J. Austral. Math. Soc. (*Series A*) **53** (1992), 1–8

THE AND/OR THEOREM FOR PERCEPTRONS

I. D. BRUCE and D. EASDOWN

(Received 7 March 1990)

Communicated by T. E. Hall

Abstract

Minsky and Papert claim that, for any positive integer n, there exist predicates of order 1 whose conjunction and disjunction have order greater than n. Their proof is amended and a stronger result obtained of which their claim is a special case.

1991 *Mathematics subject classification* (*Amer. Math. Soc.*): 68 G 99.

1. Introduction

The *perceptron* is a simple parallel computing device, and its capabilities and limitations have been studied by Minsky and Papert [2]. They claim [2, Section 1.5] to have shown that for any positive integer n there exist predicates ψ_1 and ψ_2 of order 1 for which both $\psi_1 \wedge \psi_2$ and $\psi_1 \vee \psi_2$ have order greater than n. This is called the AND/OR Theorem. Their argument [2, Chapter 4] establishes the existence of predicates ψ_1 and ψ_2 of order 1 for which $\psi_1 \wedge \psi_2$ has order greater than n. A similar argument establishes the existence of predicates ψ_1' and ψ_2' of order 1 for which $\psi_1' \vee \psi_2'$ has order greater than n. However the arguments do not seem to guarantee that $\psi_1 = \psi_1'$ and $\psi_2 = \psi_2'$.

We amend their proof by adapting their techniques to prove a more general result, of which the AND/OR Theorem is a special case.

2. Preliminaries

Let R be a finite set. A *predicate on* R is a function ϕ of subsets of R whose values, depending on the context, may be thought of as TRUE and FALSE, or 1 and 0. To pass conveniently between these two kinds of predicate values, square bracket notation is used: if $\phi(X)$ is a statement about X then

$$[\phi(X)] = \begin{cases} 1 \text{ if } \phi(X) & \text{is true.} \\ 0 \text{ if } \phi(X) & \text{is false.} \end{cases}$$

Let Φ be a family of predicates on R, which is finite since R is finite. A predicate ψ is called a *linear threshold function* with respect to Φ if there are numbers θ, and α_ϕ for each $\phi \in \Phi$ such that

$$\psi(X) = \left[\sum_{\phi\in\Phi} \alpha_\phi \phi(X) > \theta\right].$$

Let $L(\Phi)$ denote the set of linear threshold functions with respect to Φ.

The *support* $S(\phi)$ of a predicate ϕ is the smallest subset S of R for which

$$(*) \qquad (\forall X \subseteq R)\ \phi(X) = \phi(X \cap S)$$

The support exists because $(*)$ is satisfied when $S = R$ and if any (finite) collection of subsets S satisfies $(*)$ then their intersection also satisfies $(*)$. Note however that if R is allowed to be infinite then the notion of support need not make sense.)

The *order* of a predicate ψ is the smallest number k for which there is a set Φ of predicates for which $\psi \in L(\Phi)$ and $|S(\phi)| \leq k$ for all $\phi \in \Phi$.

If $A \subseteq R$ then the *mask of* A, denoted μ_A, is the predicate defined by

$$\mu_A(X) = \begin{cases} 1 & \text{if } A \subseteq X \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that $S(\mu_A) = A$. If $A = \{a\}$ denote μ_A also by μ_a.

THEOREM 1 [2, 1.5.3]. *A predicate ψ has order k if and only if k is the smallest number for which there exists a set Φ of masks such that $\psi \in L(\Phi)$ and $|S(\phi)| \leq k$ for all $\phi \in \Phi$.*

Denote the (symmetric) group of all permutations of a set X by $\mathrm{Sym}(X)$. Let Γ be a subgroup of $\mathrm{Sym}(R)$. If ψ is a predicate on R and $\gamma \in \Gamma$, then define the predicate $\psi\gamma$ by

$$\psi\gamma(X) = \psi(\gamma(X)) \quad \text{for } X \subseteq R.$$

https://doi.org/10.1017/S1446788700035333 Published online by Cambridge University Press

Call predicates ψ and ϕ Γ-*equivalent*, written $\psi \equiv \phi$, if $\psi = \phi\gamma$ for some $\gamma \in \Gamma$. The Relation $\equiv$ is an equivalence because Γ is a group. Note that, for masks, if $A, B \subseteq R$ then $\mu_A \gamma = \mu_{\gamma^{-1}(A)}$ for any $\gamma \in \Gamma$ so that

$$\mu_A \equiv \mu_B \quad \text{if and only if} \quad A = B\gamma \quad \text{for some } \gamma \in \Gamma.$$

