IDIAP RESEARCH REPORT

IDIAP

Martigny - Valais - Suisse

# On the Complexity of the Class of Regions Computable by a Two-Layered Perceptron

Eddy N. Mayoraz †

IDIAP–RR 96-03

December 1996

Dalle Molle Institute for Perceptive Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 − 27 − 721 77 11
fax +41 − 27 − 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

† IDIAP—Dalle Molle Institute for Perceptive Artificial Intelligence, P.O.Box 592, CH-1920 Martigny, Valais, Switzerland Eddy.Mayoraz@idiap.ch

IDIAP Research Report 96-03

# On the Complexity of the Class of Regions Computable by a Two-Layered Perceptron

Eddy N. Mayoraz

December 1996

**Abstract.** This work is concerned with the computational complexity of the recognition of $\mathcal{LP}_2$, the class of regions of the Euclidian space that can be classified exactly by a two-layered perceptron. Several subclasses of $\mathcal{LP}_2$ of particular interest are also considered. We show that the recognition problems of $\mathcal{LP}_2$ and of other classes considered here are intractable, even in some favorable circumstances. We then identify special cases having polynomial time algorithms.

**Acknowledgements**: This work was initiated while the author was a postdoctoral visitor at RUTCOR—Rutgers University's Center for Operations Research, supported by the Swiss National Science Foundation, by RUTCOR and by DIMACS—Center for Discrete Mathematics and Theoretical Computer Science. The author would like to thank Dr. Motakuri Ramana for his always pertinent critics and fruitful discussions.

# 1 Introduction

Several works have been lately devoted to the problem of characterizing the regions of the Euclidian space $\mathbb{R}^n$ that can be classified by a multilayer perceptron (MLP) with $n$ real inputs, linear threshold processing units and a single binary output [Lip87, GC90, ZAW91, Zwi94]. Usually, the attention is restricted to $\mathcal{U}$ denoting the set of all regions of $\mathbb{R}^n$ expressible as a finite union of finite intersections of open or closed half-spaces, since the classification of any region out of $\mathcal{U}$ would required infinitely many threshold units.

The construction of a depth-3 MLP (2 hidden layers) classifying any region of $\mathcal{U}$ is straitforward, given any expression of the latter as a union of intersections of half-spaces [Lip87]. Thus, the research in this area is mainly concerned with the characterization of the regions classifiable by a depth-2 MLP. In some works, exact classifications is studied [ZAW92, Sho93, Zwi94], while others deal with almost everywhere correct classifications [Gib96] or approximations [SGM96].

The present work focuses on the computational complexity of the decision problem "Given $V$, does $V$ belong to $\mathcal{LP}_2$ ?", where $\mathcal{LP}_2$ is the class of regions of $\mathbb{R}^n$ *exactly* classifiable with a depth-2 MLP. The same issue will be studied for some subclasses of $\mathcal{LP}_2$ and for different input formats for the specification of $V$ and thus, the boundary between tractable and untractable problems will highlight the real core of the difficulty of the general decision problem.

Several constructive training algorithms have been proposed for MLPs. The most elaborate ones (see for example [TTK93]) have at each iteration a strategy of construction based on the causes of the non-membership of a region into the class $\mathcal{LP}_2$ or some subclasses that will be considered below. The computational complexity of these algorithms is directly related to the complexity of the recognition problems studied in this paper.

# 2 Formalization of the recognition problems

The functions computed by two-layered perceptrons as considered in the present paper are defined from the $n$ dimensional Euclidian space $\mathbb{R}^n$ onto the Boolean set $\{0,1\}$ and are of the form:

$$f(\boldsymbol{x}) = g(h_1(\boldsymbol{x}), \ldots, h_m(\boldsymbol{x}))\ , \tag{1}$$
$$h_i(\boldsymbol{x}) = \mathrm{sgn}(w_{i0} + \boldsymbol{x}^\top \boldsymbol{w}_i),\ i = 1, \ldots, m\ , \tag{2}$$
$$g(\boldsymbol{b}) = \mathrm{sgn}(t_0 + \boldsymbol{b}^\top \boldsymbol{t})\ . \tag{3}$$

