

Avaliação Estatística de Métodos de Poda Aplicados em Neurônios Intermediários da Rede Neural MLP

Miriam Rodrigues Silvestre, Lee Luan Ling

Resumo—Existem diversos artigos sobre métodos de poda na área de redes neurais artificiais. Entretanto, com raras exceções, nenhum deles apresenta uma avaliação estatística adequada de tais métodos. Neste artigo, comprovamos estatisticamente a habilidade de alguns métodos para reduzir o número de neurônios da camada intermediária de uma rede neural de perceptron multicamadas (MLP), e manter o mesmo patamar de erro de classificação da rede inicial. São avaliados sete métodos de poda. A investigação experimental foi realizada sobre cinco conjuntos de dados gerados e em dois conjuntos de dados reais. Foram acompanhadas três variáveis no estudo: razão de erro aparente de classificação do conjunto de teste (REA); número de neurônios intermediários, obtido após a aplicação do método de poda; e número de épocas de treinamento/re-treinamento, para avaliar o esforço computacional. Para a análise estatística, foi utilizado o teste não-paramétrico de Friedman.

Palavras-chave—método de poda iterativo, rede neural de perceptron multicamadas, rede neural MLP, reconhecimento de padrão, teste de Friedman.

I. INTRODUÇÃO

EM [1] é feito um levantamento dos principais métodos de poda até 1993, dentre os quais existe um grupo de métodos que calcula a medida de sensibilidade de cada unidade de poda (peso sináptico ou neurônio) e retira as unidades que apresentarem valores baixos para esta medida [2], [3], etc.). Posteriormente, surgiram outros métodos na área de redes neurais artificiais: [4], [5], [6], [7] e [8]. O objetivo principal dos métodos de poda é reduzir a arquitetura da rede de forma a realizar a classificação com rapidez e economia. Neste artigo, propomos aplicar os métodos de poda para melhorar a fronteira de decisão prática, através da exclusão dos hiperplanos que não contribuem para a sua formação. Quando os métodos de poda são construídos para remover os pesos sinápticos, o resultado final é a produção de uma rede grande e esparsamente conectada, ou seja, com vários pesos nulos. Entretanto, fisicamente estes pesos continuam na rede. Para que isto não ocorra, os sete métodos de poda apresentados na seção II estão direcionados teoricamente para realizar poda de

neurônios, e serão capazes de produzir redes finais completamente conectadas, pois os neurônios que contribuem pouco ou nada para a classificação serão fisicamente excluídos da rede (ver seção III-A). Desta maneira, a rede final será bem menor e a classificação de um novo padrão será bem mais rápida.

Em [9], foi realizado um experimento comparativo envolvendo o método chamado CONTROLE, rede inicial na qual não é aplicada nenhuma técnica de poda, e vários métodos de poda:

- **Iterativos**, os quais retiram um neurônio em cada passo: Força Bruta Iterativo (FBI, [9]), Razão de Erro Aparente de Validação (REAV, [9]), Razão de Erro Aparente de Treinamento (REAT, [9]), Poda Baseado na Fronteira de Decisão Efetiva de Bayes (PEB, [9]), Saliências (SAL, [4]), Optimal Brain Damage (OBD, [3]), Poda Iterativo (PI, [7]), Poda Dinâmico (PD, [6]);
- **Não iterativos**, os quais são executados num único passo: São eles: Principal Component Analysis (ACP, [5]), Gap das Saliências (GAPSAL, [8]), Força Bruta num Único Passo (FBUP, [9]).

Os métodos foram adaptados teoricamente para a retirada de neurônios intermediários e avaliados em 7 problemas de classificação já conhecidos e obtidos em [10], são eles: Ecoli, Pima Indians, Glass, Iris, Monk3, Wine e Bupa. As variáveis estudadas no experimento foram: razão de erro de classificação no conjunto de teste (REA), número de neurônios intermediários na rede final (#neurônios) e número de épocas de treinamento/re-treinamento da rede (#épocas). Os resultados obtidos indicaram que os métodos de poda apresentam o mesmo desempenho para a variável REA que a rede CONTROLE e, embora necessitem de um número maior de épocas de treinamento/re-treinamento, somente os procedimentos de poda **iterativos** são capazes de reduzir o número de neurônios intermediários.

Neste artigo são apresentados novos resultados experimentais envolvendo somente os métodos de poda **iterativos**. O objetivo principal é avaliar a capacidade dos métodos em diferentes estruturas de dados (ver seção III-B) provenientes de distribuições normais, bem como em problemas com dados reais (ver seção III-C e D).

M. R. Silvestre pertence ao Departamento de Matemática, Estatística e Computação da Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista (UNESP), Presidente Prudente, São Paulo, CEP 19060-900 BRASIL (e-mail: miriam@fct.unesp.br) L. L. Ling pertence ao Departamento de Comunicações da Faculdade de Engenharia Elétrica e Computação da Universidade Estadual de Campinas (UNICAMP), CEP 13083-970 BRASIL (e-mail: lee@decom.fee.unicamp.br).

