

# Comparação de funções de aprendizado no método Backpropagation

Texto apresentado à disciplina Redes Neurais Artificiais

Graduação em Engenharia de Sistemas

Universidade Federal de Minas, 2018 01

Professor Antônio Pádua de Braga

Matheus Araujo, 2013066265

**Resumo**—O texto a seguir foi apresentado como trabalho final para a disciplina *Redes Neurais Artificiais*, no primeiro semestre de 2018, graduação em *Engenharia de Sistemas*, na *Universidade Federal de Minas Gerais*. Nele duas funções de aprendizado do método Backpropagation são avaliadas e comparadas: a função padrão e a função SCG, Scaled Conjugate Gradient, ou Gradiente Conjugado em Escala. A eficiência das funções foi comparada através da confusion matrix dos resultados obtidos por redes treinadas em condições equiparadas.

## I. INTRODUÇÃO

Para o presente trabalho, apresentado como trabalho final à disciplina de Redes Neurais Artificiais, graduação em Engenharia de Sistemas, foi proposto a escolha de cinco bases de dados a fim de verificar o comportamento de redes neurais multicamadas. As bases escolhidas foram: Iris, Letter Recognition, Wine, Tic Tac Toe e Balance Scale. Foi utilizada a ferramenta RStudio, com a linguagem R, bem como o pacote RSNNS.

Nos experimentos foram comparados os métodos Backpropagation Padrão e SCG, os resultados e comparativos são apresentados a seguir.

## II. REVISÃO DE LITERATURA

### A. Redes Neurais Artificiais

Redes Neurais Artificiais são sistemas paralelos distribuídos compostos por unidades de processamento simples que executam cálculos matemáticos. Essas unidades são dispostas em uma ou mais camadas e são interligadas por um grande número de conexões, normalmente unidirecionais. As conexões possuem pesos que armazenam o conhecimento adquirido pelo modelo. RNAs têm como procedimento usual uma etapa de aprendizagem, onde um conjunto de exemplos é apresentado a ela. A rede então aprende e generalização a informação [1].

### B. Backpropagation

Backpropagation é um algoritmo utilizado para o treinamento de Redes MLP, Multi-Layer-Perceptron. Esse método utiliza o gradiente descendente para calcular o erro das camadas intermediárias através de uma estimativa do efeito que elas causam na camada de saída. No algoritmo, o erro na saída é calculado e é então retroalimentado para as camadas intermediárias, possibilitando o ajuste dos pesos proporcionalmente aos valores das conexões entre as camadas.

### C. Scaled Conjugate Gradient

O algoritmo SCG, Scaled Conjugate Gradient, é um algoritmo de aprendizado supervisionado com taxa de convergência superlinear. Ele utiliza uma técnica de otimização baseada no Gradiente Conjugado. [2]

### D. Confusion Matrix

A Confusion Matrix, também conhecida como Error Matrix, é uma tabela que permite a visualização da performance de um algoritmo de classificação. Nas colunas são representadas as classes reais e nas linhas as classes calculadas pela rede. Assim, cada célula da matriz representa o cruzamento entre a quantidade real de casos e a quantidade encontrado dos casos. Por consequência, a diagonal principal da matriz irá representar a quantidade de acertos da rede.

## III. METODOLOGIA UTILIZADA

Para avaliar as duas funções a seguinte metodologia foi utilizada:

- 1) 85% dos dados disponíveis foram utilizados para treinar a rede;
- 2) Após o treinamento, os 15% dos dados restantes foram avaliados pela rede neural criada;
- 3) A confusion matrix dos resultados avaliados foi calculada;
- 4) O índice de acurácia foi calculado segundo a fórmula:

$$\text{índice acurácia} = \frac{\text{soma diagonal principal}}{\text{soma toda matriz}}$$

- 5) Os passos 1 a 4 foram repetidos 10 vezes para cada método e cada base de dados;
- 6) A média do índice de acurácia foi então calculada.

Para executar as redes foi utilizado o método `mlp` do pacote RSNNS. Nesse método o parâmetro `size` diz respeito à quantidade de neurônios na rede neural. O parâmetro `learnFuncParams` define parâmetros para as funções de aprendizado. E o parâmetro `maxit` define a quantidade máxima de épocas para a etapa de aprendizagem [3].

As bases de dados utilizadas estão apresentadas a seguir.

