Bounded Independence Fools Halfspaces

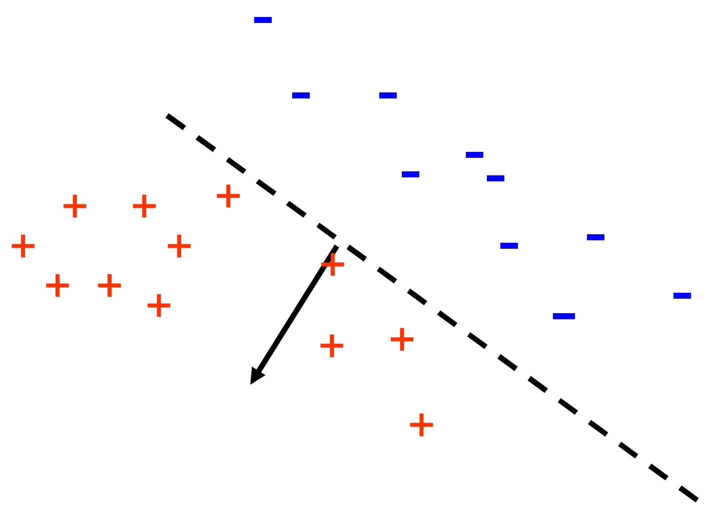
Emanuele Viola

Northeastern University
work partially done at Columbia University

Joint work with
Ilias Diakonikolas, Parikshit Gopalan,
Ragesh Jaiswal, Rocco Servedio

April 2009

Halfspaces



- Halfspace (a.k.a. Threshold)

  $h: \{-1,1\}^n \rightarrow \{-1,1\}$

  $h(x) = \text{sign}(w \cdot x - t) = \text{sign}(w_1 x_1 + \ldots + w_n x_n - t)$

  weights $w_1, \ldots, w_n, t \in R$

- Studied in complexity (NP $\subseteq$? halfspace of halfspaces)

  learning (Perceptron, Winnow, ...)

  social choice

Examples of halfspaces

- $h: \{-1,1\}^n \to \{-1,1\}$; $h(x) = \text{sign}(w_1x_1 + \ldots + w_nx_n - t)$

- $\text{Majority}(x) = \text{sign}(x_1 + \ldots + x_n)$

- $\text{AND}(x) = \text{sign}(x_1 + \ldots + x_n - n + 1/2)$

- "x > y?" $= \text{sign}(2^n(x_n - y_n) + 2^{n-1}(x_{n-1} - y_{n-1}) + \ldots + 2^1(x_1 - y_1))$

- weights $w_1, \ldots, w_n$ can be taken integer; need $w_i > 2^n$

Our results

- Def. Distribution D over $\{-1,1\}^n$ is k-wise independent if projection on any k coordinates is uniform over $\{-1,1\}^k$

- Thm: Any such D ε–fools any halfspace h: $\{-1,1\}^n \to \{-1,1\}$

$$| E_{x \in \text{uniform}} [h(x)] - E_{x \in D} [h(x)] | \leq \varepsilon$$

where $k = (1/\varepsilon)^2 \cdot \text{polylog}(1/\varepsilon)$

- Optimal up to polylog$(1/\varepsilon)$:

$D := (x_1, x_2, \ldots, x_k, \prod_{i \leq k} x_i, \ldots)$ $\quad$ $h(x) := \text{sign}(\sum_{i \leq k+1} x_i)$

Our results on generators

- k-wise independent distribution on $\{-1,1\}^n$ can be generated with $s = (\log n)\cdot k$ random bits

  [Alon Babai Itai; Chor Goldreich; '85]

- Corollary: Explicit generator $G : \{-1,1\}^s \to \{-1,1\}^n$ that ε–fools any halfspace $h: \{-1,1\}^n \to \{-1,1\}$

$$| E_{x \in \text{uniform}} [h(x)] - E_Y[h(G(Y))] | \leq \varepsilon$$

  where $s = (\log n)\cdot(1/\varepsilon)^2 \text{ polylog}(1/\varepsilon)$

- First generator for halfspaces

Our generator vs. others

- Our result: Explicit generator $G : \{-1,1\}^s \to \{-1,1\}^n$ that ε–fools halfspaces $h(x) = \text{sign}(w_1x_1 + \ldots + w_nx_n - t)$ with $s =$ (log n) poly(1/ε)