Call a predicate ψ *invariant under* Γ or Γ-*invariant* if

$$\psi = \psi\gamma \quad \text{for all } \gamma \in \Gamma.$$

THEOREM 2 (Group Invariance Theorem) [2, 2.3]. *Let R be a finite set, Γ a subgroup of* Sym(R) *and Φ a set predicates on R closed under Γ containing μ_ϕ the mask of the empty set. Suppose that $\psi \in L(\Phi)$ and that ψ is Γ-invariant. Then*

$$\psi = \left[\sum_{\phi \in \Phi} \beta_\phi \phi > 0 \right]$$

where the coefficients β_ϕ depend only on the Γ-equivalence class of ϕ, that is, if $\phi \equiv \phi'$ then $\beta_\phi = \beta_{\phi'}$.

Note (for when we apply this theorem later) that if Φ is the set of masks of support size $\leq k$ then Φ is closed under Γ and contains μ_ϕ.

Let $\psi_1, \ldots, \psi_m$ be a sequence of predicates or formulae. Define a collection AND/OR$(\psi_1, \ldots, \psi_m)$ of predicates or formulae inductively:

$$\text{AND/OR}(\psi_1) = \{\psi_1\},$$

$$\text{AND/OR}(\psi_1, \ldots, \psi_{i+1}) = \{\psi \vee \psi_{i+1}, \psi \wedge \psi_{i+1} \mid \psi \in \text{AND/OR}(\psi_1, \ldots, \psi_i)\} \quad \text{for } i = 1, \ldots, m-1.$$

Thus AND/OR$(\psi_1, \ldots, \psi_m)$ contains 2^{m-1} predicates or formulae, obtained from ψ_1 by successively conjoining or disjoining ψ_i's. For example AND/OR$(\psi_1, \psi_2) = \{\psi_1 \vee \psi_2, \psi_1 \wedge \psi_2\}$, and

$$\text{AND/OR}(\psi_1, \psi_2, \psi_3) = \{\psi \vee \psi_2 \vee \psi_3, (\psi_1 \vee \psi_2) \wedge \psi_3, \psi_1 \wedge \psi_2 \wedge \psi_3, (\psi_1 \wedge \psi_2) \vee \psi_3\}.$$

The following is a trivial but important observation: if $\psi_i(X) = \{^1_0$ for all $i = 1, \ldots, m$ then $\psi(X) = \{^1_0$ for all ψ in AND/OR$(\psi_1, \ldots, \psi_m)$.

The last result we shall need is an adaptation of [2, Lemma 1].

THEOREM 3 (Compactness). *Let $Q_1(x_1, \ldots, x_m)$, $Q_2(x_1, \ldots, x_m), \ldots$ be an infinite sequence of nonzero polynomials of m variables of degree at most N, and let $n_1, n_2, \ldots$ be an infinite increasing sequence of positive* *integers. Suppose* $\phi\colon \mathbb{Z}^m \to \{0, 1\}$ *is a function (predicate) such that for each* $i = 1, 2, \ldots$ *we have for all integers* $x_1, \ldots, x_m$ *between* $-n_i$ *and* n_i

$$\phi(x_1, \ldots, x_m) = [Q_i(x_1, \ldots, x_m) > 0].$$

Then there exists a nonzero polynomial $Q(x_1, \ldots, x_m)$ *of degree* $\leq N$ *such that for all integers* $x_1, \ldots, x_m$

$$\phi(x_1, \ldots, x_m) = \begin{cases} 1 \\ \\ 0 \end{cases} \text{implies } Q(x_1, \ldots, x_m) \quad \begin{cases} \geq 0 \\ \leq 0. \end{cases}$$

Proof. Let $\mathbf{x}$ be a vector of coordinates ranging over all products of powers of $x_1, \ldots, x_m$ in which for each product the sum of the exponents is at most N. Polynomials in $x_1, \ldots, x_m$ of degree $\leq N$ are then dot products $\mathbf{x} \cdot \mathbf{c}$ where $\mathbf{c}$ is a vector of constant coefficients. Thus for each i

$$Q_i(x_1, \ldots, x_m) = \mathbf{x} \cdot \mathbf{c}_i \quad \text{for some } \mathbf{c}_i.$$

