Separating the Vertices of N-Cubes by Hyperplanes and its Application to Artificial Neural Networks

by
R. Shonkwiler
School of Mathematics
Georgia Institute of Technology
Atlanta, GA. 30332
e-mail: shenk@math.gatech.edu

Abstract

We obtain a new sufficient condition that a region be classifiable by a 2-layer feed-forward net using threshold activation functions. Briefly, it is either a convex polytope, or that minus the removal of convex polytope from its interior, or that minus a convex polytope from its interior, or ... recursively. We refer to these sets as convex recursive deletion regions. Our proof of implementability exploits the equivalence of this problem with that of characterizing two set partitions of the vertices of a hypercube which are separable by a hyperplane for which we also obtain a new result.

§1 Introduction

By an (n-input) *neuron* N we mean a device capable first of forming the weighted sum $\sigma = \Sigma x_i w^i$ of its inputs, $x_1, \ldots, x_n$, with weights $w^1, w^2, \ldots, w^n$, and second of thresholding the resultant sum with a given value θ to produce an output: $y = 0$ if $\sigma < \theta$ or $y = 1$ if $\sigma \geq \theta$. Mathematically such a neuron evaluates the function $y = s_\theta(\mathbf{x} \cdot \mathbf{w})$ where $s_\theta(\cdot)$ is the step function (or *hard limiter*) with threshold θ, and $\sigma = \mathbf{x} \cdot \mathbf{w}$ is the dot product between the input vector $\mathbf{x} = (x_1, \ldots, x_n)^t$ and weight vector $\mathbf{w} = (w^1, \ldots, w^n)^t$.

With each n-input neuron there is associated a unique oriented hyperplane $H_{\mathbf{w},\theta}$ of n-dimensional Euclidean space $\mathbf{R}^n$ given by

$$H_{\mathbf{w},\theta} = \{\mathbf{x} : \mathbf{x} \cdot \mathbf{w} = \theta\}$$

whose *positive* side is in the direction $\mathbf{w}$ and whose distance from the origin is $d = |\theta|/\|\mathbf{w}\|$. $H_{\mathbf{w},\theta}$ decomposes $\mathbf{R}^n$ into two half-spaces, $H^1_{\mathbf{w},\theta} = \{\mathbf{x} : \mathbf{x} \cdot \mathbf{w} \geq 0\}$ and $H^0_{\mathbf{w},\theta} = \{\mathbf{x} : \mathbf{x} \cdot \mathbf{w} <$

$0\}$, and is the boundary of both. The hyperplane itself belongs fully to the positive half space. The output of the neuron is 1 for an input $\mathbf{x}$ if and only if $\mathbf{x} \in H^1_{\mathbf{w},\theta}$. Since the solutions $\mathbf{x}$ to the inequality $\mathbf{x} \cdot \mathbf{w} \geq \theta$ are invariant under its multiplication by a positive constant, distinct neurons (their weights and thresholds differing by a positive multiple) may be associated with the same half-space. But in fact this may be used to advantage by allowing scaling of the weights or threshold as necessary to meet implementation requirements. An artificial Neural Network consisting of a single such neuron is known as a *Perceptron*, [2], [5].

By an (m-neuron) *layer* $\mathcal{L}_n$ of n-input neurons we mean a list $N_1, \ldots, N_m$ of m neurons defined over the same set of n inputs. Let neuron N_j have threshold θ_j and weight vector $\mathbf{w}_j = (w^1_j, \ldots, w^n_j)^t$, $j = 1, \ldots, m$, that is w^i_j is the weight connecting the ith input to the jth neuron. Let W be the $n \times m$ matrix whose m columns are the n-vectors $\mathbf{w}_j$. Then the m-dimensional vector of weighted sums $\bar{\sigma}$ is given by the matrix product

$$\bar{\sigma}^t = \mathbf{x}^t W$$

and the m-dimensional vector *output* of the layer is given by

$$\mathbf{y} = S_\theta(\mathbf{x}^t W) = (s_{\theta_1}(\mathbf{x}^t \mathbf{w}_1), \ldots, s_{\theta_m}(\mathbf{x}^t \mathbf{w}_m))^t.$$