- Nisan ('92): fools if weights $w_i \in \{1,\ldots,\text{poly}(n)\}$, $S > \log^2 n$

- Rabani Shpilka ('08): $G : \{-1,1\}^s \to \{-1,1\}^n$ hits $h^{-1}(1)$ (when $> \varepsilon$), does not fool, $s = O(\log n/\varepsilon)$

Progress on generators [2005 - now]

- Random walks: Trevisan Vadhan Reingold
- Polynomials: Bogdanov V., Lovett
- Constant-depth circuits: Bazzi, Razborov, Braverman
- Halfspaces: Rabani Shpilka, this talk
- Challenges: (1) RL =? L

  (2) Fool width-3 read-once branching program, s = O(log n)

Outline

- Overview and our results
- Proof

Recall our result

- Def. Distribution D over $\{-1,1\}^n$ is k-wise independent if projection on any k coordinates is uniform over $\{-1,1\}^k$

- Thm: Any such D ε–fools any halfspace h: $\{-1,1\}^n \rightarrow \{-1,1\}$

$$| E_{x \in \text{uniform}} [h(x)] - E_{x \in D} [h(x)] | \leq \varepsilon$$

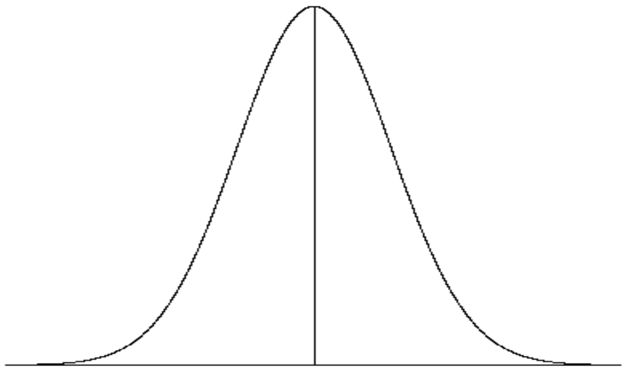
where $k = (1/\varepsilon)^2 \cdot \text{polylog}(1/\varepsilon)$

Proof overview

- Case analysis based on structure of halfspace [Servedio 2007; Rabani Shpilka 2008]

- Def. halfspace $h(x) = \text{sign}(w \cdot x - t) = \text{sign}(w_1 x_1 + \ldots + w_n x_n - t)$
  regular if every $|w_i|$ small w.r.t. $(\Sigma_i w_i^2)^{1/2}$ (at most ε frac.)

- regular $\Rightarrow w \cdot x \approx \text{Normal}(0, \Sigma_i w_i^2)$
  [Berry-Esséen]



Outline

- Overview and our results
- Proof
  - Regular halfspaces
  - Non-regular halfspaces

"Sandwich" approximation

- Lemma [Bazzi, Benjamini Gurel-Gurevich Peled]

  $h : \{-1,1\}^n \rightarrow \{-1,1\}$ is ε-fooled by k-wise ind. distributions

  $$\Leftrightarrow$$

  $\exists$ degree-k polynomials $q_u, q_l : \{-1,1\}^n \rightarrow R$:

  (1) $q_l(x) \leq h(x) \leq q_u(x)\ \forall\ x \in \{-1,1\}^n$

  (2) $E_{X \in \{-1,1\}^n}[q_u(X) - h(X)] \leq \varepsilon$, $E[h(X) - q_l(X)] \leq \varepsilon$

- Proof ($\Leftarrow$): If D k-wise independent, X uniform

  $E\,[h(D)] - E\,[h(X)]$

  $\leq E\,[q_u(D)] - E[q_l(X)]$ (1)

