Oral exam: 26th August

$$O\left(\frac{d}{\varepsilon}\ln\left(\frac{1}{\varepsilon}\right)\right)$$

<u>Proof.</u> We want to show that a random sample of size $r = \frac{cd}{\varepsilon}\ln\left(\frac{1}{\varepsilon}\right)$ has a positive probability to be an ε-net.

→ two random samples R, R' of size r

→ events:

- E: R is not an ε-net
  i.e. $\exists S \in \mathcal{S}: R \cap S = \emptyset$ with S large enough

to show: $P(E) < 1 \Rightarrow P(\bar{E}) > 0$

- E': R is not an ε-net and for some $S \in \mathcal{S}$ with $R \cap S = \emptyset$:

$$|R' \cap S| \geq \frac{\varepsilon r}{2}$$

$$P(E) \geq P(E')$$

Now consider $P(E'|R)$, so R is fixed and R' is drawn randomly from U, and R' implies E'. If R is an ε-net

$$P(E'|R) = 0 = P(E|R).$$

Otherwise, R is not an ε-net:

$$P(E'|R) \geq P\left(|R' \cap S^*| \geq \frac{\varepsilon r}{2}\right)$$

for $S^* \in \mathcal{S}$ witness for R not being an ε-net ($S^* \cap R = \emptyset$).

Using the Chernoff-Lemma, we can lower bound $P(E'|R)$ by $\frac{1}{2}$.

$$P(E'|R) \geq \frac{1}{2}$$

$$P(E|R) = 1$$

$$\Rightarrow P(E|R) \leq 2\,P(E'|R)$$

As this is true for any kind of R, we can conclude

$$P(E) \leq 2P(E').$$

It remains to show

$$2P(E') < 1.$$

Now choose a 'big' sample A with $|A| = 2r$. Then subdivide A in R and R' each of size r using randomization.

So for fixed A there are $\binom{2r}{r}$ possibilities to create R, R'. Let A be fixed and also some $S \in \mathcal{S}$. We define $P_S = P(R \cap S = \emptyset, |R' \cap S| \geq \frac{\varepsilon r}{2} \mid A)$

Obviously, if $|A \cap S| < \frac{\varepsilon r}{2}$ then $P_S = 0$.

Otherwise $P_S \leq P(R \cap S = \emptyset \mid A)$.

This term can be bounded as follows:

$$P(R \cap S = \emptyset \mid A) \le \frac{\binom{2r - \frac{\epsilon r}{2}}{r}}{\binom{2r}{r}}$$

$$\le \left(1 - \frac{\frac{\epsilon r}{2}}{2r}\right)^r$$

$$\le e^{-\left(\frac{\epsilon r/2}{2r}\right) r}$$

$$= e^{-\frac{\epsilon r}{4}}$$

So with $r = \frac{cd}{\epsilon} \ln\left(\frac{1}{\epsilon}\right)$ we get $e^{-\frac{\epsilon r}{4}} = e^{-\frac{cd}{4} \ln\left(\frac{1}{\epsilon}\right)}$

$$= \left(\frac{1}{\epsilon}\right)^{-\frac{cd}{4}}$$

Now we use the shatter lemma to determine the number of sets in $\mathcal{S}$ that A can intersect in order to compute $P(E' \mid A)$.

So we set $m = \frac{2r}{\varepsilon}$ in the lemma and get:

$$P(E' \mid A) \leq \left(\frac{\frac{2re}{\varepsilon}}{d} \right)^{d} \cdot \left(\frac{1}{\varepsilon} \right)^{-\frac{cd}{4}}$$

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

For ε small enough and $\overset{d \geq 2}{d}$ and c large enough, this term can be upper bounded by $\frac{1}{2}$.

$$\Big(\varepsilon \leq \tfrac{1}{2} : \quad P(E' \mid A) \leq \left(ec \ln\left(\frac{1}{\varepsilon}\right) \frac{1}{2^{c}} \right)^{d}$$

$$P(E' \mid A) < \frac{1}{2}.$$

Again, this is true for any A, and therefore it yields: $P(E') < \frac{1}{2}$. //

Matousek showed that such an ε-net of size $O\left(\frac{d}{\varepsilon} \ln\left(\frac{1}{\varepsilon}\right)\right)$ can be found in polynomial time (with the VC-dim. being the exponent).

So we are able to compute good ε-nets based on small VC-dim, but our final goal was to compute good hitting sets. Therefore, we want to have a closer look now at the results of Brönnimann and Goodrich.

They showed that we can compute a hitting set of size

$$d r^* \log(d r^*)$$

with r^* being the size of the optimal solution.

Their algorithm needs two ingredients:

- a net-finder and
- a verifier

The verifier has to return for a net-suggestion R if this is a valid net or give some violation, $S \in \mathcal{S} : S \cap R = \emptyset$.

Def. (Net-Finder): For a non-decreasing function s an s-net-finder for a set system $(U, \mathcal{S})$ with a weight function $w: U \to \mathbb{R}^+$, is an algorithm that for given r computes an $\varepsilon = \frac{1}{r}$-net of size $s(r)$.