In what follows, the *coefficients* $w_{ij}$ and $t_i, i = 1, \ldots, m, j = 1, \ldots, n$ and the *thresholds* $w_{i0}$ and $t_0$ are real values. The sign function is defined as $\mathrm{sgn}(r) = 1$ if $r \geq 0$ and 0 otherwise. Note that replacing this function sgn by a function $\mathrm{sgn}^d$ everywhere identical to sgn except in 0 ($\mathrm{sgn}^d(0) = 0$) has no incidence on the class $\mathcal{LP}_2$, since on the one hand equation $\mathrm{sgn}(r) = -\mathrm{sgn}^d(-r)$ provides a solution in case $\mathrm{sgn}^d$ replaces sgn in (2), and on the other hand $t_0$ can always be chosen distinct from $-\boldsymbol{b}^\top \boldsymbol{t}$ for all $\boldsymbol{b} \in \{0,1\}^n$.

**Definition 2.1** A region $V \subset \mathbb{R}^n$ is in $\mathcal{LP}_2$ if and only if its characteristic function can be written in the form of equations (1-3).

A *polyhedron* (resp. a *pseudo-polyhedron*) is a finite intersection of closed (resp. open or closed) half-spaces. The set of all half-spaces used in the expression of $V$ as a union of pseudo-polyhedron is called a *basis* of $V$. The general recognition problem can be stated as :

$\mathcal{LP}_2$-Recognition: Given $V \in \mathcal{U}$ specified by a basis $\mathcal{H}$ and a union of intersections of elements of $\mathcal{H}$, is $V$ in $\mathcal{LP}_2$ ?

There exist simple regions of $\mathbb{R}^2$ which are in $\mathcal{LP}_2$ and for which the only MLPs we know of for their classification have a number of hidden units exponential in the size of a reasonable encoding of the region. This observation leads us naturally to a simplified version of $\mathcal{LP}_2$-Recognition where each hidden unit of the MLP classifying the region has to compute the characteristic function of one element of the basis given as input of the problem :

$\overline{\mathcal{LP}}_2$-Recognition: Given $V \in \mathcal{U}$ specified by a basis $\mathcal{H}$ and a union of intersections of elements of $\mathcal{H}$, is there an expression of the form of equations (1-3) for the characteristic function of $V$, such that each $h_i$ is the characteristic function of one element of $\mathcal{H}$ ?

Based on this new recognition problem a sub-class of $\mathcal{LP}_2$ can be defined, although it requires some care in order to have a basis uniquely specified by the region. A basis $\mathcal{H}$ of $V$ is *minimal* if $V$ cannot be expressed as a union of pseudo-polyhedra based on a proper subset of $\mathcal{H}$. A pseudo-polyhedron is *fully dimensioned* in $\mathbb{R}^n$ if it contains an open ball of dimension $n$. $V \in \mathcal{U}$ is *fully dimensioned* if it is a union of fully dimensioned pseudo-polyhedra, and $\mathcal{FD}$ denotes the set of fully dimensioned regions of $\mathcal{U}$. It can be shown that $V \in \mathcal{FD}$ has a unique minimal basis which is denoted $\mathcal{H}_V$[1]. Restricting our attention to fully dimensioned regions, we can now define the class $\overline{\mathcal{LP}}_2$ as follows:

**Definition 2.2** A region $V$ is in $\overline{\mathcal{LP}}_2$ if $V \in \mathcal{FD}$ and if $\mathcal{H}_V$ together with an expression of $V$ as a union of intersections of elements of $\mathcal{H}_V$ is a positive instance of $\overline{\mathcal{LP}}_2$-Recognition.

In their attempt at characterizing the class $\mathcal{LP}_2$, two groups of researchers proposed independently a sufficient condition for a region $V$ to be in $\mathcal{LP}_2$ [ZAW92, Sho93].

