PSI 5886 – Prof. Emilio – 2018

Princípios de Neurocomputação

- Quartas feiras – das 18 hs às 21 hs
  ... ou (se combinamos) 18:30hs até 21:20hs
  18:10
- Sala B2-12

- Prof. Emilio Del Moral Hernandez
- emilio@lsi.usp.br

Disciplinas oferecidas

Disciplina PSI5886 - 1

Princípios de Neurocomputação

Unidade: Escola Politécnica

Número de vagas:

Alunos regulares	Alunos especiais	Total
20	20	40

Número mínimo de alunos: 5

Data inicial: 10/09/2018 **Data final**: 02/12/2018
Data limite de cancelamento: 30/09/2018
Número de créditos: 8

Docente(s) Ministrante(s)
Emílio Del Moral Hernandez

Horário / Local:

Terça	18:00 - 21:00	Prédio da Engenharia Elétrica

ICONE – EPUSP: Grupo de Inteligência Computacional, Modelagem e Neurocomputação Eletrônica

Prof. Dr. Emilio Del Moral Hernandez

Graduação em Engenharia Elétrica na EPUSP

Doutorado em Engenharia Elétrica pela
University of Pennsylvania (Upenn – Philadelphia)

Livre Docente da EPUSP, na àrea de
Neurocomputação Eletrônica e Sistemas Adaptativos

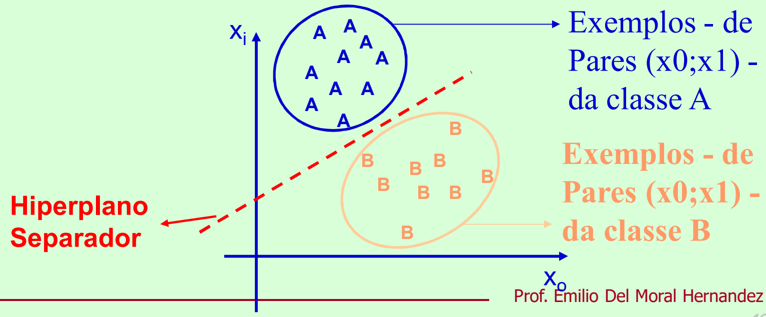
Atuante no IEEE e na IEEE - CIS

Contato: emilio.delmoral@usp.br / emilio@lsi.usp.br

***Website do Grupo de Pesquisa*: www.lsi.usp.br/ICONE**

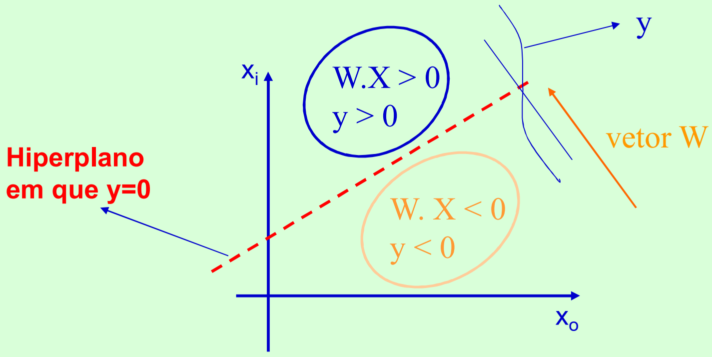
O Perceptron Digital: y=signal(Σ w_i x_i - θ) (função de transferência tipo "degrau")

- Viabiliza a classificação de padrões com separabilidade linear
- O algoritmo de aprendizado adapta os Ws de forma a encontrar o hiperplano de separação adequado
- Aprendizado por conjunto de treinamento

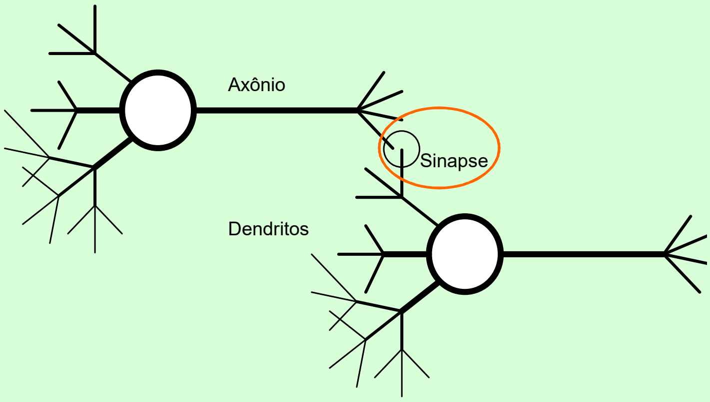


E se a saída do nosso problema não for digital? O “Perceptron Contínuo”: $y=tgh(\Sigma\ w_i\ x_i - \theta)$

