Question 1

Call a function *k-uncertifiable* if for every subset of k inputs, no partial assignment to those inputs fixes the function value to 1. Show the following:

(a) $DISTINCT(x, y) = (x_1 \neq y_1)$ OR $\cdots$ OR $(x_n \neq y_n)$ has a decision tree of size $O(2^n)$.

Solution: The following decision tree solves $DISTINCT$. Read x_1 and then y_1. If they are different output zero, if they are equal recursively solve $DISTINCT$ on $2n-2$ inputs. The size of this decision tree satisfies the recurrence $s(n) = 2s(n-1) + 2$ with initial condition $s(1) = 4$ which solves to $s(n) = 3 \cdot 2^n - 2$.

(b) If ϕ is a DNF for a k-uncertifiable function then all terms of ϕ must have width more than k.

Solution: We argue the contrapositive. If ϕ has a term of width k or less then there is an assignment to its k variables that forces the value of ϕ to 1 so ϕ is k-certifiable.

(c) A size-s DNF for a k-uncertifiable function can accept at most a $s \cdot 2^{-k-1}$-fraction of all possible inputs.

Solution: Let $\phi_1, \ldots, \phi_s$ be the terms of the DNF ϕ. By a union bound, for a random input x,

$$\mathrm{P}[\phi \text{ accepts } x] \leq \sum_{i=1}^{s} \mathrm{P}[\phi_i \text{ accepts } x].$$

By part (b), each ϕ_i has at least $k+1$ literals. The input x is not accepted unless all of them are true, which can happen with probability at most $2^{-(k+1)}$. So the probability that ϕ accepts x can be at most $s \cdot 2^{-k-1}$.

(d) The function $EQUAL =$ NOT $DISTINCT$ requires DNFs of size 2^n. Use part (c).

Solution: $EQUAL$ is $(2n-1)$-uncertifiable: Consider a partial assignment to any $2n-1$ variables, say all except for y_n. If $x_1 \ldots x_{n-1} \neq y_1 \ldots y_{n-1}$ then $f(x, y)$ is fixed to zero. Otherwise, setting y_n to the negation of x_n sets $f(x, y)$ to zero. In either case $f(x, y)$ is unfixed by the partial assignment. Since x and y are equal with probability 2^{-n}, by part (c) any size-s DNF for EQUAL must satisfy $s \cdot 2^{-2n} \geq 2^{-n}$, so $s \geq 2^n$.

(e) $DISTINCT$ requires decision trees of size 2^n. Use part (d).

Solution: The decision tree sizes of f and NOT f are equal as one can be obtained by the other by relabeling the value at each leaf with its negation. In particular, if $DISTINCT$ had a decision tree of size 2^n so would $EQUAL$. As DNF size is upper bounded by decision tree size this would contradict part (d).

Question 2

Recall the INJ function from Lecture 2:

$$INJ(x_1, \ldots, x_n) = AND_{i \neq j} DISTINCT(x_i, x_j), \qquad x_i \in \{0, \ldots, m-1\}.$$

Assume $m \geq n$, n and m are both powers of 2, and elements of $\{0, \ldots, m-1\}$ are specified by their bit representation. Show the following:

(a) INJ is $(n \log n - 1)$-uncertifiable when $m = n$. (**Optional:** Is this true when $m > n$?)

Solution: By symmetry it is enough to prove this when one of the bits of x_n is missing. If $x_1, \ldots, x_{n-1}$ are not distinct then INJ is fixed to zero. If they are distinct then the two choices for the missing bit of x_n yield two distinct values for x_n. If both of them appear among $x_1, \ldots, x_{n-1}$ the value is again fixed to zero. If not then one of them appears and the other one doesn't so the value of INJ depends on the missing bit.

The general case is modeled by the following combinatorial problem. A *bipartite covering* of the complete graph K_n on n vertices is a collection of complete bipartite graphs $X_1 \times Y_1, \dots, X_k \times Y_k$ such that their union covers all $\binom{n}{2}$ possible edges. The size of the covering is $|X_1| + |Y_1| + \dots + |X_k| + |Y_k|$. INJ is $(n \log n - 1)$-uncertifiable if and only no bipartite covering of K_n of size less than $n \log n$ exists. Proof sketch: If a bipartite covering of size s exists, then a certificate of the same size is obtained by setting the i-th bit of all items in X_i and Y_i to 0 and 1, respectively. Conversely, an s-certificate for INJ can be represented by a size-s covering. Similar problems have been studied (see these papers by Alon and by Jukna and Kulikov and references within). I couldn't work out this variant or find a reference for it; it can be a possible project if you are interested.

(b) INJ requires DNFs of size at least $n!$ when $m \geq n$.
(**Hint:** Reduce to the case $m = n$ and use part (c) of question 1.)

Solution: INJ with $m > n$ restricts to INJ with $m = n$ by setting all but say the $\log n$ least significant bits of each item to zero, so any DNF size lower bound for $m = n$ also applies to $m > n$. When $m = n$, are n^n possible inputs $(x_1, \dots, x_n)$ out of which $n!$ have all items distinct, so the fraction of satisfying inputs is $n!/n^n$. Therefore any DNF must have size at least $(n!/n^n) \cdot 2^{n \log n} = n^n$. Another way of saying this is that every term must look at all inputs so it can accept exactly one. Therefore there must be $n!$ terms to cover all of them.