But the set $\{\hat{\mathbf{c}}_i = \mathbf{c}_i/\|\mathbf{c}_i\| \mid i = 1, 2, \ldots\}$ lies on the surface of the unit hypersphere which is compact [1], so has a limit point $\mathbf{c}$ of length 1. In particular $\mathbf{c} \neq \mathbf{0}$. Let $Q(x_1, \ldots, x_m) = \mathbf{x} \cdot \mathbf{c}$, so Q is nonzero of degree $\leq N$. Let $x_1, \ldots, x_m$ be any integers. Choose n_i larger than each of $|x_1|, \ldots, |x_m|$. Suppose $\phi(x_1, \ldots, x_m) = 1$, so by our hypothesis $Q_j(x_1, \ldots, x_m) > 0$ for $j \geq i$. Thus $\mathbf{x} \cdot \mathbf{c}_j > 0$, so $\mathbf{x} \cdot \hat{\mathbf{c}}_j > 0$ for $j \geq i$. If $\mathbf{x} \cdot \mathbf{c} < 0$ then choose $j \geq i$ such that

$$\|\hat{\mathbf{c}}_j - \mathbf{c}\| < \frac{|\mathbf{c} \cdot \mathbf{c}|}{2\|\mathbf{x}\|}$$

and so

$$\begin{aligned} \mathbf{x} \cdot \hat{\mathbf{c}}_j &= \mathbf{x} \cdot \mathbf{c} + \mathbf{x} \cdot (\hat{\mathbf{c}}_j - \mathbf{c}) \leq \mathbf{x} \cdot \mathbf{c} + |\mathbf{x} \cdot (\hat{\mathbf{c}}_j - \mathbf{c})| \\ &\leq \mathbf{x} \cdot \mathbf{c} + \|\mathbf{x}\|\, \|\hat{\mathbf{c}}_j - \mathbf{c}\| \quad \text{(by the Cauchy-Schwarz inequality)} \\ &< \mathbf{x} \cdot \mathbf{c} + \frac{|\mathbf{x} \cdot \mathbf{c}|}{2} = \frac{\mathbf{x} \cdot \mathbf{c}}{2} < 0, \end{aligned}$$

a contradiction. Hence $\mathbf{x} \cdot \mathbf{c} \geq 0$.

Similarly, if $\phi(x_1, \ldots, x_m) = 0$ then $Q_j(x_1, \ldots, x_m) \leq 0$ for $j \geq i$, and in the same way one shows that $\mathbf{x} \cdot \mathbf{c} \leq 0$. This completes the proof of Theorem 3.

https://doi.org/10.1017/S1446788700035333 Published online by Cambridge University Press

3. The main theorem

The purpose of this section is to prove

THEOREM 4. *Let* N *and* m *be any positive integers,* $m \geq 2$. *There exist predicates* $\psi_1, \dots, \psi_m$ *of order* 1 *such that every predicate in*

$$\mathrm{AND/OR}(\psi_1, \dots, \psi_m)$$

has order greater than N.

Then Minsky and Papert's claim follows.

COROLLARY 5 (the AND/OR theorem). *Let* N *be any positive integer. There exist predicates* ψ_1 *and* ψ_2 *of order* 1 *such that both* $\psi_1 \wedge \psi_2$ *and* $\psi_1 \vee \psi_2$ *have order greater than* N.

PROOF OF THEOREM 4. To prove Theorem 4, let n be a positive integer and R any set containing $2nm$ elements. Express R as the disjoint union

$$R = A_1 \cup \dots \cup A_m \quad \text{where } |A_i| = 2n \text{ for each } i.$$

Define the predicate ψ_i, for $i = 1$ to m, by

$$\psi_i(X) = [|X \cap A_i| \geq n] \quad \text{for } X \subseteq R.$$

Note that each ψ_i has order 1 because

$$\psi_i(X) = \left[\sum_{a \in A_i} \mu_a(X) \geq n\right].$$

Thus for each n we have defined a sequence $\psi_1, \dots, \psi_m$ of order 1 predicates. We will show that for some n each member of $\mathrm{AND/OR}(\psi_1, \dots, \psi_m)$ has order larger than N.

Suppose to the contrary that for each n there is a predicate in

$$\mathrm{AND/OR}(\psi_1, \dots, \psi_m)$$

of order $\leq N$. There are infinitely many n and only 2^{m-1} formulae in $\mathrm{AND/OR}(\psi_1, \dots, \psi_m)$, so there exists at least one *formula* ψ in

$$\mathrm{AND/OR}(\psi_1, \dots, \psi_m)$$

such that for some infinite sequence of positive integers $n_1 < n_2 < \cdots$ the *predicates* ψ for $n = n_1, n_2, \dots$ have order $\leq N$.

