Please turn in your solution in class on Tuesday October 15. You are encouraged to collaborate on the homework and ask for assistance, but you are required to write your own solutions, list your collaborators, acknowledge any sources of help, and provide external references if you have used any.

Question 1

In this question you will improve the lower bound on the decision tree size for recursive majority by a different method. Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$, let $L(f) = \sum|\hat{f}_S|$ be the sum of the absolute values of the coefficients of its polynomial (Fourier) representation. For example, $MAJ_3(x_1,x_2,x_3) = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_1x_2x_3$ and so $L(MAJ_3) = 4 \cdot \frac{1}{2} = 2$. In this question we represent both inputs and outputs by $-1/1$ values.

(a) Let f be the AND of n literals (variables or their negations) in $-1/1$ representation. What is $L(f)$?

Solution: Let's use f_{01} for $0/1$ output representation and f for $-1/1$ output representation. They are related by the formula $f = 1 - 2f_{01}$. Then

$$AND_{01}(x_1,\ldots,x_n) = \frac{1+x_1}{2}\cdots\frac{1+x_n}{2} = \sum_{S\subseteq[n]}\frac{1}{2^n}\prod_{i\in S}x_i$$

so $L(AND_{01}) = 1$. Moving to $-1/1$ outputs, all the coefficients double except the constant term which becomes $1 - 2\cdot 2^{-n}$. Therefore $L(AND) = 2(1-2^{-n}) + (1-2\cdot 2^{-n}) = 3 - 4\cdot 2^{-n}$ (for $n > 0$).

(b) Use part (a) to show that if f has a decision tree of size at most s then $L(f) \le 3s$.

Solution: In $0/1$ output representation, a decision tree f_{01} can be written as a sum ANDs of literals, one for each path leading to a 1-leaf. Each AND of literals has L-value 1 so by the triangle inequality f_{01} can have L-value at most s. Moving to $-1/1$ representation we get that $L(f) = L(1-2f_{01}) \le 2L(f_{01}) + 1 \le 2s + 1 \le 3s$.

(c) Let $h(y_1,y_2,y_3) = f(g(y_1),g(y_2),g(y_3))$, where y_1,y_2,y_3 are sets of disjoint variables and g has no constant term (i.e. $\hat{g}_\varnothing = 0$). Show that $L(h) = F(L(g),L(g),L(g))$, where F is the polynomial obtained from f by turning all its coefficients positive (i.e., $F(x) = \sum_S|\hat{f}_S|\prod_{i\in S}x_i$).

Solution: Assume first that f is a monomial $m_I = \prod_{i\in I}x_i$. Then the coefficients of h are of the form $\hat{h}_S = \prod_{i\in I}\hat{g}_{S\cap Y_i}$, where Y_i are the indices of variables y_i. Then

$$L(h) = \sum_S|\hat{h}_S| = \prod_{i\in I}\sum_{S_i}|\hat{g}_{S_i}| = m_I(L(g),L(g),L(g)).$$

If m_I and m_J are distinct monomials, then $m_I(g(y_1),g(y_2),g(y_3))$ and $m_J(g(y_1),g(y_2),g(y_3))$ do not share any common monomials because there is some set of variables y_i one of which is present in each monomial of one but not of the other. So if f is a linear combination of monomials, each of them will contribute distinct terms in h and $L(h) = \sum|\hat{f}(I)|m_I(L(g),L(g),L(g)) = F(L(g),L(g),L(g))$.

(d) Use part (c) to show that $L(RMAJ_d) \ge L(RMAJ_{d-1})^3/2$.

Solution: Let $\ell_d = L(RMAJ_d)$. By part (c) ℓ_d satisfies the recurrence $\ell_d = F(\ell_{d-1},\ell_{d-1},\ell_{d-1})$ where $F(x_1,x_2,x_3) = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 + \frac{1}{2}x_1x_2x_3$. Therefore $\ell_d = \frac{3}{2}\ell_{d-1} + \frac{1}{2}\ell_{d-1}^3 \ge \frac{1}{2}\ell_{d-1}^3$.

(e) Use parts (b) and (d) to show that $RMAJ_d$ requires decision tree size $2^{\Omega(3^d)}$.

