Boolean Decision trees

Problems and Results,
Old and New

Michael Saks, Rutgers University

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

- Unknown graph G on vertex set $\{1, 2, 3, 4, 5\}$.

- I want to know: is G connected?

- I can ask: Is (i, j) an edge?

What is the fewest number of questions needed in worst case?

Adversary view

- You (my adversary) answer the questions.
- How many questions can you force me to ask?

Analogous problem for any graph property, e.g.:

- Is G planar?
- Is G 3-colorable?

More generally . . .

Evaluating boolean functions

- Boolean variables $x_1, x_2, \ldots, x_n$ with unknown values
- Given boolean function $f : \{0, 1\}^n \longrightarrow \{0, 1\}$
- Goal: evaluate $f(x_1, \ldots, x_n)$.
- Elementary step: What is x_i?

Evaluating boolean functions

How many questions are needed in worst case?

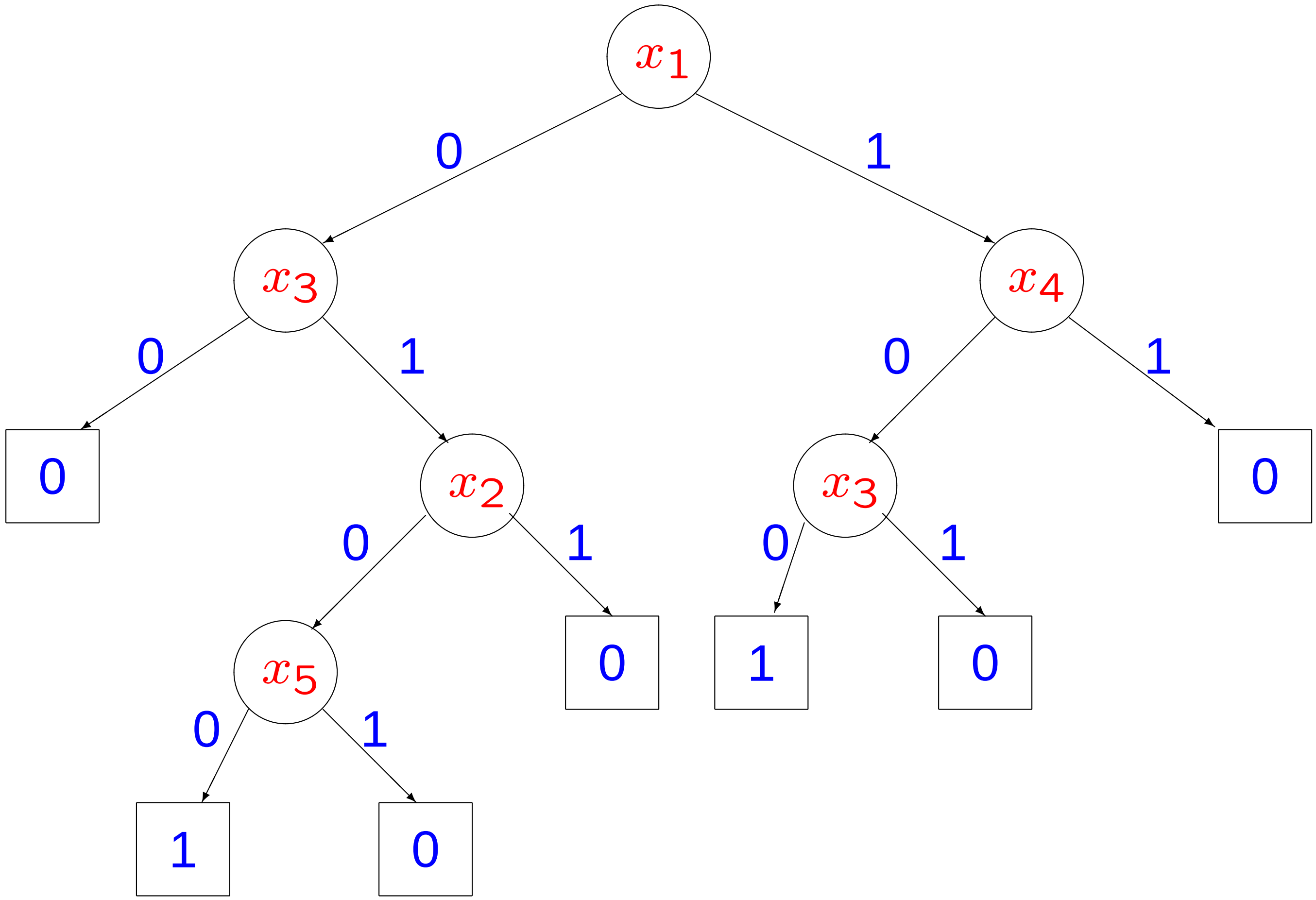
Note: All other computation is free.

