L21 Trick Imagine taking s more independent samples M from A (after sampling to get N)

Let E_0 = event that N fails to be an ε-net same

Let E_1 = event that $\left\{ \begin{array}{l} \exists\, S \in F \text{ with } N \cap S = \emptyset \\ \underline{\text{and}} \text{ at least } \tfrac{1}{2} s\varepsilon \text{ samples in } M \text{ come from } S. \end{array} \right.$

(Shorthand: $|M \cap S| \geq \frac{1}{2} s\varepsilon$)

1. $\Pr[E_1] \leq \Pr[E_0]$ clearly
2. $\Pr[E_1] \geq \frac{1}{2}\Pr[E_0]$ $\Pr[E_1 | N]$ means prob. choice of M causes E_1 given fixed N

If N is an ε-net then

$$\Pr[E_0|N] = \Pr[E_1|N] = 0$$

Otherwise let S_N be one of the sets in F missed by N

$$\Pr[E_1 | N] \geq \Pr\Big[\underbrace{|M \cap S_N|}_{\text{this is like } X_1 + X_2 + \dots + X_s} \geq \tfrac{1}{2} s\varepsilon\Big] \geq \tfrac{1}{2}$$

so we can use our tool

Thus $\Pr[E_0|N] \leq 2\Pr[E_1|N]$ for all N

Thus $\Pr[E_0] \leq 2\Pr[E_1]$

3. Bound $\Pr[E_1]$ differently

a) Choose $2s$ points from A $=$ W

b) Split W into N and M randomly: $\binom{2s}{s}$ possibilities

[a)+b) is the same as choosing N and then M]

We show $\Pr[E_1 | W]$ is small for any W.   over $\binom{2s}{s}$ splittings of W

For fixed set $S \in F$:

$$\Pr\left[N \cap S = \emptyset,\ |M \cap S| \geq \tfrac{s\varepsilon}{2} \,\middle|\, W\right] = 0 \quad \text{if } |W \cap S| < \tfrac{s\varepsilon}{2}$$

and it's $\leq \Pr[N \cap S = \emptyset | W]$ otherwise

o.w.

$\Pr[N \cap S = \emptyset | W] =$ Prob. random sample of s positions out of $2s$ positions in W avoids the $\geq \frac{s\varepsilon}{2}$ positions in W from S

$$\leq \frac{\binom{2s - \frac{s\varepsilon}{2}}{s}}{\binom{2s}{s}} \leq \left(1 - \frac{s\varepsilon/2}{2s}\right)^s = \left(1 - \frac{\varepsilon}{4}\right)^s \leq e^{-\frac{s\varepsilon}{4}} = e^{-\frac{Cd\ln 1/\varepsilon}{4}}$$

$$= 1/\varepsilon^{-Cd/4}$$

The previous derivation was for fixed $S \in F$

Now we use VC-dim

Sets of F have at most $\Phi_d(2s)$ distinct intersections with W

Since "$N \cap S = \emptyset$ and $|M \cap S| \geq \frac{s\epsilon}{2}$" depends only on $W \cap S$, we only need to consider $\Phi_d(2s)$ sets S.

$\Rightarrow$ For fixed W

$\Phi_d(2s) = \sum_{i=0}^{d} \binom{2s}{i}$

$$\Pr[E_1 \mid W] \leq \Phi_d(2s) \cdot \frac{1}{\varepsilon}^{-cd/4}$$

$$\leq \left(\frac{2es}{d}\right)^d \cdot \frac{1}{\varepsilon}^{-cd/4}$$

for $\varepsilon < 1/2$

$$= \left(\frac{2e\, c d \, \frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}}{d}\right)^d \frac{1}{\varepsilon}^{-cd/4}$$

$$= \left(2ec \left(\tfrac{1}{\varepsilon} \ln \tfrac{1}{\varepsilon}\right) \cdot \frac{1}{\varepsilon}^{-c/4}\right)^d < 1/2$$

for big enough c
and small enough ε

CPSC 516

Almost Optimal Set Covers in Finite VC-dimension

H. Brönnimann and M.T. Goodrich

Min Set Cover Given a family F of sets $S_1 S_2 \dots S_n$

Find smallest subfamily $C \subseteq F$ whose union is $X = \bigcup_i S_i$

Min Hitting Set Given a family F of sets $S_1 S_2 \dots S_n$

Find smallest set H such that $H \cap S_i \neq \emptyset$ forall i

Note: These problems are NP-hard.