Solution: We can rewrite the inequality in part (c) as $\ell_d/\sqrt{2} \geq (\ell_{d-1}/\sqrt{2})^3$. Plugging in the base case $\ell_1 = 2$ we get that $\ell_d \geq (\sqrt{2})^{3^{d-1}+1}$. By part (b) any decision tree for $RMAJ_d$ must have size at least $(\sqrt{2})^{3^{d-1}+1}/3 = 2^{\Omega(3^d)}$.

Question 2

In Lecture 4 we showed that $R_0(RMAJ_d) \leq (8/3)^d$ for the recursive majority of threes function $RMAJ_d\colon \{0,1\}^n \to \{0,1\}$, $n = 3^d$. In this question you will prove a lower bound for $R_{1/3}(RMAJ_d)$. We say a bit is ε-biased (where $-1 \leq \varepsilon \leq 1$) if it takes value 0 with probability $(1-\varepsilon)/2$ and value 1 with probability $(1+\varepsilon)/2$.

(a) Let X, Y, Z be independent ε-biased bits. Let ε' be the bias of $MAJ_3(X,Y,Z)$. What is ε' as a function of ε?

Solution: This can be calculated by hand but here is a more immediate way to derive it from the polynomial representation of MAJ_3. For $\{-1,1\}$-valued bits the bias is equal to the expectation. If we view MAJ_3 as a function from $\{-1,1\}^n$ to $\{-1,1\}$ and its inputs are independent ε-biased bits then

$$\begin{aligned}\mathrm{E}[MAJ_3(X,Y,Z)] &= \mathrm{E}\left[\tfrac{1}{2}X + \tfrac{1}{2}Y + \tfrac{1}{2}Z - \tfrac{1}{2}XYZ\right] \\ &= \tfrac{1}{2}\mathrm{E}[X] + \tfrac{1}{2}\mathrm{E}[Y] + \tfrac{1}{2}\mathrm{E}[Z] - \tfrac{1}{2}\mathrm{E}[X]\mathrm{E}[Y]\mathrm{E}[Z] \\ &= \tfrac{3}{2}\varepsilon - \tfrac{1}{2}\varepsilon^3.\end{aligned}$$

(b) Show that there exists some small constant $\varepsilon_0 > 0$ such that if $|\varepsilon| \leq \varepsilon_0$ then $\varepsilon' + \varepsilon'^2 \geq \frac{3}{2}(\varepsilon + \varepsilon^2)$.

Solution: By part (a),

$$\varepsilon' + \varepsilon'^2 = (\tfrac{3}{2}\varepsilon - \tfrac{1}{2}\varepsilon^3) + (\tfrac{3}{2}\varepsilon - \tfrac{1}{2}\varepsilon^3)^2 = \tfrac{3}{2}(\varepsilon + \varepsilon^2) + \varepsilon^2(\tfrac{3}{4} - \tfrac{1}{2}\varepsilon - \tfrac{3}{2}\varepsilon^2 + \tfrac{1}{4}\varepsilon^4).$$

The term in the second parenthesis is dominated by $\frac{3}{4}$ and so it is positive when $|\varepsilon|$ is at most say $1/2$, so the whole expression is at least $\frac{3}{2}(\varepsilon + \varepsilon^2)$.

(c) Now let $X_1, \ldots, X_n$, where $n = 3^d$ be $(2/3)^d$-biased bits. Use part (b) to show that the bias of the bit $RMAJ(X_1, \ldots, X_n)$ is lower bounded by some constant independent of d.

Solution: Let's rename the constant ε_0 from part (b) to $1/2$ because we'll use ε_0 for something else. Let ε_d be the bias of $RMAJ_d$ when the inputs are independent ε_0-biased bits. Since $\frac{3}{2}\varepsilon - \frac{1}{2}\varepsilon^3 \geq \varepsilon$ for all $\varepsilon > 0$, the sequence $\varepsilon_0, \ldots, \varepsilon_d$ is non-decreasing. Moreover, by part (b) $\varepsilon_t + \varepsilon_t^2 \geq (3/2)^t(\varepsilon_0 + \varepsilon_0^2)$ as long as $\varepsilon_{t-1} \leq 1/2$. Setting $\varepsilon_0 = (2/3)^d$ we get that $\varepsilon_d + \varepsilon_d^2$ must be at least as large as $1/2$. Then ε_d is at least $1/3$.

For the last part you will need the following theorem from statistics: If $X_1, \ldots, X_\ell$ and $Y_1, \ldots, Y_\ell$ are independent ε-biased and $(-\varepsilon)$-biased bits respectively, then $(X_1, \ldots, X_\ell)$ and $(Y_1, \ldots, Y_\ell)$ are $O(\sqrt{\varepsilon^2\ell})$-indistinguishable by all algorithms.