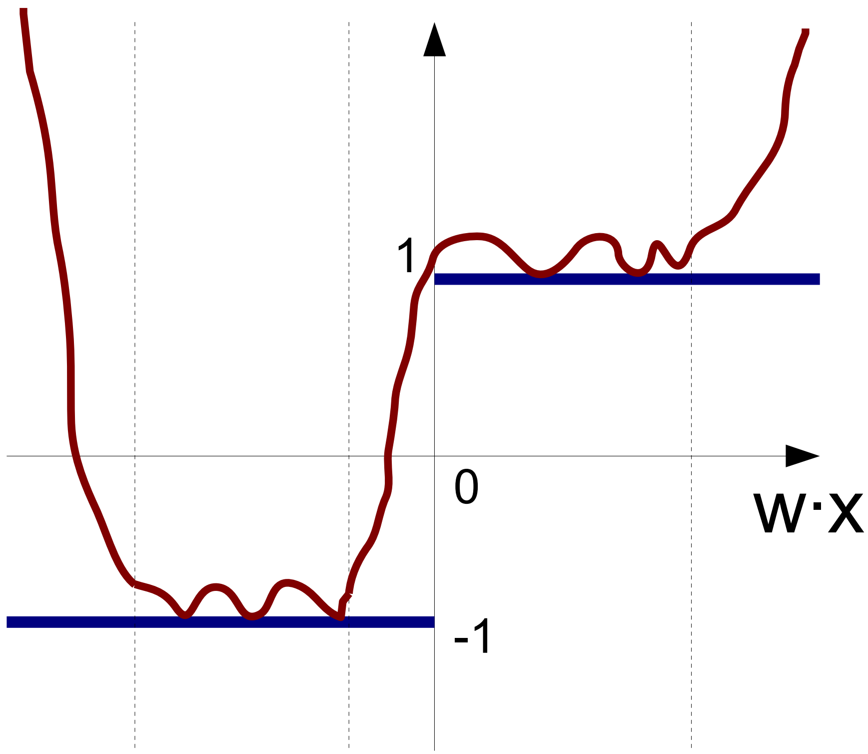
  $\leq E\,[q_u(X)] - E[q_l(X)]$ because q_u has degree k

  $\leq 2\varepsilon$ (2)

Q.e.d.

Construction of q_u

Build univariate P : R→R approximator to sign : R→{-1,1}



$$q_u(x) := P(w_1x_1 + \ldots + w_nx_n)$$

(t = 0 and ignore scaling)

Properties of P

P : degree $k \sim (1/\varepsilon)^2$

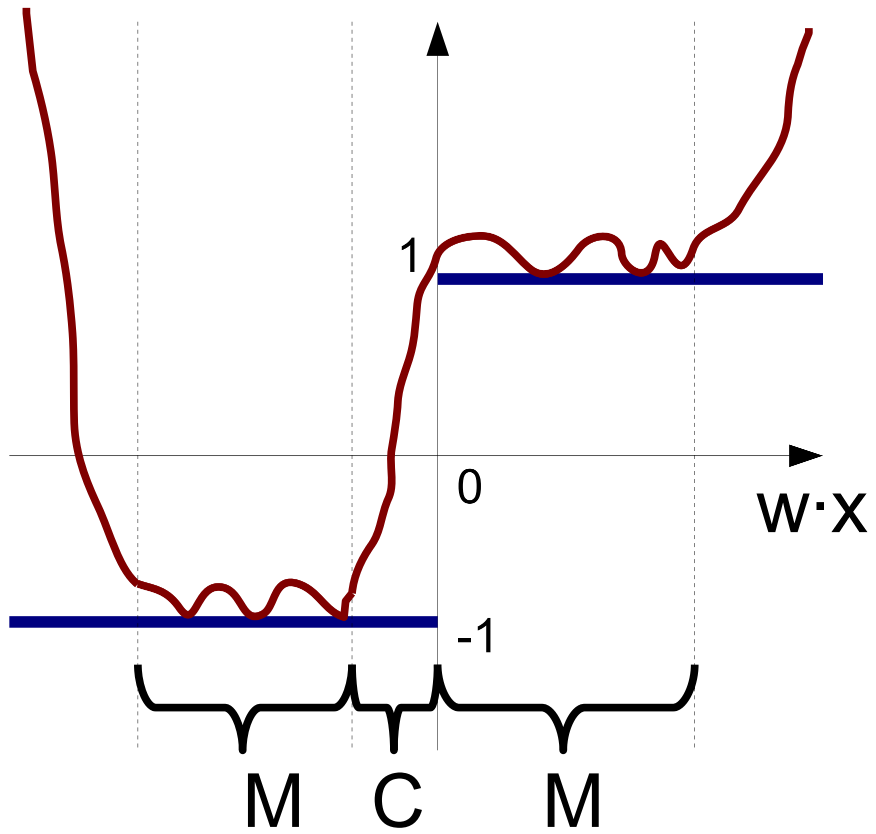
(a) $\forall\, x : P(x) \geq \text{sign}(x)$

