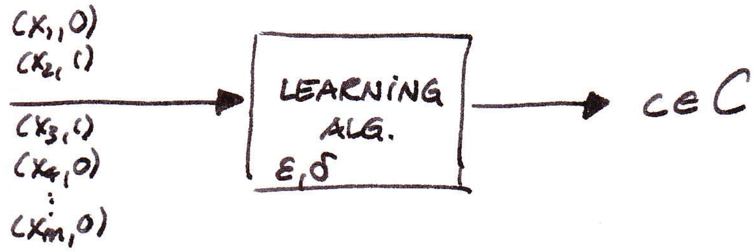
# SAMPLE COMPRESSION AND THE VC-DIMENSION

**MOTIVATION:** concept learning

- instance space: $X$
- concepts: $c$ is a concept on $X$ if $c \subseteq X$.

$$c(x) = \begin{cases} 0 & x \in c \\ 1 & x \notin c \end{cases} \quad \text{for } x \in X.$$

- concept classes: $C$ is a concept class on $X$ if $C \subseteq 2^X$.




- drawn at random wrt $\mathcal{D}$ (unknown)
- labeled wrt $c^* \in C$,
- $m$ polynomial in $\frac{1}{\varepsilon}, \frac{1}{\delta}$

- with prob. $\geq 1-\delta$, the weight of $c \triangle c^*$ under $\mathcal{D}$ is at most $\varepsilon$

"PAC-learning" (Valiant 1984)

- questions:
  - when can we learn every $c \in C$?
  - how large does $m$ have to be to learn every $c \in C$?
  - how large an $m$ is sufficient — " —

- the big answer: the Vapnik-Chervonenkis Dimension (VCD)
  - a combinatorial notion, comb. measure for a concept class, easy to define
  - it determines whether or not a class $C$ is PAC-learnable
  - if $C$ is PAC-learnable, $VCD(C)$ gives us upper and lower bounds on $m$

## THE VC-DIMENSION

**DEFINITION.** Let $C$ be a concept class on $X$ and $Y \subseteq X$, $|Y| < \infty$.
$C$ shatters $Y$ if $|C|_Y| = 2^{|Y|}$.

$X = \{x_1, x_2, x_3, x_4\}$, $C = \{c_1, \dots, c_7\}$

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $c_1$ | 0 | 0 | 0 | 0 |
| $c_2$ | 1 | 0 | 0 | 0 |
| $c_3$ | 0 | 1 | 0 | 0 |
| $c_4$ | 0 | 0 | 1 | 0 |
| $c_5$ | 0 | 0 | 0 | 1 |
| $c_6$ | 1 | 1 | 0 | 0 |
| $c_7$ | 1 | 0 | 1 | 0 |

$C$ shatters $\{x_1, x_2\}$, $\{x_1, x_3\}$
$\{x_1\}$
$\{x_2\}$
$\{x_3\}$
$\{x_4\}$

$VCD(C) = 2$

(Vapnik 1982) (VC, 1971)
**DEFINITION.** $VCD(C) := \max\{|Y| \mid Y \subseteq X,\ Y \text{ finite},\ C \text{ shatters } Y\}$
($VCD(C) = \infty$ if sets $Y$ of arbitrary size are shattered by $C$)

$C = \{c \subseteq X \mid |c| \leq d\} \Rightarrow VCD(C) = d$

$C = \{[a,b] \subseteq \mathbb{R} \mid a \leq b\} \Rightarrow VCD(C) = 2$

$C = \{[a,b] \times [c,d] \subseteq \mathbb{R}^2 \mid a \leq b, c \leq d\} \Rightarrow VCD(C) = 4$

$C = 2^{\mathbb{N}} \Rightarrow VCD(C) = \infty$ ($C = 2^X$ for some infinite $X$)

(Blumer et al. 1989)
**THEOREM.** $C$ is PAC-learnable iff $VCD(C) < \infty$.

**THEOREM.**
- If $VCD(C) = d < \infty$ then $C$ can be PAC-learned with parameters $\varepsilon, \delta$, using
$$m \in O\left(\frac{1}{\varepsilon} \ln \frac{1}{\delta} + \frac{d}{\varepsilon} \ln \frac{1}{\varepsilon}\right)$$
(Blumer et al. 1989)
- If $VCD(C) \geq d\ (< \infty)$ then any PAC-learner for $C$ needs at least $\Omega\left(\frac{1}{\varepsilon} \ln \frac{1}{\delta} + \frac{d}{\varepsilon}\right)$ many examples.
(Ehrenfeucht et al., 1988)

**LEMMA.** If $d = VCD(C)$ then $2^d \underset{\text{easy}}{\le} |C| \left( \le (|X|+1)^d. \right)$

"Sauer–Shelah" (Sauer 1972):

**LEMMA.** If $d = VCD(C)$ and $|X| = m < \infty$ then $|C| \le \sum_{i=0}^{d} \binom{m}{i}$

**PROOF.** Just prove that $|C| \le |\{Y \subseteq X \mid C \text{ shatters } Y\}|$ ⊛
because $|\{Y \subseteq X \mid C \text{ shatters } Y\}| \le \sum_{i=0}^{d} \binom{m}{i}$, since only subsets of size at most $d$ are shattered by $C$.