II. MÉTODOS DE PODA ITERATIVOS

Considere uma rede MLP, contendo uma camada intermediária. Denotamos o n -ésimo ($n=1, \dots, N$) par de exemplos de treinamento como $[\mathbf{x}(n), \mathbf{t}(n)]$, tal que o vetor de entrada $\mathbf{x}(n) = [x_1(n), \dots, x_d(n)]^T$ é aplicado aos neurônios da camada de entrada e o vetor de respostas desejadas $\mathbf{t}(n) = [t_1(n), \dots, t_c(n)]^T$ é aplicado aos neurônios da camada de saída, onde d representa o número de variáveis de entrada e c o número de classes do problema. Seja $\mathbf{W}^{(1)} = \{w_{ji}^{(1)}(n)\}$ a matriz de pesos sinápticos que conecta a camada de entrada e intermediária, e $\mathbf{W}^{(2)} = \{w_{kj}^{(2)}(n)\}$, a matriz de pesos sinápticos que conecta a camada intermediária e de saída. O procedimento de otimização para classificação é minimizar a soma dos erros quadráticos médios entre a saída da rede $y_k(n)$ e a resposta desejada $t_k(n)$, definida por:

$$E = \sum_{n=1}^N E(n), \text{ onde } E(n) = \frac{1}{2} \sum_{k=1}^c [y_k(n) - t_k(n)]^2. \quad (1)$$

Adotando $g^{(1)}$ e $g^{(2)}$, funções de ativação do tipo sigmóide logística, denota-se a k -ésima saída da rede por:

$$y_k(n) = g^{(2)} \left(\sum_{j=0}^M w_{kj}^{(2)}(n) g^{(1)} \left(\sum_{i=0}^d w_{ji}^{(1)}(n) x_i(n) \right) \right). \quad (2)$$

Vários métodos de poda descritos nesta seção irão utilizar a saída do j -ésimo neurônio da camada intermediária ($j=1, \dots, M$), definido como:

$$z_j(n) = g^{(1)} \left(\sum_{i=0}^d w_{ji}^{(1)}(n) x_i(n) \right). \quad (3)$$

Note que, todos os métodos descritos abaixo, necessitam de um retreinamento após cada passo de poda. Os métodos FB e OBD foram construídos para executar poda de pesos sinápticos, porém, serão apresentados com as modificações necessárias para que se faça poda de neurônios intermediários. Uma breve discussão de cada um destes métodos é dada a seguir.

O método denominado Força Bruta (FB), [1], [2] e [3], tem como objetivo detectar e anular os pesos sinápticos que provocam pouca alteração na soma dos erros quadráticos médios (1). Este método foi repensado em [9] para detectar e remover neurônios pouco influentes e consiste em definir um neurônio com valor de saída zero e avaliar a alteração provocada no erro E , utilizando a medida $S_j = \sum_{j=1}^N S_j(n)$, tal que:

$$S_j(n) = E(z_j(n)=0) - E(z_j(n)). \quad (4)$$

Quando definimos $E(z_j(n)=0)$, todos os pesos sinápticos que conectam o neurônio j às camadas de entrada e saída são zerados, e um valor aproximado para E é obtido para uma rede sem a presença daquele neurônio. Já $E(z_j(n))$ será o valor de E com todos os pesos da rede associados a este neurônio. Se S_j for pequeno, então o neurônio j pode ser removido da rede.

Teoricamente, isto pode ser realizado para todos os possíveis subconjuntos de neurônios intermediários da rede, é o chamado procedimento ótimo de Busca Exaustiva (BE), onde todas as possíveis combinações (2^M-1) dos neurônios que restarem na rede devem ser avaliadas. De acordo com [2] e [3] este procedimento é considerado ideal, porém, é computacionalmente impraticável para redes grandes. Então, foi introduzido em [9] um melhoramento no procedimento de busca, o qual será descrito a seguir.

1) Força Bruta Iterativo (FBI): no método FB descrito acima, ao invés de utilizar BE, sugerimos adotar o procedimento sub-ótimo de busca para trás (BPT), o qual é capaz de reduzir o número de diferentes redes que deveriam ser investigadas de 2^M-1 para $M(M+1)/2$. A idéia principal é iniciar com uma rede grande e calcular S_j somente para os neurônios que restarem na rede em cada passo de poda. Por exemplo, para $M=10$ neurônios intermediários, seriam 1023 redes a serem investigadas com BE, utilizada no método FB, contra somente 55 redes com o procedimento BPT, para o método FBI proposto.

Alguns outros métodos existentes na literatura foram modificados quanto à parte a ser removida, e todos foram implementados utilizando o procedimento BPT. Tais métodos serão descritos a seguir.

2) Razão de Erro Aparente de Treinamento (REAT): este método, também proposto em [9], utiliza a razão de observações que foram classificadas incorretamente pela rede (razão de erro aparente: REA), que neste caso, será calculada sobre o conjunto de treinamento. A variável

$$S_j = REA(z_j = 0) - REA(z_j) \quad (5)$$

é adotada como uma medida da importância de cada neurônio. A razão de erro é expressa em percentual, a qual, em geral é facilmente calculada a partir da matriz de confusão. O processo de poda é iterativo e retira-se o neurônio com menor valor de S_j , definida em (5), em cada passo de poda.

3) Optimal Brain Damage (OBD): neste método, proposto em [3], a medida de importância de cada neurônio é dada por

$$S_j = \sum_{j=1}^N \Delta E(z_j(n)), \text{ onde}$$

$$\Delta E(z_j(n)) = \frac{1}{2} \frac{\partial^2 E(n)}{(\partial z_j(n))^2} [\Delta(z_j(n))]^2. \quad (6)$$

As derivadas são obtidas pela regra da cadeia e $\Delta(z_j(n)) = -z_j(n)$.