(b) $\forall\, x \in M : P(x) - \text{sign}(x) < \varepsilon$

(c) $|M| > 1, \; |C| < \varepsilon$

$h(x) := \text{sign}(w \cdot x)$

$q_u(x) := P(w \cdot x)$



Assuming P, we show how to fool regular h

Correctness of q_u

P : degree $k \sim (1/\varepsilon)^2$

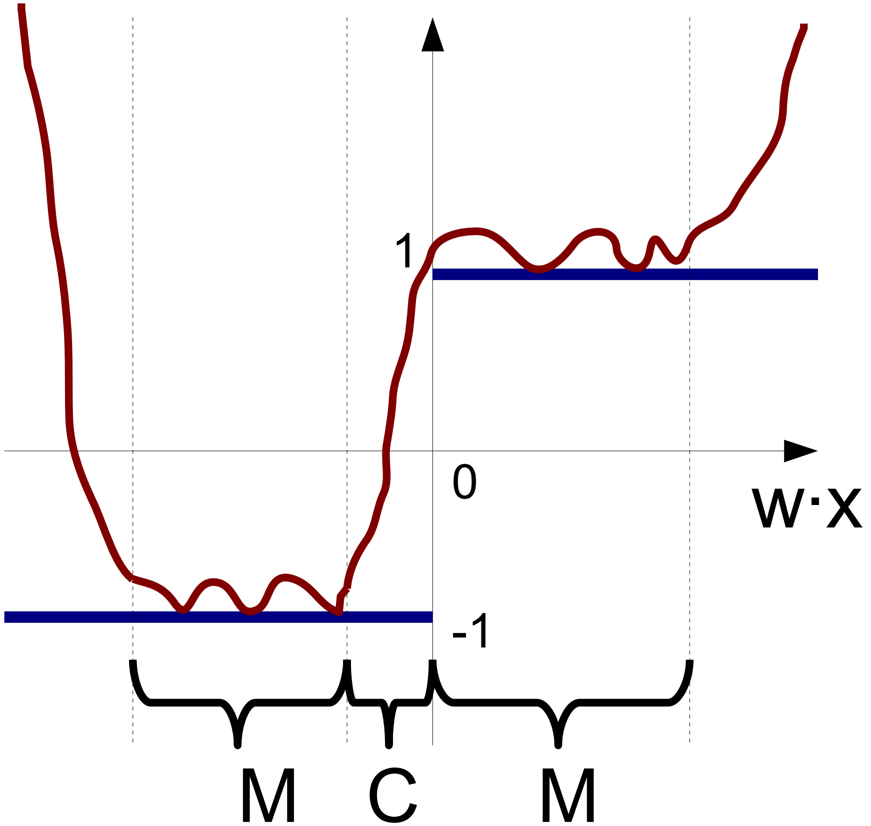
(a) $\forall\, x : P(x) \geq \mathrm{sign}(x)$

(b) $\forall\, x \in M : P(x) - \mathrm{sign}(x) < \varepsilon$

(c) $|\, M \,| > 1, \;\; |\, C \,| < \varepsilon$

$h(x) := \mathrm{sign}(w \cdot x)$

$q_u(x) := P(w \cdot x)$



Want:

