

Perceptrons Morfológico de Camada Única

Marcos Eduardo R. Valle Mesquita
E-mail: mevalle@gmail.com
RA: 971216

Márcio Annibal Pimenta
E-mail: marcioapimenta@gmail.com
RA: 040799

Resumo—Neste trabalho discutimos a rede neural artificial introduzida por Ritter e Urcid conhecida como Perceptrons Morfológico de Camada Única (SLMP). Este modelo considera o papel dos dendritos no processamento das informações. Analisamos a capacidade computacional da rede SLMP e verificamos que ela resolve o problema do Ou-Exclusivo com um único dendrito. Apresentamos o algoritmo construtivo de Ritter-Urcid usado para problemas de classificação e introduzimos uma versão alternativa. Realizamos três experimentos computacionais com as redes SLMP e as redes MLP. Terminamos o artigo com uma breve conclusão.

I. INTRODUÇÃO

A neurocomputação é fortemente baseada na maneira biológica do cérebro agir. Devido a esse fato, nos estudos de computação neural, a importância das estruturas periféricas dos neurônios (dendritos, axônios, etc), vêm sendo levadas cada vez mais em consideração. Essas estruturas agora são vistas como unidades de computação primárias do neurônio, capazes de realizar operações lógicas e não somente a transferência das mesmas. Com base em recentes descobertas, Ritter e Urcid desenvolveram um novo paradigma para a computação envolvendo neurônios onde o processo é realizado no dendrito, tornando mais realista o modelo matemático de um neurônio numa rede neural artificial. Neste artigo apresentaremos o modelo neural proposto por Ritter e Urcid que é referido na literatura como *perceptrons morfológico* [8].

Na próxima seção apresentamos a motivação biológica para o modelo neural de Ritter-Urcid. Na seção III apresentamos o modelo matemático do perceptrons morfológico, introduzimos as chamadas redes SLMP. Na seção IV discutimos a capacidade computacional das redes SLMP, verificamos que uma rede SLMP com um único neurônio é capaz de resolver o problema do Ou-Exclusivo e apresentamos uma interpretação geométrica da computação realizada nesta rede neural. Na seção V apresentamos o algoritmo construtivo de Ritter-Urcid e uma versão alternativa. Estes algoritmos podem ser usados para treinar uma rede SLMP para problemas de classificação binária. Na seção VI apresentamos experimentos computacionais com as redes SLMP e as redes MLP. Terminamos o artigo com a conclusão na seção VII.

II. MOTIVAÇÃO BIOLÓGICA

A *célula nervosa* ou *neurônio* é a unidade estrutural do sistema nervoso. Um neurônio é constituído de um corpo celular, chamado *soma* e alguns processos. Esses processos são de dois tipos e são chamados *dendritos* e *axônios*. Os dendritos, que são múltiplos, conduzem impulsos de entrada

para o corpo da célula. Os dendritos geralmente possuem muitos ramos que criam longas e complicadas árvores. O axônio, que usualmente se localiza no lado oposto da célula, num ponto chamado de tronco axonal, é constituído de uma longa fibra com ramos que formam a árvore axonal. Para alguns neurônios, o axônio pode possuir ramos em intervalos ao longo de seu comprimento além da terminação arborizada. As pontas desse ramos são chamadas de terminações nervosas. O axônio é o principal ramo do neurônio como um todo ao que diz respeito da transmissão de impulsos nervosos. Entretanto, é importante notar que cada neurônio possui um único axônio. Conseqüentemente, um neurônio artificial deve produzir uma única saída.

Um impulso nervoso propaga-se através do tronco axonal do neurônio para as terminações nervosas. Essas terminações fazem contato com os muitos dendritos de outras células nervosas. As regiões de contato entre axônio e dendrito são as regiões sinápticas, pois ali ocorrem as sinápses. A *sinápsis* é a estrutura especializada em realizar as comunicações entre os neurônios. O conhecimento exato de como é realizada a sinápsis ainda é desconhecido, mas sabemos que existem dois tipos de sinápses: as *sinápses excitantes*, que tendem a despolarizar a membrana pós-sináptica facilitando a ativação final do neurônio, e as *sinápses inibidoras* que tentam prevenir a ativação final do neurônio [10]. O número de sinápses num neurônio pode chegar a 200.000 no córtex cerebral. A maioria das sinápses ocorrem nos dendritos e, conseqüentemente, os dendritos ocupam a maior área e volume de nosso cérebro. Além disso, estudos recentes mostraram que a informação transmitida é processada nos dendritos [9], [10]. Por estas razões, os dendritos não podem ser desconsiderados na neurocomputação se essa tem inspiração biológica.

III. MODELO NEURAL MORFOLÓGICO

O modelo neural artificial é constituído de três elementos [3]:

- 1) Um conjunto de pesos w_i que ponderam as entradas x_i . Este conjunto de pesos representa as sinápses do neurônio ou dendrito.
- 2) Um combinador que pode ser linear ou não linear. Este combinador representa a computação realizada no corpo celular ou no dendrito.
- 3) Uma função de ativação φ usada para restringir a amplitude da saída do neurônio. A função (sigmoidal)

logística

$$s(x) = \frac{1}{1 + e^{-x}}, \quad (1)$$

e a função degrau

$$f(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (2)$$

são exemplos comuns de função de ativação. A função de ativação representa o axônio que transmite a informação processada pelo neurônio para outras células nervosas.

Alguns modelos neurais artificiais incluem um bias θ_j aplicado externamente. Entretanto, podemos interpretar o bias como um peso sináptico conectado a uma entrada constante. Por esta razão, não acrescentaremos este termo nos nossos modelos neurais.

No modelo neural clássico ou convencional, o mapeamento de entrada e saída de um neurônio é descrito pelo par de equações:

$$\tau(\mathbf{x}) = \sum_{i=1}^n w_i x_i, \quad (3)$$

e

$$y(\mathbf{x}) = \varphi(\tau_j(\mathbf{x})), \quad (4)$$

onde $\mathbf{x} \in \mathbb{R}^n$ é a entrada, $\tau(\mathbf{x})$ é a ativação do neurônio, φ é a função de ativação e $y(\mathbf{x})$ é o sinal de saída do neurônio. Por exemplo, o *Perceptron* é um modelo neural descrito pelas Equações 3 e 4 com a função degrau dada pela Equação 2 como função de ativação. Este modelo neural pode ser usado em problemas de classificação de padrões.