(The last member defines $S_\theta(\cdot)$.) Each component y_j of $\mathbf{y}$ is either 0 or 1 depending on the output of the j^{th} neuron and so the possible outputs are the vertices of the m-dimensional unit cube Q_m,

$$Q_m = \{(y_1, \ldots, y_m) : y_j \in \{0, 1\},\ 1 \leq j \leq m\}.$$

For a layer of n-input neurons, define the function $q : \mathbf{R}^n \to Q_m$ by

$$q(\mathbf{x}) = S_\theta(\mathbf{x}^t W).$$

Since the m-cube has exactly 2^m vertices, many inputs $\mathbf{x}$ will have the same q value. Given a vertex $\mathbf{y} \in Q_m$, the $\mathbf{y}^{th}$ *atom* or *cell* $a_{\mathbf{y}} \subset \mathbf{R}^n$ is the inverse image

$$a_{\mathbf{y}} = \{\mathbf{x} \in \mathbf{R}^n : q(\mathbf{x}) = \mathbf{y}\}.$$

Each such atom is the intersection of half-spaces,

$$a_{\mathbf{y}} = \bigcap_{j=1}^{m} H^{\pm_j}_{\mathbf{w}_j, \theta_j},$$

where $\pm_j = 1$, if $y_j = 1$, and $\pm_j = 0$, if $y_j = 0$. Therefore each atom is a convex polytope or the empty set. The set of all (non-empty) atoms $\{a_{\mathbf{y}} : \mathbf{y} \in Q_m\}$ forms a partition of $\mathbf{R}^n$ into mutually exclusive, exhaustive convex polytopes. Atoms may be empty, bounded, unbounded, open, closed, or contain only part of their boundary. In general there must necessarily be empty atoms. Fig. 1 illustrates these concepts for a 2-input 3-neuron example.

Network

3-Cube Representation

Input Space

fig. 1

Now let the output $\mathbf{y}$ of the layer $\mathcal{L}^1_n$ be taken as the input to a second layer $\mathcal{L}^2_m$ consisting of a single m-input neuron O with weight matrix (vector) U and threshold η. We will refer to such a two layer feed forward net as a *Two-layer Perceptron.* As above, O corresponds to an m-dimensional hyperplane $K_{U,\eta}$ which exists along with the hypercube Q_m. The two may intersect. In the event that they do, the vertices of the cube are partitioned into two disjoint sets,

$$F = K^1_{U,\eta} \cap Q_m, \qquad \text{and} \qquad G = K^0_{U,\eta} \cap Q_m.$$

In turn, the set of vertices F correspond to a set of atomic convex regions of input space; let

$$\mathcal{F} = \bigcup_{\mathbf{y} \in F} a_{\mathbf{y}} \subset \mathbf{R}^n.$$

We say the region $\mathcal{F}$ is *implemented* by the two layer net $\mathcal{L}^1_n$ and $\mathcal{L}^2_m$ because the output of neuron O is 1 if and only if $x \in \mathcal{F}$; see fig. 2. (Note in general only 7 regions result from the intersection of 3 lines in the plane while there are 8 vertices of the 3-cube. Vertex E in the figure is a "don't care" as it corresponds to no actual region.)

fig. 2

Alternatively such a region $\mathcal{F}$ may be referred to as a region *classifiable* by a Two-layer Perceptron. The Two-layer Perceptron classification problem is that of finding a characterization of those regions of n-dimensional space which can be implemented by a two layer neural net. As we've seen, such a collection $\mathcal{F}$ of convex polytopes arising from the decomposition of the input space $\mathbf{R}^n$ by hyperplanes will be two-layer classifiable if and only if their corresponding set of vertices in net-space can be separated by a hyperplane from the vertices corresponding to the complementary region to $\mathcal{F}$.

The complete solution to the Two-layer Perceptron classification problem is not known. However it is known that a region which is the arbitrary union of convex polytopes can be classified by three layer net, [2], and as a result there has been less interest in the two layer problem. Nevertheless there continues to be work done on the two layer problem [3],[7],[8], culminating in a body of known sufficient conditions.

