$.
- $(-1,1,1)$: choose $w = (-0.2,1,0)$.
- $(1,-1,-1)$, $(-1,1,-1)$, $(-1,-1,-1)$: appropriate choices of w exist.

Thus, 6 labelings are achievable.

4. Two impossible labelings (key step)

(1) $(1,1,-1)$

The conditions

$$\langle w, e_1 \rangle > 0, \qquad \langle w, e_2 \rangle > 0$$

imply

$$\langle w, (1,1,0)\rangle = \langle w, e_1\rangle + \langle w, e_2\rangle > 0.$$

However, the required label for $x_3 = (1,1,0)$ is -1, which is impossible.

(2) $(-1,-1,1)$

Similarly,

$$\langle w, e_1\rangle < 0, \qquad \langle w, e_2\rangle < 0 \Rightarrow \langle w, (1,1,0)\rangle < 0.$$

But the required label is $+1$, again impossible.

5. Conclusion for this set A

Out of 8 possible labelings, 2 are impossible. Hence,

$$|\mathcal{H}_A| = 6.$$

6. Counting shattered subsets

- Every singleton subset is shattered.
- Every subset of size 2 is shattered.
- The full set A is not shattered.

Therefore,

$$\big|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}\big| = 7.$$

7. Why the Sauer bound equals 8

The shattered set witnessing $\mathrm{VCdim}(\mathcal{H}) \geq 3$ is *not* the set A. For example,

$$\{e_1, e_2, e_3\}, \quad e_3 = (0,0,1),$$

is fully shattered by homogeneous halfspaces.

Thus,

$$\mathrm{VCdim}(\mathcal{H}) \geq 3,$$

and we may use $d = 3$ in Sauer's inequality:

$$\sum_{i=0}^{3} \binom{3}{i} = 8.$$

8. Final comparison

$$|\mathcal{H}_A| = 6 < 7 = \left|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}\right| < 8 = \sum_{i=0}^{d} \binom{|A|}{i}.$$

Result. This example realizes the $(<,<)$ case exactly.

- $(=,=)$: Let $d = 1$, and consider the class

$$\mathcal{H} = \left\{\mathbf{1}_{\{x \geq t\}} : t \in \mathbb{R}\right\}$$

of thresholds on the line. We have seen that every singleton is shattered by $\mathcal{H}$, and that every set of size at least 2 is not shattered by $\mathcal{H}$. Choose any finite set $A \subset \mathbb{R}$. Then each of the three terms in Sauer's inequality equals $|A| + 1$. This is the general case.