(c) INJ has CNFs (ANDs of ORs of literals) of size $\binom{n}{2}m$.

Solution: $DISTINCT(x, y)$ can be written as an OR over all possible assignments $t \in [m]$ of the predicate $(x = t)$ AND $(y = t)$ giving a CNF representation of size m. (For a slightly larger representation one can use the decision tree of problem 1(a)). Therefore INJ can be represented as an AND of $\binom{n}{2}$ ANDs of m ORs giving a CNF of size $\binom{n}{2}m$.

(d) INJ requires CNFs of size at least m for any $n \geq 2$. (**Hint:** Use part (d) of question 1.)

Solution: When $n = 2$, INJ is $DISTINCT$ on $2 \log m$ input bits so it requires CNF size 2^n by 1(d) (a CNF for f can be converted to a DNF for NOT f of the same size using de Morgan's laws so minimum CNF size for f equals minimum DNF size for NOT f).

I don't know how to solve this problem when $n > 2$. In fact, the claim is false when $n > m$ as INJ is then always false. I was hoping that other values of n can be reduced to $n = 2$ by restriction, but restricting in this question changes the value of m affecting the bound. It may still be possible to prove that NOT INJ is k-uncertifiable for sufficiently large k and use 1(c). At first I thought that NOT INJ is $(2 \log m - 1)$-uncertifiable for every $m = n$ but this turns out to be false when m and n are small: If $n = m = 4$ then NOT INJ has a 3-certificate, namely "the first bits of x_1, x_2, and x_3 are all zero." As there are only four possible values this forces two of them to be equal, i.e., NOT INJ to evaluate to one. This can be another research project...

Question 3

Let $f(x_1, \dots, x_m)$ be a boolean function and $y_1, \dots, y_n$ be another set of varaibles. A *projection* of f is a function obtained by replacing each x_i with one of the literals $y_1, \dots, y_n, \overline{y_1}, \dots, \overline{y_n}$ or one of the constants $0, 1$. We say f projects to g if there exists a projection of f that equals the function $g(y_1, \dots, y_n)$.

(a) The AND-OR tree on n inputs is defined by the recursive formula

$$AOT(x, y, z, w) = (AOT(x) \text{ OR } AOT(y)) \text{ AND } (AOT(z) \text{ OR } AOT(w)),$$

where $x, y, z, w \in \{0,1\}^{n/4}$ and n is a power of four. The base case is $AOT(x) = x$. Show that AOT on n^2 inputs projects to PARITY on n inputs for every n that is a power of two. (**Hint:** Use induction.)

Solution: AOT on four inputs projects to $PARITY$ of two bits because $x \oplus y$ equals $(x$ OR NOT $y)$ AND (NOT x OR $y)$. Assuming the claim is true for n, we can represent $PARITY(x, y)$ for a $2n$-bit string (x, y) as

$$PARITY(x) \oplus PARITY(y) = (PARITY(x) \text{ OR NOT } PARITY(y)) \text{ AND } (\text{NOT } PARITY(x) \text{ OR } PARITY(y)).$$

By inductive assumptions, $PARITY(x)$ and $PARITY(y)$ are projections of AOT on n^2 inputs, so $PARITY(x, y)$ is a projection of AOT on $4n^2 = (2n)^2$ inputs as desired.

(b) Use part (a) and facts about PARITY from class to show that AOT requires depth-d AND/OR circuits of size $2^{\Omega(n^{1/2(d-1)})}$.

Solution: If AOT on $n = m^2$ inputs had such circuits of size s by part (a) so would $PARITY$ on m inputs, so s would have to be at least $2^{\Omega(m^{1/(d-1)})} = 2^{\Omega(n^{1/2(d-1)})}$.

(c) Valiant's Theorem states that there exists a constant $c > 1$ for which recursive majority on at most n^c inputs projects to MAJORITY on n inputs. Recursive majority is the function

$$RMAJ(x, y, z) = MAJORITY\big(RMAJ(x), RMAJ(y), RMAJ(z)\big),$$

where $x, y, z \in \{0,1\}^{n/3}$ and n is a power of 3. The base case is $RMAJ(x) = x$. Assuming Valiant's theorem, show that AOT requires depth-d AND/OR/PARITY circuits of size $2^{\Omega(n^{\varepsilon/d})}$ for some constant $\varepsilon > 0$.

Solution: We first show that AOT on 16 inputs projects to majority on 3 inputs. One way to obtain this projection is to first represent $MAJORITY(x, y, z)$ as the CNF $(x \text{ OR } y)$ AND $(y \text{ OR } z)$ AND $(z \text{ OR } x)$ (at least two out of three must be true). This can be written as, for example

$$AOT(x, y, y, z) \text{ AND } AOT(z, x, 1, 1) = AOT(AOT(x, y, y, z), AOT(0, 0, 0, 0), AOT(z, x, 1, 1), AOT(0, 0, 0, 0)).$$

By the same inductive argument as in part (a), AOT on $n = 16^d$ inputs projects to $RMAJ$ on $3^d = n^{(\log_1 63)d}$ inputs, which itself projeccts to $MAJORITY$ of n^ϵ inputs for some $\epsilon > 0$. As this function requires AND/OR/PARITY circuits of size at least $\Omega(2^{n^\epsilon/4d})$ so must AOT.