Graph properties

For a graph property over graphs on vertex set V of size v:

- Variables are $x_{\{i,j\}}$ for $i, j \in V$.
- $n = \binom{v+1}{2}$.

Boolean decision trees



Function computed by BDT T

Function f_T computed by decision tree T:

- Input x determines a root-to-leaf path.
- Output $f_T(x)$ = label of leaf.

Cost of decision tree

The cost of T on input x:

$$D_x(T) = \text{depth of path followed by } x$$
$$= \text{number of queries.}$$

Worst case cost of T: $D(T) = \max_x D_x(T)$.

Deterministic DT complexity of f:

$$D(f) = \min\{D(T) : T \text{ computes } f\}$$

.

Evasive functions

Trivial upper bound:

$$D(f) \leq n.$$

f is evasive if

$$D(f) = n.$$

Lower bounds by adversary arguments

Adversary strategy: strategy for answering questions

- OR
- MAJ
- Symmetric functions
- Graph connectivity (Adversary strategy: Just say No)

Can $D(f)$ ever be less n?

Irrelevant variables

- $f(x_1, \ldots, x_n) = x_7$.

What if all variables are relevant?

The addressing function

Variables $x_1, \ldots, x_k, y_1, \ldots, y_{2^k}$:

$$f(x_1, \ldots, x_k, y_1, \ldots, y_{2^k}) = y_{x_1, \ldots, x_k}.$$

- $n = 2^k + k$
- $D(f) = k + 1$.

f is highly asymmetric...

Automorophism group of f

$\Gamma(f)$ is Automorphism group of f:

Set of permutations of variables that leave f unchanged.

- For symmetric functions,

  $\Gamma(f)$ is the full symmetric group

- For graph properties on graphs with vertex set V

  $\Gamma(f)$ is subgroup induced by symmetric group on V.

Weakly symmetric functions

f is weakly symmetric if $\Gamma(f)$ is transitive:

for any two variables some σ in $\Gamma(f)$ maps one to the other.

E.g. any function coming from a graph property.

- Is every non-constant weakly symmetric functions evasive?
- Is every non-constant graph property evasive?

An example

Digraph property: Does G have a supersink?

- With $v-1$ questions narrow to one candidate.

- With $2v-3$ more questions check the candidate.

$D(f) \leq 3v-4$ compared to $v(v-1)$ variables.

(There are also examples for undirected graph properties.)

Aanderaa-Rosenberg conjecture

A graph property is monotone if it is preserved under addition of edges (e.g. connectivity).

> (1973) Any non-trivial monotone graph property on v vertices has complexity $\omega(v^2)$.

Stengthening (attributed to (but denied by) Karp):

> Any non-trivial monotone graph property is evasive.

Progress on AR conjecture

Lower bounds for monotone graph properties on v vertices:

$\frac{v^2}{16}$ Rivest-Viullemin (1975) settling AR

$\frac{v^2}{9}$ Kleitman-Kwiatkowski (1980)

$\frac{v^2}{4} + o(v^2)$ Kahn-Saks-Sturtevant (1984)

Progress on evasiveness

Specific properties	Best-van Emde Boas- Lenstra (1974)
	Milner-Welsh (1975)
	Bollobás (1976)
Any GP for v a prime power	Kahn-Saks-Sturtevant (1984)
Analog for bipartite graph properties	Yao (1988)
Additional classes of graph properties	Triesch (1994)
	Chakrabarti, Khot, Shi (2001)

General lower bounds on $D(f)$

$D(f) \geq deg(f)$ (Fourier degree) (Best, van Emde Boas, Lenstra)

Implications:

- Almost all functions f are evasive
- Together with elegant counting argument: When n is a prime power, every weakly symmetric n-variate f satisfying $f(0^n) \neq f(1^n)$ is evasive. (Rivest-Viullemin 1975).

A topological connection

Associate f to a collection of subsets of $\{1, \ldots, n\}$:

$$\Delta(f) = \{S \subseteq \{1, \ldots, n\} : f(\chi_S) = 0\}.$$

If f is monotone then $\Delta(f)$ is an abstract simplicial complex. Then:

> $D(f) < n$ implies $\Delta(f)$ is collapsible (and thus contractible). (Kahn-Saks-Sturtevant 1984)

Contrapositive: if $\Delta(f)$ is not contractible then f is evasive.

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

Randomized decision trees

Questioner may flip coins when choosing next query.