No modelo neural morfológico consideramos o fato das informações serem processadas nos dendritos. Considere um neurônio com K dendritos e suponha que todos os dendritos recebem como entrada o vetor $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$. O peso da conexão sináptica da entrada x_i do dendrito k será denotado por w_{ik}^l , onde $l \in \{0, 1\}$. O índice l é usado para distinguir as sinapses excitantes das sinapses inibidoras. Em particular, w_{ik}^1 representa o fato da entrada x_i produzir uma excitação e w_{ik}^0 o fato da entrada x_i induzir uma inibição. O efeito final produzido pela entrada x_i no dendrito k será $+(w_{ik}^1 + x_i)$ para as sinapses excitantes e $-(w_{ik}^0 + x_i)$ para as sinapses inibidoras. Note que quanto maior o valor de x_i , maior será o efeito de uma sinapse excitante e menor será o efeito de uma sinapse inibidora. Simplificadamente, $(-1)^{l+1}(w_{ik}^l + x_i)$, para $l \in \{0, 1\}$, representa o efeito da conexão sináptica w_{ij}^l quando apresentamos x_i como entrada.

No modelo neural morfológico permitimos que uma entrada x_i produza ambas excitação e inibição no dendrito. Uma competição entre o efeito de todas as conexões sinápticas é realizada e somente a sinapse de menor valor é considerada como ativação do dendrito. A computação realizada no k -ésimo dendrito para o vetor de entrada \mathbf{x} é descrita pela equação

$$\tau_k(\mathbf{x}) = p_k \bigwedge_{i=1}^n \bigwedge_{l \in L} (-1)^{l+1} (x_i + w_{ik}^l), \quad (5)$$

onde $L = \{0, 1\}$ e $p_k \in \{-1, 1\}$ denota uma resposta excitante ou inibidora do k -ésimo dendrito. Note que $w_{ik}^1 = \infty$ e $w_{ik}^0 = -\infty$ representam ausência de conexão sináptica excitante ou inibidora, respectivamente.

O valor τ_k é transferido para o corpo celular do neurônio onde uma nova competição é realizada. A ativação do neurônio é tomada como sendo o menor valor entre as excitações de todos os dendritos. Matematicamente, a ativação do neurônio é dada pela equação

$$\tau(\mathbf{x}) = \bigwedge_{k=1}^K \tau_k(\mathbf{x}), \quad (6)$$

onde K representa o número de dendritos. A saída é determinada por uma função de ativação aplicada a ativação do neurônio. Neste trabalho usaremos somente a função degrau apresentada na Equação 2 como função de ativação do modelo neural morfológico. Note que neste caso, o neurônio será ativado se e somente se todos os dendritos produzirem saídas positivas, isto é, se $\tau_k(\mathbf{x}) \geq 0$ para todo $k = 1, \dots, K$.

Resumindo, a saída de um neurônio morfológico é dado por

$$y(\mathbf{x}) = f \left(\bigwedge_{k=1}^K \left[p_k \bigwedge_{i=1}^n \bigwedge_{l \in L} (-1)^{l+1} (w_{ik}^l + x_i) \right] \right). \quad (7)$$

O modelo neural morfológico com a função degrau dada pela Equação 2 será chamado *perceptrons morfológico*. Uma rede de camada única com os perceptrons morfológicos será chamada de Perceptrons Morfológico de Camada Única (SLMP, Single Layer Morphological Perceptrons).

Diferente das redes neurais tradicionais que utilizam operações de soma e multiplicação, no modelo neural morfológico utilizamos somente as operações de mínimo e soma. A operação de mínimo é uma operação não linear e portanto esse modelo neural induz operadores que tratam as transformações não lineares de outra perspectiva. As operações não lineares realizadas por um neurônio morfológico podem ser vistas como operações de *dilatação* e *erosão* da *morfologia matemática*, daí o nome *neurônio morfológico* [4], [7]. Há algumas vantagens em substituir as operações aritméticas usadas nas redes tradicionais pelas operações da morfologia matemática. Por exemplo, as operações de máximo e soma possibilitam rapidez computacional e fácil implementação em hardware [8]. Uma rede neural baseada nos neurônios morfológicos é referida como rede neural morfológica (RNM). As redes neurais morfológicas encontram aplicações em problemas de reconhecimento automático de alvos, reconhecimento de padrões escritos à mão, detecção de minas terrestres, controle de suspensão de veículos e predição de mercados financeiros [1], [2], [5], [6].

IV. CAPACIDADE COMPUTACIONAL DAS REDES SLMP

A capacidade computacional do perceptron morfológico é bem diferente do perceptron clássico. Um bom exemplo dessa distinção está no fato de um único perceptron morfológico resolver o problema do OU-Exclusivo (XOR) [3].

Dendrito	w_{1k}^1	w_{2k}^1	w_{1k}^0	w_{2k}^0	p_k
D_1	-0.5	$+\infty$	$-\infty$	-0.5	+1
D_2	$+\infty$	-0.5	-0.5	$-\infty$	+1

TABELA I

PESOS SINÁPTICOS DO PERCEPTON MORFOLÓGICO QUE RESOLVE O PROBLEMA XOR.

Exemplo 1 (Problema do OU-Exclusivo). O problema do *OU-Exclusivo* (XOR) pertence à classe de problemas que envolvem padrões não linearmente separáveis. Neste problema consideramos os quatro vértices do quadrado unitário que correspondem aos padrões de entrada $(0, 0)$; $(0, 1)$; $(1, 0)$; $(1, 1)$. Os padrões $(0, 0)$ e $(1, 1)$ estão em vértices opostos mas pertencem à classe \mathcal{C}_0 . Os padrões $(0, 1)$ e $(1, 0)$ pertencem à classe \mathcal{C}_1 . O problema XOR é um problema conhecido na literatura de redes neurais porque um único perceptron clássico não consegue resolver este problema. Entretanto, um único perceptron morfológico com dois dendritos é capaz de resolver o problema XOR com os pesos sinápticos da tabela I.