Question 4

A function $f\colon \{0,1\}^n \to \{0,1\}$ has *rational degree* at most d (over $\mathbb{F}_2$) if

$$\text{there exist polynomials } p \text{ and } q \text{ of degree at most } d \text{ such that } pf = q \text{ and } p \neq 0. \tag{1}$$

($a = b$ means $a(x) = b(x)$ for all x.) Show that for any given f,

(a) (1) implies

$$\begin{array}{l}\text{there exists } r_0 \neq 0 \text{ of degree at most } d \text{ such that for every } x,\ f(x) = 0 \text{ implies } r_0(x) = 0, \text{ or}\\ \text{there exists } r_1 \neq 0 \text{ of degree at most } d \text{ such that for every } x,\ f(x) = 1 \text{ implies } r_1(x) = 0.\end{array} \tag{2}$$

Solution: If $q \neq 0$ set $r_0 = q$. As $pf = q$, the zeros of q must contain all the zeros of p. If $q = 0$, set $r_1 = p$. As $pf = 0$, $p(f + 1) = p$ so whenever f equals one pf must equal zero and so must p.

(b) (2) implies (1).

Solution: In the first case $r_0(f + 1)$ must always equal zero so we can choose $p = q = r_0$. In the second case $r_1 f$ must always equal zero so we can choose $p = r_1$ and $q = 0$.

(c) (1) implies

$$\text{there exists a function } g \text{ such that for all polynomials } p \text{ and } q \text{ of degree less than } n-d,\ g \neq pf+q. \tag{3}$$

(**Hint:** Try a proof by contradiction.)

Solution: Assume (3) is false, namely every g has a representation of the form $pf + q$ for some p, q of degree less than $n-d$, but (1) is true so $sf = t$ for some degree-d polynomials s and t with $s \neq 0$. Combining the two equations we obtain that for every g there exists p, q with the given degrees so that $gs = pfs + qs = pt + qs$. The right-hand side is a polynomial of degree strictly less than n. However we can always choose a g so that the left-hand side has degree n. For example, we can take g to be the monomial consisting of all the variables that do not appear in some highest-degree term in s. Then gs must contain the monomial $x_1 x_2 \cdots x_n$ so it has degree n.

(d) (**Optional**, requires some $\mathbb{F}_2$-linear algebra) (3) implies (1).
(**Hint:** $g = pf + q$ is a system of linear equations whose variables are the coefficients of p and q.)

Solution: If the system of linear equations $g(x) \neq p(x)f(x)+q(x)$ as x ranges over $\{0,1\}^n$ has no solution in the coefficients of p and q for some g then some linear combination of the equations must give a contradiction. Namely, there must exist a (nonzero) function r such that $\sum g(x)r(x) \neq \sum(p(x)f(x) + q(x))r(x)$ for all p and q of degree less than $n-d$. The summation is over all $x \in \{0,1\}^n$. We will show that both r and fr can have degree at most d giving the representation $r \cdot f = fr$ of the desired form.

As the left-hand side does not depend on p or q the right-hand side must take the same value for all p, q. By setting $p = q = 0$ we get that this value must be zero, namely

$$\sum(p(x)f(x) + q(x))r(x) = 0 \qquad \text{for all } p, q \text{ of degree less than } n-d.$$

Setting p to zero we get that $\sum q(x)r(x) = 0$ for all q of degree less than $n-d$. The only monomial m for which $\sum m(x) = 1$ is the degree-n monomial $x_1 \cdots x_n$, so $q(x)r(x)$ cannot contain this monomial for every choice of q. Therefore r cannot contain any monomials of degree greater than d because q could then be chosen to equal the complementary monomial. In conclusion, r can have degree at most d. By the same argument, choosing $q = 0$ gives the constraint $\sum p(x)(f(x)r(x)) = 0$ for all p of degree less than $n-d$, so the degree of fr is also at most d.

(e) f has rational degree at most $\lceil n/2 \rceil$. Use part (d).

Solution: By part (d) it is sufficient to show there exists a g that cannot be represented as $pf + q$ for p and q of degree strictly less than $n - \lceil n/2 \rceil = \lfloor n/2 \rfloor$. To do this we count the number of possible representations of this form. A degree-d polynomial is specified by its coefficients, which correspond to subsets of $\{1, \ldots, n\}$ of size at most d. When $d < \lfloor n/2 \rfloor$ this number is strictly less than $2^n/2$ because the subsets of size at most d and their complements do not cover all possible sets. Therefore there are fewer than $2^{2^n/2}$ choices for each of p and q and so fewer than 2^{2^n} representations of type $pf + q$. At least one of the 2^{2^n} functions g does not have a representation.