- Que problemas de entradas contínuas conseguimos atacar usando uma função de transferência tangente hiperbólica)



Cômputos mais complexos … são realizados pelo encadeamento de vários nós



A conexão entre um axônio de um neurônio e um dendrito de outro é denominada **Sinapse**

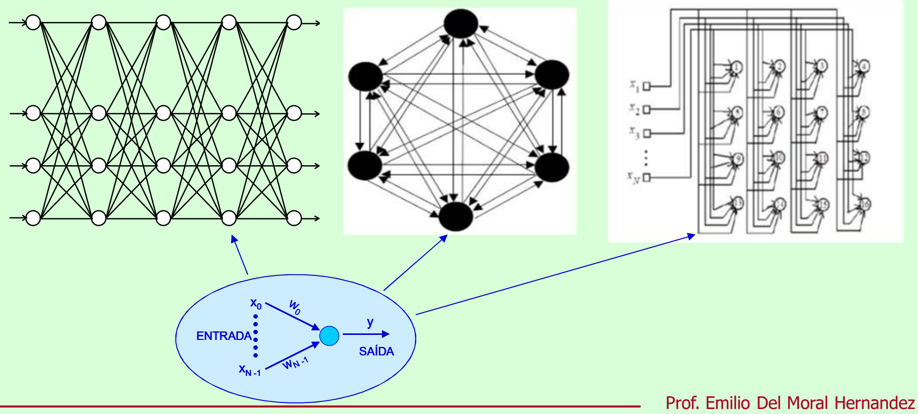
Três arquiteturas neurais importantes:

ou recorrentes em geral

1) MLP - Multi Layer Perceptron

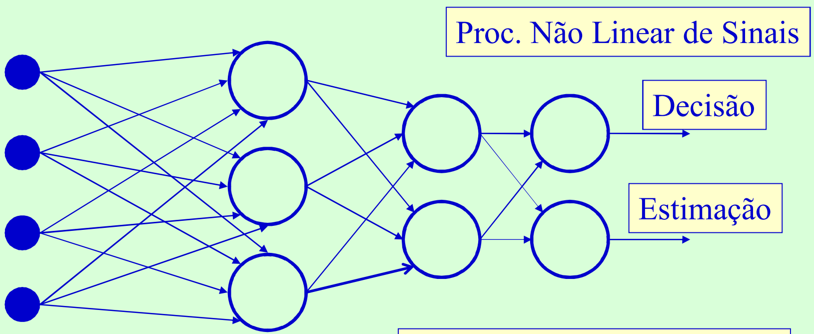
2) ~~Memória Associativa de Hopfield~~

3) Mapas Auto-Organizáveis de Kohonen



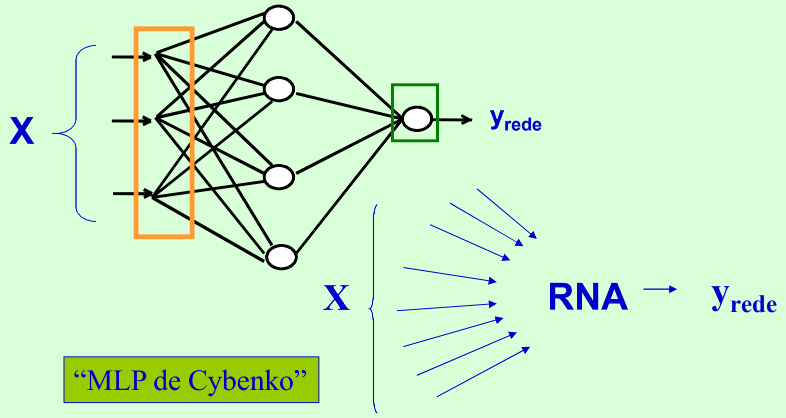
O Multi Layer Perceptron (MLP)