Lembramos que uma rede neural com uma camada oculta com dois neurônios (perceptrons) também é capaz de resolver o problema XOR. Esta arquitetura de rede neural é chamada *Perceptrons de Múltiplas Camadas* (MLP, Multi-Layer Perceptrons). Repare na semelhança entre a rede MLP e a SLMP que resolve o problema do XOR. Na rede MLP temos dois neurônios ocultos, no modelo morfológico temos dois dendritos. Podemos pensar que os dendritos do modelo morfológico atuam como os neurônios ocultos da rede MLP. Entretanto, nos neurônios ocultos temos a função de ativação, nos dendritos não temos nenhuma função de ativação.

Considere uma rede SLMP com um neurônio morfológico com K dendritos que recebe uma entrada $\mathbf{x} \in \mathbb{R}^n$. Se as conexões sinápticas do k -ésimo dendrito são $p_k = 1$, $\mathbf{w}_k^1 = -\boldsymbol{\alpha}_k \in \mathbb{R}^n$ e $\mathbf{w}_k^0 = -\boldsymbol{\beta}_k \in \mathbb{R}^n$ então

$$\tau_k(\mathbf{x}) = p_k \bigwedge_{i=1}^n \bigwedge_{l \in L} (-1)^{l+1} (w_{ik}^l + x_i) \geq 0 \quad (8)$$

$$\Leftrightarrow -\alpha_{ik} + x_i \geq 0 \quad \text{e} \quad \beta_{ik} - x_i \geq 0 \quad \forall i \quad (9)$$

$$\Leftrightarrow \boldsymbol{\alpha}_k \leq \mathbf{x} \leq \boldsymbol{\beta}_k. \quad (10)$$

Se $p_k = -1$, então teremos $\tau_k(\mathbf{x}) \geq 0$ se e somente se $\mathbf{x} \leq \boldsymbol{\alpha}_k$ ou $\boldsymbol{\beta}_k \leq \mathbf{x}$. Note que o conjunto dos pontos $\mathbf{x} \in \mathbb{R}^n$ tais que $\boldsymbol{\alpha}_k \leq \mathbf{x} \leq \boldsymbol{\beta}_k$ representa um paralelepípedo no \mathbb{R}^n e o conjunto dos pontos $\mathbf{x} \in \mathbb{R}^n$ tais que $\mathbf{x} \leq \boldsymbol{\alpha}_k$ ou $\boldsymbol{\beta}_k \leq \mathbf{x}$ representa o conjunto complementar do interior de um paralelepípedo.

Sabemos que um perceptron morfológico com K dendritos é ativado se

$$\tau(\mathbf{x}) = \bigwedge_{k=1}^K \tau_k(\mathbf{x}) \geq 0, \quad (11)$$

ou seja, se todos os dendritos forem ativados. Se A_k é a região de ativação do k -ésimo dendrito, a região de ativação

do perceptrons morfológico será

$$X = \bigcap_{k=1}^K A_k, \quad (12)$$

isto é, X é a interseção de A_k . Lembre-se que A_k pode ser um paralelepípedo ou o complementar do interior de um paralelepípedo. Se A_k é o complementar do interior de um paralelepípedo, digamos B_k , então quando computamos a interseção de A_k com uma região de ativação X , estamos excluindo B_k de X . Esta interpretação será usada para a elaboração do algoritmo de aprendizado da rede SLMP.

Num problema de classificação binária, queremos sintetizar uma rede neural artificial cuja região de ativação aproxima um conjunto compacto $X \subseteq \mathbb{R}^n$. No caso das redes SLMP, aproximamos X através de intersecções de paralelepípedos. O seguinte teorema apresentado por Ritter e Urcid garante que podemos aplicar uma rede SLMP em problemas de classificação de padrões binários [8]:

Teorema 1. *Qualquer conjunto compacto (fechado e limitado) em \mathbb{R}^n pode ser representado por uma rede SLMP com precisão ϵ .*

V. APRENDIZADO CONSTRUTIVO

Nesta seção apresentaremos um procedimento construtivo para o treinamento de uma rede SLMP aplicada à um problema de classificação binária. Seja $T = \{(\mathbf{x}^\xi, c_\xi) : \xi = 1, \dots, M\}$ o conjunto de treinamento onde $\mathbf{x}^\xi \in \mathbb{R}^n$ e $c_\xi \in \{0, 1\}$. Neste problema dizemos que $\mathbf{x}^\xi \in \mathcal{C}_1$ se $c_\xi = 1$ e $\mathbf{x}^\xi \in \mathcal{C}_0$ se $c_\xi = 0$. O objetivo do algoritmo de aprendizado é encontrar pesos sinápticos \mathbf{w}_k^1 , \mathbf{w}_k^0 e p_k , para $k = 1, \dots, K$ tais que a saída da rede SLMP descrita pela Equação 7 satisfaça $y(\mathbf{x}^\xi) = c_\xi$ para todo $\xi = 1 \dots, M$. Com base nos resultados apresentados na seção anterior, no treinamento da rede SLMP devemos encontrar conjuntos

$$A_k = \{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{\alpha}_k \leq \mathbf{x} \leq \boldsymbol{\beta}_k, \quad \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k \in \mathbb{R}^n\} \quad (13)$$

ou

$$A_k = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \leq \boldsymbol{\alpha}_k, \quad \boldsymbol{\beta}_k \leq \mathbf{x}, \quad \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k \in \mathbb{R}^n\}, \quad (14)$$

tais que $\mathbf{x}^\xi \in \bigcap_{k=1}^K A_k$ para todo $\xi \in \mathcal{C}_1$. Desta forma, os pesos sinápticos podem ser obtidos através dos conjuntos A_k .

Utilizamos um algoritmo construtivo para obter os conjuntos A_k dados pelas Equações 13 e 14. Do ponto de vista das redes neurais, iniciamos a rede SLMP com um único dendrito e acrescentamos novos dendritos até atingirmos nosso objetivo. Isso é feito da seguinte maneira.