(1) $h(x) \leq q_u(x) \;\forall\, x$

(2) $E_X[q_u(X) - h(X)] \leq \varepsilon$

Correctness of q_u

P : degree $k \sim (1/\varepsilon)^2$

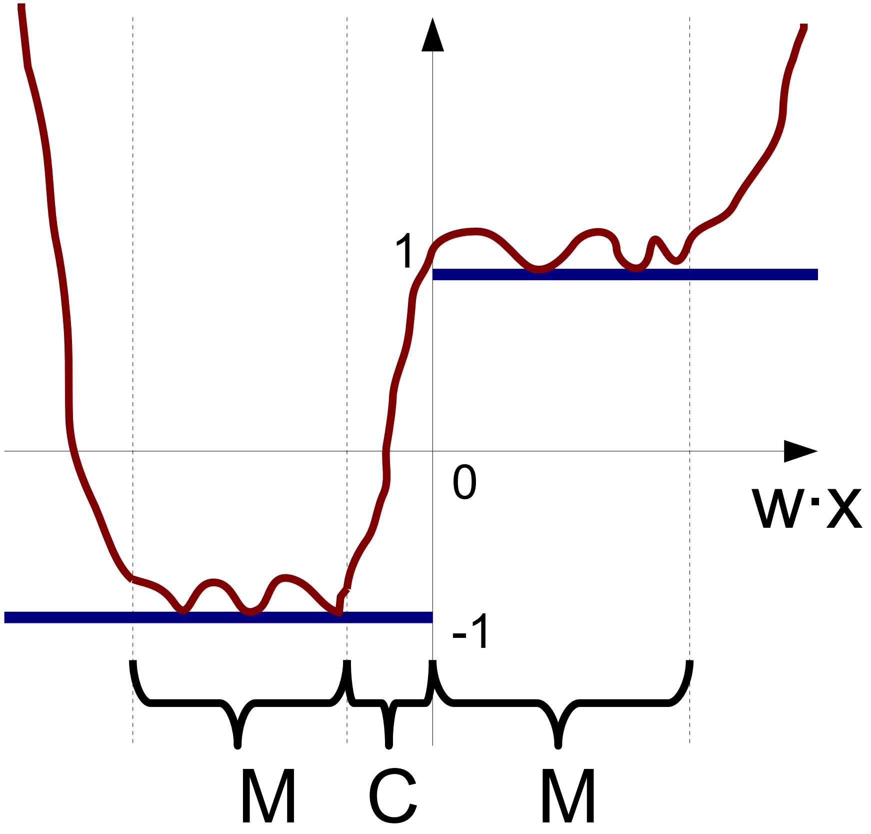
(a) $\forall\, x : P(x) \geq \text{sign}(x)$

(b) $\forall\, x \in M : P(x) - \text{sign}(x) < \varepsilon$

(c) $| M | > 1, \;\; | C | < \varepsilon$

$h(x) := \text{sign}(w \cdot x)$

$q_u(x) := P(w \cdot x)$



Want:

(1) $h(x) \leq q_u(x) \;\forall\, x$

Given by (a)

Q.e.d.

Correctness of q_u

P : degree $k \sim (1/\varepsilon)^2$

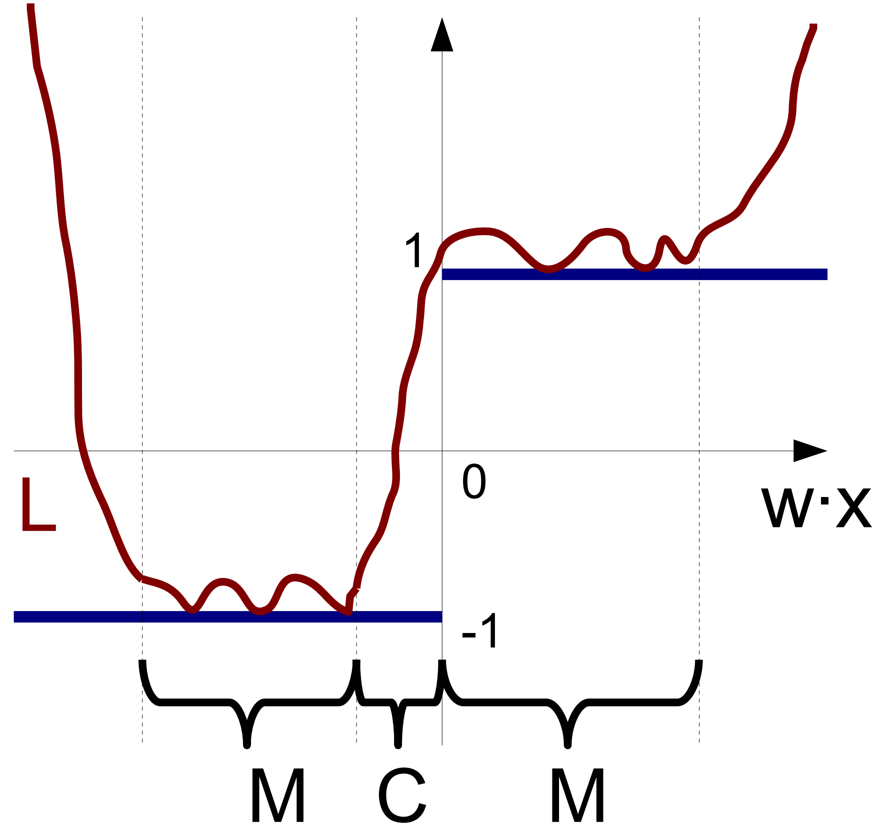
(a) $\forall\, x : P(x) \geq \text{sign}(x)$