In this paper we give new sufficient conditions on a region in order that it be 2-layer implementable. These conditions subsume all those known to us. The test is easy to apply to two dimensional regions given graphically and many interesting regions are decided by the conditions; see fig. 3. We obtain this result as an application of a new sufficient condition for the hypercube separation problem (Main Lemma §3).

(a) Decided (b) Decided (c) Not Decided

Regions implementable by a 2-layer net, some decided by the Main Theorem.

fig. 3

§2 CoRD Regions.

Let $C_1, C_2, \ldots, C_p$ be a nest of convex polytopes

$$C_1 \supset C_2 \supset \ldots \supset C_p. \tag{2.1}$$

We assume p is even, otherwise put $C_{p+1} = \emptyset$. By a *convex recursive deletion*, or CoRD, region we mean a set S of the form

$$S = (C_1 \cap C_2') \cup (C_3 \cap C_4') \cup \ldots \cup (C_{p-1} \cap C_p'). \tag{2.2}$$

where C' denotes the complement of the region C. We allow the possibility that $C_1 = \mathbf{R}^n$. Some examples of CoRD regions in $\mathbf{R}^2$ are illustrated in fig. 3a,b above. The representation is not unique as seen by the example illustrated in fig. 4. Nevertheless our arguments follow from the CoRD representation and so the results apply to any region capable of at least one such representation.

$$(C_1 \cap C_2') \cup C_3 = C_1 \cap K'$$

fig. 4

Theorem. The class of CoRD regions is closed under complementation and closed under intersection with convex polytopes.

Proof. Let S be as in (2.1) and (2.2) and consider the complement

$$\begin{aligned} S' = [(C_1 \cap C_2')\cup \ldots \cup (C_{p-1} \cap C_p')]' = \\ (C_1' \cup C_2) \cap (C_3' \cup C_4) \cap \ldots \cap (C_{p-1}' \cup C_p) \end{aligned} \tag{2.3}$$

In general suppose for sets A, B, C, and D that $A' \subset C'$ and $B \supset D$, then

$$\begin{aligned} (A' \cup B) \cap (C' \cup D) &= (A' \cap (C' \cup D)) \cup (B \cap (C' \cup D)) \\ &= A' \cup (B \cap C') \cup D. \end{aligned} \tag{2.4}$$

The latter is a reassociation of the first member. Since $C_1' \subset C_3'$ and $C_2 \supset C_4$, the first two terms of the right member of (2.3) becomes

$$[C_1' \cup (C_2 \cap C_3') \cup C_4] \cap (C_5' \cup C_6) \cap \ldots \cap (C_{p-1}' \cup C_p). \tag{2.5}$$

Now taking $A' = C_1' \cup (C_2 \cap C_3')$, $B = C_4$, $C = C_5$, $D = C_6$, it is easy to see that $A' \subset C'$ and $B \supset D$ so that the reassociation can continue; (2.5) becomes

$$[C_1' \cup (C_2 \cap C_3') \cup (C_4 \cap C_5') \cup C_6] \cap \ldots \cap (C_{p-1}' \cup C_p).$$

Continuing inductively we obtain

$$\begin{aligned}(C_1' \cup C_2) \cap (C_3' \cup C_4) \cap \ldots \cap (C_{p-1}' \cup C_p) \\ &= C_1' \cup (C_2 \cap C_3') \cup (C_4 \cap C_5') \cup \ldots \cup C_p \\ &= (\mathbf{R}^n \cap C_1') \cup (C_2 \cap C_3') \cup \ldots \cup (C_p \cap \emptyset')\end{aligned}$$

which is in the form of a CoRD region. Next let S be a CoRD region as above and C a convex polytope. Then

$$C \cap S = [(C_1 \cap C) \cap C_2'] \cup \ldots \cup [(C_{n-1} \cap C) \cap C_n'].$$

Since $C_i \cap C$ is also a convex polytope, it follow that $C \cap S$ is a CoRD region. This completes the proof. ∎

Definition. Let $\mathcal{C}$ denote the intersection of all classes $\mathcal{P}$ of subsets of input space, $\mathbf{R}^n$, containing the convex polytopes and closed under complementation and intersections with convex polytopes.

From what we've done so far it follows that the class of CoRD regions contains $\mathcal{C}$. Actually the two are the same.

Theorem. The class of CoRD regions is identical with $\mathcal{C}$.