Iniciamos a rede SLMP com um único dendrito que representa um paralelepípedo A_1 que inclui todos os pontos $\mathbf{x}^\xi \in \mathcal{C}_1$. O conjunto A_1 pode conter pontos $\mathbf{x} \in \mathcal{C}_0$ que serão classificados de forma incorreta pela rede SLMP. No caso positivo, escolhemos aleatoriamente um ponto $\mathbf{x} \in \mathcal{C}_0$ tal que $y(\mathbf{x}) = 1$ e adicionamos um novo dendrito que representa o complementar do interior de um paralelepípedo que contém \mathbf{x} , isto é, o novo dendrito representa um conjunto A_k dado pela equação 14. Este novo dendrito impede a ativação da

rede SLMP após a apresentação do padrão $\mathbf{x} \in \mathcal{C}_0$. Os pesos sinápticos do novo dendrito são definidos através das equações

$$p_k = -1, \quad \mathbf{w}_k^1 = -\alpha_k, \quad \mathbf{w}_k^0 = \beta_k, \quad (15)$$

onde α_k e β_k são os vetores que definem o conjunto A_k dado pela Equação 14. O processo é repetido adicionando novos dendritos até que todos os pontos sejam classificados corretamente pela rede SLMP. Na figura 1 apresentamos a região de ativação da rede SLMP com 9 dendritos treinada com o aprendizado construtivo de Ritter e Urcid. Esta rede classificou corretamente todos os padrões do conjunto de treinamento. Todavia, existem pontos isolados da classe \mathcal{C}_1 .

O teorema 1 garante que o processo de treinamento termina com um número dendritos. Entretanto, podemos terminar o processo antes para evitar um aprendizado excessivo acarretando na perda da capacidade de generalização da rede neural. Na prática, usaremos um conjunto de validação para definir o número de dendritos da rede SLMP. A rede SLMP adotada será aquela que produz o menor erro de validação.

Para concluir o algoritmo construtivo de Ritter-Urcid, precisamos especificar como é feita a escolha dos conjuntos A_k dados pelas equações 13 e 14, para $k = 1, \dots, K$. Na proposta original escolhemos A_1 como sendo o menor paralelepípedo que inclui todos os pontos da classe \mathcal{C}_1 . Este paralelepípedo pode ser encontrado tomando

$$p_1 = 1, \quad \mathbf{w}_1^1 = - \bigwedge_{\mathbf{x} \in \mathcal{C}_1} \mathbf{x}, \quad \mathbf{w}_1^0 = - \bigvee_{\mathbf{x} \in \mathcal{C}_1} \mathbf{x}. \quad (16)$$

Para $k \geq 2$, A_k representa o complementar do interior de um paralelepípedo $B_k = \{\mathbf{x} \in \mathbb{R}^n : \alpha_k \leq \mathbf{x} \leq \beta_k, \alpha_k, \beta_k \in \mathbb{R}^n\}$. Nesta proposta, escolhemos o maior paralelepípedo B_k que inclui o ponto $\mathbf{x} \in \mathcal{C}_0$ que está sendo classificado de maneira incorreta e que não contém nenhum ponto da classe \mathcal{C}_1 . Os pesos sinápticos do k -ésimo dendrito que representa o conjunto A_k para $k \geq 2$ são definidos da seguinte maneira:

- 1) Seja $\mathbf{x}^\gamma \in \mathcal{C}_0$ tal que $y(\mathbf{x}^\gamma) = 1$.
- 2) Tome $D = \{\mathbf{x}^\xi : \mathbf{x}^\xi \in \mathcal{C}_1\}$.
- 3) Enquanto $D \neq \emptyset$ faça:
 - a) Encontre $\mathbf{x}^\xi \in D$ que minimiza $\|\mathbf{x}^\gamma - \mathbf{x}^\xi\|_\infty$.
 - b) Encontre a coordena i tal que $|x_i^\gamma - x_i^\xi| = \|\mathbf{x}^\gamma - \mathbf{x}^\xi\|_\infty$.
 - c) Defina:

$$w_{ik}^1 = -x_i^\xi \quad \text{se} \quad x_i^\xi < x_i^\gamma, \quad (17)$$

$$w_{ik}^0 = -x_i^\xi \quad \text{se} \quad x_i^\xi > x_i^\gamma. \quad (18)$$

- d) Atualize D seguindo a equação:

$$D = \{\mathbf{x}^\xi \in D : -w_{ik}^1 < x_i^\xi \quad \text{e} \quad x_i^\xi < -w_{ik}^0\}.$$

Note que no passo 3b) tomamos $w_{ik}^1 = -x_i^\xi$ ou $w_{ik}^0 = -x_i^\xi$. Isso significa que $\mathbf{x} \in D$ pertence à borda do paralelepípedo B_k que está sendo construído para excluir o ponto \mathbf{x}^γ da região de ativação da rede SLMP. Como foi dito anteriormente, B_k é o maior paralelepípedo que contém $\mathbf{x}^\gamma \in \mathcal{C}_0$. Neste algoritmo, o paralelepípedo B_k é escolhido de modo a diminuir o número de dendritos da rede SLPM visto que B_k representa o conjunto

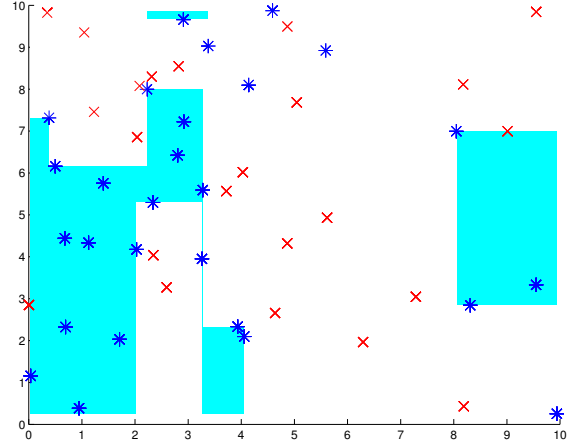


Fig. 1. Região de ativação da rede SLMP com nove dendritos treinada usando o aprendizado construtivo de Ritter e Urcid.

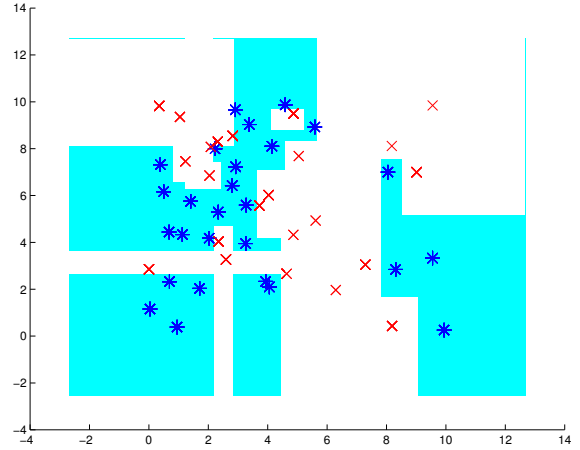


Fig. 2. Região de ativação da rede SLMP com dezesseis dendritos treinada usando o aprendizado construtivo alternativo.

com maior probabilidade de conter outros pontos $\mathbf{x} \in \mathcal{C}_0$ que estão sendo classificados incorretamente. Entretanto, impomos que $\mathbf{x}^\xi \in \mathcal{C}_1$ pertença a fronteira de decisão entre as classes \mathcal{C}_1 e \mathcal{C}_2 . Este fato pode produzir um efeito negativo no que diz respeito à capacidade de generalização da rede SLMP. Por esta razão, apresentamos a seguinte alternativa para o algoritmo de aprendizado construtivo de Ritter-Urcid.