(b) $\forall\, x \in M : P(x) - \text{sign}(x) < \varepsilon$

(c) $|\, M \,| > 1$, $|\, C \,| < \varepsilon$

$h(x) := \text{sign}(w \cdot x)$

$q_u(x) := P(w \cdot x)$



Want: (2) $E_X[q_u(X) - h(X)] \leq \varepsilon$

- $\Pr[\, w \cdot x \in C \,] < \varepsilon$ (c) + h regular ($w \cdot x \approx$ normal)
- $w \cdot x \in M \Rightarrow q_u(X) - h(X) < \varepsilon$ (b)
- $\Pr[\, |w \cdot x| = L \,] < \varepsilon / q_u(L)$

Q.e.d.

Construction of P: Approximation Theory

P : degree $k \sim (1/\varepsilon)^2$

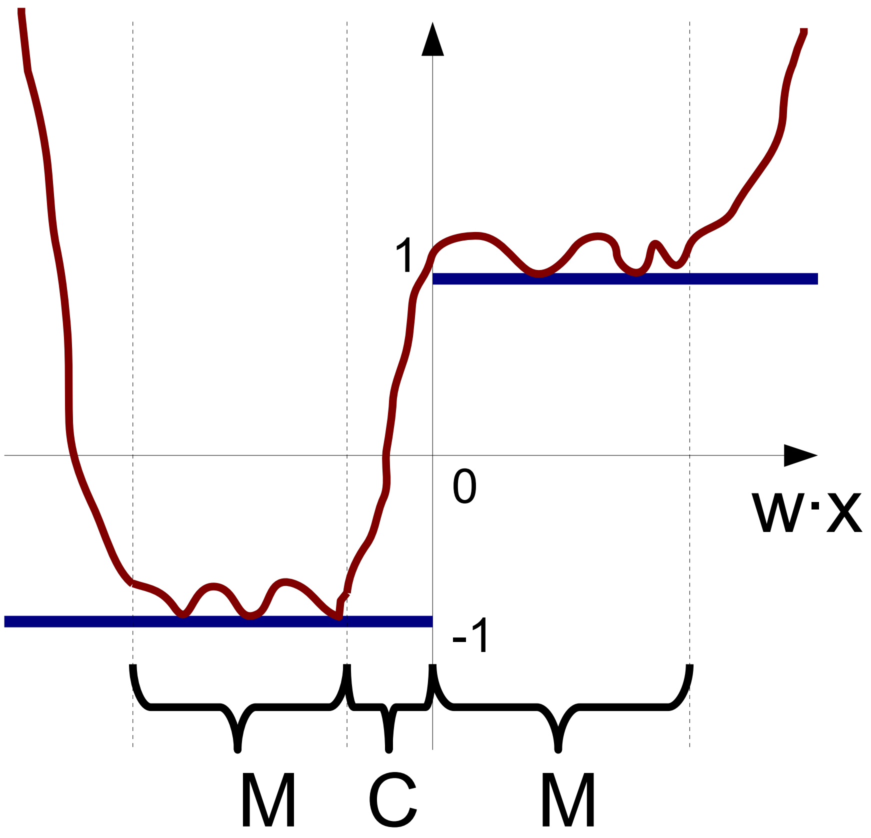
(a) $\forall\, x : P(x) \geq \text{sign}(x)$