### A. Iris

O primeiro problema estudado foi o *Iris* [4].

- **Número de instâncias:** 150
- **Número de atributos:** 4
- **Número de classes:** 3

Essa é uma base de dados muito conhecida no ramo de Redes Neurais Artificiais. Ela avalia dimensões físicas de espécimes de três classes de plantas iris: Iris Setosa, Iris Versicolour e Iris Virginica.

### B. Letter Recognition

O segundo problema estudado foi o *Letter Recognition* [5].

- **Número de instâncias:** 20000
- **Número de atributos:** 16
- **Número de classes:** 26

Nesse problema as 26 letras do alfabeto inglês são identificadas em imagens preto e branco de onde foram extraídos 16 atributos numéricos.

### C. Wine

O terceiro problema estudado foi o *Wine* [6].

- **Número de instâncias:** 178
- **Número de atributos:** 13
- **Número de classes:** 3

Nesse problema resultados de análises químicas são usados para determinar a origem de vinhos.

### D. Tic Tac Toe

O quarto problema estudado foi o *Tic Tac Toe* [7].

- **Número de instâncias:** 958
- **Número de atributos:** 9
- **Número de classes:** 2

Nesse problema, todas as possibilidades de um tabuleiro do *Jogo da Velha* onde  $x$  começou são mostradas e classificadas quanto à possibilidade do  $x$  vencer ou não.

### E. Balance Scale

O quinto problema estudado foi o *Balance Scale* [8].

- **Número de instâncias:** 625
- **Número de atributos:** 4
- **Número de classes:** 3

Nesse problema um modelo físico de uma balança é avaliado. Na balança há dois diferentes pesos em diferentes distâncias do centro. Em cada configuração a balança pode estar pendendo para a esquerda, ou para a direita ou equilibrada.

## IV. RESULTADOS

A seguir são apresentados os resultados das execuções para as bases de dados avaliados.

### A. Iris

1) *Backpropagation padrão*: O backpropagation padrão foi executado para a base de dados Iris com os seguintes parâmetros:

- **size:** 5
- **learnFuncParams:** 0.1
- **maxit:** 50

A confusion matrix de uma execução é apresentada na Figura 1.

	1	2	3
1	6	0	0
2	0	9	1
3	0	0	7

Figura 1. Confusion Matrix - Iris - Backpropagation padrão

Os resultados para as 10 execuções do backpropagation padrão com as taxas de acurácia são apresentados na Tabela I. A média de acurácia foi de 0.9609, com desvio padrão de 0.0432.

Tabela I  
RESULTADOS - IRIS - BACKPROPAGATION PADRÃO

	Acurácia
Execução 1	0.8696
Execução 2	0.9130
Execução 3	1.0000
Execução 4	0.9565
Execução 5	1.0000
Execução 6	0.9565
Execução 7	1.0000
Execução 8	1.0000
Execução 9	0.9565
Execução 10	0.9565
<b>Média</b>	<b>0.9609</b>
<b>Desvio Padrão</b>	<b>0.0432</b>

2) *SCG*: O backpropagation com função de aprendizado SCG foi executado para a base de dados Iris com os seguintes parâmetros:

- **size:** 5
- **learnFuncParams:** (0, 0, 0, 0)
- **maxit:** 50

A confusion matrix de uma execução é apresentada na Figura 2.

	1	2	3
1	10	0	0
2	0	4	2
3	0	0	7

Figura 2. Confusion Matrix - Iris - Backpropagation SCG

Os resultados para as 10 execuções do backpropagation SCG com as taxas de acurácia são apresentados na Tabela II. A média de acurácia foi de 0.9739, com desvio padrão de 0.0367.

Tabela II  
RESULTADOS - IRIS - BACKPROPAGATION SCG

	Acertividade
Execução 1	1.0000
Execução 2	1.0000
Execução 3	0.9565
Execução 4	0.9565
Execução 5	1.0000
Execução 6	1.0000
Execução 7	0.9130
Execução 8	1.0000
Execução 9	1.0000
Execução 10	0.9130
<b>Média</b>	<b>0.9739</b>
<b>Desvio Padrão</b>	<b>0.0367</b>

## B. Letter Recognition

1) *Backpropagation padrão*: O backpropagation padrão foi executado para a base de dados Letter Recognition com os seguintes parâmetros:

- size: 10
- learnFuncParams: 0.1
- maxit: 100

A confusion matrix de uma execução é apresentada na Figura 3.