Example. Does 0-1 matrix M have an all 1 row?

On worst case input, expected number of queries is $n/2$.

RDTs: Model

Two models (equivalent for our purposes):

- Decision tree may have random nodes which are not counted towards computation cost.

- Randomized Algorithm = probability distribution $\tilde{T}$ over deterministic decision trees.

RDTs: Model

Las Vegas algorithm:

- Algorithms must always answer correctly
- Algorithm to compute f is a distribution over DDTs that each compute f.
- Goal: Minimize the expected number of queries used on worst case input.
- Adversary view: Adversary chooses the input knowing the algorithm, but not the results of the coin flips.

Iterated majority of 3

- $\mathsf{I3M}_1$ is majority of 3

- $\mathsf{I3M}_k$ has 3^k variables, split into 3 groups x^k, y^k, z^k.

- $\mathsf{I3M}_k(x^k, y^k, z^k) = \mathsf{MAJ}(\mathsf{I3M}_{k-1}(x^k), \mathsf{I3M}_{k-1}(y^k), \mathsf{I3M}_{k-1}(z^k))$.

$\mathsf{I3M}_k$ is evasive. (By easy adversary argument, also R-V theorem.

Randomized DT for I3M

To evaluate I3M$_k$

- Choose 2 groups of variables (out of 3) at random
- Recursively Evaluate I3M$_{k-1}$ on 2 selected groups.
- Evaluate I3M$_{k-1}$ on remaining group if needed

Randomized DT for I3M

Expected cost: $\left(\frac{8}{3}\right)^k \approx n^{.893}$

Upper bound can be improved (Saks-Wigderson 1986).

Best lower bound: $\left(\frac{7}{3}\right)^k$ (Jayram, Kumar, Sivakumar 2003 via information theory)

Best upper and lower bounds don't match.

Another example: Iterated $\vee$-$\wedge$

Function F_k has

- F_1 has 4 variables, $F_1(a, b, c, d) = (a \wedge b) \vee (c \wedge d)$
- F_k has 4^k variables and is obtained by iterating F_1.

Best upper bound (Snir 1983): $O(n^{.754})$

Matching lower bound (Saks-Wigderson 1986).

How much does randomization help?

- For all n, $\exists$ n-variate evasive f with $R(f) \leq n^{.754}$.

- For any f, $R(f) \geq D(f)^{1/2}$.

Open problems

- Find largest α such that $R(f) \geq D(f)^{\alpha}$ for all f

- Conjecture (Saks-Wigderson)

$$\alpha = .754.$$

- Conjecture (Yao): For monotone graph properties, randomness does not help, i.e.

$$R(f) = \Omega(n) = \Omega(v^2)$$

.

Lower bounds on $R(f)$ for monotone graph properties

v	easy
$v(\log v)^c$	Yao 1987
$v^{5/4}$	King 1988
$v^{4/3}$	Hajnal 1991

Lower bounds on $R(f)$ for monotone graph properties

$v^{4/3}(\log v)^{1/3}$	Chakrabarti and Khot 2001
$\min\{\frac{v}{p}, \frac{v^2}{\log v}\}$	Friedgut, Kahn, Wigderson 2002
$\frac{v^{4/3}}{p^{1/3}}$	O'Donnell, Saks, Schramm, Servedio 2005.

Critical probability p for monotone f is unique p so that:

if $x_i = 1$ independently with probability p, then **Prob**$[f = 1]$ is $\frac{1}{2}$.

Lower bound for $R(f)$ for weakly symmetric f

Theorem. For any monotone weakly symmetric n-variate f,

$$R(f) = \frac{n^{2/3}}{p^{1/3}}.$$

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

Influences

How do we measure the influence of a variable on a boolean function?

$$f : \{-1,1\}^n \longrightarrow \{-1,1\}$$

Product distribution on $\{-1,1\}^n$:

$$\mu = \prod \mu_i \qquad (\mu_i \text{ is 1 w.p. } p_i)$$

$p_i = 1/2$ for all i: uniform distribution

p-biased influence of variable i on f

	1	2	3	4	5	6	7	8	9	10	11	12
x												
y												

Inf$_i^p(f)$ is **Prob**$[f(x) \neq f(y)]$

when x, y generated by ...

p-**biased influence** of variable i on f

	1	2	3	4	5	6	7	8	9	10	11	12
x	x_1	x_2	x_3	x_4	x_5	x_6		x_8	x_9	x_{10}	x_{11}	x_{12}
y	x_1	x_2	x_3	x_4	x_5	x_6		x_8	x_9	x_{10}	x_{11}	x_{12}