**Definition 2.3** A region $V \in \mathbb{R}^n$ is in $\mathcal{D}$ (resp. $\tilde{\mathcal{D}}$) if it can be expressed as an iterated difference of finitely many polyhedra (resp. pseudo-polyhedra) $P_1, \ldots, P_k$ as follows:

$$V \;=\; P_1 \backslash (P_2 \backslash (\ldots P_k) \ldots).$$

Clearly $\mathcal{D}$ is a proper subset of $\tilde{\mathcal{D}}$. The proof that $\tilde{\mathcal{D}} \subset \mathcal{LP}_2$ is established with the following straitforward lemma :

**Lemma 1** $\forall$ pseudo-polyhedron $P$ and $\forall$ $V \in \mathcal{LP}_2$, $P \backslash V \in \mathcal{LP}_2$.

Several examples are proposed in the literature showing that $\tilde{\mathcal{D}} \subsetneq \mathcal{LP}_2$. Recently, M. Ramana and the author proved that if $V \in \tilde{\mathcal{D}}$ has a basis $\mathcal{H}$, there exist polyhedra $P_i$, $i, \ldots, k$, all based on $\mathcal{H}$, such that $V = P_1 \backslash (P_2 \backslash (\ldots P_k) \ldots)$, which implies that $\tilde{\mathcal{D}} \cap \mathcal{FD} \;\subset\; \overline{\mathcal{LP}}_2$.

## 3 From geometry to combinatorics

An ordered set of half-spaces $\mathcal{H} = \{H_1, \ldots, H_h\}$ of $\mathbb{R}^n$, called an *arrangement*, defines a mapping $\Phi_{\mathcal{H}} : \mathbb{R}^n \rightarrow \{0,1\}^h$, where $(\Phi_{\mathcal{H}})_i$ is the characteristic function of $H_i$. The set $D_{\mathcal{H}} = \Phi_{\mathcal{H}}(\mathbb{R}^n)$ is called the *domain* of the arrangement $\mathcal{H}$. Given an arrangement $\mathcal{H}$, there is a one-to-one correspondence between regions of $\mathcal{U}$ having $\mathcal{H}$ as basis and partial Boolean functions defined on $D_{\mathcal{H}}$ : the function corresponding $V \in \mathcal{U}$ and denoted $f_{V,\mathcal{H}}$ is such that $f_{V,\mathcal{H}} \circ \Phi_{\mathcal{H}}$ is the characteristic function of $V$. In this correspondence, an expression of $V$ as a union of pseudo-polyhedra based on $\mathcal{H}$ coincides with a disjunctive normal form (DNF) of $f_{V,\mathcal{H}}$. For two arrangements $\mathcal{H} \subseteq \mathcal{G}$ and a function $f : D_{\mathcal{H}} \rightarrow \{0,1\}$, there is a unique function $g : D_{\mathcal{G}} \rightarrow \{0,1\}$ satisfying $f \circ \Phi_{\mathcal{H}} \;=\; g \circ \Phi_{\mathcal{G}}$. Function $g$ is called the *expansion* of $f$ from $\mathcal{H}$ to $\mathcal{G}$.

Recall that a partial Boolean function $f : X \subset \{0,1\}^h \rightarrow \{0,1\}$ is *threshold* if there exist $\boldsymbol{t} \in \mathbb{R}^h$ and $t_0 \in \mathbb{R}$ such that $f(\boldsymbol{b}) = \mathrm{sgn}(t_0 + \boldsymbol{b}^\top \boldsymbol{t})$ $\forall \boldsymbol{b} \in X$. Existence of MLPs is reduced to thresholdness of functions of the type $f_{V,\mathcal{H}}$. Exploiting this transformation of the issues from geometrical terms into a combinatorial language, the development of threshold function theory will bring some light on the complexity issues addressed in the present work.

Our recognition problems can be reformulated as follows :

$\overline{\mathcal{LP}}_2$-Recognition: Given an arrangement $\mathcal{H}$ and a DNF for a function $f : D_{\mathcal{H}} \rightarrow \{0,1\}$, is $f$ threshold ?

$\mathcal{LP}_2$-Recognition: Given an arrangement $\mathcal{H}$ and a DNF for a function $f : D_{\mathcal{H}} \rightarrow \{0,1\}$, is there an arrangement $\mathcal{G} \supseteq \mathcal{H}$ such that the expansion of $f$ from $\mathcal{H}$ to $\mathcal{G}$ is threshold ?