(d) Show that $R_{1/3}(RMAJ_d) \geq \Omega((9/4)^d)$.

Solution: First we argue that for any deterministic decision tree T of depth ℓ that queries i.i.d. coin flips, the acceptance probability doesn't change if the i-th query is always answered by the i-th coin flip, so that T's input is effectively ℓ bits long. From the perspective of the decision tree, each query is answered by a coin flip independent of all the bits queried so far, so the acceptance probability should not be affected as long as a fresh coin flip is used to answer every query. Therefore $\mathrm{E}[T(X)] - \mathrm{E}[T(Y)] = O(\sqrt{\varepsilon^2\ell})$. If you don't find this convincing, read the next paragraph.

Proof. Write T as a sum of juntas $J_p(X_{p(1)}, \ldots, X_{p(\ell)})$, one for each path p leading to a 1-leaf that queries inputs $p(1), \ldots, p(\ell)$ in that order. Since the inputs are i.i.d. (and therefore exchangeable), $\mathrm{E}[J_p(X_{p(1)}, \ldots, X_{p(\ell)})] = \mathrm{E}[J_p(X_1, X_2, \ldots, X_\ell)]$. By linearity of expectation, $\mathrm{E}[T(X)]$ is therefore equal to $\mathrm{E}[T'(X_1, \ldots, X_\ell)]$, where T' is the decision tree whose i-th query is always the i-th variable. In particular, $\mathrm{E}[T(X)] - \mathrm{E}[T(Y)] = \mathrm{E}[T'(X_1, \ldots, X_\ell)] - \mathrm{E}[T'(Y_1, \ldots, Y_\ell)] = O(\sqrt{\varepsilon^2 \ell})$ by the theorem from statistics. □

Set $\varepsilon = (2/3)^d$. Then $\mathrm{E}[T(X)] - \mathrm{E}[T(Y)]$ is at most $O(\sqrt{(2/3)^{2d}\ell})$, which is less than $1/9$ when $\ell = c(3/2)^{2d}$ for a sufficiently small constant $c > 0$. Suppose for contradiction that there is a randomized decision tree, that is a distribution T over deterministic decision trees, with error $1/9$ for $RMAJ_d$. Then $|\mathrm{E}[RMAJ_d(X)] - \mathrm{E}[T(X)]|$ and $|\mathrm{E}[RMAJ_d(Y)] - \mathrm{E}[T(Y)]|$ are both at most $1/9$. By the triangle inequality $|\mathrm{E}[RMAJ_d(X)] - \mathrm{E}[RMAJ_d(Y)]|$ is less than $1/3$. But by part (c) this difference is at least $(1+1/3)/2-(1-1/3)/2 = 1/3$, a contradiction.

(e) (**Extra credit:**) In conclusion, $(9/4)^d \preceq R_{1/3}(RMAJ_d) \preceq R_0(RMAJ_d) \leq (8/3)^d$. Can you improve any of these bounds?

Solution: Here is a reference. There is still plenty of room for improvement!

Question 3

In Lecture 5 we claimed there exist $\varepsilon\sqrt{n}$-wise indistinguishable distributions μ and ν on $\{-1,1\}^n$ such that μ assigns probability at least 0.99 to the all-ones string and ν assigns probability at least 0.62 to all strings with exactly one -1 for some $\varepsilon > 0$. Let $\phi\colon \{-1,1\}^n \to \mathbb{R}$ be the function

$$\phi(x) = \left(\prod\nolimits_{i=1}^{n} x_i\right) \cdot \mathrm{E}_S\left[\prod\nolimits_{i \in S} x_i\right]^2,$$

where S is a random subset of $\{1, \ldots, n\}$ of size at most $(n-d)/2$ (chosen uniformly among all such subsets), and E_S is expected value.

(a) Show that if $p\colon \{-1,1\}^n \to \mathbb{R}$ is a polynomial of degree less than d then $\sum_{x \in \{-1,1\}^n} p(x)\phi(x) = 0$. (**Hint:** Look at the monomials of p and ϕ.)

Solution: All monomials in ϕ are of degree at least d: The monomial inside the expectation are of degree at most $(n-d)/2$. After squaring their degree does not exceed $n-d$, so after multiplying by the leading term the degree will be at least d. Therefore p and ϕ do not have any monomials in common. Multiplying any two of these monomials is therefore a non-constant monomial, which averages out to zero when x ranges over all n-bit $\{-1,1\}$ strings.

