

Relatório de Classificação de Padrões

Victor Carreira

Resumo—Este relatório apresenta o resultado de sete testes conduzidos em uma rede neural artificial *Perceptron multi-camadas* em um banco de dados de uma instituição financeira com o intuito de se fazer uma análise de crédito bancário. A base de dados contém 2077 exemplos de créditos concedidos. Possui 11 atributos de entrada e 2 classes de saída. A saída da rede indica se o cliente pagou o empréstimo ou não. Os resultados dos testes estão apresentados em formato de tabela. Os melhores resultados de classificação foram obtidos ao se realizar a união de atributos semelhantes dentro das classes NDEP e ESTC, associada com a normalização e codificação binária dos atributos categóricos de entrada da rede atingindo um índice de classificação correta de 92%.

I. INTRODUÇÃO

As Redes Neurais Artificiais (RNA) são inspiradas em modelos sensoriais do processamento de tarefas realizadas pelo cérebro (Hagan et al., 1996). Uma RNA, portanto pode ser criada através da aplicação de algoritmos matemáticos que imitem a tarefa realizada por um neurônio (Nedjah et al., 2016). Uma rede neural artificial possui semelhanças com a rede biológica presente no sistema nervoso central, neste o cômputo de informações realizado do cérebro é feito através de uma vasta quantidade de neurônios interconectados (Feldman et al., 1988; Poulton, 2002). A comunicação entre essas células é realizada através de impulsos elétricos. Estes são transmitidos e recebidos por meio de sinapses nervosas entre axônios e dendritos. As sinapses são estruturas elementares e uma unidade funcional localizada entre dois neurônios (Krogh, 2008).

McCulloch and Pitts (1943) redigem o trabalho pioneiro onde foi modelado um neurônio cuja resposta dependia do *input*¹ que provinha de outros neurônios e do peso utilizado. Já Rosenblatt (1962) cria a teoria de convergência do *Perceptron* onde ele prova que modelos de neurônios possuem propriedades similares ao cérebro humano (Kanal, 2001). Neste sentido as rede neuronais artificiais podem realizar performances sofisticadas no reconhecimento de padrões, mesmo se alguns neurônios forem destruídos (Levy, 1997). Minsky and Papert (1969) demonstraram que um único *Perceptron* somente resolve uma classe muito limitada de problemas que podem ser linearizados.

Neste relatório são apresentados os resultados do Trabalho 01 MLP de Classificação da disciplina ELE 2394, Redes Neurais I, da Engenharia elétrica.

II. OBJETIVO E METODOLOGIA

Uma instituição financeira possui uma base de dados com o histórico de crediário oferecido aos seus clientes. Baseado

neste histórico, a instituição deseja inferir se um novo cliente pagará ou não a dívida contraída.

A base de dados possui 2077 exemplos, com 11 atributos cada, de créditos concedidos aos seus clientes. A base informa ainda se o cliente honrou ou não o pagamento do empréstimo.

A partir da base original, foram criadas 3 bases de treinamento, com 1500 exemplos cada escolhidos aleatoriamente a partir da base original, e 3 bases de testes com 577 exemplos cada, representando, respectivamente, 72, 2% e 27, 8% do total de cada sub- grupos de dados. Estas bases estão nos arquivos treino01.txt, treino02.txt, treino03.txt, teste01.txt, teste02.txt e teste03.txt. Utilizar o software WEKA, para criar um classificador, baseado em redes neurais, capaz de informar se um novo cliente será potencialmente adimplente ou não.

Abra cada um dos arquivos no WEKA e grave-os em formato .arff. Através de um editor de textos (por exemplo, o WordPad), altere o tipo associado às variáveis categóricas conforme o exemplo *@attribute NDEP numeric => @attribute NDEP 0,1,2,3,4,5,6,7*

Para cada uma das configurações abaixo, apresente os resultados para cada par de conjuntos de treino e de teste, assim como a média e o desvio padrão dos 3 pares: I, II, III, IV, V, VI.

- Sem normalização dos atributos de entrada;
- Com normalização dos atributos de entrada e SEM codificação binária dos atributos categóricos;
- Com normalização e codificação binária dos atributos categóricos de entrada e com 2 números diferentes de neurônios na camada escondida;
- Com normalização dos atributos de entrada e variando o número de épocas durante a fase de treinamento. Escolha 3 durações de treino diferentes (por exemplo: 1, 100 e 1000);
- Com normalização dos atributos de entrada e utilizando um conjunto de validação;
- Tente obter melhores resultados (se possível) agrupando algumas categorias das variáveis ESTC e NDEP. Para isto, utilize o filtro não supervisionado MergeTwoValues.

Para os itens I, II, IV, V e VI, indique para cada um dos casos o número de neurônio na camada escondida e explique a sua escolha. Para todos os itens, não varie a taxa de aprendizagem nem o termo de momento.