A. Aprendizado Construtivo Alternativo

Primeiramente, não impomos que A_1 seja o menor paralelepípedo que contém todos os pontos da classe \mathcal{C}_1 . Nesta versão do algoritmo tomamos $p_1 = 1$,

$$\mathbf{w}_1^1 = - \left(\bigwedge_{\mathbf{x} \in \mathcal{C}_1} \mathbf{x} - \boldsymbol{\mu} \right), \quad \mathbf{w}_1^0 = - \left(\bigvee_{\mathbf{x} \in \mathcal{C}_1} \mathbf{x} + \boldsymbol{\mu} \right), \quad (19)$$

onde $\boldsymbol{\mu}$ é o desvio padrão dos padrões $\mathbf{x} \in \mathcal{C}_1$.

Seguimos a seguinte proposta para o passo 3b) do algoritmo de aprendizado construtivo para $k \geq 2$:

3b') Defina:

$$w_{ik}^1 = -(x_i^\xi + x_i^\gamma)/2 \quad \text{se } x_i^\xi < x_i^\gamma, \quad (20)$$

$$w_{ik}^0 = -(x_i^\xi + x_i^\gamma)/2 \quad \text{se } x_i^\xi > x_i^\gamma. \quad (21)$$

Na figura 2 apresentamos um exemplo da região de ativação de uma rede SLMP com 16 dendritos e treinada com o aprendizado construtivo alternativo. Note que a rede SLMP treinada usando o aprendizado construtivo alternativo requer um número maior de dendritos. Foram necessários 16 dendritos para classificar corretamente todos os padrões do conjunto de treinamento. Entretanto não temos pontos isolados como aconteceu após o treinamento com o aprendizado de Ritter-Urcid. O teorema 1 continua válido e o treinamento da rede pelo aprendizado construtivo alternativo termina após adicionarmos um número finito de dendritos. Novamente podemos terminar o processo antes para evitar um aprendizado excessivo. O critério de parada pode ser baseado num conjunto de validação.

VI. EXPERIMENTOS COMPUTACIONAIS

Nesta seção apresentaremos três problemas de classificação que foram resolvidos pela rede SLMP e pela rede MLP.

Neste trabalho consideramos somente redes MLP com uma única camada oculta com n neurônios, todos com a função tangente hiperbólico como função de ativação. A camada de saída é constituída por um único neurônio com a função logística como função de ativação. Foi aplicado um limiar de nível 0.5 na saída da rede MLP para obter uma resposta binária. Deste modo, um padrão é classificado como sendo da classe \mathcal{C}_1 se a resposta da rede for maior ou igual a 0.5. Utilizamos o algoritmo de retropropagação com o método de otimização de Levenberg-Marquardt e os parâmetros padrões da toolbox de redes neurais artificiais do software MATLAB. O critério de parada foi baseado no erro de validação.

A. Experimento com Duas Espirais

O primeiro experimento consiste na classificação dos padrões em espirais como apresentados nas figuras 3 e 4. Neste problema temos um conjunto de treinamento com 472 amostras, um conjunto de validação com 236 amostras e um conjunto de teste com 236 amostras. Os conjuntos de treinamento, validação e teste não possuem nenhum elemento em comum.

Na figura 5 apresentamos gráfico da porcentagem de padrões classificados incorretamente pela rede SLMP treinada com o algoritmo de Ritter-Urcid. Note que o número de padrões classificados incorretamente foi para zero no conjunto de treinamento. A menor porcentagem de padrões classificados incorretamente no conjunto de validação foi 7.2% atingida com 8 dendritos. A rede com 8 dendritos classificou incorretamente 6.4% dos padrões do conjunto de teste. Estes resultados estão resumidos na tabela II. Na figura 3 apresentamos a região de ativação da rede SLMP com 8 dendritos e o conjunto de teste. Note que a rede com 9 dendritos produziu um erro de validação maior.

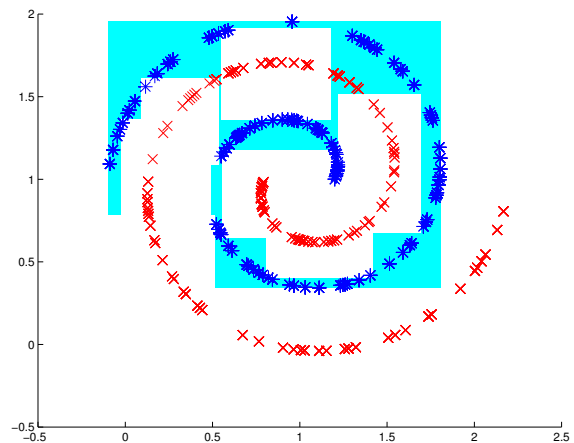


Fig. 3. Região de ativação da rede SLMP com 8 dendritos treinada com o aprendizado construtivo de Ritter-Urcid e os dados do conjunto de teste.

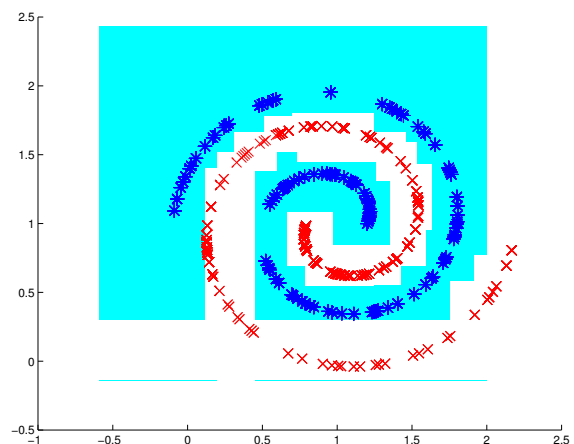


Fig. 4. Região de ativação da rede SLMP com 18 dendritos treinada com o aprendizado construtivo alternativo e os dados do conjunto de teste.