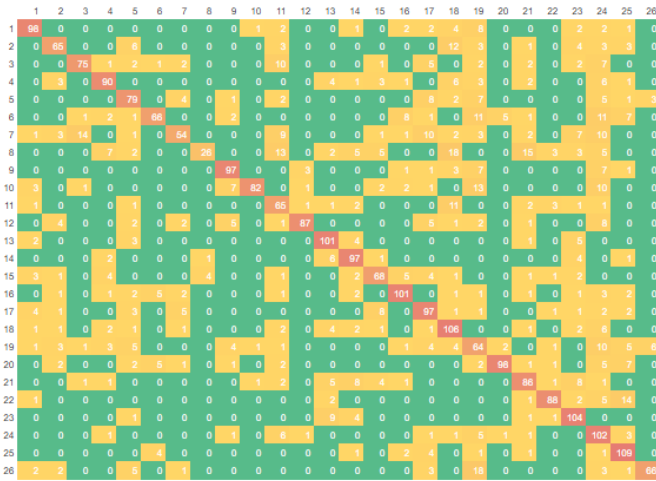


Figura 3. Confusion Matrix - Letter Recognition - Backpropagation padrão

Os resultados para as 10 execuções do backpropagation padrão com as taxas de acertividade são apresentados na Tabela III. A média de acertividade foi de 0.7391, com desvio padrão de 0.0062.

2) *SCG*: O backpropagation com função de aprendizado SCG foi executado para a base de dados Letter Recognition com os seguintes parâmetros:

- size: 10
- learnFuncParams: (0, 0, 0, 0)
- maxit: 100

A confusion matrix de uma execução é apresentada na Figura 4.

Os resultados para as 10 execuções do backpropagation SCG com as taxas de acertividade são apresentados na Tabela

Tabela III  
RESULTADOS - LETTER RECOGNITION - BACKPROPAGATION PADRÃO

	Acertividade
Execução 1	0.7363
Execução 2	0.7370
Execução 3	0.7380
Execução 4	0.7527
Execução 5	0.7360
Execução 6	0.7353
Execução 7	0.7367
Execução 8	0.7477
Execução 9	0.7380
Execução 10	0.7330
<b>Média</b>	<b>0.7391</b>
<b>Desvio Padrão</b>	<b>0.0062</b>

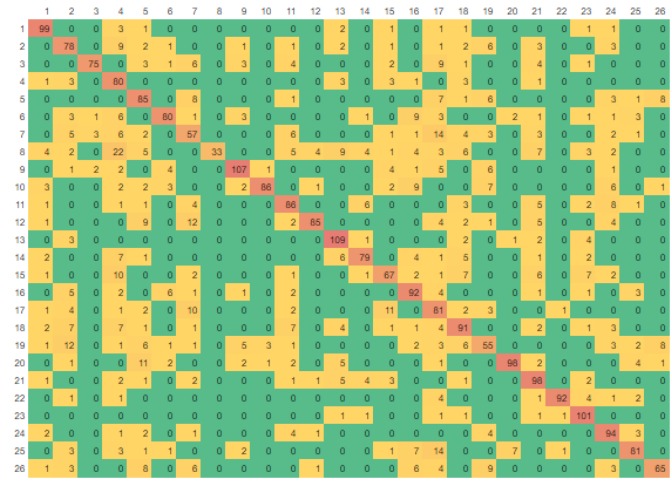


Figura 4. Confusion Matrix - Letter Recognition - Backpropagation SCG

IV. A média de acertividade foi de 0.7262, com desvio padrão de 0.0112.

Tabela IV  
RESULTADOS - LETTER RECOGNITION - BACKPROPAGATION SCG

	Acertividade
Execução 1	0.7430
Execução 2	0.7280
Execução 3	0.7187
Execução 4	0.7117
Execução 5	0.7203
Execução 6	0.7327
Execução 7	0.7163
Execução 8	0.7257
Execução 9	0.7207
Execução 10	0.7453
<b>Média</b>	<b>0.7262</b>
<b>Desvio Padrão</b>	<b>0.0112</b>

## C. Wine

1) *Backpropagation padrão*: O backpropagation padrão foi executado para a base de dados Wine com os seguintes parâmetros:

- size: 5
- learnFuncParams: 0.1

- maxit: 50

A confusion matrix de uma execução é apresentada na Figura 5.