4) Saliências (SAL): este método, definido em [4], acrescenta a primeira derivada à medida definida em (6), tal que:

$$\Delta E(z_j(n)) = \frac{\partial E(n)}{\partial z_j(n)} \Delta(z_j(n)) + \frac{1}{2} \frac{\partial^2 E(n)}{(\partial z_j(n))^2} [\Delta(z_j(n))]^2. \quad (7)$$

O método é iterativo e retira da rede o neurônio j com o menor valor de S_j em cada passo.

5) Poda Dinâmica (PD): este método, proposto em [6], utiliza a menor medida

$$S_j = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^c (w_{kj}^{(2)})^2 z_j^2(n), \quad (8)$$

para detectar o neurônio que deve ser retirado em cada passo de poda do processo iterativo deste método.

6) Poda Iterativo (PI): este método, definido em [7], utiliza a mesma medida do método PD, equação (8), para selecionar o neurônio que deve ser retirado da rede, porém, o reajustamento dos pesos sinápticos restantes é feito pelo algoritmo do gradiente conjugado da equação normal preconditionada [13].

7) Poda Baseado na Fronteira de Decisão Efetiva de Bayes (PEB): neste método proposto em [9], a medida de saliência utilizada é similar à medida adotada no método PD, em (8), exceto pela exclusão dos termos quadráticos, desta forma, a medida S_j representa toda a atividade sináptica, excitatória e inibitória, relacionada ao neurônio j na forma exata como ela

$$\text{ocorre:} \quad S_j = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^c w_{kj}^{(2)} z_j(n). \quad (9)$$

III. INVESTIGAÇÃO EXPERIMENTAL

Nesta seção, serão apresentados quatro experimentos. O primeiro realizado com o problema XOR, onde se conhece o número mínimo de neurônios intermediários que uma rede neural necessitaria para classificar todos os elementos da amostra corretamente. O segundo um estudo envolvendo dados gerados para avaliar os métodos de poda iterativos. O terceiro e o quarto experimentos realizados com dois conjuntos de dados reais contendo um número grande de variáveis (características).

A. Problema XOR

O principal objetivo deste experimento é exemplificar como um procedimento de poda é capaz de manter na rede os neurônios mais importantes para a construção da região de decisão, definimos um experimento com o problema XOR (ou-exclusivo). Este conjunto de dados contém $N=4$ amostras de treinamento, $d=2$ características e $c=2$ classes. Os padrões de entrada são $(0,0)$ e $(1,1)$ os quais produzem saída 0 e portanto são atribuídos à classe 0; e $(0,1)$ e $(1,0)$ que produzem saída 1 e constituem a classe 1. O problema XOR pode ser resolvido por uma rede neural com uma camada escondida e $M=2$ neurônios intermediários [11, p.158]. Definimos $M=4$ neurônios na camada intermediária. É importante observar que este número é maior que o necessário para a solução adequada do problema XOR.

Após o treinamento da rede inicial, obteve-se a matriz de pesos sinápticos $\mathbf{W}^{(1)}$, e para cada linha (neurônio) $j=1, \dots, 4$ da matriz, construiu-se um hiperplano (ver Fig. 1): $w_{j1}^{(1)} x_1 + w_{j2}^{(1)} x_2 + w_{j0}^{(1)} (-1) = 0$. Desta forma, temos 4 hiperplanos, sendo um para cada neurônio intermediário.

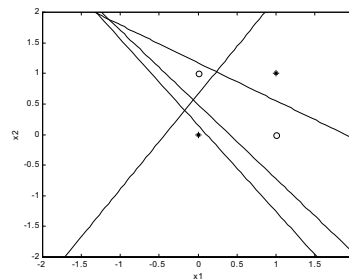


Fig. 1. Hiperplanos correspondentes aos 4 neurônios intermediários, para o problema XOR.

Para exemplificar, foi aplicado o método de poda FBI após o treinamento inicial da rede. Na Tabela I são apresentados os resultados obtidos após cada passo de poda, para as variáveis: número de neurônios retirados, número de épocas de treinamento, erro quadrático médio (EQM) e percentual de erro de classificação (EC).

TABELA I
RESULTADOS OBTIDOS PARA O PROBLEMA XOR E MÉTODO FBI

#Neurônios Retirados	#Épocas Trein.	Erro Quadrático Médio (EQM)	Perc. Erro Classificação (EC)
0	12516	0,000360	0
1	3592	0,000346	0
2	8833	0,000319	0
3	8633	0,672823	25

Analisando as colunas EQM e EC da Tabela I, nota-se que houve um grande aumento da segunda para a terceira etapa de poda, e isto significa que foi retirado um neurônio importante, e sua ausência ocasionou uma piora no desempenho da rede resultante. Desta forma, a rede final (melhor subrede) conterá dois hiperplanos, os quais serão capazes de separar as duas classes com 100% de classificação correta (ver Fig. 2).

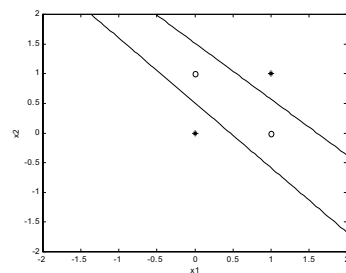


Fig. 2. Hiperplanos correspondentes aos 2 neurônios intermediários restantes na rede, após o segundo passo de poda do método FBI, para o problema XOR.

