PROBLEM SET 3

Due: Monday, Oct. 1, beginning of class

Homework policy: Please work on the homework by yourself; it isn't intended to be too difficult. Questions about the homework or other course material can be asked on Piazza.

---

1. Let $p \in [1,\infty)$. Given $f : \{-1,1\}^n \to \mathbb{R}$ we define its *p-norm* to be $\|f\|_p = \mathbf{E}_{\boldsymbol{x}}[|f(\boldsymbol{x})|^p]^{1/p}$ and its *Fourier p-norm* to be $\hat{\|} f \hat{\|}_p = (\sum_S |\widehat{f}(S)|^p)^{1/p}$. We also define $\|f\|_\infty = \max_x\{|f(x)|\}$ and $\hat{\|} f \hat{\|}_\infty = \max_S\{|\widehat{f}(S)|\}$.

   (a) For $1 \le p \le q \le \infty$, show that $\|f\|_p \le \|f\|_q$. (Hint: one solution uses the fact that the function $t \mapsto t^{q/p}$ is convex on $[0,\infty)$.) Conversely, show $\hat{\|} f \hat{\|}_p \ge \hat{\|} f \hat{\|}_q$.

   (b) Show that $\hat{\|} f \hat{\|}_\infty \le \|f\|_1$ and $\|f\|_\infty \le \hat{\|} f \hat{\|}_1$.

2. In this exercise you are asked to prove some fancily-named properties of the noise operator T_ρ.

   (a) Show that T_ρ is "positivity-preserving" for all $\rho \in [-1,1]$, meaning $f \ge 0 \Rightarrow \mathrm{T}_\rho f \ge 0$.
   Show also that it is "positivity-improving" for all $\rho \in (-1,1)$, meaning $f \ge 0, f \not\equiv 0 \Rightarrow \mathrm{T}_\rho f > 0$.

   (b) Show the "semigroup property": $\mathrm{T}_{\rho_1} \circ \mathrm{T}_{\rho_2} = \mathrm{T}_{\rho_1\rho_2}$ for all $\rho_1, \rho_2, \in [0,1]$.
   (If you like, prove it even for $\rho_1, \rho_2 \in [-1,1]$.)

   (c) Show that T_ρ is a "contraction on L^p" for all $p \ge 1$ and $\rho \in [-1,1]$; i.e., $\|\mathrm{T}_\rho f\|_p \le \|f\|_p$.

3. For functions $f : \mathbb{F}_2^n \to \mathbb{R}$, sometimes it is more natural to index the Fourier coefficients not by subsets $S \subseteq [n]$ but by elements $\gamma \in \mathbb{F}_2^n$; here we identify a subset with its indicator vector. In this case we would write the Fourier expansion as

   $$f = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{f}(S)\chi_\gamma, \qquad \text{where } \chi_\gamma(x) = (-1)^{\gamma \cdot x}$$

   and $\gamma \cdot x$ is the dot-product of γ and x in the vector space $\mathbb{F}_2^n$. Note that for $\beta, \gamma \in \mathbb{F}_2^n$ we have $\chi_\beta \chi_\gamma = \chi_{\beta+\gamma}$.

   (a) Let H be a vector subspace of $\mathbb{F}_2^n$. Let $H^\perp$ be its "perpendicular subspace"; i.e., $H^\perp = \{\gamma \in \mathbb{F}_2^n : \gamma \cdot x = 0 \text{ for all } x \in H\}$. Show that the indicator function $1_H : \mathbb{F}_2^n \to \{0,1\}$ of H has the Fourier expansion $1_H = \sum_{\gamma \in H^\perp} 2^{-k} \chi_\gamma$, where $k = \dim(H^\perp)$. (Remark: $k = n - \dim(H)$ is sometimes denoted $\mathrm{codim}(H)$.)

   (b) Given the subspace H and also $y \in \mathbb{F}_2^n$, the set $H + y = \{h + y : h \in H\}$ is called an "affine subspace" of $\mathbb{F}_2^n$. Show that the indicator function $1_{H+y} : \mathbb{F}_2^n \to \{0,1\}$ of this affine subspace has the Fourier expansion $1_{H+y} = \sum_{\gamma \in H^\perp} 2^{-k} \chi_\gamma(y) \chi_\gamma$, where again $k = \dim(H^\perp)$.

4. Suppose the Fourier spectrum of $f : \{-1,1\}^n \to \mathbb{R}$ is ϵ_1-concentrated on $\mathscr{F}$ and that $g : \{-1,1\}^n \to \mathbb{R}$ satisfies $\|f - g\|_2^2 \le \epsilon_2$. Show that the Fourier spectrum of g is $2(\epsilon_1 + \epsilon_2)$-concentrated on $\mathscr{F}$.

5. Given $s \in \mathbb{N}^+$, let $\mathscr{C}$ be the class of all functions $f : \{-1,1\}^n \to \{-1,1\}$ expressible as $f(x) = g(h_1(x), \ldots, h_s(x))$, where $h_1, \ldots, h_s : \{-1,1\}^n \to \{-1,1\}$ are weighted majority functions and $g : \{-1,1\}^s \to \{-1,1\}$ is any function. Show that $\mathscr{C}$ is learnable from random examples to error ϵ in time $n^{O(s^2/\epsilon^2)}$. You may use Peres's Theorem, that $\mathbf{NS}_\delta[h] \le 2\sqrt{\delta}$ for all $\delta \in [0, \frac{1}{2}]$ and all weighted majorities h. (Hint: how can you bound $\mathbf{NS}_\delta[f]$?)

6. (a) Let $k \in \mathbb{N}^+$ and let $\mathscr{C} = \{f : \{-1,1\}^n \to \{-1,1\} \mid \deg(f) \le k\}$. (In particular, $\mathscr{C}$ contains all functions computable by depth-k decision trees.) Show that $\mathscr{C}$ is learnable from random examples with error 0 in time $n^k \cdot \mathrm{poly}(n, 2^k)$. You may use the following "Degree/Granularity Fact": for every $f \in \mathscr{C}$ and every $S \subseteq [n]$, the Fourier coefficient $\widehat{f}(S)$ is an integer multiple of 2^{1-k}.

   (b) (Bonus.) Prove the Degree/Granularity Fact.