	1	2	3
1	9	0	0
2	2	7	0
3	0	1	8

Figura 5. Confusion Matrix - Wine - Backpropagation padrão

Os resultados para as 10 execuções do backpropagation padrão com as taxas de acurácia são apresentados na Tabela V. A média de acurácia foi de 0.9630, com desvio padrão de 0.0349.

Tabela V  
RESULTADOS - WINE - BACKPROPAGATION PADRÃO

	Acurácia
Execução 1	1.0000
Execução 2	0.9630
Execução 3	0.9630
Execução 4	0.9630
Execução 5	0.9630
Execução 6	1.0000
Execução 7	0.9259
Execução 8	0.9630
Execução 9	0.8889
Execução 10	1.0000
<b>Média</b>	<b>0.9630</b>
<b>Desvio Padrão</b>	<b>0.0349</b>

2) SCG: O backpropagation com função de aprendizado SCG foi executado para a base de dados Wine com os seguintes parâmetros:

- size: 5
- learnFuncParams: (0, 0, 0, 0)
- maxit: 50

A confusion matrix de uma execução é apresentada na Figura 6.

	1	2	3
1	8	0	0
2	1	9	1
3	0	0	8

Figura 6. Confusion Matrix - Wine - Backpropagation SCG

Os resultados para as 10 execuções do backpropagation SCG com as taxas de acurácia são apresentados na Tabela VIII. A média de acurácia foi de 0.9593, com desvio padrão de 0.0273.

#### D. Tic Tac Toe

1) *Backpropagation padrão*: O backpropagation padrão foi executado para a base de dados Tic Tac Toe com os seguintes parâmetros:

- size: 5

Tabela VI  
RESULTADOS - WINE - BACKPROPAGATION SCG

	Acurácia
Execução 1	1.0000
Execução 2	0.9630
Execução 3	0.9630
Execução 4	0.9630
Execução 5	0.9259
Execução 6	0.9630
Execução 7	0.9259
Execução 8	0.9630
Execução 9	0.9259
Execução 10	1.0000
<b>Média</b>	<b>0.9593</b>
<b>Desvio Padrão</b>	<b>0.0273</b>

- learnFuncParams: 0.1
- maxit: 50

A confusion matrix de uma execução é apresentada na Figura 7.

	1	2
1	28	31
2	6	79

Figura 7. Confusion Matrix - Tic Tac Toe - Backpropagation Padrão

Os resultados para as 10 execuções do backpropagation padrão com as taxas de acurácia são apresentados na Tabela VII. A média de acurácia foi de 0.7833, com desvio padrão de 0.0401.

Tabela VII  
RESULTADOS - TIC TAC TOE - BACKPROPAGATION PADRÃO

	Acurácia
Execução 1	0.7500
Execução 2	0.7986
Execução 3	0.7500
Execução 4	0.7778
Execução 5	0.7500
Execução 6	0.7500
Execução 7	0.8194
Execução 8	0.8403
Execução 9	0.7500
Execução 10	0.8472
<b>Média</b>	<b>0.7833</b>
<b>Desvio Padrão</b>	<b>0.0401</b>

2) SCG: O backpropagation com função de aprendizado SCG foi executado para a base de dados Tic Tac Toe com os seguintes parâmetros:

- size: 5
- learnFuncParams: (0, 0, 0, 0)
- maxit: 50

A confusion matrix de uma execução é apresentada na Figura 8.

Os resultados para as 10 execuções do backpropagation SCG com as taxas de acurácia são apresentados na Tabela

	1	2
1	16	21
2	13	94

Figura 8. Confusion Matrix - Tic Tac Toe - Backpropagation SCG

VIII. A média de acurácia foi de 0.7868, com desvio padrão de 0.0656.

Tabela VIII  
RESULTADOS - WINE - BACKPROPAGATION SCG

	Acertividade
Execução 1	0.7986
Execução 2	0.7014
Execução 3	0.9097
Execução 4	0.7431
Execução 5	0.6875
Execução 6	0.8264
Execução 7	0.8264
Execução 8	0.7639
Execução 9	0.8056
Execução 10	0.8056
<b>Média</b>	<b>0.7868</b>
<b>Desvio Padrão</b>	<b>0.0656</b>

### E. Balance Scale

1) *Backpropagation padrão*: O backpropagation padrão foi executado para a base de dados Balance Scale com os seguintes parâmetros:

- size: 5
- learnFuncParams: 0.1
- maxit: 50

A confusion matrix de uma execução é apresentada na Figura 9.