- Múltiplas entradas / Múltiplas saídas / Múltiplas camadas
- Variáveis (internas e externas) analógicas ou digitais
- Relações lineares ou não lineares entre elas



Kolmogorov, Cybenko (~1990)

Caso particular de uma rede de uma única camada de neurônios escondidos (e um neurônio na saída)



"MLP de Cybenko"

Prof. Emilio Del Moral Hernandez

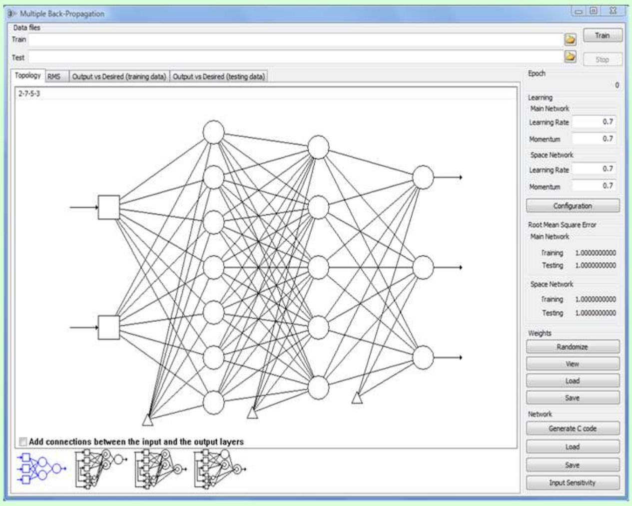
MBP – uma plataforma didática para redes neurais gratuita, de fácil uso e com 12 excelentes tutoriais

site http://mbp.sourceforge.net/

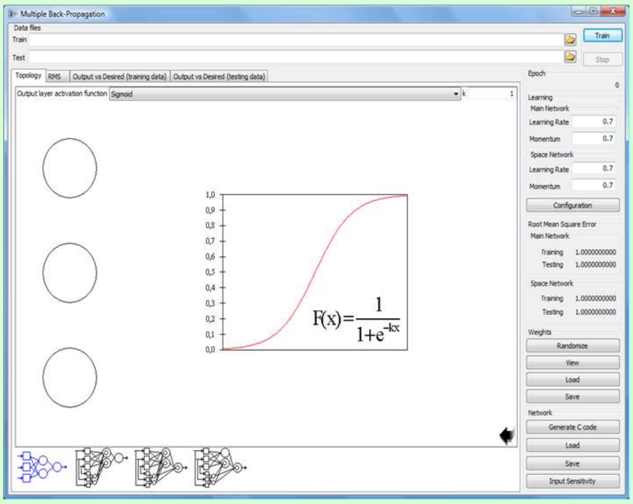
Ambiente desenvolvido pelo Prof. Noel Lopes e colaboradores – Instituto Politécnico da Guarda – Portugal

Prof. Emilio Del Moral Hernandez

Exemplo de tela do ambiente MBP definindo uma Rede Neural

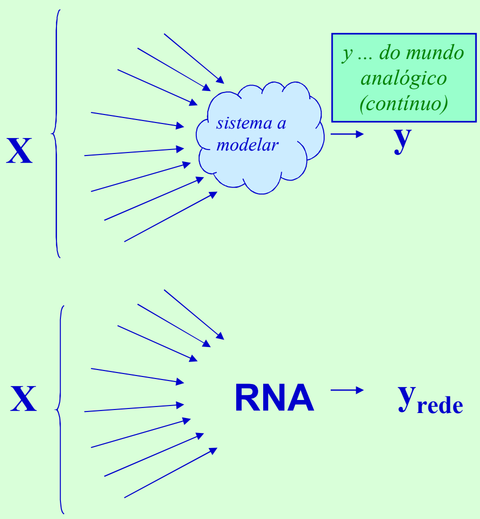


Algumas Telas do MBP Mudando a função do nó neural



Modelagem com o MLP

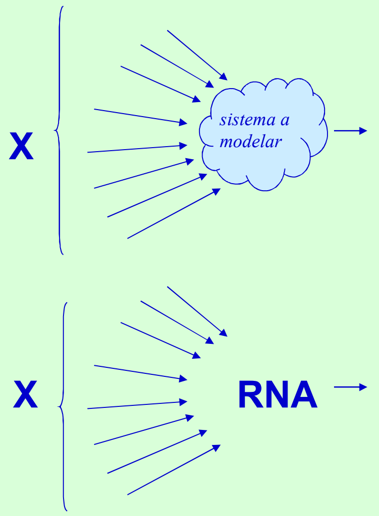
Modelagem de um sistema por função de mapeamento X -> y (a RNA como regressor contínuo não linear multivariável)