Na figura 6 apresentamos gráfico da porcentagem de padrões classificados incorretamente pela rede SLMP treinada com o aprendizado alternativo. Note que o número de padrões classificados incorretamente foi para zero no conjunto de treinamento, validação e teste. Entretanto, foram necessários 18 dendritos, mais que o dobro do número de dendritos obtidos pelo aprendizado de Ritter-Urcid. Os resultados obtidos pela rede SLMP com 18 dendritos estão apresentados na tabela II. Na figura 4 temos a região de ativação da rede SLMP com 18 dendritos e o conjunto de teste. Note que todos os padrões do conjunto de teste foram classificados corretamente.

Na figura 7 apresentamos o erro quadrático médio da rede MLP com 8 neurônios na camada oculta. Na tabela II apresentamos a porcentagem de padrões classificados incorretamente do conjunto de treinamento, validação e teste.

Neste experimento, a rede SLMP treinada com algoritmo construtivo alternativo produziu o melhor resultado. O segundo melhor resultado foi obtido pela rede MLP.

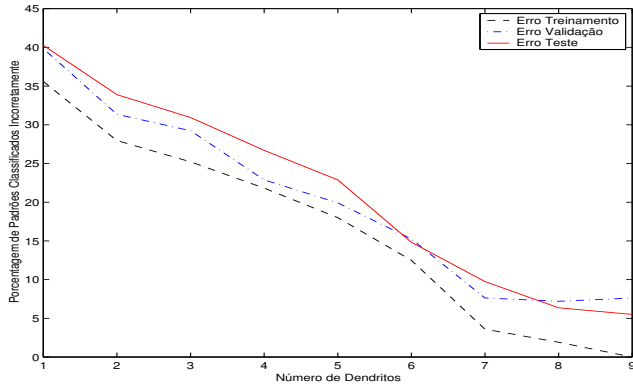


Fig. 5. Porcentagem de padrões classificados incorretamente pelo número de dendritos da rede SLMP treinada com o algoritmo construtivo de Ritter e Urcid no problema das duas espirais.

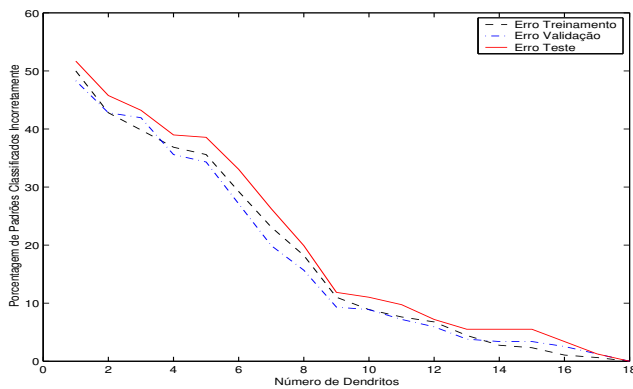


Fig. 6. Porcentagem de padrões classificados incorretamente pelo número de dendritos da rede SLMP treinada com o algoritmo construtivo alternativo no problema das espirais.

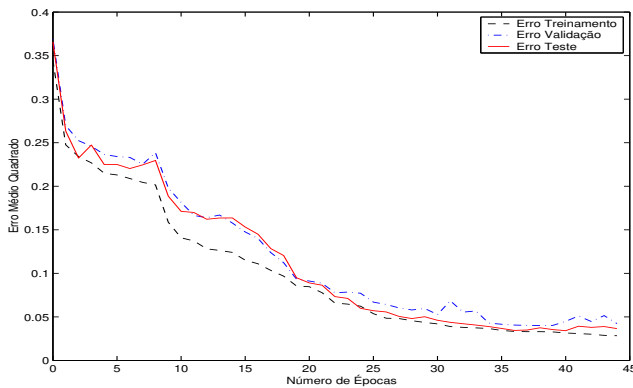


Fig. 7. Erro quadrático médio pelo número de épocas da rede MLP com 8 neurônios ocultos treinada pelo algoritmo de retropropagação no problema das duas espirais.

B. Experimento com as Duas Gaussianas

Nesta seção consideramos o experimento computacional apresentado em [3]. Neste experimento devemos classificar duas classes de padrões bidimensionais com distribuição gaussiana. Este é um problema de difícil resolução visto que existe uma sobreposição das gaussianas. O conjunto de treinamento, validação e teste considerados contém 500, 250 e 250 amostras, respectivamente.

A rede SLMP treinada com o aprendizado construtivo de Ritter e Urcid classificou corretamente todos os padrões do conjunto de treinamento com 55 dendritos. Entretanto, devido ao erro do conjunto de validação, consideramos uma rede com 31 dendritos. Os resultados obtidos por esta rede estão apresentados na tabela II. Na figura 8 apresentamos a porcentagem de padrões classificados incorretamente pelo número de dendritos da rede. Note que o critério de parada baseado no conjunto de validação foi essencial neste problema.

A rede SLMP treinada com o aprendizado construtivo alternativo classificou corretamente todos os padrões do conjunto de teste com 90 dendritos. Foram necessários muito mais dendritos que o aprendizado de Ritter e Urcid. Entretanto, o número de padrões classificados incorretamente no conjunto de teste foi menor. Na figura 9 apresentamos a porcentagem de padrões classificados incorretamente pelo número de dendritos. O menor erro de validação foi obtido pela rede com 42 dendritos. Na tabela II apresentamos a porcentagem de padrões classificados incorretamente nos conjuntos de treinamento, validação e teste.

Na figura 10 apresentamos o erro quadrático médio da rede MLP com dois neurônios na camada oculta. Na tabela II apresentamos os resultados obtidos pela rede.

A rede MLP apresentou o melhor resultado para este problema. O segundo melhor resultado foi obtido pela rede SLMP treinada com o algoritmo alternativo. Note que ambas as redes SLMP apresentaram um erro de treinamento menor que a rede MLP. Isso mostra que a rede SLMP possui uma capacidade de generalização inferior a capacidade de generalização da rede MLP.