[1] $H \in \mathcal{H}_V$ only if $B(H) \cap B(V)$ contains an open ball of dimension $n-1$ in the subspace $B(H)$, where $B(X)$ denotes the *border* of $X$ according to the usual topology on $\mathbb{R}^n$.

Several results established in various works on the characterization of $\mathcal{LP}_2$ [GC90, ZAW92, Sho93] have a simple explanation in threshold function theory. For example, Lemma 1 is a direct consequence of two facts : (i) the negation of a threshold function is threshold; (ii) the conjunction between a single variable and a threshold function is threshold.

As second example is related to the class $\mathcal{M} \supset \mathcal{LP}_2$ defined by :

$$\begin{aligned} \mathcal{M} \;=\; \{ \;& V \in \mathcal{U} \mid \text{ there is no half-space } H, \text{ and balls } B_1, B_2 \text{ in } I\!\!R^n, \text{ s.t.} \\ & (H \cap B_1) \cup (H^c \cap B_2) \subset V \;\;, \text{ and} \\ & (H \cap B_2) \cup (H^c \cap B_1) \subset V^c \;\; \}^{\,2} \end{aligned}$$

and introduced in Proposition 8 in [GC90] (or Theorem 1 in [ZAW92]) to express a necessary condition for $V$ to be in $\mathcal{LP}_2$. The condition expressed in the definition of $\mathcal{M}$ is translated into *monotonicity* of Boolean function [Mur71], which is the simplest necessary condition for a function to be threshold.

In the next section, we will take advantage on results in Boolean function theory to discuss the complexity of each class of regions introduced above :

$$\mathcal{D} \subsetneq \tilde{\mathcal{D}} \subsetneq \mathcal{LP}_2 \subsetneq \mathcal{M}, \qquad \tilde{\mathcal{D}} \cap \mathcal{FD} \subsetneq \overline{\mathcal{LP}}_2 \subsetneq \mathcal{LP}_2 \cap \mathcal{FD}.$$

# 4 Intractability results

If a lot is known in threshold function theory about completely defined Boolean function, several questions remain open when partial functions are considered. For example, thresholdness of a complete function given by a monotonic[3] DNF can be tested in polynomial time [PS85], but nothing is known for partial function. On the other hand, even if $V \subset I\!\!R^n$ based on $\mathcal{H}$ is given by a partial function defined on the very structured domain $D_{\mathcal{H}}$, it still generalizes the case of complete Boolean function. Indeed, for any arbitrary complete function $f : \{0,1\}^h \to \{0,1\}$, there is a region $V \subset I\!\!R^h$ of basis $\mathcal{H} = \{H_1, \ldots, H_h\}$, where $H_i = \{\boldsymbol{x} \mid x_i \geq 0\}$, such that $f = f_{V,\mathcal{H}}$. Therefore, all the questions we might ask in our setting will be at least as hard as their equivalent formulation for complete Boolean functions. In particular, since it is co-$NP$-Complete to decide whether an arbitrary DNF represents a threshold function, $\overline{\mathcal{LP}}_2$-Recognition is $NP$-Hard. The proof that checking thresholdness is co-$NP$ can be generalized to prove the following lemma :

**Lemma 2** $\overline{\mathcal{LP}}_2$-Recognition is co-$NP$.

For a region $V \in \tilde{\mathcal{D}}$ given by a basis $\mathcal{H}$ and a DNF expression for $f_{V,\mathcal{H}}$, it can be shown that (i) there is a specification of $V$ as an iterated difference of $k$ pseudo-polyhedra with $k$ bounded by $|\mathcal{H}|$, (ii) the equality between $V$ and the iterated difference can be checked in polynomial time. Thus, we have :

**Lemma 3** $\tilde{\mathcal{D}}$-Recognition is $NP$.

Using the fact that within $\mathcal{FD}$, $\mathcal{D} \subset \tilde{\mathcal{D}} \subset \overline{\mathcal{LP}}_2 \subset \mathcal{LP}_2$, a single construction will be used to prove the intractability of each of these classes :

**Theorem 4** $\mathcal{LP}_2$-Recognition and $\mathcal{D}$-Recognition are $NP$-Hard, $\overline{\mathcal{LP}}_2$-Recognition is co-$NP$-Complete, and $\tilde{\mathcal{D}}$-Recognition is $NP$-Complete.

