Nearest neighbor representations of Boolean functions

(Joint work with György Turán and Zhihao Liu)

Peter Hajnal

Bolyai Institute, University Szeged

18th of September, 2020.

The main notion

The main notion

Definition

A *nearest neighbor (NN)* representation of a Boolean function

$$f : \{0,1\}^n \to \{0,1\}$$

is a pair of disjoint subsets (P, N) of $\mathbb{R}^n$ such that for every $a \in \{0,1\}^n$

- if a is positive/$f(a) = 1$ then there exists $b \in P$ such that for every $c \in N$ it holds that $d(a,b) < d(a,c)$,
- if a is negative/$f(a) = 0$ then there exists $b \in N$ such that for every $c \in P$ it holds that $d(a,b) < d(a,c)$.

Language of learning and complexity

- The points in P (resp., N) are called positive (resp., negative) *prototypes*.

Language of learning and complexity

- The points in P (resp., N) are called positive (resp., negative) *prototypes*.

- The size of the representation is $|P \cup N|$.

Language of learning and complexity

- The points in P (resp., N) are called positive (resp., negative) *prototypes*.

- The size of the representation is $|P \cup N|$.

- The nearest neighbor complexity, $NN(f)$, of f is the minimum of the sizes of the representations of f.

Language of learning and complexity

- The points in P (resp., N) are called positive (resp., negative) *prototypes*.

- The size of the representation is $|P \cup N|$.

- The nearest neighbor complexity, $NN(f)$, of f is the minimum of the sizes of the representations of f.

- A nearest neighbor representation is *Boolean* if $P \cup N \subseteq \{0, 1\}^n$, i.e., if the prototypes are Boolean vectors.

Language of learning and complexity

- The points in P (resp., N) are called positive (resp., negative) *prototypes*.

- The size of the representation is $|P \cup N|$.

- The nearest neighbor complexity, $NN(f)$, of f is the minimum of the sizes of the representations of f.

- A nearest neighbor representation is *Boolean* if $P \cup N \subseteq \{0,1\}^n$, i.e., if the prototypes are Boolean vectors.

- The minimum of the sizes of the Boolean nearest neighbor representations is denoted by $BNN(f)$.

Symmetric functions and their complexity

Definition

A Boolean function is *symmetric* if its value depends only on the weight of its input.

Symmetric functions and their complexity

Definition

A Boolean function is *symmetric* if its value depends only on the weight of its input.

- A symmetric function f can be specified by a set $I_f \subseteq \{0, \ldots, n\}$ such that $f(a) = 1$ iff $|a| \in I_f$.

Symmetric functions and their complexity

Definition

A Boolean function is *symmetric* if its value depends only on the weight of its input.

• A symmetric function f can be specified by a set $I_f \subseteq \{0, \ldots, n\}$ such that $f(a) = 1$ iff $|a| \in I_f$.

Proposition

a) For every n-variable symmetric function f it holds that $NN(f) \leq n + 1$.

b) $BNN(x_1 \oplus x_2 \oplus \ldots \oplus x_n) = 2^n$.

Threshold functions

Threshold functions

Definition

A Boolean function f is a *threshold* function if there are weights $w_1, \ldots, w_n \in \mathbb{R}$ and a threshold $t \in \mathbb{R}$ such that for every $x \in \{0,1\}^n$ it holds that $f(x) = 1$ iff $w_1 x_1 + \ldots + w_n x_n \geq t$.

Threshold functions

Definition

A Boolean function f is a *threshold* function if there are weights $w_1, \ldots, w_n \in \mathbb{R}$ and a threshold $t \in \mathbb{R}$ such that for every $x \in \{0,1\}^n$ it holds that $f(x) = 1$ iff $w_1 x_1 + \ldots + w_n x_n \geq t$.

- The special case when $w_1 = \ldots = w_n = 1$ is denoted by TH_n^t.

Threshold functions

Definition

A Boolean function f is a *threshold* function if there are weights $w_1, \ldots, w_n \in \mathbb{R}$ and a threshold $t \in \mathbb{R}$ such that for every $x \in \{0,1\}^n$ it holds that $f(x) = 1$ iff $w_1 x_1 + \ldots + w_n x_n \geq t$.

- The special case when $w_1 = \ldots = w_n = 1$ is denoted by TH_n^t.

- In particular, when $t = \frac{n}{2}$, we get the n-variable majority function $MAJ_n(x)$.