	1	2	3
1	0	3	1
2	0	50	1
3	0	3	36

Figura 9. Confusion Matrix - Balance Scale - Backpropagation padrão

Os resultados para as 10 execuções do backpropagation padrão com as taxas de acurácia são apresentados na Tabela IX. A média de acurácia foi de 0.8713, com desvio padrão de 0.0267.

2) *SCG*: O backpropagation com função de aprendizado SCG foi executado para a base de dados Balance Scale com os seguintes parâmetros:

- size: 5
- learnFuncParams: (0, 0, 0, 0)
- maxit: 50

A confusion matrix de uma execução é apresentada na Figura 10.

Tabela IX  
RESULTADOS - BALANCE SCALE - BACKPROPAGATION PADRÃO

	Acertividade
Execução 1	0.8617
Execução 2	0.8617
Execução 3	0.8511
Execução 4	0.9255
Execução 5	0.8723
Execução 6	0.8298
Execução 7	0.8723
Execução 8	0.9043
Execução 9	0.8617
Execução 10	0.8723
<b>Média</b>	<b>0.8670</b>
<b>Desvio Padrão</b>	<b>0.0267</b>

	1	2	3
1	4	1	3
2	3	47	0
3	1	0	35

Figura 10. Confusion Matrix - Balance Scale - Backpropagation SCG

Os resultados para as 10 execuções do backpropagation SCG com as taxas de acurácia são apresentados na Tabela X. A média de acurácia foi de 0.9234, com desvio padrão de 0.0425.

Tabela X  
RESULTADOS - WINE - BACKPROPAGATION SCG

	Acertividade
Execução 1	0.9149
Execução 2	0.9043
Execução 3	0.9574
Execução 4	0.9255
Execução 5	0.9574
Execução 6	0.8191
Execução 7	0.9574
Execução 8	0.9468
Execução 9	0.9468
Execução 10	0.9043
<b>Média</b>	<b>0.9234</b>
<b>Desvio Padrão</b>	<b>0.0425</b>

## V. CONCLUSÕES

A Tabela XI apresenta os índices de acurácia médios de todas as execuções.

Nela é possível perceber que o Backpropagation Padrão teve melhores resultados em duas das bases de dados, enquanto o Backpropagation SCG teve melhor desempenho nas outras três. Na média, a função SCG tem um melhor desempenho em comparação com o Backpropagation Padrão. Nos experimentos realizados, há uma melhora de cerca de 1,13%.

## VI. BIBLIOGRAFIA

### REFERÊNCIAS

- [1] Braga, A. de P., "Redes neurais artificiais: teoria e aplicações" LTC Editora, 2007.

Tabela XI  
ÍNDICES DE ACERTIVIDADE

Base de Dados	Padrão	SCG	Comparação
Iris	0.9609	0.9739	1.0135
Letter Recognition	0.7391	0.7262	0.9825
Wine	0.9630	0.9593	0.9962
Tic Tac Toe	0.7833	0.7868	1.0045
Balance Scale	0.8713	0.9234	1.0598
<b>Média</b>	<b>0.8635</b>	<b>0.8739</b>	<b>1.0113</b>

- [2] Moller, Martin Fodslette. A scaled conjugate gradient algorithm for fast supervised learning.
- [3] <https://github.com/cran/RSNNS/>
- [4] <https://archive.ics.uci.edu/ml/datasets/Iris>
- [5] <https://archive.ics.uci.edu/ml/datasets/Letter+Recognition>
- [6] <https://archive.ics.uci.edu/ml/datasets/Wine>
- [7] <https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>
- [8] <https://archive.ics.uci.edu/ml/datasets/Balance+Scale>
- [9] Fisher, R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions to Mathematical Statistics" (John Wiley, NY, 1950).
- [10] P. W. Frey and D. J. Slate (Machine Learning Vol 6 #2 March 91): "Letter Recognition Using Holland-style Adaptive Classifiers".
- [11] Forina, M. et al, PARVUS - An Extendible Package for Data Exploration, Classification and Correlation. Institute of Pharmaceutical and Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy.
- [12] Matheus, C. J., & Rendell, L. A. (1989). Constructive induction on decision trees. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 645–650). Detroit, MI: Morgan Kaufmann.