C. Diagnóstico de Pacientes com Diabetes

O último experimento consiste no diagnóstico de pacientes que apresentam sinais de diabetes de acordo com o critério imposto pela Organização Mundial de Saúde (World Health Organization) [11]. O banco de dados consiste de 768 amostras que foram divididas em três conjuntos: Conjunto de treinamento com 384 amostras, conjunto de validação com 192 amostras e conjunto de teste com 192 amostras.

A rede SLMP treinada com o algoritmo de Ritter-Urcid com critério de parada baseado no conjunto de validação classificou corretamente 68.22% do conjunto de teste. Na figura 11 apresentamos o gráfico da porcentagem de padrões classificados incorretamente pelo número de dendritos da rede. A rede SLMP com o aprendizado alternativo classificou corretamente 70.83% e a rede MLP com 5 neurônios na camada oculta classificou corretamente 78.81% dos padrões do conjunto de teste. Nas figuras 12 e 13 apresentamos a porcentagem de

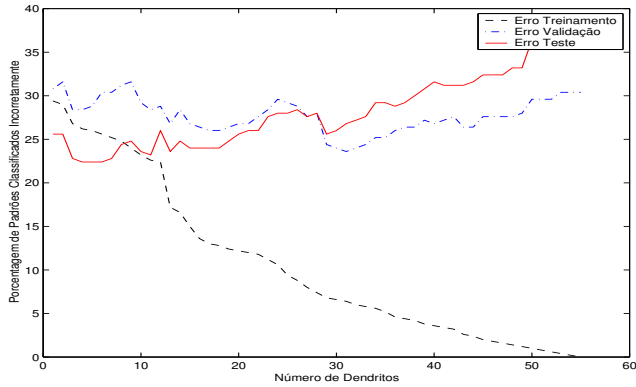


Fig. 8. Porcentagem de padrões classificados incorretamente pelo número de dendritos da rede SLMP treinada com o algoritmo construtivo de Ritter e Urcid no problema das duas gaussianas.

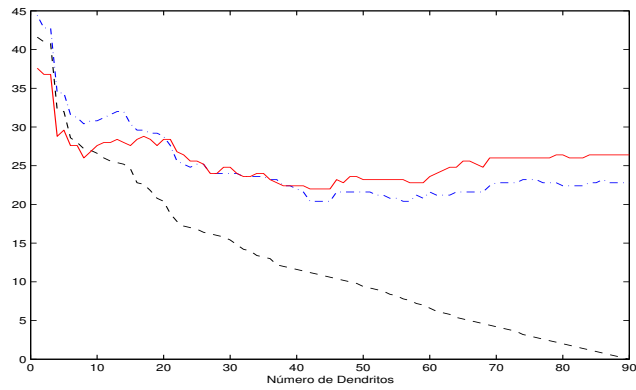


Fig. 9. Porcentagem de padrões classificados incorretamente pelo número de dendritos da rede SLMP treinada com o algoritmo construtivo alternativo no problema das duas gaussianas.

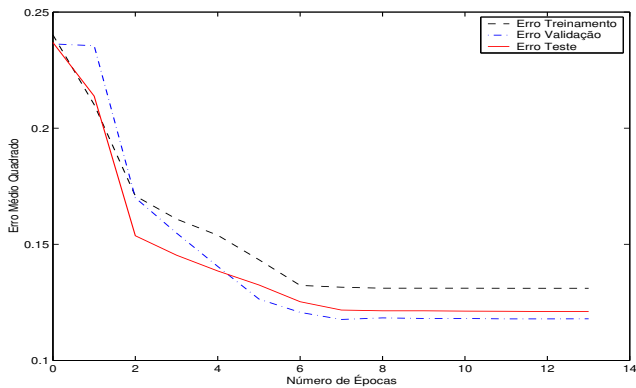


Fig. 10. Erro quadrático médio pelo número de épocas da rede MLP com 2 neurônios ocultos treinada pelo algoritmo de retropropagação no problema das gaussianas.

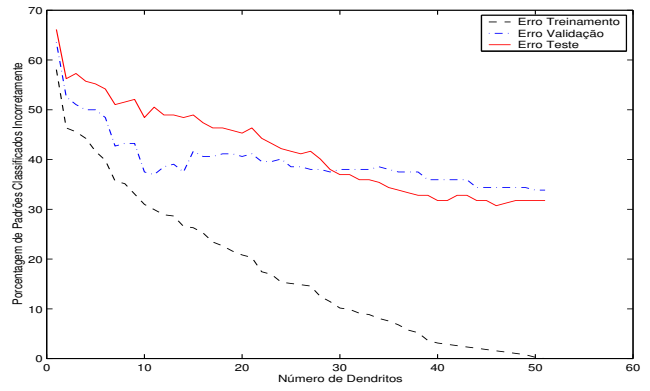


Fig. 11. Porcentagem de padrões classificados incorretamente pelo número de dendritos da rede SLMP treinada com o algoritmo construtivo de Ritter e Urcid no problema de diagnóstico de pacientes com diabetes.

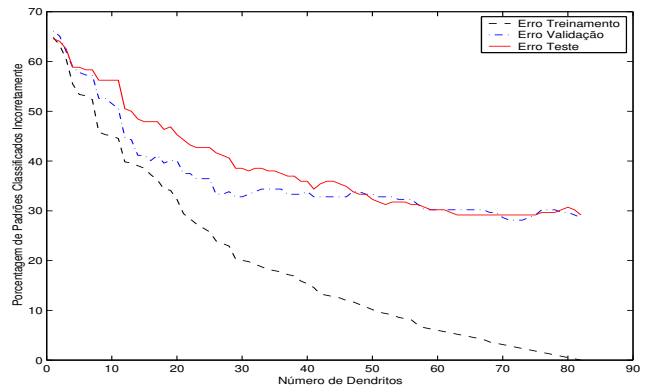


Fig. 12. Porcentagem de padrões classificados incorretamente pelo número de dendritos da rede SLMP treinada com o algoritmo construtivo alternativo no problema de diagnóstico de pacientes com diabetes.

padrões classificados incorretamente pelo número de dendritos e pelo número de épocas do algoritmo de retropropagação para a rede SLMP e MLP, respectivamente. Os resultados foram apresentados na tabela II

Os melhores resultados foram obtidos pelas redes MLP e SLMP treinada com o algoritmo alternativo, respectivamente. Novamente verificamos o fato da redes SLMP não possuírem uma boa capacidade de generalização pois o erro de treinamento é muito menor que o erro de validação e teste.