B. Estudo envolvendo dados gerados

Nesta seção são apresentados 5 casos envolvendo situações diferentes em problemas com duas classes: C_1 e C_2 . Para todos os experimentos desenvolvidos foram geradas $N=3000$ amostras, sendo $n_1=n_2=1500$ observações para cada classe. Cada conjunto de dados gerado foi dividido em dois diferentes grupos: o conjunto de treinamento e de teste, contendo $N_{\text{trei}}=2250$ e $N_{\text{tst}}=750$ observações, respectivamente. Ao todo

foram gerados 7 conjuntos de dados diferentes para cada um dos 5 casos. Considerou-se $P(C_1)=P(C_2)=0,5$.

Foram construídas 7 redes com matrizes de pesos iniciais diferentes, e avaliadas com relação ao Erro de Classificação no conjunto de teste, estas redes iniciais foram denominadas redes CONTROLE. A cada uma delas foram aplicados os 7 métodos de poda descritos na seção II. As redes neurais foram construídas usando a estrutura de camadas: entrada-intermediária-saída. O número de neurônios na camada de entrada foi definido como sendo o número de características do problema, na camada intermediária este número foi denotado por M e definido segundo o método de ramo-e-folhas [9], e na camada de saída foi considerado um número de neurônios igual ao número de classes do problema. Foi utilizada a função de ativação logística, o parâmetro de razão de aprendizagem com o valor 0,9, a constante de momentum nula, o critério de parada como a norma do vetor gradiente menor ou igual a 0,05 e empregou-se o algoritmo back-propagation sem regularização. Nos experimentos C e D, também utilizou-se as mesmas definições.

Para os métodos de poda foram realizadas $M-1$ etapas de poda para cada rede, sendo que a rede ótima foi considerada aquela que produziu o menor Erro de Classificação no conjunto de treinamento. Para esta rede anotou-se a razão de erro de classificação no conjunto de teste (REA), o número de neurônios intermediários (#neurônios) e o número total de épocas de treinamento e retreinamento desta rede (#épocas). Para a rede CONTROLE o #neurônios foi definido como sendo igual a M , pois neste método não foi aplicado nenhum método de poda.

Os Casos 1 ao 4 possuem $d=2$ características, $c=2$ classes. As características foram denominadas X_1 e X_2 . O Caso 5 apresenta $d=3$ características e $c=2$ classes. Neste caso, as características foram definidas como X_1 , X_2 e X_3 . Em todos os casos, os vetores de média e matrizes de covariância serão definidos por μ_i e Σ_i , para as classes $i=1,2$ respectivamente.

Caso 1 (Normais com médias iguais, e região de decisão circular).

Os parâmetros são: $\mu_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$,

$\mu_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ e $\Sigma_2 = \begin{bmatrix} 0,81 & 0 \\ 0 & 0,81 \end{bmatrix}$ (ver Fig. 3).

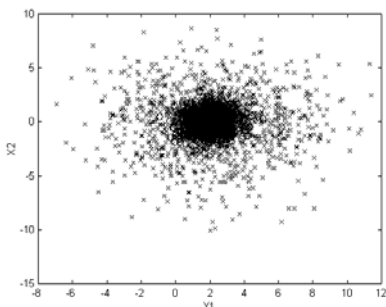


Fig. 3. Exemplo de conjunto de treinamento para o Caso 1 ($x=C_1$, $o=C_2$).

Para uma comparação visual dos resultados obtidos pelos métodos, foi construído o gráfico disposto na Fig.4, envolvendo a estatística soma dos postos (R_j) separadamente para cada uma das variáveis: REA e #neurônios. A construção desta estatística é feita da seguinte forma: para cada rede, os resultados obtidos são ordenados do menor para o maior, os quais recebem posto r_i valendo de 1 a 8. No caso de empates, todos os elementos do grupo com empates recebem a média dos postos empatados. A seguir, é realizada a operação de soma dos postos $R_j = \sum_{i=1}^7 r_i$, para as 7 redes de cada método ($j=1, \dots, 8$).

Os métodos que se concentrarem no canto inferior esquerdo do gráfico (Fig.4) serão considerados os melhores por apresentarem baixo REA (menor erro de classificação no conjunto de teste) e rede final com menor #neurônios (rede mais enxuta).

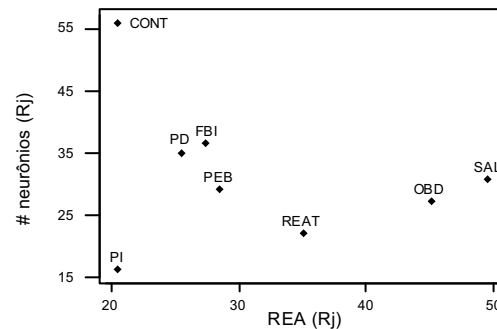


Fig. 4. Gráficos de REA (R_j) x #neurônios (R_j), para o Caso 1.

Analisando a Fig. 4, notamos que o método PI pode ser considerado visualmente como o melhor dos métodos. Comparando os resultados obtidos, nota-se que, enquanto a rede CONT (rede CONTROLE) necessitou de $32,57 \pm 1,134$ (média \pm desvio-padrão) neurônios em média na camada intermediária para resolver o problema, no método PI, este número caiu para $5,43 \pm 1,618$; o que representa uma economia de 83,33% de neurônios nesta camada, e com o mesmo REA da rede CONT.

Caso 2 (Normais com médias diferentes, com região de decisão parabólica). Os parâmetros são: $\mu_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$,

$\Sigma_1 = \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$ e $\Sigma_2 = \begin{bmatrix} 0,81 & 0 \\ 0 & 0,81 \end{bmatrix}$ (ver Fig. 5).