(Even if every S_i has size 2 = Vertex Cover)

$\boxed{\text{Alg}}$ to find an approxi Min Hitting Set $\boxed{\text{given size } c \text{ of Min H.S. of } F}$

contains an element from every set with weight $> \varepsilon\, w(X)$

total weight of elements

① Put weight $w(x)$ on all elements x in $X = \bigcup S_i$

[initially $w(x) = 1$ for all x]

② Select a $\boxed{\varepsilon\text{-net}}$ N for weighted (X, F) $\boxed{\varepsilon = 1/2c}$

③ If N misses some set S_i, double the weights of elements in S_i and goto ②

④ Output N.

Hitting Sets for Finite VC-dim

<u>Lemma H</u> If $\exists$ hitting set H of size c for (X,F) algorithm loops $\leq 4c \log(n/c)$ times and the total weight of X will be $\leq n^4/c^3$

<u>proof</u> If S_i is missed by N, $w(S_i) \leq w(X)/2c$, so doubling weights of all $x \in S_i$ increases $w(X)$ by factor $\leq 1 + 1/2c$ in all iterations.

But some element in S_i is in H.

So after k iterations if each $h \in H$ has been doubled z_h times:

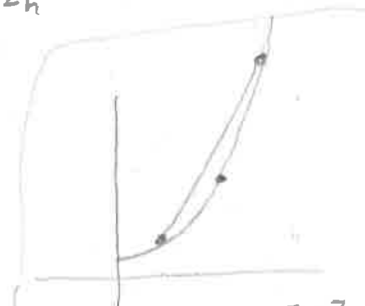
$$w(X) \leq n\left(1 + \tfrac{1}{2c}\right)^k \leq n e^{k/2c}$$

and

$$w(H) = \sum_{h \in H} 2^{z_h} \quad \text{where} \quad \sum_{h \in H} z_h \geq k$$

by convexity of 2^x:

$$w(H) \geq c\, 2^{k/c}$$



$$\frac{2^{z_1} + 2^{z_2}}{2} > 2^{\frac{z_1 + z_2}{2}}$$

so

$$c\, 2^{k/c} \leq w(H) \leq w(X) \leq n e^{k/2c} \leq n\, 2^{3/2 \cdot k/2c} = n^{3k/4c}$$

$$\Rightarrow \quad k \leq 4c \ln(n/c)$$

$\square$

Hitting Sets continued

Lemma H implies

If weight-doubling <u>fails</u> to produce hitting set after more than $4c \ln(n/c)$ iterations

Then no hitting set of size $\leq c$ exists.

So use Alg with $c = 1, 2, 4, \ldots$ until it <u>succeeds</u>

If $|\text{Min Hit Set}| = c^*$ then $c \leq 2c^*$

$$\Rightarrow \frac{1}{2c}\text{-net has size } \frac{2cd}{\varepsilon} \ln 1/\varepsilon < \frac{4c^* d}{\varepsilon} \ln 1/\varepsilon$$

Since $\varepsilon = 1/2c \geq 1/4c^* \Rightarrow 1/\varepsilon < 4c^*$

This yields a $O(d \log c^*)$ approximation to the minimum hitting set in time $O\left(c^* \log\left(\frac{n}{c^*}\right)\right)(T_N + T_W)$

T_N: find candidate net

T_W: find witness that candidate is bad or is good

[provided we can find an ε-net <u>quickly</u> ($= T_N$) and can find, given H, a set S_i not hit by H or confirm that H is a hitting set <u>quickly</u> ($= T_W$)]

This is poly time.