Inf$_i^p(f)$ is **Prob**$[f(x) \neq f(y)]$

For $j \neq i$, select $x_j = y_j$ by p-biased coin

p-biased influence of variable i on f

	1	2	3	4	5	6	7	8	9	10	11	12
x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
y	x_1	x_2	x_3	x_4	x_5	x_6	y_7	x_8	x_9	x_{10}	x_{11}	x_{12}

Inf$_i^p(f)$ is **Prob**$[f(x) \neq f(y)]$

Select x_i and y_i independently by p-biased coin

Max influence and Total influence

$\mathbf{Inf}_i^p(f)$	p-biased influence of i on f
$\mathbf{Inf}_{\mathrm{max}}^p(f)$	Maximum of $\mathbf{Inf}_i^p(f)$ over i
$\mathbf{Inf}_{\Sigma}^p(f)$	Sum of $\mathbf{Inf}_i^p(f)$ over i (Total influence).

KKL lower bound on max influence

Assume uniform distribution.

f is *balanced* if critical probability is 1/2.

- Elementary: There is always a variable of influence at least $1/n$.
- Kahn-Kalai-Linial 1988: $\ldots$ always a variable of influence $\Omega(\frac{\log n}{n})$.

This is best possible (tribes function).

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

Influences in functions of small decision tree depth

If f has (randomized) decision tree depth d, then f depends at most 2^d variables.

By KKL, f has a variable of influence $\Omega(\frac{d}{2^d})$.

Theorem.(O'Donnell-Saks-Schramm-Servedio 2005)

For balanced f:

There is a variable of influence at least $\frac{1}{R(f)}$.

(A family of examples shows that this is best possible.)

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

Consequence for $R(f)$

Theorem says:

For balanced f, $R(f) \geq \frac{1}{\mathbf{Inf}_{\mathsf{max}}(f)}$.

Generalization to arbitrary f with critical probability p:

Theorem. (OSSS 2005)

$$R(f) \geq \frac{1}{\mathbf{Inf}^{p}_{\mathsf{max}}(f)}.$$

Consequence for $R(f)$

$$R(f) \geq \frac{1}{\mathbf{Inf}^p_{\max} f}.$$

Theorem.(O'Donnell-Servedio 2005) For monotone f:

$$R(f) \geq \frac{(\mathbf{Inf}^p_{\Sigma}(f))^2}{4p(1-p)}$$

Corollary.(OSSS 2005) For monotone, weakly-symmetric f,

$$R(f) \geq \frac{n^{2/3}}{(4p(1-p))^{1/3}}.$$

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

Proof of influence bound

Goal: f has a variable of p-biased influence at least $1/R(f)$.

A Strengthening

.

Fix:

- n-variate deterministic decision tree T for f
- p-biased probability distribution on variables.
- $\delta_i =$ **Prob**$[T$ reads variable $i]$

Theorem.

$$\sum_{i=1}^{n} \delta_i \mathbf{Inf}_i \geq \mathbf{Var}[f]$$

Corollary. There is a variable with influence at least $\frac{\mathbf{Var}[f]}{R(f)}$.

If p is a critical probability for f, then $\mathbf{Var}[f] = 1$.

Theorem.

$$\sum_{i=1}^{n} \delta_i \mathbf{Inf}_i \geq \mathbf{Var}[f]$$

Two (related) proofs:

- Combinatorial (injective) (more intuitive)
- Analytic (gives more general results)

Combinatorial proof

$$\textbf{Var}[f] \leq \sum_{i=1}^{n} \delta_i \textbf{Inf}_i.$$

(For this talk: assume $p = \frac{1}{2}$.)

Multiply both sides by 2^{2n-1}:

Combinatorial formulation

LHS counts:

Pairs x, y

x	x_1	x_2	x_3	x_4	x_5	x_6

$$f(x) \neq f(y)$$

y	y_1	y_2	y_3	y_4	y_5	y_6

RHS counts:

Triples u, v, i

u	u_1	u_2	u_3	u_4	u_5	u_6

$T(u)$ reads i

i=4

$f(v)$ sensitive at i

v	v_1	v_2	v_3	v_4	v_5	v_6

Proof strategy: Construct an injection

Input

	1	2	3	4	5	6	7	8	9	10	11	12
x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
y	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}

$f(x)$ is not equal to $f(y)$.