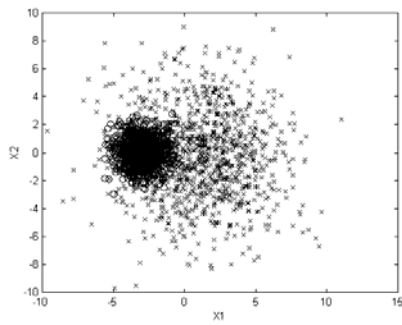


Fig. 5. Exemplo de conjunto de treinamento para o Caso 2 ($x=C_1, o=C_2$).

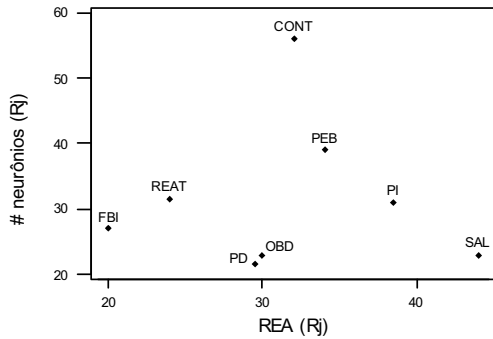


Fig. 6. Gráficos de REA (Rj) x #neurônios (Rj) para o Caso 2.

Visualmente, o método mais à esquerda no gráfico da Fig.6 é o método FBI. Este método apresentou uma redução de 70,00% de neurônios intermediários, quando comparado à rede CONT (32,71±1,604 versus 9,86±6,12 neurônios, em média). Quanto ao REA, obteve-se 7,410±1,236 para CONT versus 7,086±1,106 com FBI, uma redução de 4,37%.

Caso 3 (Problema Linearmente Separável). Os parâmetros são: $\mu_1 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ e $\Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (ver Fig. 7).

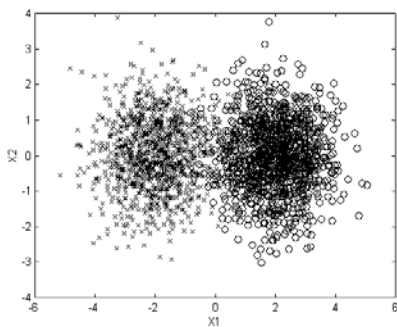


Fig. 7. Exemplo de conjunto de treinamento para o Caso 3 ($x=C_1, o=C_2$).

Na Fig. 8 pode-se observar que o método SAL aparece mais à esquerda que os demais métodos. Este método apresentou uma redução de aproximadamente 87,31% de neurônios na camada intermediária, quando comparado à rede CONT (28,143±0,90 versus 3,57±3,41 neurônios, em média), e REA (em média 2,495±0,665) 9,67% inferior que o da rede CONT (em média 2,762±0,679).

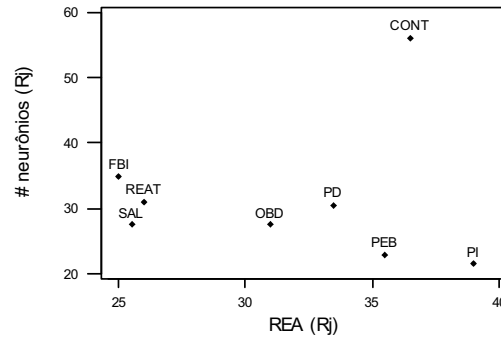


Fig. 8. Gráficos de REA (Rj) x #neurônios (Rj) para o Caso 3.

Caso 4 (Problema com Distribuição Multi-Modal e região de decisão linear dupla). Os parâmetros são: $\mu_{11} = \begin{bmatrix} -8 \\ -8 \end{bmatrix}$,

$$\Sigma_{11} = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}, \quad \mu_{12} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}, \quad \Sigma_{12} = \begin{bmatrix} 7 & 2 \\ 2 & 7 \end{bmatrix}, \quad \text{e} \quad \mu_{21} = \begin{bmatrix} -7 \\ 5 \end{bmatrix},$$

$$\Sigma_{21} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}, \quad \mu_{22} = \begin{bmatrix} -3 \\ -12 \end{bmatrix}, \quad \Sigma_{22} = \begin{bmatrix} 7 & 4 \\ 4 & 7 \end{bmatrix} \quad (\text{ver Fig.9}).$$

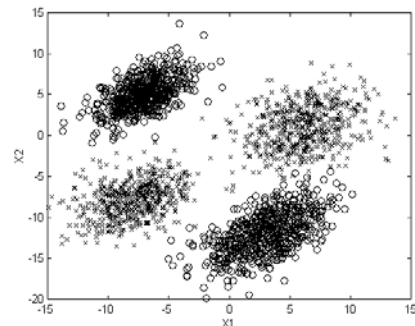


Fig. 9. Exemplo de conjunto de treinamento para o Caso 4 ($x=C_1, o=C_2$).

A Fig.10 a seguir, apresenta os métodos FBI e PI como os mais promissores para este tipo de problema. Porém, este último exigiu menor esforço computacional (ver Tabela III). Assim, com a aplicação do método PI houve uma diminuição de 70,18% no número de neurônios (média 12,14±6,74 para PI, versus 40,71±1,11 da rede CONT), embora apresente um

pequeno aumento na variável REA quando (em média 0,419±0,2098 versus 0,3810±0,2235) comparada à CONT.

