CS 294-92 Analysis of Boolean Functions
Problem Set 4

Due: May 5, 11:59 PM.

You are encouraged to discuss the problems and solve them in groups (over Zoom, Slack, Skype, Hangouts, etc.). However, the solutions are to be written up alone, listing all the collaborators.

1. **The $2/\pi$ Theorem:** Recall that the Majority is Stablest theorem states that any balanced[1] Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ with $\mathbf{Inf}_i[f] \le \varepsilon$ for all $i \in [n]$ (think of ε as small) satisfies
$$\mathbf{Stab}_\rho[f] \le \mathbf{Stab}_\rho[\mathrm{MAJ}_n] + o_\varepsilon(1).^{2}. \qquad (1)$$
We consider the special case of the "Majority is Stablest theorem" in the limit when $\rho \to 0$. Recall that $\mathbf{Stab}_\rho[f] = \mathbf{W}^0[f] + \rho\mathbf{W}^1[f] + \rho^2\mathbf{W}^2[f] + \ldots + \rho^n\mathbf{W}^n[f]$. When f is balanced the first term is zeroed out, and thus for small ρ the noise stability is roughly $\rho \cdot \mathbf{W}^1[f]$ (up to an $O(\rho^2)$ error). Recall also that $\mathbf{W}^1[\mathrm{MAJ}_n] = \frac{2}{\pi} \pm o(1)$. Thus, in this regime, under the same assumptions (f is balanced and have influences at most ε), the question boils down to showing that
$$\mathbf{W}^1[f] \le \frac{2}{\pi} + o_\varepsilon(1). \qquad (2)$$
This result was proved by Talagrand a decade before the "Majority is Stablest theorem".
In this exercise, we will reprove Talagrand's result (i.e., Eq. (2)).
   (a) Let $L(x) = f^{=1}(x) = \sum_{i=1}^n \widehat{f}(\{i\})x_i$ be the linear part of the Fourier expansion of f (since $\widehat{f}(\emptyset) = 0$ this also equals $f^{\le 1}(x)$). Show that $\mathbf{W}^1[f] = \langle f, L\rangle = \langle L, L\rangle$.
   (b) Let $\ell(x) = L(x)/\sqrt{\mathbf{W}^1[f]}$. Note that $L(x)$ is a linear function of the form $L(x) = \sum_{i=1}^n a_i x_i$ where $a_i = \widehat{f}(\{i\})/\sqrt{\mathbf{W}^1[f]}$. Show that $\sum_{i=1}^n a_i^2 = 1$ and that
$$\mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x})\ell(\mathbf{x})] = \sqrt{\mathbf{W}^1[f]}.$$
   (c) Deduce that $\mathbf{E}_{\mathbf{x}\sim\{-1,1\}^n}[|\ell(\mathbf{x})|] \ge \sqrt{\mathbf{W}^1[f]}$.
   (d) Let $\mathbf{Z} \sim N(0,1)$ be a standard (univariate) Gaussian. Show that if $\mathbf{W}^1[f] \ge 1/2$ then $\mathbf{E}_{\mathbf{x}\sim\{-1,1\}^n}[|\ell(\mathbf{x})|] \le \mathbf{E}[|\mathbf{Z}|] + O(\varepsilon)$.
   You may use the following variant of the Berry-Esseen's Theorem: Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random bits (i.e., $\mathbf{x}_i \sim \{-1,1\}$). Let $\mathbf{S} = \sum_{i=1}^n a_i\mathbf{x}_i$, where $a_i \in \mathbb{R}$ and $\sum_{i=1}^n a_i^2 = 1$. Let $\mathbf{Z}$ be a standard Gaussian. Then $|\,\mathbf{E}[|\mathbf{Z}|] - \mathbf{E}[|\mathbf{S}|]| \le \sum_{i=1}^n |a_i|^3$.
   (e) Finish the proof (points to address: What happens if $\mathbf{W}^1[f] < 1/2$? What is $\mathbf{E}[|\mathbf{Z}|]$?)

[1] i.e., a function f with $\mathbf{Pr}_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x}) = 1] = 1/2$.
[2] $o_\varepsilon(1)$ means that it is a function of ε that tends to 0 as ε tends to 0

2. **Block-sensitivity versus Decision Tree Complexity:** Recall that for a given function $f : \{-1,1\}^n \to \{-1,1\}$, and an input $x \in \{-1,1\}^n$, the block-sensitivity of f at x, denoted $bs(f,x)$, is the maximal number of disjoint blocks $B_1,\ldots,B_m \subseteq [n]$ such that $f(x) \neq f(x^{\oplus B_i})$ for all $i \in [m]$. The block-sensitivity of f is simply $bs(f) = \max_{x\in\{-1,1\}^n} bs(f,x)$. Let $\mathcal{T}$ be a randomized decision tree of depth $d = R(f)$ computing f with probability at least 2/3. Our first goal is to show that $R(f) \geq \Omega(bs(f))$. Our second goal would be to show that $D(f) \leq O(R(f)^4)$.

   Recall that $\mathcal{T}$ is a distribution over deterministic decision trees $T_1,\ldots,T_m$ where we apply the tree T_i with probability p_i, and $p_1 + \ldots + p_m = 1$.

   (a) Show that for any distribution μ on $\{-1,1\}^n$ (i.e., a distribution on the inputs to f) there exists a deterministic decision tree from the above trees, i.e., T_i for $i \in [m]$, that computes f correctly with probability at least 2/3 over μ.

   (b) Consider the following distribution induced by the block-sensitivity of f on x. Let $B_1,\ldots,B_m$ be disjoint sensitive block for f on x. The distribution μ will assign probability mass 1/2 to the string x, and probability $1/(2m)$ to any string $x^{\oplus B_i}$ for $i \in [m]$.

   Show that any decision tree that solves f with probability at least 2/3 over μ must query at least one variable from at least $m/3$ of the blocks. (In other words, this shows that $bs(f)/3 \leq R(f)$.)

   (c) Let $b \in \{-1,+1\}$. Recall the a b-certificate for f is a partial assignment $\rho : \{1,\ldots,n\} \to \{-1,*,+1\}$ such that $f|_\rho \equiv b$. Intuitively, revealing the coordinates that are assigned by ρ convinces a verifier that $f(x) = b$. Alternatively, a b-certificate is a partial assignment such that any $x \in \{-1,1\}^n$ consistent with this partial assignment must satisfy $f(x) = b$. Let ρ^1 and ρ^{-1} be -1 and $+1$ certificates for f, respectively. Prove that there exists a coordinate i such that $\rho^1(i)$ and $\rho^{-1}(i)$ differ and both of them are not equal to $*$.

   (d) Deduce that if we query all coordinates in a -1 certificate and a $+1$ certificate for f, then after receiving the answers, either:

   - one of the certificates is consistent with the answers, in which case we know $f(x)$.
   - or, the remaining function, f', defined as the function restricted to inputs that are consistent with the queries so far, satisfies $C(f') \leq C(f) - 1$.

   Conclude that there exists a deterministic decision tree for f of depth $O(C(f)^2)$.

   (e) Deduce that $D(f) = O(bs(f)^4)$ and that $D(f) = O(R(f)^4)$ (hint: recall the connection between $C(f)$ and $bs(f)$ from the lecture, and use Item (b)).