For the time being fix $n = n_j$. Let Φ be the set of masks of support size $\leq N$. Then $\psi \in L(\Phi)$ by Theorem 1. Consider the group

$$\Gamma = \{\gamma \in \mathrm{Sym}(R) \mid \gamma(A_i) = A_i \quad \text{for } i = 1, \ldots, m\}.$$

Then each predicate ψ_i is Γ-invariant since

$$|X \cap A_i| = |\gamma(X \cap A_i)| = |\gamma(X) \cap \gamma(A_i)| = |\gamma(X) \cap A_i|$$

for each $\gamma \in \Gamma$ and $X \subseteq R$. By a simple induction all members of

$$\mathrm{AND/OR}(\psi_1, \ldots, \psi_m)$$

are Γ-invariant. Hence ψ is Γ-invariant, so by Theorem 2,

$$\psi = \left[\sum_{\phi \in \Phi} \beta_\phi \phi > 0\right]$$

for some coefficients β_ϕ which depend only on the Γ-equivalence class of ϕ. Let

$$V = \{\mathbf{v} = (v_1, \ldots, v_m) \in \mathbb{Z}^m \mid 0 \leq v_i \leq 2n \quad \text{for } i = 1 \text{ to } m \text{ and } v_1 + \cdots + v_m \leq N\}.$$

Then

$$\Phi = \bigcup_{\mathbf{v} \in V} \Phi_{\mathbf{v}}$$

where $\Phi_{\mathbf{v}} = \{\text{masks } \mu_A \mid |A \cap A_i| = v_i \text{ for } i = 1 \text{ to } m\}$. Since $\Gamma|_{A_i} = \mathrm{Sym}(A_i)$, the $\Phi_{\mathbf{v}}$ are the equivalence classes of Φ under Γ as $\mathbf{v}$ ranges over V. Hence

$$\psi(X) = \left[\sum_{\phi \in \Phi} \beta_\phi \phi(X) > 0\right] = \left[\sum_{\mathbf{v} \in V} \beta_{\mathbf{v}} N_{\mathbf{v}}(X) > 0\right]$$

where $\beta_{\mathbf{v}} = \beta_\phi$ for any $\phi \in \Phi_{\mathbf{v}}$ and $N_{\mathbf{v}}(X)$ is the number of masks in $\phi_{\mathbf{v}}$ whose support is contained in X. But then

$$N_{\mathbf{v}}(X) = \binom{|X \cap A_1|}{v_1} \cdots \binom{|X \cap A_m|}{v_m}$$

where $\binom{m}{r}$ is by definition $m(m-1)\cdots(m-r+1)/r!$ so that $\binom{m}{r} = 0$ if $m < r$.

Hence each $N_{\mathbf{v}}(X)$ is a polynomial in the m variables $y_1 = |X \cap A_1|, \ldots, y_m = |X \cap A_m|$ of degree $v_1 + \cdots + v_m \leq N$. But a linear combination of polynomials of degree $\leq N$ is also a polynomial of degree $\leq N$, so

$$\psi(X) = [P(y_1, \ldots, y_m) > 0]$$

for some polynomial $P(y_1, \ldots, y_m)$ of degree $\leq N$. Note that $0 \leq y_i \leq 2n$ since $|A_i| = 2n$ for each i. Put $x_i = y_i - n$, so that $-n \leq x_i \leq n$ for each i. Now put

$$Q_j(x_1, \ldots, x_m) = P(x_1 + n, \ldots, x_m + n)$$

which is a polynomial in $x_1, \ldots, x_m$ of degree $\leq N$. The subscript j is to remind us that $n = n_j$. Thus

$$\psi(X) = [Q_j(x_1, \ldots, x_m) > 0]$$

where $-n_j \leq x_1, \ldots, x_m \leq n_j$.

Define functions $\phi_i\colon \mathbb{Z}^m \to \{0, 1\}$ for $i = 1$ to m by

$$\phi_i(z_1, \ldots, z_m) = \begin{cases} 1 & \text{if } z_i \geq 0 \\ 0 & \text{if } z_i < 0. \end{cases}$$

Then $\psi(X) = \phi_i(x_1, \ldots, x_m)$ where as before $x_i = |X \cap A_i| - n$ for $i = 1$ to m. Let $\phi \in \text{AND/OR}(\phi_1, \ldots, \phi_m)$ be built using $\wedge$ and $\vee$ in exactly the same way as ψ. Then $\phi\colon \mathbb{Z}^m \to \{0, 1\}$ and for the x_i defined above $\phi(x_1, \ldots, x_m) = \psi(X)$. Thus

$$\phi(x_1, \ldots, x_m) = [Q_j(x_1, \ldots, x_m) > 0]$$

where $-n_j \leq x_1, \ldots, x_m \leq n_j$.