Proof. It remains to show that the CoRD regions are contained in $\mathcal{C}$. We do this by showing that any class $\mathcal{P}$ containing the convex polytopes and closed as required, contains all CoRD regions. So let S be a CoRD region

$$S = (C_1 \cap C_2') \cup \ldots (C_{p-1} \cap C_p').$$

It is easy to see that $C_{p-1} \cap C'_p \in \mathcal{P}$. By closure under complementation, $(C_{p-1} \cap C'_p)' \in \mathcal{P}$, therefore $C_{p-2} \cap (C_{p-1} \cap C'_p)' \in \mathcal{P}$. Again by complementation, $C'_{p-2} \cup (C_{p-1} \cap C'_p) \in \mathcal{P}$. Next C_{p-3} is a convex polytope and contains $(C_{p-1} \cap C'_p)$, therefore

$$C_{p-3} \cap [C'_{p-2} \cup (C_{p-1} \cap C'_p)] = (C_{p-3} \cap C'_{p-2}) \cup (C_{p-1} \cap C'_p) \in \mathcal{P}.$$

Continue by induction obtaining $S \in \mathcal{P}$. ∎

Remark It is shown in [4] that the sets of $\mathcal{C}$ are 2-layer implementable by directly constructing an implementing network.. In the next section we show that the class of CoRD regions is 2-layer implementable by considering their representing hypercubes.

§3 Cubes Corresponding to CoRD Regions

By the j^{th} ($m-1$ dimensional) face R^i_j, $i = 0, 1$ of the cube Q_m we mean the set of vertices of Q_m whose jth component is i, $j = 1, \ldots, m$. The opposite face is $R^{i'}_j$ where $i' = 1$ if $i = 0$ and $i' = 0$ if $i = 1$.

More generally, by the *facet* $R^{i_1 \cdots i_r}_{j_1 \cdots j_r}$ we mean the vertices $(b_1, b_2, \ldots, b_m)$ of Q_m such that $b_{j_k} = i_k$, $k = 1, \ldots, r$. Evidently $R^{i_1 \cdots i_r}_{j_1 \cdots j_r}$ is itself an $m - r$ dimensional cube.

Definition. Let A and B be two sets of vertices of a cube Q_m. We say these vertices can be *separated* if there exits an oriented hyperplane K in m-dimensional space such that $A \subset K^1$ and $B \subset K^0$.

Main Lemma. Let A and B be separable sets of vertices of an $m-1$ face R^i_j of an m cube Q_m. If C is any subset of the opposite face $R^{i'}_j$, then $A \cup C$ and B are separable subsets of Q_m.

Proof. Let co(A) and co(B) denote the convex hulls of the sets A and B respectively. Then co(A), co$(B) \subset R^i_j$ and

$$\text{co}(A) \cap \text{co}(B) = \emptyset$$

because the sets A and B can be separated. Let $\mathrm{co}(A \cup C)$ denote the convex hull of $A \cup C$. Then $\mathrm{co}(A \cup C) \cap R_j^i = \mathrm{co}(A)$ and therefore $\mathrm{co}(B) \cap \mathrm{co}(A \cup C) = \emptyset$. It follows by the Hahn-Banach Theorem, [1,p47], that $A \cup C$ and B can be separated. ∎

Remark In particular, if either A or B is empty, then the result holds (trivially if B is empty).

Main Theorem. Let S be a CoRD region of input space, then S is two layer implementable.

Proof. Let S be a CoRD region, (2.1), (2.2), where $i_0 = 0$ and

$$C_k = \bigcap_{j=i_{k-1}+1}^{i_k} H_j^{\pm_j}, \qquad k = 1, 2, \ldots, p,$$

and Q_m the corresponding hypercube of $m = i_p$ dimensional space. Let F be the set of vertices corresponding to S as above, i.e.

$$F = \{\mathbf{y} \in Q_m \, : \, q(\mathbf{y}) \subset S, \, q(\mathbf{y}) \neq \emptyset\}$$

and let

$$G = \{\mathbf{y} \in Q_m \, : \, q(\mathbf{y}) \subset S', \, q(\mathbf{y}) \neq \emptyset\}$$

Let E be the set of vertices of Q_m corresponding to no convex polytope of input space. We show that F and G are separable by induction on the sequence of hyperplanes. There are two cases, for odd k, C_k is an *included* polytope in that $(C_k \cap C'_{k+1}) \subset S$. For an even k, C_k is an *excluded* region in that $(C_{k-1} \cap C'_k) \subset S$. We may start without loss of generality with the including case, i.e. with $C_1 \neq \mathbf{R}^n$.

