1 Definitions and Notations, again

Dichotomies
For any class of representations, $\mathcal{H}$ over $\mathcal{X}$, $S \subseteq \mathcal{X}$, the dichotomies of $S$ realized by $\mathcal{H}$ are $\Pi_{\mathcal{H}}(S)$ defined as $\{\{h(x_1), \ldots, h(x_m)\} : h \in \mathcal{H}, S = \{x_1, \ldots, x_m\}\}$ or identifying $h \in \mathcal{H}$ with the set $\{x \in \mathcal{X} : h(x) = 1\}$, $\Pi_{\mathcal{H}}(S) = \{h \cap S : h \in \mathcal{H}\}$.

Shattering
If $\Pi_{\mathcal{H}}(S)$ is all subsets of $S$, we say $S$ is shattered by $\mathcal{H}$.

VC-Dimension
The VC-dimension of $\mathcal{H}$ is the size of the largest set $S$ shattered by $\mathcal{H}$.

Error Regions
For any target property $c$, the class of error regions of $h$ wrt $c$ is $\Delta_{\mathcal{H},c}(c) \overset{\text{def}}{=} \{\{h \neq c\}(x) : h \in \mathcal{H}\}$.
The error regions of probability $\geq \epsilon$ under a distribution $\mathcal{D}$ is denoted $\Delta_{\mathcal{H,c},\epsilon}(c)$.

$\epsilon$-net
For any $\epsilon > 0$, we say $S \subseteq \mathcal{X}$ is an $\epsilon$-net for $\Delta_{\mathcal{H,c},\epsilon}$(c) w.r.t $\mathcal{D}$ if for every $r \in \Delta_{\mathcal{H,c},\epsilon}$(c) there is some $x \in r \cap S$.

Growth Function
The growth function of $\mathcal{H}$ is $\Pi_{\mathcal{H}}(m) \overset{\text{def}}{=} \max_{S \subseteq \mathcal{X}} |\Pi_{\mathcal{H}}(S)|$.

2 Finish VC-dimension sample complexity bound

Theorem: Let $c$ be any property over $\mathcal{X}$, and $\mathcal{H}$ is any class of representations of VC-dimension $d$. Then any solution $h$ to Consis$_{\mathcal{H}}$ on a data set of $m \geq c_0 \frac{1}{\epsilon} \left(d \log \frac{1}{\delta} + \log \frac{1}{\delta}\right)$ examples drawn independently from any distribution $\mathcal{D}$ on labeled examples from $\mathcal{X}$, labeled by $c$, satisfies $Pr_{x \in \mathcal{D}}[h(x) = c(x)] \geq 1 - \epsilon$ with probability $1 - \delta$.

From last time, we proved:

Lemma 1: If $S_1$ and $S_2$ are sets of $m$ examples drawn independently from $\mathcal{D}$ and labeled by $c$, the events $A = "S_1$ is not an $\epsilon$-net for $\Delta_{\mathcal{H,c},\epsilon}(c)"$ and $B = "\exists r \in \Delta_{\mathcal{H,c},\epsilon}(c) : |S_1 \cap r| = 0$ and $|S_2 \cap r| \geq \frac{\epsilon m}{2}"$ satisfy $Pr[A] \leq 2Pr[B]$.

Note if $A$ does not occur, we can output any solution to Consis$_{\mathcal{H}}(S_1)$. So, it is enough to show $Pr[B] \leq \frac{\delta}{2}$.
Lemma 2: \( \Pi_{H_{\mathcal{H}}(c)}(2m) \leq \Pi_{H_{\mathcal{H}}(e)}(2m) = \Pi_{H}(2m) \)

Sauer's Lemma: If \( \mathcal{H} \) has VC-dimension \( d \), then \( \Pi_{H}(m) \leq \sum_{i=0}^{d} \binom{m}{i} \), which is \( (\frac{em}{d})^{d} \).

Suppose we have drawn \( S = (S_1, S_2) \), we can write \( B|S_1, S_2 \) as
\[
B|S_1, S_2 = \bigvee_{r \in \Pi_{H_{\mathcal{H}}(c)}(S_1, S_2)} [S_1 \cap r = \emptyset] \land [S_2 \cap r \geq \frac{em}{d}]
\]

\[
\Pr[B] = \mathbb{E}_S[\mathbb{E}[B|S]] \leq \mathbb{E}_S \left( \sum_{r \in \Pi_{H_{\mathcal{H}}(c)}(S_1, S_2)} \Pr[[S_1 \cap r = \emptyset] \land [S_2 \cap r \geq \frac{em}{d}]] \right)
\]

\[
\leq \mathbb{E}_S[\Pi_{H_{\mathcal{H}}(c)}(S)] + 2^{-\frac{em}{d}} \leq \Pi_{H}(2m) + 2^{-\frac{em}{d}} \leq (\frac{em}{d})^{d} * 2^{-\frac{em}{d}}
\]

This follows since each example in \( S_1 \) and \( S_2 \) is an independent draw from the same distribution \( \mathcal{D} \). So, we could first draw \( S \) of size \( 2m \) and choose \( m \) of them uniformly at random to be \( S_1 \) (the rest are \( S_2 \)). E.g. we randomly reorder \( S \), and call the first \( m \) elements \( S_1 \), the second \( m \) elements \( S_2 \).