Assumimos que a variável y do sistema a modelar é uma função (normalmente desconhecida e possivelmente não linear) de diversas outras variáveis desse mesmo sistema

A RNA , para ser um bom modelo do sistema, deve reproduzir essa relação entre X e y, tão bem quanto possível

RNAs como reconhecedor de padrões ...



X é um "padrão" notável?

Sim ( ) Não ( )

y ... mundo binário (sim x não)

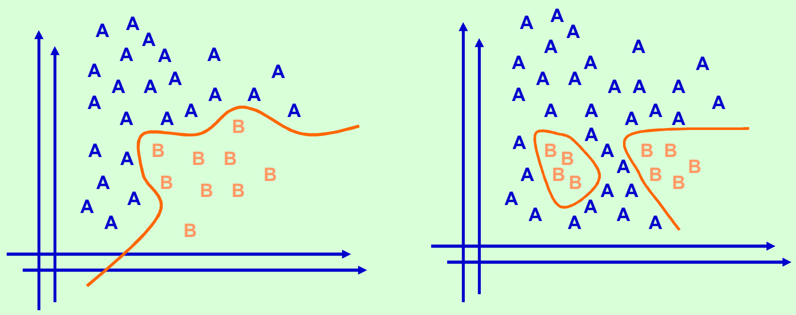
y_{rede} +1 ( ) - 1 ( )

... ou alternativamente

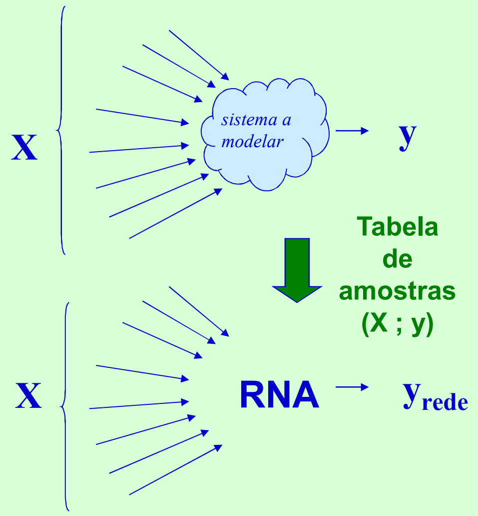
$y_{rede} > 0$ ( ) $y_{rede} < 0$ ( )

Capacidade de reconhecimento de padrões em casos complexos NÃO LINEARES

Com as RNAs, a hipersuperfície de separação entre classes vai muito além dos hiperplanos



O Conjunto de treinamento como amostragem do universo (infinito) de possibilidades (X ; y)

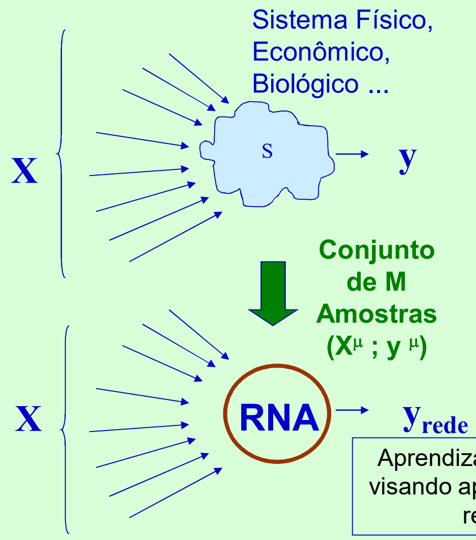


O sistema a modelar na prática não é observável em todas as infinitas situações possíveis para as entradas X.

Somente o observamos em "casos" , ou seja, amostras isoladas de X e dos correspondentes y.

Esses casos isolados, supostamente muitos e representativos, formam o conjunto de treinamento, ou seja, um "retrato amostrado" do sistema real.