**Proof**: By Lemmas 2 and 3, it remains to show that the four recognition problems are $NP$-Hard. This will be done by reduction to SAT.

An instance of SAT is given as a DNF of a Boolean function $a : \{0,1\}^n \to \{0,1\}$ with the question "is $a$ a tautology ?". Consider the Euclidian space $I\!\!R^n$ and the arrangement $\mathcal{H}$ containing the following $2n$ closed half-spaces :

$$H_i = \{\boldsymbol{x} \mid x_i \geq 0\}, \quad H_{n+i} = \{\boldsymbol{x} \mid x_i \leq -1\} \;\; \forall i = 1, \ldots, n.$$

[2]For $X \subset I\!\!R^n$, $X^c$ denotes the complement $I\!\!R^n \backslash X$.

[3]A DNF is monotonic if it does not contain positive and negative occurrences of the same variable.

Let $b_i$ denotes the Boolean variable which is the characteristic function of $H_i$ for all $i = 1, \ldots, 2n$.

Consider the region $V \subset I\!R^n$ based on $\mathcal{H}$ defined by

$$f_{V,\mathcal{H}}(b_1, \ldots, b_{2n}) \;=\; \begin{cases} a(b_1, \ldots, b_n) & \text{if } (b_{n+1}, \ldots, b_{2n}) = (0, \ldots, 0), \\ 0 & \text{if } (b_{n+1}, \ldots, b_{2n}) = (1, \ldots, 1), \\ 1 & \text{otherwise.} \end{cases} \tag{4}$$

The 4 complexity results are proved if the two following statements hold:

(i) $V \in \mathcal{D}$ if $a$ is a tautology;

(ii) $V \notin \mathcal{LP}_2$ if $a^{-1}(0) \neq \emptyset$. $a(\boldsymbol{b}) = 0$.

The proof of (i) is easy, since if $a$ is a tautology, $\boldsymbol{b} = (0^n, 1^n)$ is the only point for which $f_{V,\mathcal{H}}$ is 0. Thus, $V = I\!R^n \setminus \cap_{i=1}^n H_{n+i}$, which is clearly in $\mathcal{D}$.

The proof of (ii) goes as follows. Assume that $a$ is not a tautology and let distinguish two cases. If $\exists \boldsymbol{b} \in a^{-1}(0)$, with $b_j = 0$ for at least one $j$, we can show that $V \notin \mathcal{M}$. On the other hand, if $a^{-1}(0) = \{(1, \ldots, 1)\}$, we have to show that the region $I\!R^n \setminus (\cap_{i=1}^n H_i \cup \cap_{i=1}^n H_{n+i})$ is not in $\mathcal{LP}_2$. This situation is a generalization to $n$ dimensions of the example pictured in Figure 5 in [Gib96] and the proof is based on another property of threshold function known as *assumability* [Mur71]. $\triangle$

Deciding whether a Boolean function given by an arbitrary DNF is monotonic or not is *NP*-Complete. From this we can establish the following result :

**Proposition 5** Given a region $V$ specified by a basis $\mathcal{H}$ and its expression as a union of intersections of elements of $\mathcal{H}$, deciding whether $V \in \mathcal{M}$ is *NP*-Complete.

# 5 Tractable cases

The complexity results of Theorem 4 rely essentially on the fact that it is hard to find a DNF expression characterizing the complement of a region defined by a DNF, since it is already hard to decide whether this complement is empty or not (SAT). The first tractable cases listed in this section consist of situations where both $V$ and $V^c$ are available as input.

**Proposition 6** If $V$ is given by a basis $\mathcal{H}$ and two DNFs, one for $f_{V,\mathcal{H}}$ and the other for $f_{V^c,\mathcal{H}}$, then it can be decided in polynomial time whether $V$ is in $\mathcal{D}$ (similarly in $\tilde{\mathcal{D}}$).

