15-859(B) Machine Learning Theory

Lecture 02/26/02

- Finish SQ $\rightarrow$ PAC + Noise from last time

- Fourier analysis to characterize what's learnable by SQs

[Start with notes from last time]

SQ-dimension

- Functions f and g are uncorrelated w.r.t $\mathcal{D}$ if

$$\Pr_{x \in \mathcal{D}}[f(x) = g(x)] = 1/2$$

- For a given $C, \mathcal{D}$, let S be the largest set of "pairwise nearly uncorrelated" functions in C.

  For all $f, g \in S$,

$$|\Pr_{x \in \mathcal{D}}[f(x) = g(x)] - \frac{1}{2}| \leq 1/|S|.$$

If $|S| \leq poly(\ldots)$ then you can weak-learn over $\mathcal{D}$.

If $|S| > poly(\ldots)$ then you can't weak-learn over $\mathcal{D}$.

(using an SQ algorithm)

Example: parity

- Let $\mathcal{D}$ be the uniform dist over $\{0,1\}^n$.
- Theorem: Any two parity functions are uncorrelated.
- So, $\text{PARITY} = \{\text{all } 2^n \text{ parity fns over } \{0,1\}^n\}$ is not SQ-learnable over the uniform dist.

Parity, contd.

- Let $\text{PARITY}_{\log n}$ be the set of parities of $\log n$ variables. Also not SQ-learnable.
- So, DNF and decision trees not SQ-learnable.

  (i.e., no SQ algorithm can guarantee to learn in time poly in n and the size of the target function)
- Can anyone think of a non-SQ alg to learn parity?

  Learnability with noise is big open crypto and coding-theory problem.