Conjunto de treino em arquiteturas supervisionadas (ex. clássico: MLP com Error Back Propagation)



A computação desejada da rede pode ser definida simplesmente através de amostras / exemplos do comportamento requerido

$$Eqm = \frac{1}{M}\sum_{\mu=1}^{M}\left(y_{rede}(\vec{X}^{\mu},\vec{W}) - y^{\mu}\right)^2$$

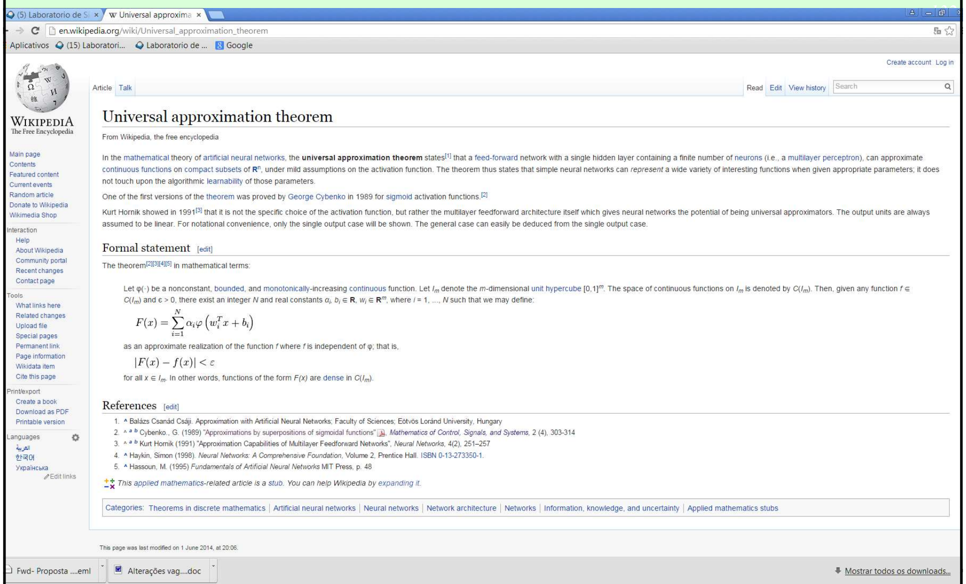
... em loop ...

$$\Delta\vec{W} = -\eta.\vec{\nabla}Eqm$$

Aprendizado: Espaço de pesos W é explorado visando aproximar ao máximo a computação da rede da computação desejada

De onde vem o grande poder do MLP?

Cybenko – Enunciado da Prova ... (premissas + resultado)



en.wikipedia.org/wiki/Universal_approximation_theorem

Universal approximation theorem

From Wikipedia, the free encyclopedia

In the mathematical theory of artificial neural networks, the **universal approximation theorem** states[1] that a feed-forward network with a single hidden layer containing a finite number of neurons (i.e., a multilayer perceptron), can approximate continuous functions on compact subsets of $\mathbf{R}^n$, under mild assumptions on the activation function. The theorem thus states that simple neural networks can *represent* a wide variety of interesting functions when given appropriate parameters; it does not touch upon the algorithmic learnability of those parameters.

One of the first versions of the theorem was proved by George Cybenko in 1989 for sigmoid activation functions.[2]

Kurt Hornik showed in 1991[3] that it is not the specific choice of the activation function, but rather the multilayer feedforward architecture itself which gives neural networks the potential of being universal approximators. The output units are always assumed to be linear. For notational convenience, only the single output case will be shown. The general case can easily be deduced from the single output case.

Formal statement [edit]

The theorem[2][3][4][5] in mathematical terms:

Let $\varphi(\cdot)$ be a nonconstant, bounded, and monotonically-increasing continuous function. Let I_m denote the m-dimensional unit hypercube $[0,1]^m$. The space of continuous functions on I_m is denoted by $C(I_m)$. Then, given any function $f \in C(I_m)$ and $\varepsilon > 0$, there exist an integer N and real constants $\alpha_i, b_i \in \mathbf{R}$, $w_i \in \mathbf{R}^m$, where $i = 1, \ldots, N$ such that we may define:

$$F(x) = \sum_{i=1}^{N} \alpha_i \varphi\left(w_i^T x + b_i\right)$$

as an approximate realization of the function f where f is independent of φ; that is,

$$|F(x) - f(x)| < \varepsilon$$

for all $x \in I_m$. In other words, functions of the form $F(x)$ are dense in $C(I_m)$.