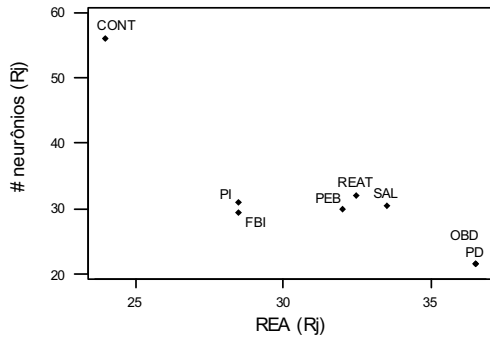


Fig. 10. Gráficos de REA (Rj) x #neurônios (Rj) para o Caso 4.

Caso 5 (Classificação Multi-Dimensional e região de decisão

planar). Os parâmetros são: $\mu_1 = \begin{bmatrix} -6 \\ -4 \\ -8 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 9 & 4 & 5 \\ 4 & 8 & 4 \\ 5 & 4 & 9 \end{bmatrix}$ e

$\mu_2 = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 8 & 2 & 6 \\ 2 & 9 & 2 \\ 6 & 2 & 10 \end{bmatrix}$ (ver Fig. 11). E os resultados

obtidos com o treinamento de 7 redes encontra-se na Fig. 12.

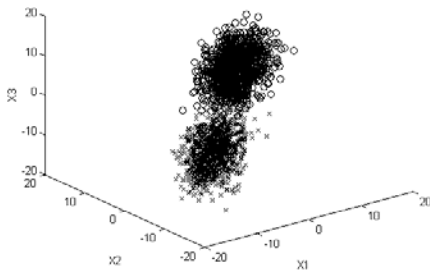


Fig. 11. Exemplo de conjunto de treinamento para o caso 5 ($x=C_1$, $o=C_2$).

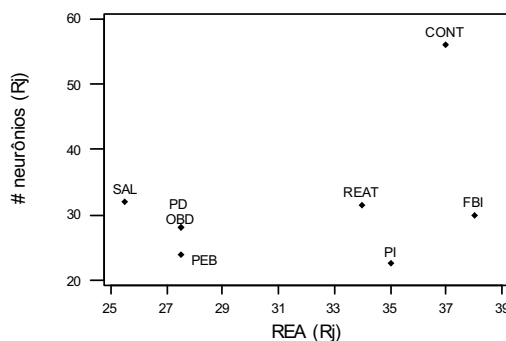


Fig. 12. Gráficos de REA (Rj) x #neurônios (Rj) para o Caso 5.

A Fig. 12 apresenta o método SAL com o menor REA(Rj) (em média, as redes apresentam REA 0,914±0,373 versus 1,067±0,422) com relação à rede CONT, que representa uma melhora de 14,34% no desempenho de classificação. Quanto aos neurônios, apresenta uma economia de quase 87,00% (5±6,38 versus 38,43±8,77).

C. Experimento 3 – Conjunto de dados sobre Tiróide

Nesta parte da investigação experimental, é testado o conjunto de dados Thyroid [10], o qual apresenta $N=7200$ observações divididas em: $N_{trei}=3772$ para treinamento e $N_{tst}=3428$ para teste. Neste problema foram treinadas 7 redes neurais iniciais com os seguintes parâmetros: $N=3772$, $d=21$ variáveis, $M=35$ neurônios intermediários e $c=3$ classes.

Os resultados podem ser visualizados na Fig. 13, na qual observa-se que o método de poda SAL apresentou os melhores resultados para este problema. A economia no número de neurônios foi de 73,46% (em média 9,29±1,98 para SAL versus 35,00 para a rede CONT). Quanto ao REA, a diminuição foi de 8,23% (em média 1,9045±0,2301 versus 2,0753±0,1184 para CONT) com o método SAL.

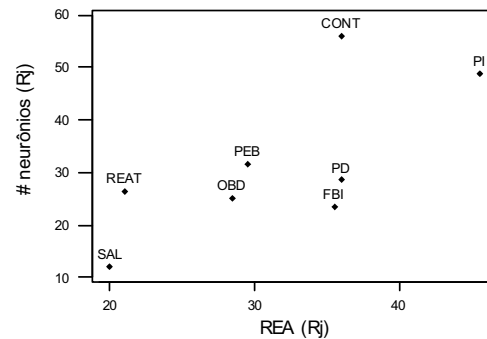


Fig. 13. Gráficos de REA (Rj) x #neurônios (Rj) para o problema Tiróide.

D. Experimento 4 – Conjunto de dados de Assinaturas

Com o objetivo de investigar os métodos em dados com dimensionalidade alta, apresentamos um problema de classificação de assinaturas. Este conjunto contém: $N=470$ assinaturas, $d=284$ características, $M=83$ neurônios intermediários e $c=10$ classes ou assinaturas de 10 indivíduos diferentes. O conjunto de treinamento contém 348 assinaturas e o conjunto de teste, 122 assinaturas.

Para a obtenção das características foram utilizadas técnicas de processamento digital de imagens: enquadramento da imagem, normalização do tamanho e divisão em 4 quadros. Após foram construídos 7 vetores de características para representar cada imagem. Foram construídas 5 redes com matrizes de pesos iniciais diferentes, e avaliadas com relação ao Erro de Classificação no Conjunto de Teste. A Fig.14 apresenta a soma dos postos para REA e #neurônios.