Prove ⊛ by induction on $m$.

$m=1$. $X = \{x\}$. $|C| \in \{1,2\}$ and $|\{Y \subseteq X \mid C \text{ shatters } Y\}| \in \{1,2\}$

$|C| = 1$ ✓

$|C| = 2 \Rightarrow$ there are $c, c' \in C$ with $c(x) \neq c'(x)$
$\Rightarrow \emptyset$ and $\{x\}$ are shattered by $C$
$\Rightarrow |\{Y \subseteq X \mid C \text{ shatters } Y\}| = 2$

$m \to m+1$. $|X| = m+1$, $X = \{x_1, \dots, x_{m+1}\}$

$$|C| = \underbrace{\left|C|_{\{x_1,\dots,x_m\}}\right|}_{Q_1} + \underbrace{\left|\{c \in C \mid \exists c' \in C : c(x_{m+1}) = 1,\ c'(x_{m+1}) = 0,\ c|_{\{x_1,\dots,x_m\}} = c'|_{\{x_1,\dots,x_m\}}\}\right|}_{Q_2}$$

i.h. $\Rightarrow Q_1 \le |\{Y \subseteq \{x_1, \dots, x_m\} \mid C \text{ shatters } Y\}|$
$= |\{Y \subseteq X \mid x_{m+1} \notin Y \text{ and } C \text{ shatters } Y\}|$

$Q_2 = \left|C'|_{\{x_1,\dots,x_m\}}\right|$ where $C' = \{c \in C \mid \exists c' \in C : c(x_{m+1}) = 1,\ c'(x_{m+1}) = 0,\ c|_{\{x_1,\dots,x_m\}} = c'|_{\{x_1,\dots,x_m\}}\}$

i.h. $\Rightarrow Q_2 \le |\{Y' \subseteq \{x_1,\dots,x_m\} \mid C' \text{ shatters } Y'\}|$
$= |\{Y' \subseteq \{x_1,\dots,x_m\} \mid C \text{ shatters } Y' \cup \{x_{m+1}\}\}|$
$= |\{Y \subseteq X \mid x_{m+1} \in Y,\ C \text{ shatters } Y\}|$

$|C| \le Q_1 + Q_2$
$= |\{Y \subseteq X \mid C \text{ shatters } Y\}|$

# SAMPLE COMPRESSION

**DEFINITION.** Let $k \in \mathbb{N}$. A sample compression scheme of size $k$ for $C$ is a pair $(f, g)$ with

(1) $f, g : 2^{X \times \{0,1\}} \longrightarrow 2^{X \times \{0,1\}}$

(2) If $S \subseteq \{(x, c(x)) \mid x \in X\}$ for some $c \in C$ and $S$ finite then

- $f(S) \subseteq S$
- $|f(S)| \leq k$
- $S \subseteq g(f(S))$

**COOL FACT:** Sample compression schemes of size $\leq k$ can be turned into PAC-learners using a bound on $m$ that is linear in $k$.

**EXAMPLE.**

- axis-aligned rectangles $\rightarrow k = 4$
- concepts of size at most $k$
- intersection-closed classes of VCD $k$ (every $c \in C$ has a subset $A \subseteq c$ with $|A| \leq k$ and $c = \bigcap \{c' \in C \mid A \subseteq c'\}$)

**THE BIG QUESTION:**

Does every $C$ have an SCS of size $\leq VCD(C)$?
—"— in $O(VCD(C))$?
} **OPEN!**

Known results:

- If $VCD(C) = 1$ then $C$ has an SCS of size 1.
- If $C$ is $\boxed{\text{maximum of } VCD(C) = d \text{ over } X \text{ with } |X| = m}$, i.e., $|C| = \sum_{i=0}^{d} \binom{m}{i}$, then $C$ has an SCS of size $\leq d$.

Can we prove that every <u>maximal</u> class of VC-dimension $d$ over $m$ instances is contained in a <u>maximum</u> class of VC-dimension $d$ over $m' > m$ many instances?

COMPACTNESS LEMMA. (Ben-David, Litman 1998)

Let $C$ be any concept class, $n \in \mathbb{N}$. If every finite subclass of $C$ has an SCS of size $n$ then $C$ has an SCS of size $n$.

The proof uses the (same method as that in the) compactness theorem of predicate logic.

$\leadsto$ we need to talk only about the case that $X$ is finite.