References [edit]

1. ^ Balázs Csanád Csáji. Approximation with Artificial Neural Networks; Faculty of Sciences; Eötvös Loránd University, Hungary
2. ^ a b Cybenko., G. (1989) "Approximations by superpositions of sigmoidal functions", *Mathematics of Control, Signals, and Systems*, 2 (4), 303-314
3. ^ a b Kurt Hornik (1991) "Approximation Capabilities of Multilayer Feedforward Networks", *Neural Networks*, 4(2), 251–257
4. ^ Haykin, Simon (1998). *Neural Networks: A Comprehensive Foundation*, Volume 2, Prentice Hall. ISBN 0-13-273350-1.
5. ^ Hassoun, M. (1995) *Fundamentals of Artificial Neural Networks* MIT Press, p. 48

This applied mathematics-related article is a stub. You can help Wikipedia by expanding it.

Categories: Theorems in discrete mathematics | Artificial neural networks | Neural networks | Network architecture | Networks | Information, knowledge, and uncertainty | Applied mathematics stubs

This page was last modified on 1 June 2014, at 20:06.

Kurt Hornik showed in 1991[?] that it is not the specific choice of the a
assumed to be linear. For notational convenience, only the single out

Formal statement [edit]

The theorem[2][3][4][5] in mathematical terms:

Let φ(·) be a nonconstant, bounded, and monotonically-increas
$C(I_m)$ and ε > 0, there exist an integer N and real constants α_i,

$$F(x) = \sum_{i=1}^{N} \alpha_i \varphi \left(w_i^T x + b_i \right)$$

as an approximate realization of the function f where f is indepe

$$|F(x) - f(x)| < \varepsilon$$

for all $x \in I_m$. In other words, functions of the form $F(x)$ are den

Kurt Hornik showed in 1991[?] that it is not the specific choice of the a
assumed to be linear. For notational convenience, only the single out

Formal statement [edit]

The theorem[2][3][4][5] in mathematical terms:

$y_{rede}(X)$

X

Let φ(·) be a nonconstant, bounded, and monotonically-increas
$C(I_m)$ and ε > 0, there exist an integer N and real constants α_i,

$$F(x) = \sum_{i=1}^{N} \alpha_i \varphi \left(w_i^T x + b_i \right)$$

viés$_i$: viés do nó escondido i

sigmoidal

número de nós escondidos

W_i : vetor de pesos do nó escondido i

elementos do vetor de pesos do nó linear de saída W_s

as an approximate realization of the function f where f is indepe

$$|F(x) - f(x)| < \varepsilon$$

for all $x \in I_m$. In other words, functions of the form $F(x)$ are den

Kurt Hornik showed in 1991[4] that it is not the specific choice of the a
assumed to be linear. For notational convenience, only the single out

Formal statement [edit]

The theorem[2][3][4][5] in mathematical terms:

Let φ(·) be a nonconstant, bounded, and monotonically-increas
$C(I_m)$ and ε > 0, there exist an integer N and real constants α_i,

$$F(x)=\sum_{i=1}^{N} v_i\varphi\left(w_i^T x+b_i\right)$$

$y_{rede}(X)$

Fescondida_sistema(X)

as an approximate realization of the function f where f is indepe

$$|F(x)-f(x)|<\varepsilon$$

Limite de erro

for all $x \in I_m$. In other words, functions of the form $F(x)$ are den

---

Cybenko – a prova matemática, disponível para download na internet, é bastante complexa

Math. Control Signals Systems (1989) 2: 303–314

Mathematics of Control, Signals, and Systems

Approximation by Superpositions of a Sigmoidal Function*

G. Cybenko†

Abstract. In this paper we demonstrate that finite linear combinations of compositions of a fixed, univariate function and a set of affine functionals can uniformly approximate any continuous function of n real variables with support in the unit hypercube; only mild conditions are imposed on the univariate function. Our results settle an open question about representability in the class of single hidden layer neural networks. In particular, we show that arbitrary decision regions can be arbitrarily well approximated by continuous feedforward neural networks with only a single internal, hidden layer and any continuous sigmoidal nonlinearity. The paper discusses approximation properties of other possible types of nonlinearities that might be implemented by artificial neural networks.