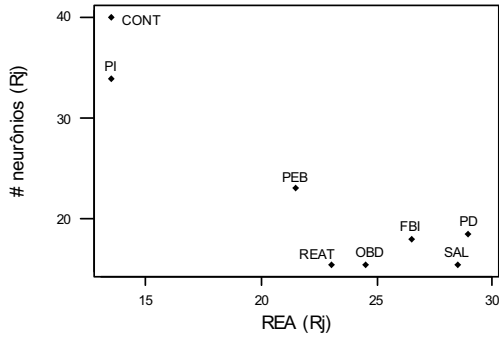


Fig. 14. Gráficos de REA (Rj) x #neurônios (Rj) para o problema de Assinaturas.

Na Fig.14, o método PI aparece mais à esquerda, entretanto, os testes realizados na próxima seção mostram que estatisticamente este método não conseguiu reduzir o número de neurônios. Desta forma, escolhemos o método REAT como mais indicado, com economia de 95,18% no número de neurônios (em média $4 \pm 0,00$ versus 83 da rede CONT) e pequeno aumento no REA (em média $0,492 \pm 0,449$ versus 0,00 da rede CONT).

IV. ANÁLISE ESTATÍSTICA

Para verificar se havia diferença significativa entre a rede CONT (nesta seção será denotada por C) e os demais 7 métodos de poda, foram realizados 21 testes de Friedman [12]. Este teste é indicado para dados que não apresentam distribuição normal, como por exemplo: dados de proporção e contagem. Foram testadas as seguintes hipóteses:

$$\begin{cases} H_0 : t_1 = t_2 = \dots = t_8 \\ H_a : \text{pelo menos dois métodos diferem entre si} \end{cases} \quad (10)$$

Para a variável REA, somente o conjunto de dados 1 apresentou resultado significativo. Então, é de interesse realizar comparações múltiplas envolvendo pares de Métodos versus Método C, para verificar qual par provocou a rejeição da hipótese nula, H_0 , em (10). As comparações múltiplas em que estamos interessados estão descritas em (11):

$$\begin{cases} H_0 : t_i = t_1 \\ H_a : t_i \neq t_1 \end{cases} \quad (11)$$

Em (11), t_1 representa a rede C e t_i representa qualquer um dos demais métodos de poda. Os resultados obtidos indicaram que os métodos SAL e OBD apresentaram REA superiores à rede C, com nível de significância global de 5%. Enquanto os demais pares foram não significativos. Desta forma, concluiu-se que os métodos SAL e OBD não são indicados para este tipo de problema, pois aumentaram o Erro de Classificação.

Quanto às variáveis #neurônios e #épocas, todos os casos apresentaram resultados significativos para os testes de Friedman. Então, podem ser feitas comparações múltiplas para detectar o(s) par(es) responsável(is) por esta significância. Os

resultados das comparações múltiplas podem ser visualizados nas Tabelas II e III.

Primeiramente, foi avaliada a variável #neurônios, e comparando os métodos com a rede C, esperava-se que os métodos fossem capazes de reduzir o número de neurônios intermediários, ou seja, esperava-se rejeitar H_0 em favor de H_a :

$$\begin{cases} H_0 : t_i = t_1 \\ H_a : t_i < t_1 \end{cases} \quad (12)$$

Analisando a Tabela II, os resultados que apresentarem NS (não significativos) indicam que o método que está sendo comparado à rede C não foi capaz de reduzir o #neurônios intermediários, e isto ocorreu para os métodos FBI e PD no Caso 1, PEB no Caso 2, FBI no Caso 3, e PI nos problemas Tiróide e Assinatura. Consequentemente, todos os demais pares foram significativos indicando a habilidade dos métodos em reduzir o número de neurônios nos diferentes problemas abordados. Entretanto, é conveniente lembrar que os métodos SAL e OBD no Caso 1 embora tenham reduzido neurônios, aumentaram o Erro de Classificação.

Na Tabela III, estamos testando a hipótese alternativa de que os métodos de poda apresentam #épocas maior que o método C:

$$\begin{cases} H_0 : t_i = t_1 \\ H_a : t_i > t_1 \end{cases} \quad (13)$$

Então, levam vantagem os métodos que apresentarem resultado NS, pois quando comparados ao método C, necessitam de um #épocas de treinamento/retreinamento igual ao deste.

Analisando a Tabela III, nota-se que os métodos mais rápidos foram: SAL nos Casos 3, 5 e Assinaturas; OBD para os Casos 3 e Assinaturas (excluiu-se o Caso 1 porque houve aumento de REA); PI nos Casos 1, 2, 4 e Tiróide; e PD no Caso 3 e Assinaturas.

