

3/9/2015 Russell's course lec 3.

Linear, Mansour, Nissen

Algorithmic uses of switching lemma:

lower bound on approx using Fourier representation [LML]

Exercise show that a read-once setting, 2/3 of inputs means bottom bit in  $O(\log S)$  bits.

Consequences of switching lemma:

Let  $d$  be depth of tree of size  $S$ ,  $D \geq \log S$ ,  $p = \Omega(\frac{1}{D})$ .

$\exists$  a read-once setting with  $n$  inputs.  $\text{Prob}[\exists \text{ tree } T \text{ of size } S \text{ with } \text{err} \leq \frac{1}{2} - p]$

Prob  $\leq 2^{-\Omega(p \cdot S)}$

Branch access to  $d$ -training:  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$  where  $\mathcal{D}_i$  are random  $n$ -bit inputs.

Learn hypothesis  $h: \{0,1\}^n \rightarrow \{0,1\}$ ; approx of success  $\geq 1 - \epsilon$

- know  $d$  is depth and  $S, \epsilon, n$  inputs

Fourier representation:

- represent true by  $-1$ , and false by  $1$ .

Over  $0,1$ ,  $\sum \mathcal{D}_i \bmod 2$  is parity. Over  $\{-1,1\}$ , parity is  $\prod x_i$

- function as a vector:  $F(x) = \langle F(x_1, \dots, x_1), F(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n), \dots, F(x_1, \dots, x_n) \rangle$

-  $F(1, \dots, 1)$ . This is  $2^n$  dim vector, embedded in  $2^n$ -dim space.

$\|F\|^2 = \sum_x (F(x))^2 = 2^{n+1}$

Normalizing const  $1/2^{n/2}$ ,  $1 \neq -1$  if  $\neq$ .

$\vec{F}, \vec{G}$ .  $\vec{F} \cdot \vec{G} = \frac{1}{2^{n/2}} \sum_x F(x) \cdot G(x) =$

$= \frac{1}{2^n} \cdot (\sum_x F(x) \cdot G(x)) = \text{Prob}[F(x) = G(x)] - \text{Prob}[F(x) \neq G(x)] \in \{-1, 1\}$

measure of correlation between  $F$  and  $G$

Geometrically, work under diff. bases - one basis  $(1, 0, \dots)$

Orthogonal basis: Set of  $n$  bases  $F_1, \dots, F_n$  s.t.  $\langle F_i, F_j \rangle = 0$  if  $i \neq j$ .

$\langle F_i, F_j \rangle = 0$ . So  $\text{Prob}[F_i = F_j] = \text{Prob}[F_i \neq F_j]$

Ideal subsets  $S \subseteq \{1, \dots, n\}$ .  $\mathcal{D}_S = \{ \prod_{i \in S} x_i : (x_i \in \{-1, 1\}) \}$

$\mathcal{D}_S \perp \mathcal{D}_T$ . Let  $S, T \subseteq [n]$ ,  $S \neq T$ . When does  $\mathcal{D}_S(x) = \mathcal{D}_T(x)$ ?

$\mathcal{D}_S \perp \mathcal{D}_T$ . Let  $S, T \subseteq [n]$ ,  $S \neq T$ . When does  $\mathcal{D}_S(x) = \mathcal{D}_T(x)$ ?

$\text{Prob}_{x \in \mathcal{D}_{S \Delta T}}(\mathcal{D}_S(x) = \mathcal{D}_T(x)) = 1/2$

$\Phi_s$  are  $2^n$  orthonormal vectors in  $2^n$  dim space: basis.

Any bool fn  $\vec{F} = \sum \hat{d}_s \Phi_s$

Fourier coeffs.

$\langle \vec{F}, \Phi_s \rangle = \sum \hat{d}_s \langle \Phi_s, \Phi_s \rangle = \sum \hat{d}_s = \text{Prob}_{x_i} [F(x) = 1] - \text{Prob}_{x_i} [F(x) \neq 1]$

$\frac{1}{2^n} \sum \hat{d}_s \Phi_s(x) = \sum \hat{d}_s \Phi_s(x) = \sum \hat{d}_s \prod_{i \in S} x_i$

- multilinear poly.

Dec. tree for  $P$ :  $P|_{x_i=0} = 1$   $P|_{x_i=1} = 0$

$P(x) = \frac{1-x_i}{2} \cdot F|_{x_i=1} + \frac{(1+x_i)}{2} F|_{x_i=0}$  - multilinear rep. of def.

$\Rightarrow$  any Fourier coeff. of  $F$  of size  $\geq \frac{1}{2}$  is 0.

Idea [LNU] this means large Fourier coeffs or  $\vec{F}$  are small.

Parsival's identity:  $\sum (\hat{d}_s)^2 = 1$

$\|\vec{F}\| \geq 1$ .  $\vec{F} = \langle \hat{d}_\emptyset, \hat{d}_{x_1}, \dots, \hat{d}_{x_n} \rangle$ .  $\|\vec{F}\| = \sqrt{1} = 1$

LNU claim:  $\sum_{|S| \geq \frac{n}{2}} (\hat{d}_S)^2 \leq \epsilon \text{ vol}(\mathcal{D})$ ,  $\mathcal{D} \gg \log S$ .

$P = \sum \hat{d}_s \prod_{i \in S} x_i$ .  $F|_{x_i=1} = \sum_{S \subseteq \{2, \dots, n\}} (\hat{d}_S + \hat{d}_{S \cup \{1\}}) \prod_{i \in S} x_i$

$P|_{x_i=1} = \sum_{S \subseteq \{2, \dots, n\}} (\hat{d}_S - \hat{d}_{S \cup \{1\}}) \prod_{i \in S} x_i$

$\mathbb{E} \left( \sum \hat{d}_S \right)^2 = \frac{1}{2} (\hat{d}_S + \hat{d}_{S \cup \{1\}})^2 + \frac{1}{2} (\hat{d}_S - \hat{d}_{S \cup \{1\}})^2 = \hat{d}_S^2 + \hat{d}_{S \cup \{1\}}^2$

So we set all  $x_i$  in  $T$  at random. Let  $F^*|_T = P|_T$ . Let  $S$  be subset of vars. Then  $\mathbb{E} \left( \sum_{T \cap S} \hat{d}_S \right)^2 < \dots$

Lemma if  $f$  is multilinear poly  $p$  proc. of vars in set, then

$\mathbb{E} \left( \sum_{S \subseteq \text{vars}, T \cap S} \hat{d}_S \right)^2 \leq \sum_{u \in T} (\hat{d}_u)^2$

Work on the set  $T$  of set vars,  $u$  contributes  $\hat{d}_u$  to  $\left( \sum \hat{d}_S \right)^2$ . Part of sum is  $|u-T| \geq 0$ .  $\mathbb{E}[|u-T|] = p_u |u| \geq 1$ .  $u$ . const prob  $\geq 1$ .