Key words. Neural networks, Approximation, Completeness.

1. Introduction

A number of diverse application areas are concerned with the representation of general functions of an n-dimensional real variable, $x \in \mathbb{R}^n$, by finite linear combinations of the form

$$\sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j), \tag{1}$$

where $y_j \in \mathbb{R}^n$ and $\alpha_j, \theta \in \mathbb{R}$ are fixed. (y^T is the transpose of y so that $y^T x$ is the inner product of y and x.) Here the univariate function σ depends heavily on the context of the application. Our major concern is with so-called sigmoidal σ's:

$$\sigma(t) \to \begin{cases} 1 & \text{as } t \to +\infty, \\ 0 & \text{as } t \to -\infty. \end{cases}$$

Such functions arise naturally in neural network theory as the activation function of a neural node (or *unit* as is becoming the preferred term) [L1], [RHM]. The main result of this paper is a demonstration of the fact that sums of the form (1) are dense in the space of continuous functions on the unit cube if σ is any continuous sigmoidal

* Date received: October 21, 1988. Date revised: February 17, 1989. This research was supported in part by NSF Grant DCR-8619103, ONR Contract N000-86-G-0202 and DOE Grant DE-FG02-85ER25001.

† Center for Supercomputing Research and Development and Department of Electrical and Computer Engineering, University of Illinois, Urbana, Illinois 61801, U.S.A.

4. Results for Other Activation Functions

In this section we discuss other classes of activation functions that have approximation properties similar to the ones enjoyed by continuous sigmoidals. Since these other examples are of somewhat less practical interest, we only sketch the corresponding proofs.

There is considerable interest in discontinuous sigmoidal functions such as hard limiters ($\sigma(x) = 1$ for $x \geq 0$ and $\sigma(x) = 0$ for $x < 0$). Discontinuous sigmoidal functions are not used as often as continuous ones (because of the lack of good training algorithms) but they are of theoretical interest because of their close relationship to classical perceptrons and Gamba networks [MP].

Assume that σ is a bounded, measurable sigmoidal function. We have an analog of Theorem 2 that goes as follows:

Theorem 4. *Let σ be bounded measurable sigmoidal function. Then finite sums of the form*

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

are dense in $L^1(I_n)$. In other words, given any $f \in L^1(I_n)$ and $\varepsilon > 0$, there is a sum, $G(x)$, of the above form for which

$$\|G - f\|_{L^1} = \int_{I_n} |G(x) - f(x)|\, dx < \varepsilon.$$

The proof follows the proof of Theorems 1 and 2 with obvious changes such as replacing continuous functions by integrable functions and using the fact that $L^\infty(I_n)$ is the dual of $L^1(I_n)$. The notion of being discriminatory accordingly changes to the following: for $h \in L^\infty(I_n)$ the condition that

$$\int_{I_n} \sigma(y^T x + \theta) h(x)\, dx = 0$$

for all y and θ implies that $h(x) = 0$ almost everywhere. General sigmoidal functions are discriminatory in this sense as already seen in Lemma 1 because measures of the form $h(x)\,dx$ belong to $M(I_n)$.

Since convergence in L^1 implies convergence in measure [A], we have an analog of Theorem 3 that goes as follows:

Theorem 5. *Let σ be a general sigmoidal function. Let f be the decision function for any finite measurable partition of I_n. For any $\varepsilon > 0$, there is a finite sum of the form*

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

and a set $D \subset I_n$, so that $m(D) \geq 1 - \varepsilon$ and

$$|G(x) - f(x)| < \varepsilon \quad \text{for} \quad x \in D.$$

Questões intrigantes, p/ esta aula e p/ pensar em casa ...

- ***No que impacta escolhermos o "epslon" de Cybenko de alto valor? O que muda na estrutura de Cybenko com isso?***
- ***No que impacta escolhermos o "epslon" de Cybenko de baixo valor?***
- ***Como definimos o número de nós da primeira camada do MLP ? Isto pode ser definido a priori, antes de testar o seu desempenho? (por exemplo com base no número de entradas da rede e/ou com base no número de exemplares de treino M?)***
- ***O que ganhamos e o que perdemos se escolhermos usar POUCOS nós na construção rede neural?***
- ***O que ganhamos e o que perdemos se escolhermos usar MUITOS nós na construção da rede neural?***