Now $C_1 \subset H_1^{\pm_1}$ and $S \subset C_1$, therefore, if $\pm_1 = 1$ say, then $y_1 = 1$ for all vertices $\mathbf{y} \in F$, i.e. $F \subset R_1^1$. Hence each vertex of the opposite face, R_1^0, is either in G or in E. It follows by force of the lemma that if the sets

$$F_1 = F \cap R_1^1 \, (= F) \qquad \text{and} \qquad G_1 = G \cap R_1^1$$

are separable in the $m-1$ dimensional face R_1^1, then F and G will be separable in Q_m. Let Q_{m-1} be the $m-1$ dimensional hypercube consisting of the face R_1^1, i.e.

$$\{1\} \times Q_{m-1} = R_1^1.$$

Note that every vertex of Q_{m-1} corresponds to a convex polytope contained in $H_1^{\pm_1}$.

For induction assume the theorem is true provided it can be shown that in the $m-j$ dimensional cube, Q_{m-j}, $1 \leq j < m$,

$$\{\pm_1\} \times \ldots \times \{\pm_j\} \times Q_{m-j} = R_{1,\ldots,j}^{\pm_1 \ldots \pm_j},$$

the vertices

$$F_j = F \cap R_{1,\ldots,j}^{\pm_1 \ldots \pm_j} \quad \text{and} \quad G_j = G \cap R_{1,\ldots,j}^{\pm_1 \ldots \pm_j}$$

can be separated. The vertices of Q_{m-j} correspond to convex polytopes of input space lying within $I = H_1^{\pm_1} \cap \ldots \cap H_j^{\pm_j}$, i.e. if $\mathbf{y}^{m-j} \in Q_{m-j}$ then $q(\pm_1, \ldots, \pm_j, \mathbf{y}^{m-j}) \subset I$.

Let the next hyperplane H_{j+1} be an edge of C_k, i.e. $i_{k-1}+1 \leq j+1 \leq i_k$. Then $I \subset C_{k-1}$. There are two cases, k odd or even. Suppose the former, then C_k is an including region. Since $C_k \subset H_{j+1}^{\pm_{j+1}}$, it follows that every vertex $\mathbf{y}$ of Q_m corresponding to a polytope of S contained in I will in fact be contained in C_k and must have its $(j+1)^{st}$ component equal to $\pm_{j+1}$. Such a vertex will belong to the 1-face of Q_{m-j} whose first components are also $\pm_{j+1}$. Therefore the opposite face, $\Pi_1^{\pm'_{j+1}}$ of Q_{m-j}, consists of vertices corresponding either to convex polytopes contained in C_{k-1} but not in C_k and hence are G type vertices, or to no convex polytope of input space, i.e. E type vertices. Hence, by the lemma, if the $m-j-1$ dimensional cube Q_{m-j-1},

$$\{\pm_j\} \times Q_{m-j-1} = \Pi_1^{\pm_{j+1}}$$

is separable, then so is Q_m. Note that the vertices of Q_{m-j-1} correspond to convex polytopes contained in $I \cap H_{j+1}^{\pm_{j+1}}$. Induction is complete in this case.

Now suppose that k is even, then C_k is an excluding region. Since $C_k \subset H_{j+1}^{\pm_{j+1}}$, each vertex of the 1-face $\Pi_1^{\pm'_{j+1}}$ corresponds either to a convex polytope contained in C_{k-1}, an including region, but not in C_k or to no region of input space. Hence such a vertex is either an F or an E vertex. Therefore again if the $m-j-1$ dimensional cube Q_{m-j-1},

$$\{\pm_j\} \times Q_{m-j-1} = \Pi_1^{\pm_{j+1}}$$

is separable, then so is Q_m. Note that the vertices of Q_{m-j-1} correspond to convex polytopes contained in $I \cap H_{j+1}^{\pm_{j+1}}$. Induction is complete in this case.