TABELA II
COMPARAÇÕES MÚLTIPLAS ENVOLVENDO MÉTODOS X CONTROLE, PARA A VARIÁVEL #NEURÔNIOS, E ALFA=0,05

Par	Conjunto de Dados						
	1	2	3	4	5	Tiro	Ass
FBI x C	NS	*	NS	*	*	*	*
REATx C	*	*	*	*	*	*	*
SAL x C	*	*	*	*	*	*	*
OBD x C	*	*	*	*	*	*	*
PI x C	*	*	*	*	*	NS	NS
PD x C	NS	*	*	*	*	*	*
PEB x C	*	NS	*	*	*	*	*

TABELA III
COMPARAÇÕES MÚLTIPLAS ENVOLVENDO MÉTODO X CONTROLE (C) PARA A
VARIÁVEL #ÉPOCAS, E ALFA=0,05

Par	Conjunto de Dados					Tir	Ass
	1	2	3	4	5		
FBI x C	*	*	*	*	*	*	*
REATx C	*	*	*	*	*	*	*
SAL x C	*	*	NS	*	NS	*	NS
OBD x C	NS	*	NS	*	*	*	NS
PI x C	NS	NS	*	NS	*	NS	*
PD x C	*	*	NS	*	*	*	NS
PEB x C	*	*	*	*	*	*	*

V. CONCLUSÃO

Neste artigo, foi comprovado estatisticamente que os métodos de poda iterativos aplicados aos neurônios intermediários de redes grandes e com bom desempenho são capazes de manter a mesma acurácia de classificação, porém, com um número reduzido de neurônios na camada intermediária. Entretanto, houve algumas exceções: para o Caso 1, os métodos SAL e OBD aumentaram o erro de classificação, não sendo adequados para fronteiras de decisão do tipo circular; o método PI não foi capaz de reduzir o número de neurônios intermediários nos problemas com dimensionalidade mais elevada (Tiróide e Assinaturas).

Quanto a implementação dos métodos, os mais simples são os métodos: FBI, REAT, PD e PEB. Já os métodos SAL, OBD e PI apresentam implementações com nível de dificuldade maior, por serem mais complexos teoricamente.

Infelizmente, não é possível sugerir um único método de poda como sendo o mais indicado em todos os casos, pois verificou-se que o desempenho também depende do problema abordado. Desta forma, em aplicações práticas, sugere-se implementar mais de um método e, mediante os resultados obtidos, optar pelo melhor.

VI. BIBLIOGRAFIA

- [1] R. Reed. Pruning algorithms – a survey. *IEEE Trans. Neural Networks*, v. 4, n. 5, p. 740-747, 1993.
- [2] E. D. Karnin. A simple procedure for pruning back-propagation trained neural networks. *IEEE Trans. Neural Networks*, v. 1, n. 2, p. 239-242, 1990.
- [3] Y. Lecun, J. S. Denker, and S. A. Solla. Optimal brain damage. In *Advances in Neural Information Processing 2*, D.S. Touretzky, Ed. Denver, pp. 598-605, 1990.
- [4] J. Mao, K. Mohiuddin, and A. K. Jain. Parsimonious network design and feature selection through node pruning. In *Proc. International Conf. Pattern Recognition*, pp. 622-624, 1994.
- [5] Y. R. Park, T. J. Murray, and C. Chen. Predicting sun spots using a layered perceptron neural network. *IEEE Trans. Neural Networks*, v. 7, n. 2, p. 501-505, 1996.

- [6] K. Murase, Y. Matsunaga, and Y. Nakade. A backpropagation algorithm which automatically determines the number of association units. In *Proc. International Conf. Neural Networks*, pp. 783-788, 1997.
- [7] G. Castellano, A. M. Fanelli, and M. Pelillo. An iterative pruning algorithm for feedforward neural networks. *IEEE Trans. Neural Networks*, v. 8, n. 3, p. 519-531, 1997.
- [8] M. R. Silvestre, and L. L. Ling. Reduzindo a arquitetura de uma rede via gap das saliências dos neurônios. In *Anais V Simpósio Brasileiro de Redes Neurais*, pp. 91-96, 1998.
- [9] M. R. Silvestre, and L. L. Ling. Optimization of neural classifiers based on Bayesian decision boundaries and idle neurons pruning. In *Proc. International Conf. Pattern Recognition*, pp. 387-390, 2002.
- [10] P. M. Murphy, and D. W. Aha, (1994). UCI Repository of machine learning databases. 1994. Irvine, CA: University of California, Department of Information and Computer Science [Online]. Available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [11] S. Haykin, *Neural Networks: a comprehensive foundation*. New Jersey: Wiley, p. 696, 1994
- [12] A. C. F. Pontes. *Obtenção dos níveis de significância para os testes de Kruskal-Wallis, Friedman e Comparações Múltiplas não-paramétricas*. Dissertação (Mestrado) – Dep. de Ciências Exatas, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2002.
- [13] A. Björck and T. Elfving. Accelerated projection methods for computing in artificial neural networks. In *BIT*, vol. 19, pp. 145-163, 1979.

VII. BIOGRAFIAS



Miriam Rodrigues Silvestre obteve o título Bacharel em Estatística pela FCT-UNESP em 1987, mestre em Estatística pelo IMECC-UNICAMP, em 1992. É aluna do programa de doutorado da FEEC-UNICAMP. Tem sido Professora Assistente do Departamento de Matemática, Estatística e Computação da FCT-UNESP, em Presidente Prudente, desde 1993. Atualmente atua nas áreas de Reconhecimento de Padrões e Análise Multivariada.



Lee Luan Ling obteve o título de engenheiro eletricitista pela USP em 1980, mestre em Engenharia Elétrica pela UNICAMP em 1984 e PhD. Engenharia Elétrica pela Universidade de Cornell em 1991. Foi fundador e tem sido o coordenador do Laboratório de Reconhecimento de Padrões e Redes de Comunicações (LRPRC) da FEEC-UNICAMP desde 1994. Foi o chefe do Departamento de Comunicações em 2000. Desde 2002 ele tem sido Professor Titular pela FEEC-UNICAMP. Atualmente ele atua intensamente em duas áreas de pesquisa: Reconhecimento de Padrões e Redes de Comunicações.