The similar question about $\overline{\mathcal{LP}_2}$ was solved only with an additional assumption :

**Proposition 7** If $V$ is given by a basis $\mathcal{H}$ and two DNFs $D_1$ and $D_2$ for $f_{V,\mathcal{H}}$ and $f_{V^c,\mathcal{H}}$ such that $D_1(\boldsymbol{b}) = D_2(\boldsymbol{b}) = 0 \; \forall \boldsymbol{b} \notin D_{\mathcal{H}}$, then it can be decided in polynomial time whether $V$ is in $\overline{\mathcal{LP}_2}$.

In many practical situations, the size $n$ of the Euclidian space is small, while the number of half-spaces can be very large. Therefore, it is worth highlighting that some of the recognition problems become easy when $n$ is a constant.

**Proposition 8** $\overline{\mathcal{LP}_2}$-Recognition, $\tilde{\mathcal{D}}$-Recognition and $\mathcal{D}$-Recognition are polynomial in the size of the basis and in the size of the DNF given as input.

# 6 Open questions

It can be decided in polynomial time whether a Boolean function given by a monotonic DNF is threshold. Thus, the intractability of $\overline{\mathcal{LP}_2}$-Recognition might be connected to the intractability of class $\mathcal{M}$ (Proposition 5). We addressed the same question as in $\overline{\mathcal{LP}_2}$-Recognition with the additional assumption that the inputs are given in a monotonic form, *i.e.* the basis does not contain two half-spaces complement of each other and the DNF is monotonic. The complexity of this question is still open.

As mentioned in section 2, there exist some regions in $\mathcal{LP}_2$ for which all the known expressions of their characteristic functions in the form of equations (1-3) require $m$ exponential in the size of the input. An important and challenging problem would be to demonstrate some lower bounds on $m$ which could be exponential or even super-polynomial in the size of the input.

This work was initiated while the author was a postdoctoral visitor at RUTCOR—Rutgers University's Center for Operations Research, supported by the Swiss National Science Foundation, by RUTCOR and by DIMACS—Center for Discrete Mathematics and Theoretical Computer Science. The author would like to thank Dr. Motakuri Ramana for his always pertinent critics and fruitful discussions.

## References

[GC90] Gavin J. Gibson and Colin F. N. Cowan. On the decision regions of multilayer perceptrons. *Proceedings of the IEEE*, 78(10):1590–1594, 1990.

[Gib96] Gavin J. Gibson. Exact classification with 2-layered nets. *J. Computer and System Sciences*, 1996.

[Lip87] R. P. Lippmann. An introduction to computing with neural nets. *IEEE, Acoustic, Speech and Signal Processing*, 4:4–22, 1987.

[Mur71] Saburo Muroga. *Threshold Logic and Its Applications*. John Wiley & Sons, New York, 1971.

[PS85] U. N. Peled and B. Simeone. Polynomial-time algorithms for regular set-covering and threshold synthesis. *Discrete Applied Mathematics*, 12:57–69, 1985.

[SGM96] Catherine Z. W. Hassell Sweatman, Gavin J. Gibson, and Bernard Mulgrew. Exact classification with two-layer neural nets in $n$ dimensions. submitted, 1996.

[Sho93] Ron Shonkwiler. Separating the vertices of n-cubes by hyperplanes and its application to artificial neural networks. *IEEE Transactions on Neural Networks*, 4(2):343–347, March 1993.

[TTK93] Haruhisa Takahashi, Etsuji Tomita, and Tsutomu Kawabata. Separability of internal representations in multilayer perceptrons with application to learning. *Neural Networks*, 6:689–703, 1993.

[ZAW91] P. J. Zwietering, E. H. L. Aarts, and J. Wessels. The design and complexity of exact multilayered perceptrons. *International Journal of Neural Systems*, 2:185–199, 1991.

[ZAW92] P. J. Zwietering, E. H. L. Aarts, and J. Wessels. Exact classification with two-layered perceptrons. *International Journal of Neural Systems*, 3(2):143–156, 1992.

[Zwi94] Patrick J. Zwietering. *The Complexity of Multi-Layered Perceptrons*. PhD thesis, University of Eindhoven, 1994.