Lemma 3:
\[
\Pr_{S_1, S_2 \sim (S_1, S_2)}[[S_1 \cap r = \emptyset] \land [S_2 \cap r \geq \frac{em}{d}]] \leq 2^{-\frac{em}{d}}
\]

Proof: Note that \( S \cap r \) must be between \( m \) and \( \frac{em}{d} \), or else, respectively, the first or second condition cannot hold. The probability above is at most \( \Pr[S_1 \cap r = \emptyset \mid S \cap r \geq \frac{em}{d}] \)

which is \( (let \ l = S \cap r)\)

Number of ways all \( l \) elements of \( S \cap r \) end up in \( S_2 \)
\[
\frac{m!}{(m-l)!} \frac{l!(2m-l)!}{2m!}
\]

Since \( l \geq \frac{em}{d} \), the lemma follows.

Q.E.D.

So, we only need to show \( (\frac{em}{d})^{d} * 2^{-\frac{em}{d}} \leq \frac{\delta}{2} \iff d \log((\frac{em}{d})^{d}) - \frac{em}{d} \leq \log \frac{\delta}{2} \).

Theorem follows from Lemma 4.

Lemma 4: If \( m \geq \frac{12}{\varepsilon} (d \log(\frac{1}{\alpha}) + \log(\frac{1}{\beta})) \), then \( \frac{em}{d} \geq d \log(\frac{em}{d}) + \log \frac{\delta}{2} \)

Proof: Since \( 1 + x \leq e^{x} \), for any \( a, b > 0 \), taking \( x = ab - 1 \) gives \( \ln a + \ln b \leq ab - 1 \).

Suppose we put \( b = m \) and \( a = \frac{\epsilon \ln 2}{\delta} \)

\( \ln m + \ln m \leq \frac{\epsilon \ln 2}{\delta} \)

\( d \log_{2} \frac{\epsilon \ln 2}{\delta} + d \log_{2} me \leq \frac{\epsilon \ln 2}{\delta} \)

So, it is enough to take \( \frac{e \ln 2}{\delta} \geq d \log_{2} \frac{1}{\alpha} + \log_{2} \frac{2}{\beta} + d \log_{2} \frac{4d}{\epsilon \ln 2} \)

Since then \( \frac{e \ln 2}{\delta} \geq d \log_{2} \frac{\epsilon \ln 2}{\delta} + \frac{e \ln 2}{\delta} + d \log_{2} me + d \log_{2} \frac{1}{\alpha} + \log_{2} \frac{2}{\beta} + d \log_{2} \frac{4d}{\epsilon \ln 2} \)

So it suffices to have \( m \geq \frac{1}{d} \left[ d \log_{2} \frac{1}{\alpha} + d \log_{2} \frac{2}{\beta} + d \log_{2} \frac{4d}{\epsilon \ln 2} \right] \) where \( \log_{2} \frac{2}{\beta} \geq \log_{2} \frac{1}{\alpha} + 1 \) and \( d \log_{2} \frac{1}{\alpha} + d \log_{2} \frac{4d}{\epsilon \ln 2} \leq 2d \)

So...if
\[
m \geq 4d \log(\frac{1}{\alpha}) + 3 \log(\frac{1}{\beta}) \text{ where } 3d \log(\frac{1}{\alpha}) \geq d \log(\frac{1}{\beta}) + 2d
\]

Q.E.D.

Note:
1. Hanneke showed \( m \geq \frac{1}{d} \left[ d + \log \frac{1}{\alpha} \right] \) examples suffice. So the number of examples we need to learn class \( \mathcal{H} \) of VC-dimension \( d \) is \( O(\frac{1}{d} \left[ d + \log \frac{1}{\alpha} \right]) \) (HW5...)
2. Although it can be easier computationally to learn a more expressive \( \mathcal{H} \), it costs more data.
3 Learning halfspaces

HW4: Linear halfspaces in $\mathbb{R}^n$ have VC-dimension $n + 1$.

Linear halfspaces functions $[\sum_{i=1}^{n} w_i x_i \geq t]$ (boolean functions) for $w_1, ..., w_n \in \mathbb{R}, t \in \mathbb{R}$

So, by our theorem: it suffices to find a halfspace consistent with $\frac{12}{\varepsilon}( (n+1) \log \frac{1}{\varepsilon} + \log \frac{1}{\delta})$ examples to learn a halfspace.

Find $w_1, ..., w_n, t$ in $\mathbb{R}$ such that:

$w_1 x_1^{(i)} + ... + w_n x_n^{(i)} + (-1)t \geq 0$ for $b^{(i)} = 1$

$-w_1 x_1^{(i)} + ... + -w_n x_n^{(i)} + t \geq 0$ for $b^{(i)} = 0$

This is a homogeneous linear program, meaning of the form $A(\overrightarrow{w}, t) \leq 0$...can solve in polynomial time.

Notice...given any homogeneous linear program $A(\overrightarrow{w}, t) \leq 0$, by taking the $i$th example to be the $i$th row of $A$ ("$A_i$"), $b^{(i)} = 0$ and odd($\overrightarrow{0}$, 1), any solution to ConsistencyHalfspace on this set must satisfy

(1) $\overrightarrow{0} \overrightarrow{w} - t \geq 0 \Rightarrow t \leq 0$

(2) for every $i$th row: $A_i \overrightarrow{w} + t \geq 0$ and $A_i \overrightarrow{w} - t \leq 0 \Rightarrow A_i \overrightarrow{w} \leq 0$

So: Proper PAC-learning of halfspaces is equivalent to solving homogeneous linear programs.

Example ↔ Constraint in linear program

4 Start learning vs. cryptography

What we will see next:

Classes like Boolean Circuits cannot be learned by polynomial-time algorithms.

Their NP-Completeness is not known! Instead we will show learning would break cryptographic assumptions. Learning would allow us to solve some problems used by cryptography - such as factoring, shortest lattice vector, discrete, roots, etc. - and break cryptographic systems.