Arriving by finite induction to the last hyperplane H_{i_p}, note that the resulting cube $Q_{m-i_p} = Q_0$ is a single point. If the final polytope C_p is including, then this vertex is an F type, otherwise it is a G type. Either way, by the Remark following the Main Lemma above, this 0 dimensional facet is separable and the proof is complete. ∎

We gather together some facts which emerged in the course of the above proof.

Corollary A. Running through the indices in their natural order 1,2,..., the sequence of facets

$$R_1^{\pm'_1}, R_{1,2}^{\pm_1 \pm'_2}, \ldots, R_{1,2,\ldots,i_p-1,i_p}^{\pm_1 \pm_2 \ldots \pm_{i_p-1} \pm'_{i_p}},$$

have the properties: (assuming $C_1 \neq \mathbf{R}^n$)

(a) $R_{1,\ldots,j-1,j}^{\pm_1 \ldots \pm_{j-1} \pm'_j}$ is of dimension $m-j$, and

(b) all veritices of this facet are G's or E's if $i_{k-1} < j \leq i_k$ and k is odd and are F's or E's if k is even.

If $C_1 = \mathbf{R}^n$ then the above holds with even and odd interchanged.

This is illustrated in the cube diagram accompanying fig. 2 where $S = H_1^1 \cap H_2^0 \cap H_3^0$ and R_1^0 consists of G's and an E, $R_{1,2}^{1,1}$ consists of F's, and $R_{1,2,3}^{1,0,1}$ is an F.

Corollary B. Let A and B be a two set partition of the vertices of an m-cube Q_m. Suppose there exists a permutation π of the first m natural numbers and a choice $\pm_{\pi_j} = 0$ or 1, for $j = 1, \ldots, m$, so that for each subfacet $Q_{m-j} = R_{\pi_i \pi_2 \ldots \pi_j}^{\pm_{\pi_1} \pm_{\pi_2} \ldots \pm_{\pi_m}}$ of the chain

$$Q_m \supset R_{\pi_1}^{\pm_{\pi_1}} \supset R_{\pi_1 \pi_2}^{\pm_{\pi_1} \pm_{\pi_2}} \supset \ldots \supset R_{\pi_1 \pi_2 \ldots \pi_m}^{\pm_{\pi_1} \pm_{\pi_2} \ldots \pm_{\pi_m}}$$

either

$$Q_{m-j} \subset A \qquad \text{or} \qquad Q_{m-j} \subset B.$$

Then A and B can be separated by a hyperplane.

Proof. This follows by the construction in the proof of the Theorem.

References

[1] Balakrishnan, "Applied Functional Analysis," *Springer-Verlag*, New York (1976)

[2] R. P. Lippmann. "An Introduction to Computing with Neural Nets,"*IEEE ASSP Magazine*, 3(4), 4—22, (1987)

[3] J. Makhoul, A. El-Jaroudi, R. Schwartz, "Formation of disconnected decision regions with a single hidden layer," *Proceedings of the International Joint Conference on Neural Networks* **I**, 455—460, IEEE TAB Neural Network Committee (1989)

[4] R. McCurley, K. Miller, R. Shonkwiler, "Classification Power of Multiple-layer Artifical Neural Networks," to appear in the*Proceedings of SPIE's 1990 Technical Symposium on Optical Engineering and Photonics in Aerospace Sensing: Program on Optical/Neural Image and Information Processing.* (1990)

[5] M. L. Minsky, S. A. Papert, "Perceptrons (Expanded Edition)," *MIT Press*, (1988)

[6] F. Rosenblatt, "Perceptrons and the Theory of Brain Mechanism," *Spartan Books* (1962)

[7] P. Rujan, M. Marchand, "A Geometric Approach to Learning in Neural Networks," *Proceedings of the International Joint Conference on Nerual Networks* **II**, 105—109, IEEE TAB Neural Network Committee (1989)

[8] A. Wieland, R. Leighton, "Geometric Analysis of Neural Network Capabilities," *Proceedings of the Second IEEE International Conference on Nerual Networks* **III**, 385—392, IEEE Computer Society (1988)