CS 599B: Math for TCS, Spring 2022

Exercise Set 2

Due: 10:00PM, Monday Feburary 7, 2022 on Gradescope.

Instructions. You're encouraged to discuss these problems with other students as you solve them, but the written solutions you hand in must be in your own words. Don't worry about polishing the presentation of your solutions; these are primarily intended to keep you thinking about and engaged with the material.

Problem 1 (Degree 1 variance)**.** For a function $f : \{-1,1\}^n \to \mathbb{R}$, let $f^{=k} = \sum_{|S|=k} \hat{f}(S)\chi_S$ be the part of f at degree k. Suppose $f : \{-1,1\}^n \to \mathbb{R}$ is a function such that $f = f^{=1}$, i.e., all of f's nonzero Fourier coefficients are at level 1. Show that

$$\mathrm{Var}\left[f^2\right] = 2\sum_{i\neq j} \hat{f}(i)^2 \hat{f}(j)^2.$$

Problem 2 (Noise operator)**.** For $1 \leq q < \infty$, define the q-norm $\|\cdot\|_q$ on real-valued functions over the hypercube by $\|f\|_q = (\mathbb{E}_{x\sim\{-1,1\}^n}[|f(x)|^q]^{1/q})$. Show that for every $\rho \in [-1,1]$ and every $q \geq 1$, we have $\|T_\rho f\|_q \leq \|f\|_q$. This shows that the noise operator is a *contraction*.

Problem 3 (Trimming trees)**.** The *size* of a decision tree is the total number of leaves. In this exercise, you'll show that small-size decision trees are also efficiently learnable under the uniform distribution.

a) Let T be a decision tree of size s. Show that for $\varepsilon > 0$, there exists a decision tree T' of depth $\log(s/\varepsilon)$ such that $\mathrm{dist}(T, T') \leq \varepsilon$. Hint: Truncate the tree to this depth.

b) Show that size-s decision trees are concentrated on degree up to $\log(s/\varepsilon)$.

c) Conclude that the class of size-s decision trees is learnable under the uniform distribution in time $n^{O(\log(s/\varepsilon))}$.

Problem 4 (Estimating Fourier coefficients)**.** Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Let $Z = ((x^1, f(x^1)), \ldots, (x^m, f(x^m)))$ where $x^1, \ldots, x^m \sim \{-1,1\}^n$ are uniform and i.i.d. Let $c_S = \frac{1}{m}\sum_{i=1}^m f(x^i)\chi_S(x^i)$. Use Hoeffding's inequality (below) to show that for some $m = O(\log(1/\delta)/\varepsilon^2)$, we have

$$\Pr_Z[|c_s - \hat{f}(S)| \leq \varepsilon] \geq 1 - \delta.$$

Hoeffding's Inequality: Let $X_1, \ldots, X_m$ be independent random variables such that all $X_i \in [0,1]$. Let $\overline{X} = \frac{1}{m}(X_1 + \cdots + X_m)$. Then for all $t > 0$,

$$\Pr[|\overline{X} - \mathbb{E}[\overline{X}]| > t] \leq 2\exp(-2nt^2).$$