VII. CONCLUSÃO

Neste trabalho discutimos as redes SLMP que consideram o fato biológico da informação ser processada nos dendritos. Precisamente apresentamos o modelo matemático e analisamos a capacidade computacional das redes SLMP apresentando uma interpretação geométrica da região de ativação dos neurônios da rede. Verificamos também que uma rede SLMP com um único dendrito resolve o problema do XOR. Revisamos o algoritmo construtivo introduzido por Ritter-Urcid para problemas de classificação binária e apresentamos uma

Conjunto de Dados	Rede Utilizada	D/N	Conj. Treinamento	Conj. Validação	Conj. Teste
Espirais	SLMP Ritter-Urcid	8	1.9 %	7.2 %	6.4 %
Espirais	SLMP Alternativo	18	0 %	0 %	0 %
Espirais	MLP	8	3.8 %	5.5 %	5.9 %
Duas Gaussianas	SLMP Ritter-Urcid	31	6.4 %	23.6 %	26.8 %
Duas Gaussianas	SLMP Alternativo	90	11.2 %	20.4 %	22 %
Duas Gaussianas	MLP	2	22.5 %	22 %	19 %
Diabetes	SLMP Ritter-Urcid	50	0.3 %	33.9 %	31.8 %
Diabetes	SLMP Alternativo	71	2.9 %	28.1 %	29.2 %
Diabetes	MLP	5	16.3 %	17.8 %	21.2 %

TABELA II

TABELA COMPARATIVA CONTENDO A PORCENTAGEM DE PADRÕES CLASSIFICADOS INCORRETAMENTE. A COLUNA D/N REFERE-SE AO NÚMERO DE DENDRITOS DA REDE SLMP OU O NÚMERO DE NEURÔNIOS NA CAMADA OCULTA DA REDE MLP.

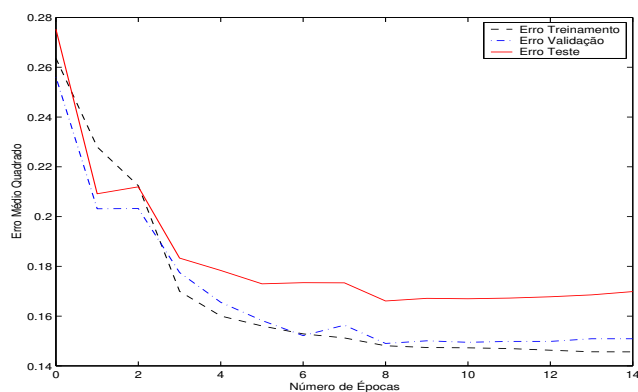


Fig. 13. Porcentagem de padrões classificados incorretamente pelo número de épocas do algoritmo de retropropagação no problema de diagnóstico de pacientes com diabetes.

versão alternativa que tende a não deixar os padrões de treinamento na fronteira de decisão entre as classes. Finalmente realizamos vários testes comparando as redes SLMP com as redes MLP. Primeiramente verificamos que o aprendizado alternativo requer um número maior de dendritos, entretanto a rede treinada com este algoritmo tende a produzir melhores resultados.

Com base nos resultados obtidos, não podemos informar qual das duas arquiteturas, redes SLMP ou redes MLP, é a melhor para problemas de classificação. Verificamos que a rede SLMP apresentam uma capacidade de generalização menor que rede MLP. Todavia, nas redes MLP precisamos especificar o número de neurônios, as funções de ativação dos neurônios e outros parâmetros para o algoritmo de treinamento. Por outro lado, as redes SLMP não requerem nenhum conhecimento prévio sobre o problema nem a especificação de parâmetros.

Em resumo, as redes SLMP apresentam uma nova forma de abordar problemas de classificação de padrões. Sua deficiência na capacidade de generalização talvez possa ser superada usando ensemble de redes neurais contendo redes SLMP e outras

arquiteturas como MLP, SVM, etc. A teoria de redes neurais morfológicas é recente e aguarda novas descobertas. Por exemplo, o algoritmo de aprendizado pode ser aperfeiçoado considerando melhores escolhas para os paralelepípedos ou aplicando transformações no conjunto de dados de modo a obter regiões que não são mais paralelepípedos.

REFERÊNCIAS

- [1] ARMSTRONG, W., AND THOMAS, M. Adaptive logic networks. In *Handbook of Neural Computation*, E. Fiesler and R. Beale, Eds. IOP Publishing and Oxford University Press, 1997, pp. C1.8:1–14.
- [2] GADER, P. D., KHABOU, M., AND KOLDOBSKY, A. Morphological regularization neural networks. *Pattern Recognition, Special Issue on Mathematical Morphology and Its Applications* 33, 6 (June 2000), 935–945.
- [3] HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ, 1999.
- [4] HEIJMANS, H. *Morphological Image Operators*. Academic Press, New York, NY, 1994.
- [5] KHABOU, M., AND GADER, P. Automatic target detection using entropy optimized shared-weight neural networks. *IEEE Transactions on Neural Networks* 11, 1 (Jan. 2000), 186–193.
- [6] PESSOA, L., AND MARAGOS, P. Neural networks with hybrid morphological/rank/linear nodes: a unifying framework with applications to handwritten character recognition. *Pattern Recognition* 33 (2000), 945–960.
- [7] RITTER, G. X., SUSSNER, P., AND DE LEON, J. L. D. Morphological associative memories. *IEEE Transactions on Neural Networks* 9, 2 (1998), 281–293.
- [8] RITTER, G. X., AND URCID, G. Lattice algebra approach to single-neuron computation. *IEEE Transactions on Neural Networks* 14, 2 (March 2003), 282–295.
- [9] SEGEV, I. Dendritic processing. In *The Handbook of Brain Theory and Neural Networks*, M. Arbib, Ed. Cambridge, MA: MIT Press, 1998, pp. 282–289.
- [10] SEJNWSKI, T., AND QIAN, N. Synaptic integration by electro-diffusion in dendritic spines. In *Single Neuron Computation*, T. McKenna, J. Davis, and S. Zornetzer, Eds. New York: Academic, 1992, pp. 117–139.
- [11] Uci machine learning - pima indians diabetes database (<ftp://ftp.ics.uci.edu/pub/machine-learning-databases/pima-indians-diabetes/>).