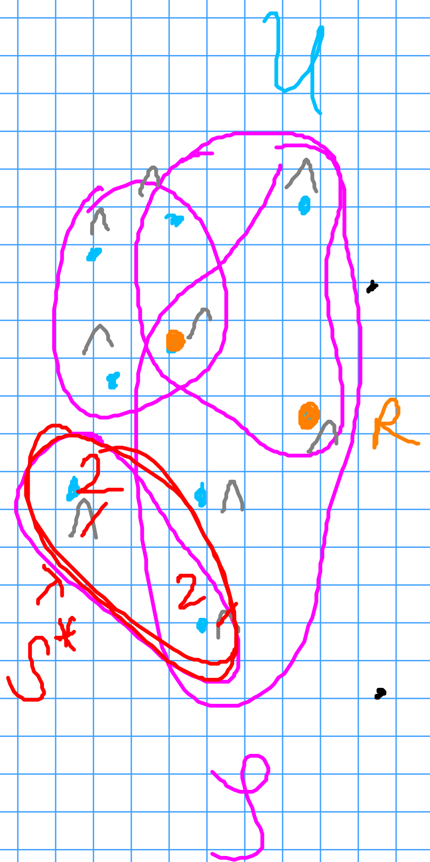
$s(x) = x$

$s(x) = \log(x)$

Theorem: Given a set system $(U, \mathcal{S})$, an s-net-finder and a verifier, then there exists an algorithm which computes a hitting set of size $s(4r^*)$ in polytime.

Algorithm:

- at the beginning all elements in $\mathcal{U}$ get a weight of 1 $(\forall u \in \mathcal{U}: w(u) = 1)$. and so $w(S) = |S|$.



- Now we apply the net finder with $\varepsilon = \frac{1}{2r^*}$ (we will see why it is ok to assume r^* to be known in the end)
- Then we run the verifier. If the returned net is a Hittingset, we are done. Otherwise we have a witness $S^* \in \mathcal{S}$ for which yields $S^* \cap R = \emptyset$. We double all weights of elements in S^*.
- Then the net-finder and the verifier are invoked again as long as no Hitting Set was found.

Note, that the total weight of the universe can not increase too much by weight doubling, as we always know: $\frac{1}{2r^*} > w(S^*)$.

In fact, we can ensure to find a hitting set in polynomial many rounds.

Lemma: With r^* being the optimal size of a HittingSet, no more than $6r^* \log\left(\frac{n}{r^*}\right)$ iterations are possible before a HittingSet is found. The total weight of the universe can not exceed $\frac{n^4}{r^{*3}}$.

Proof. Let k be the number of iterations. Observe again, that the set S^* returned as violator by the verifier satisfies

$$w(S^*) < \frac{w(U)}{2r^*}.$$

Therefore the total weight of the universe is multiplied in each round by at most $\left(1+\frac{1}{2C}\right)$.

So we can bound the weight of the universe as follows.

$$w(U) \leq n\left(1+\frac{1}{2C}\right)^{U}$$

$$\leq n\, e^{\frac{U}{2C}}$$

The weight of the computed solution L can be estimated as

$w(L) = \sum\limits_{u \in L} w(u)$. As in each

round at least the weight of a single element is doubled, we conclude that

$$w(L) \geq \underline{r}^{*}\, 2^{\frac{U}{r^*}} \quad \text{as}$$

L has to contain at least r^* elements, and as we perform U doubling rounds

We have to include at least one element with weight $\geq 2, \geq 4, \geq 8 \ldots$
$\ldots 2^{k}$.

So we get the following inequality:

$$r^* 2^{\frac{k}{r^*}} \leq w(L) \leq w(U) \leq n e^{\frac{k}{2c}}$$

$$k \left(2 \log(2) - \log(e)\right) \leq 2r^* \left(\log(n) - \log(r^*)\right)$$

and so $\quad k \leq 6 r^* \log\left(\frac{n}{r^*}\right)$.

If we insert it in the formula to bound $w(U)$ we get $w(U) \leq n e^{3 \log\left(\frac{n}{r^*}\right)}$

$$= n \cdot \left(\frac{n}{r^*}\right)^3 = \frac{n^4}{r^{*3}}$$

$\hookrightarrow$ polytime alg. to compute good hittingset

$\Rightarrow$ What about r^*? This is unknown, but we can easily find $r \leq r^*$ (e.g. $r = 1$).

Then we run the algorithm with r instead of r^*, the verifier will tell us after $6r\log\left(\frac{n}{r}\right)$ rounds, that we do not have a hitting set. So we use $2r$ now, and run the algorithm again. As soon as $2^i r \geq r^*$, we will find a hitting set with the alg. and obviously the final guess is at most a factor of 2 away from the real r^*.

<u>Exercise</u> What is the run time of the complete alg. if r^* is unknown? What is the overhead compared to knowing r^*?

Brönnimann & Goodrich:

- $\log(r^*)$ – APX alg for Hitting Set (for all sets with constant VC-dimension)
- for some selected problems (geometry), the algorithm per-

forms even better, in fact it gave a constant approximation.

Theorem of Clarkson gives some generalization when a constant HS approximation is possible.

Theorem: If for a set system $(U, \mathcal{S})$ and arbitrary $\varepsilon \in [0,1]$ there exists an ε-net of size $\frac{c}{\varepsilon}$ with c being independent of ε, then we can find an approximate solution for the hitting set of $(U, \mathcal{S})$ in poly-time which is at most a factor of c away from the optimum.

If you solve the golden exercise, it would mean, that hitting sets on shortest path systems exhibit a constant approximation.

We will applications of this