(b) $\forall\, x \in M : P(x) - \text{sign}(x) < \varepsilon$

(c) $|\,M\,| > 1, \;\; |\,C\,| < \varepsilon$

$h(x) := \text{sign}(w \cdot x)$

$q_u(x) := P(w \cdot x)$



P := best uniform approximation to sign on M

(a) Chebychev alternation theorem

(b,c) Jackson theorem

Outline

- Overview and our results
- Proof
  - Regular halfspaces
  - Non-regular halfspaces

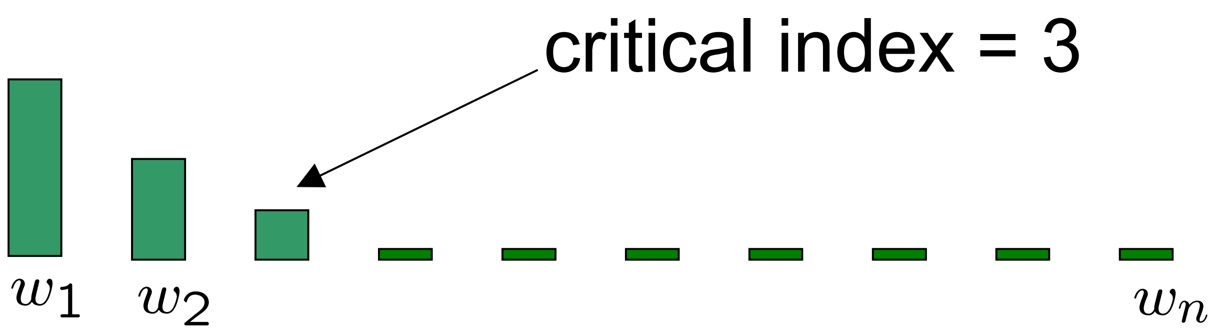
Non-regular halfspaces

Halfspace $h(x) = \text{sign}(w_1x_1+\ldots+w_nx_n - t)$

Recall regular : $\forall$ $|w_i|$ small w.r.t. $(\Sigma_i\, w_i^2)^{1/2}$ (at most ε fraction)

- Definition: critical index := minimum number of variables to fix to make halfspace regular

- Example:



- Case analysis based on critical index vs. $J := (1/\varepsilon)^2$

Case: small critical index

- If h has small critical index < $J = (1/\varepsilon)^2$

- recall fool regular halfspace with ($k \sim (1/\varepsilon)^2$)-wise indep.

- Claim: $(J + k)$-wise independent fools h

- Proof: Fixing J variables $\Rightarrow$ halfspace regular and
  still k-wise independence

Q.e.d.

Case: large critical index

- If h has large critical index > $J = (1/\varepsilon)^2$

  $\Rightarrow \exists$ J large weights $w_1, \ldots, w_J$

- Claim: $\forall$ (J + 2)-independent distribution with high probability $x_1, \ldots, x_J$ determine outcome

- Proof:
  Uniform $x_1, \ldots, x_J \Rightarrow w_1 x_1 + \ldots + w_J x_J$ large
  Other variables $x_{>J}$ still 2-wise independent
  $\Rightarrow w_{>J} x_{>J}$ concentrated $\Rightarrow$ rarely changes outcome

Q.e.d.

Conclusion

- Thm: k-wise indep. D ε–fool halfspaces $h : \{-1,1\}^n \to \{-1,1\}$:

  $| E[h(X)] - E [h(D)] | \leq \varepsilon$ for $k = (1/\varepsilon)^2$ polylog$(1/\varepsilon)$

- Tight up to polylog$(1/\varepsilon)$

- Corollary: First generator $G : \{-1,1\}^s \to \{-1,1\}^n$ that

  ε–fools halfspaces seed $s = (\log n)\,(1/\varepsilon)^2$ polylog$(1/\varepsilon)$

- Open: Improve dependence on ε. $s = O(\log (n/\varepsilon))$?

  Higher degree? E.g. $h(x) := \mathrm{sign}(\sum w_{i,j} x_i x_j - t)$

- $\sum\prod\surd\cap\notin \cup\supset\supseteq\not\subset\subset\subseteq\in \notin\Downarrow\Rightarrow\Uparrow\Leftarrow\Leftrightarrow\vee\wedge\geq\leq\forall\exists$