Complexity of threshold functions

Complexity of threshold functions

Theorem

a) For every threshold function f it holds that $NN(f) = 2$.

b) If n is odd then $BNN(MAJ_n) = 2$ and if n is even then $BNN(MAJ_n) \le \frac{n}{2} + 2$.

c) $BNN\left(TH_n^{\lfloor n/3 \rfloor}\right) = 2^{\Omega(n)}$.

Upper bound for an arbitrary function

Upper bound for an arbitrary function

Theorem

For every n-variable Boolean function it holds that

$$NN(f) \leq (1 + o(1))\frac{2^{n+2}}{n}.$$

Lower bound for a generic function

Lower bound for a generic function

Theorem

For almost all n-variable Boolean functions

$$NN(f) > \frac{2^{n/2}}{n}.$$

Explicit functions

Explicit functions

The mod 2 inner product function of $2n$ variables is defined by

$$IP_n(x_1, \ldots, x_n, y_1, \ldots, y_n) = (x_1 \wedge y_1) \oplus \ldots \oplus (x_n \wedge y_n).$$

Explicit functions

The mod 2 inner product function of $2n$ variables is defined by

$$IP_n(x_1, \ldots, x_n, y_1, \ldots, y_n) = (x_1 \wedge y_1) \oplus \ldots \oplus (x_n \wedge y_n).$$

Theorem

a) $NN(IP_n) \geq 2^{n/2}$.
b) $NN(x_1 \oplus \cdots \oplus x_n) \geq n + 1$.

Nearest neighbor problem and sign-representation of Boolean functions

Nearest neighbor problem and sign-representation of Boolean functions

Lemma

If a Boolean function has a nearest neighbor representation with m prototypes then it has a sign-representation over $\{1, 2\}$ having m terms.

Sign-representation

Sign-representation

Definition

A multivariate polynomial $p(x_1, \ldots, x_n)$ is a *sign-representation* of a Boolean function $f(x_1, \ldots, x_n)$ if for every $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ it holds that $p(x) \geq 0$ iff $f(x) = 1$.

Sign-representation

Definition

A multivariate polynomial $p(x_1, \ldots, x_n)$ is a *sign-representation* of a Boolean function $f(x_1, \ldots, x_n)$ if for every $x = (x_1, \ldots, x_n) \in \{0, 1\}^n$ it holds that $p(x) \geq 0$ iff $f(x) = 1$.

Definition

A multivariate polynomial $p(\tilde{x}_1, \ldots, \tilde{x}_n)$ is a $\{1, 2\}$*-sign-representation* of a Boolean function $f(x_1, \ldots, x_n)$ if for every $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_n) \in \{1, 2\}^n$ it holds that $p(\tilde{x}) \geq 0$ iff $\tilde{f}(\tilde{x}) = f(x) = 1$.

k-nearest neighbor representation

k-nearest neighbor representation

A k-nearest neighbor (k-NN) representation of f is a pair of disjoint subsets (P, N) of $\mathbb{R}^n$, such that for every $a \in \{0, 1\}^n$ it holds that

k-nearest neighbor representation

A k-nearest neighbor (k-NN) representation of f is a pair of disjoint subsets (P, N) of $\mathbb{R}^n$, such that for every $a \in \{0,1\}^n$ it holds that

- a is positive iff at least $\frac{k}{2}$ of the k points in $P \cup N$ closest to a belong to P.

k-nearest neighbor representation

A k-nearest neighbor (k-NN) representation of f is a pair of disjoint subsets (P, N) of $\mathbb{R}^n$, such that for every $a \in \{0,1\}^n$ it holds that

- a is positive iff at least $\frac{k}{2}$ of the k points in $P \cup N$ closest to a belong to P.

It is assumed that for every a, the k smallest distances of a from the prototypes are all smaller than the other $|P \cup N| - k$ distances from the prototypes. Thus the case $k = 1$ is the same as the nearest neighbor representation. The size of the representation is again $|P \cup N|$. The k-nearest neighbor complexity, k-$NN(f)$, of f is the minimum of the sizes of the k-nearest neighbor representations of f.

Nearest neighbor problem and linear decision trees

Lemma

For every k and every Boolean function f it holds that $LDT(f) \leq (3 + o(1)) \cdot k\text{-}NN(f)$.

Bounds

Bounds

Theorem

For every k it holds that

$$k\text{-}NN(IP_n) \geq \frac{n}{6+o(1)}.$$

Thank you for your attention!