(b) Let $\mu(x) = \max\{2\phi(x)/Z, 0\}$ and $\nu(x) = \max\{-2\phi(x)/Z, 0\}$. Use part (a) to show that for some choice of Z, μ and ν are probability distributions that are $(d-1)$-wise indistinguishable.

Solution: Setting $p(x) = 1$ in part (a) we get that $\sum \phi(x) = 0$, which means that the positive and negative values of ϕ must add up to the same value. So if they are both normalized by the same constant $Z/2$ they both become probability distributions μ and ν. Writing $\phi(x) = (Z/2)(\mu(x) - \nu(x))$, it follows from part (a) that $\sum p(x)\mu(x) = \sum p(x)\nu(x)$ for every p of degree $d-1$ or less. In particular this is true when p is a $(d-1)$-junta. Therefore all $(d-1)$-juntas average out to the same value under μ and ν and so the two are $(d-1)$-indistinguishable.

(c) Show that $Z = \sum_{x \in \{-1,1\}^n} \mathrm{E}_S\left[\prod_{i \in S} x_i\right]^2$. Calculate $2/Z$ in terms of n and d. **Solution:** The

identity follows from

$$\sum_{x\in\{-1,1\}^n} \mathrm{E}_S\big[\textstyle\prod_{i\in S} x_i\big]^2 = \sum_{x\in\{-1,1\}^n} |\phi(x)| = \frac{Z}{2}\cdot\Big(\sum_{x\in\{-1,1\}^n}\mu(x) + \sum_{x\in\{-1,1\}^n}\nu(x)\Big) = Z$$

because μ and ν are probability distributions. To calculate Z we can expand the squared expectation as a product of expectations over independent choices of sets S, T so that

$$\begin{aligned} Z &= \sum_{x\in\{-1,1\}^n} \mathrm{E}_S\Big[\textstyle\prod_{i\in S} x_i\Big]\,\mathrm{E}_T\Big[\textstyle\prod_{i\in S} x_i\Big] \\ &= \sum_{x\in\{-1,1\}^n} \mathrm{E}_{S,T}\Big[\textstyle\prod_{i\in S\oplus T} x_i\Big] \\ &= \mathrm{E}_{S,T} \sum_{x\in\{-1,1\}^n} \textstyle\prod_{i\in S\oplus T} x_i, \end{aligned}$$

where $\oplus$ is symmetric set difference. The terms $S \neq T$ vanish because the product averages out to zero and take value 2^n when $S = T$. Therefore $Z = 2^n \Pr[S = T]$ and

$$\frac{Z}{2} = 2^{n-1}\Pr[S=T] = 2^{n-1}/\big(\tbinom{n}{0} + \cdots + \tbinom{n}{(n-d)/2}\big).$$

(d) Let $d = \varepsilon\sqrt{n}$. Calculate $\lim_{\varepsilon\to 0}\lim_{n\to\infty}\mu(1^n)$, where 1^n is the all-ones string.

Solution: The reciprocal $2/Z$ is the probability that a random 0/1 string has at most $(n-d)/2$ ones conditioned on it having at most $n/2$ ones. Since a Binomial$(n, 1/2)$ random variable has mean $n/2$ and standard deviation $\sqrt{n}/2$, by the Central Limit Theorem when $d = \varepsilon\sqrt{n}$, $2/Z$ approaches the probability that a Normal$(0, 1)$ random variable is greater than ε conditioned on it being positive. When ε tends to zero this probability tends to 1.

(e) Let $W = (n - 2|S|)/\sqrt{n}$. Show that $\sum_{x\in N}\nu(x) = 2\,\mathrm{E}[W]^2/Z$, where N is the set of strings with exactly one -1.

Solution: Let x be a string with exactly one 1. Conditioned on the size of S being s (and S being otherwise random), $\prod_{i\in S} x_i$ takes value -1 with probability s/n and 1 with the remaining probability, so $\mathrm{E}_S\big[\prod_{i\in S} x_i\big]^2 = \mathrm{E}[1 - 2|S|/n]^2 = \mathrm{E}[W]^2/n$. Since $\sum_{x\in N}\nu(x)$ is $2/Z$ times the sum of n such terms it equals $2\,\mathrm{E}[W]^2/Z$.