This holds for each j, so by Theorem 3 there exists a nonzero polynomial $Q(x_1, \ldots, x_m)$ of degree $\leq N$ such that

$$\phi(x_1, \ldots, x_m) = \begin{cases} 1 \\ 0 \end{cases} \text{implies } Q(x_1, \ldots, x_m) \begin{cases} \geq 0 \\ \leq 0 \end{cases}$$

for *all* integers $x_1, \ldots, x_m$.

Put $Q(x_1, \ldots, x_m) = x_m^d q(x_1, \ldots, x_{m-1}) + r(x_1, \ldots, x_m)$ where $q(x_1, \ldots, x_{m-1})$ is a nonzero polynomial in $x_1, \ldots, x_{m-1}$ and $r(x_1, \ldots, x_m)$ is either the zero polynomial or a nonzero polynomial such that the highest power of x_m appearing is less than d. Note that $\phi = \phi' \wedge \phi_m$ or $\phi = \phi' \vee \phi_m$ for some $\phi' \in \text{AND/OR}(\phi_1, \ldots, \phi_{m-1})$, and that for all x_m

$$\phi'(x_1, \ldots, x_m) = \begin{cases} 1 & \text{if } x_1, \ldots, x_{m-1} \geq 0 \\ 0 & \text{if } x_1, \ldots, x_{m-1} < 0. \end{cases}$$

If $\phi = \phi' \wedge \phi_m$ choose $x_1, \ldots, x_{m-1} \geq 0$, whilst if $\phi = \phi' \vee \phi_m$ choose $x_1, \ldots, x_{m-1} < 0$ for which $q(x_1, \ldots, x_{m-1}) \neq 0$. In both cases

$$\phi(x_1, \ldots, x_m) = \begin{cases} 1 & \text{if } x_m \geq 0 \\ 0 & \text{if } x_m < 0 \end{cases}$$

so that

$$Q(x_1, \ldots, x_m) = \begin{cases} \geq 0 & \text{if } x_m \geq 0 \\ \leq 0 & \text{if } x_m < 0. \end{cases}$$

https://doi.org/10.1017/S1446788700035333 Published online by Cambridge University Press

If d is even then

$$0 \le \lim_{x_m \to \infty} Q(x_1, \dots, x_m) = \lim_{x_m \to \infty} x_m^d q(x_1, \dots, x_{m-1}$$

$$= \lim_{x_m \to -\infty} x_m^d q(x_1, \dots, x_{m-1}) = \lim_{x_m \to -\infty} Q(x_1, \dots, x_m) \le 0$$

so that $\lim_{x_m \to \infty} Q(x_1, \dots, x_m) = 0$, which contradicts that $q(x_1, \dots, x_{m-1}) \neq 0$.

Hence d is odd. Now, if $\phi = \phi' \wedge \phi_m$ choose $x_1, \dots, x_{m-1} < 0$, whilst if $\phi = \phi' \vee \phi_m$ choose $x_1, \dots, x_{m-1} \ge 0$ for which $q(x_1, \dots, x_{m-1}) \neq 0$. In the first case $\phi(x_1, \dots, x_m) = 0$ so that $Q(x_1, \dots, x_m) \le 0$ for all x_m, whilst in the second case $\phi(x_1, \dots, x_m) = 1$ so that $Q(x_1, \dots, x_m) \ge 0$ for all x_m. But, since d is odd,

$$\lim_{x_m \to \infty} Q(x_1, \dots, x_m) = -\lim_{x_m \to -\infty} Q(x_1, \dots, x_m)$$

so that in both cases $\lim_{x_m \to \infty} Q(x_1, \dots, x_m) = 0$, which again contradicts that $q(x_1, \dots, x_{m-1}) \neq 0$. This completes the proof of Theorem 4.

References

[1] Jerrold E. Marsden, *Elementary Classical Analysis* (W. H. Freeman, San Francisco 1974).

[2] Marvin Lee Minsky and Seymour Papert, *Perceptrons: an introduction to computational geometry* (MIT Press, Cambridge, Massachusetts 1988).

Curtin University
GPO Box U1987
Perth WA 6001
Australia

University of Sydney
NSW 2006
Australia

https://doi.org/10.1017/S1446788700035333 Published online by Cambridge University Press