To construct $u, v, i \ldots$

	1	2	3	4	5	6	7	8	9	10	11	12
u	x_1	y_2	x_3	y_4	y_5	y_6	x_7	x_8	x_9	x_{10}	y_{11}	y_{12}
v	y_1	x_2	y_3	x_4	x_5	x_6	y_7	y_8	y_9	y_{10}	x_{11}	x_{12}

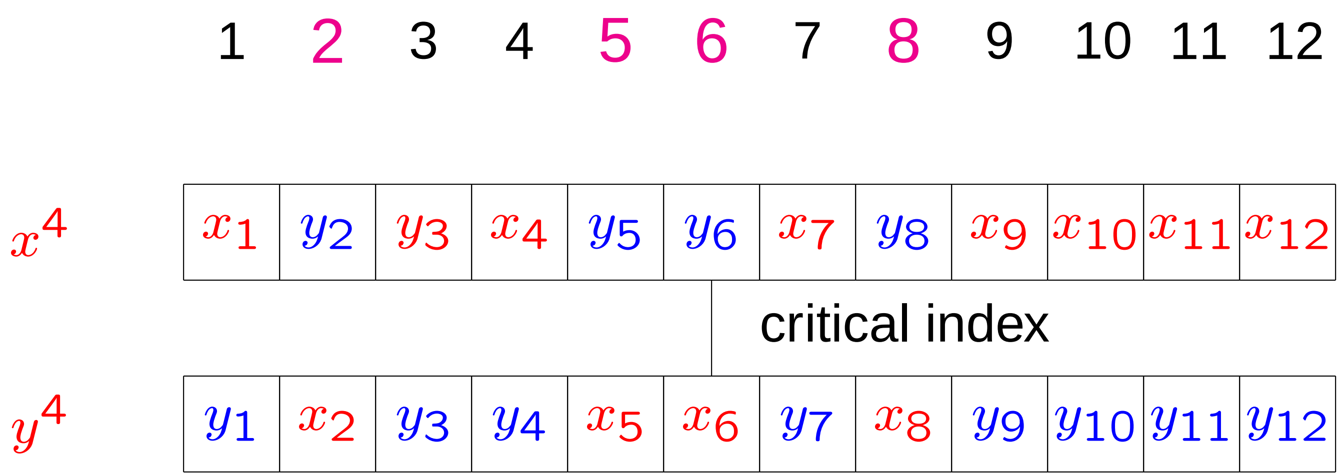
Swap coordinates on some set S and choose i from S.

Swap the first variable read by T on x.

	1	2	3	4	5	6	7	8	9	10	11	12
x^1	x_1	x_2	y_3	x_4	x_5	x_6	x_7	y_8	x_9	x_{10}	x_{11}	x_{12}
y^1	y_1	y_2	y_3	y_4	y_5	y_6	y_7	x_8	y_9	y_{10}	y_{11}	y_{12}

. . . to produce x^1 and y^1

Continue swapping variables read by T on x . . .



. . . until $f(y^j)$ changes to $f(x)$.

Are we done? f is sensitive to i at v but . . .

	1	2	3	4	5	6	7	8	9	10	11	12
u?	x_1	y_2	y_3	x_4	y_5	y_6	x_7	y_8	x_9	x_{10}	x_{11}	x_{12}
v?	y_1	x_2	y_3	y_4	x_5	x_6	y_7	x_8	y_9	y_{10}	y_{11}	y_{12}

critical index

. . . T may not read i on u.

First swap all variables read by T on x.

	1	2	3	4	5	6	7	8	9	10	11	12
x^6	x_1	y_2	y_3	y_4	y_5	y_6	x_7	y_8	x_9	y_{10}	x_{11}	x_{12}
y^6	y_1	x_2	y_3	x_4	x_5	x_6	y_7	x_8	y_9	x_{10}	y_{11}	y_{12}

$f(y^6)$ is not equal to $f(y)$.

Then swap back variables read by T on y^6 until f changes back to $f(y)$.

	1	2	3	4	5	6	7	8	9	10	11	12
u	x_1	x_2	y_3	y_4	y_5	x_6	x_7	x_8	x_9	y_{10}	x_{11}	x_{12}
v	y_1	y_2	y_3	x_4	x_5	y_6	y_7	y_8	y_9	x_{10}	y_{11}	y_{12}

critical index i

$f(u) = f(y)$ T reads i on u f is sensitive to i at v

- Deterministic Decision Trees
- Randomized Decision Trees (RDTs) and a new lower bound
- Proof of lower bound
  - Influences of boolean functions
  - Influences–Decision Tree connection theorem
  - Deducing the lower bound on RDT complexity
  - Proof of Influences-Decision Tree connection theorem
- Final remarks

What was left out . . .

- Property Testing
- Fault tolerant decision trees
- Learning theory
- Quantum query complexity
- Jointly computing many independent instances
- . . .