(f) Use part (e) and the Central Limit Theorem to calculate $\lim_{\varepsilon\to 0}\lim_{n\to\infty}\sum_{x\in N}\nu(x)$.

Solution: $|S|$ is a Binomial$(n, 1/2)$ random variable B conditioned on it having value at most $(n - d)/2$. Its normalized CDF is:

$$\begin{aligned} \Pr[W \le w] &= \Pr[|S| \le (n - w\sqrt{n})/2] \\ &= \Pr[B \le (n - w\sqrt{n})/2 \mid B \le (n - \varepsilon\sqrt{n})/2] \\ &= \frac{\Pr[B \le (n - w\sqrt{n})/2]}{\Pr[B \le (n - \varepsilon\sqrt{n})/2]}. \end{aligned}$$

By the Central Limit Theorem, the numerator and denominator converge to $\Pr[N \le w]$ and $\Pr[N \le \varepsilon]$ respectively, for a Normal$(0, 1)$ random variable N. Therefore

$$\lim_{n\to\infty}\Pr[W \le w] = \Pr[N \le w \mid N \le \varepsilon].$$

We would therefore expect that

$$\lim_{n\to\infty}\mathrm{E}[W] = \mathrm{E}[N \mid N \le \varepsilon], \tag{1}$$

from where by continuity of the function $\varepsilon \to \mathrm{E}[N \mid N \leq \varepsilon]$ we can calculate

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \mathrm{E}[W] = \mathrm{E}[N \mid N \leq 0] = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{0} x e^{-x^2/2} dx = -\sqrt{\frac{2}{\pi}},$$

which, together with parts (d) and (e), gives $\lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{x \in N} \nu(x) = 2/\pi$.

To be completely formal we need to prove (1). It is a bit easier to work with $-W$ instead of W as $-W$ is non-negative. Let us also write N_ε for N conditioned on $N \geq \varepsilon$. By the Central Limit Theorem $\Pr[-W \geq w]$ and $\Pr[N_\varepsilon \geq w]$ are δ_n-close for some δ_n that goes to zero as n increases.[1] Using the formula $\mathrm{E}[X] = \int_0^\infty \Pr[X \geq w] dw$ which holds for all non-negative X we can bound the difference in expectations by

$$\begin{aligned}
|\mathrm{E}[-W] - \mathrm{E}[N_\varepsilon]| &= \left| \int_0^\infty (\Pr[-W \geq w] - \Pr[N_\varepsilon \geq w]) dw \right| \\
&\leq \int_0^B |\Pr[-W \geq w] - \Pr[N_\varepsilon \geq w]| dw \\
&\quad + \int_B^\infty \Pr[-W \geq w] dw + \int_B^\infty \Pr[N_\varepsilon \geq w] dw.
\end{aligned}$$

Set $B = 1/\sqrt{\delta_n}$. By the Central Limit Theorem the integrand in the first term is at most δ_n so the value of the integral is at most $\sqrt{\delta_n}$. The other two terms can be handled using large deviation bounds. First, $\Pr[N_\varepsilon \geq w] = O(\Pr[N \geq w]) = O(e^{-w^2/2})$. For $\Pr[-W \geq w]$ we can apply for instance the Chernoff bound to conclude that it is also $O(e^{-w^2/2})$. Therefore, up to a constant, both integrals are at most $O(e^{-B^2/2}) = O(e^{-1/\delta_n})$. In the limit as δ_n goes to zero, the right-hand terms go to zero and so the expectations approach one another.

(g) Use parts (b), (d), and (f) to prove the claim.

Solution: Let n be sufficiently large. By part (b) μ for $d - 1 = \varepsilon\sqrt{n}$, μ and ν are $\varepsilon\sqrt{n}$-wise indistinguishable. By parts (d) and (f) in the limit $\varepsilon \to 0$, $\mu(1^n)$ approaches 1 and $\nu(N)$ approaches $2/\pi \geq 0.63$. By continuity for some $\varepsilon > 0$ we get that $\mu(1^n) \geq 0.99$ and $\nu(N) \geq 0.62$ as desired.

(h) **(Research project:)** Can you come up with $\varepsilon\sqrt{n}$-wise indistinguishable μ and ν for which both the limits in part (d) and part (f) are 1? I know that they exist but I don't know a "nice" formula for them.

[1] This assumes the convergence in the Central Limit Theorem is uniform in w, which is true.