III. PRINCÍPIO TEÓRICO

O neurônio de McCulloch and Pitts (1943) propõe um limite binário para a criação de um modelo. Este neurônio artificial registra uma soma de pesos de n sinais de entrada, x_j , $j = 1, 2, 3, \dots, n$, e fornece um *output*² de 1 caso esta

Professora: Marley. Eng. Elétrica. PUC-RIO

¹Valor de entrada

²Valor de saída

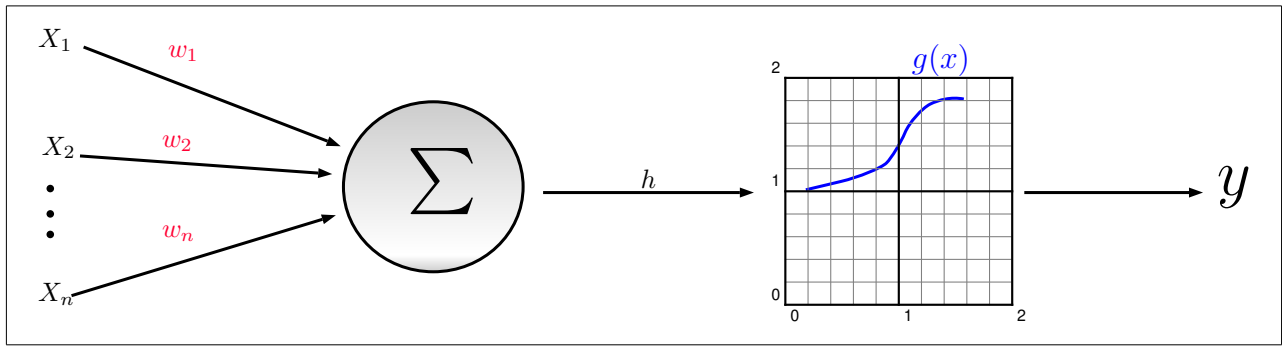


Figura 1. Modelo esquemático de um neurônio de McCulloch-Pitts. Onde x_1, x_2, \dots, x_n são os *inputs*, w_1, w_2, \dots, w_n são os pesos, h é o treino, $g(x)$ é a função de ativação, e y é o *output*.

soma esteja acima do limite u . Caso contrário o *output* é 0. Matematicamente essa relação pode ser descrita de acordo com a Eq. 1:

$$y = \theta \left(\sum_{j=1}^n w_j x_j - u \right) \quad (1)$$

Onde θ é o passo dado na posição 0, w_j é chamada sinapse-peso associado a um j -ésimo *input*. A título de simplificação a função limite³ u é considerada um outro peso $w_0 = -u$ anexado a um neurônio com um *input* constante $x_0 = 1$. Pesos positivos correspondem a uma sinapse **excitatória**, enquanto pesos negativos correspondem a uma sinapse **inibitória**. Este modelo contém uma série de simplificações que não refletem o verdadeiro comportamento dos neurônios biológicos (Mao, 1996).

Derivações do neurônio de McCulloch and Pitts (1943) na escolha das funções de ativação. Uma função largamente utilizada é a função sigmoide, que exibe uma suavização dos *outputs* a medida que o valor da função diminui (Mao, 1996; Misra and Saha, 2010). Essa função de ativação pode ser expressa de acordo com a Eq. 2:

$$g(x) = 1/(1 + e^{-\beta x}) \quad (2)$$

Onde β é o parâmetro de inclinação. A Fig. 1 ilustra a sequência lógica da operação de uma RNA para um neurônio simples de McCulloch-Pitts.

As redes alimentadas diretamente são aquelas redes cujos grafos orientados se distinguem pela presença de um ou mais camadas ocultas e cujos nós são chamados de neurônios ocultos. A função do neurônio é intervir entre a camada externa e a saída da rede de maneira útil. Adicionando-se camadas ocultas a rede torna-se capaz de realizar estatísticas de ordem elevada (Haykin, 1999).

IV. RESULTADOS

A tabela I refere-se ao conjunto de 6 testes realizados no primeiro banco de dados que é uma composição dos dados de treinamento e testes 1. Os testes utilizados neste trabalho

foram em sua totalidade compostos por 2 camadas ocultas. O número de neurônios ocultos estão indicados nas tabelas, como Neurônios na CO1 (camada oculta 1) e como Neurônios na CO2 (camada oculta 2).

Tabela I
GRUPO 01

Configuração	I	II	III	IV	V	VI
Neurônios na CO1	4	4	4	-	4	4
Neurônios na CO2	4	4	2	-	4	4
Cl. Correta (%)	52.3	91.3	91.2	-	52.3	92
Cl. Incorreta (%)	47.7	8.7	8.8	-	47.7	8
MAE (%)	0.49	0.13	0.14	-	0.49	0.1363
RMSE (%)	0.50	0.26	0.27	-	0.49	0.2586
RAE (%)	99.68	26.23	28.3	-	100.2	27.3239
RRSE(%)	100.23	52.23	53.75	-	100.1	51.7764
Dados Totais	1500	1500	1500	-	1500	1500

A tabela II refere-se os testes de variações de épocas do conjunto de dados 1.

Tabela II
GRUPO 01: VARIANDO AS ÉPOCAS NO TESTE IV

Configuração	IV ₁	IV ₁₀₀	IV ₁₀₀₀
Neurônios na CO1	4	4	4
Neurônios na CO2	4	4	4
Cl. Correta (%)	52.3	90.8	91.53
Cl. Incorreta (%)	47.7	9.2	8.47
MAE (%)	0.49	0.16	0.14
RMSE (%)	0.51	0.27	0.26
RAE (%)	99.63	32.34	27.49
RRSE(%)	100.23	54.49	52.89
Dados Totais	1500	1500	1500

A Fig. 2 apresenta a topologia da rede quando se unem os atributos semelhantes. Neste caso, o número de entradas é reduzido. Na etapa VI, nos três grupos analisados, foram unidos por meio do filtro two merge values os atributos semelhantes das classes ESTC e NDEP. Nesta exercício a classe ESTC passa a ter somente duas categorias (casado ou solteiro). E a classe NDEP passa a ter duas classes (com filhos ou sem filhos).

A tabela III apresenta o conjunto de testes para o grupo 2.

Na tabela IV são apresentados os subconjuntos de testes das épocas do grupo 2.

Na tabela V são apresentados os conjunto de testes para o grupo 3.

³Genericamente chamada de função de ativação

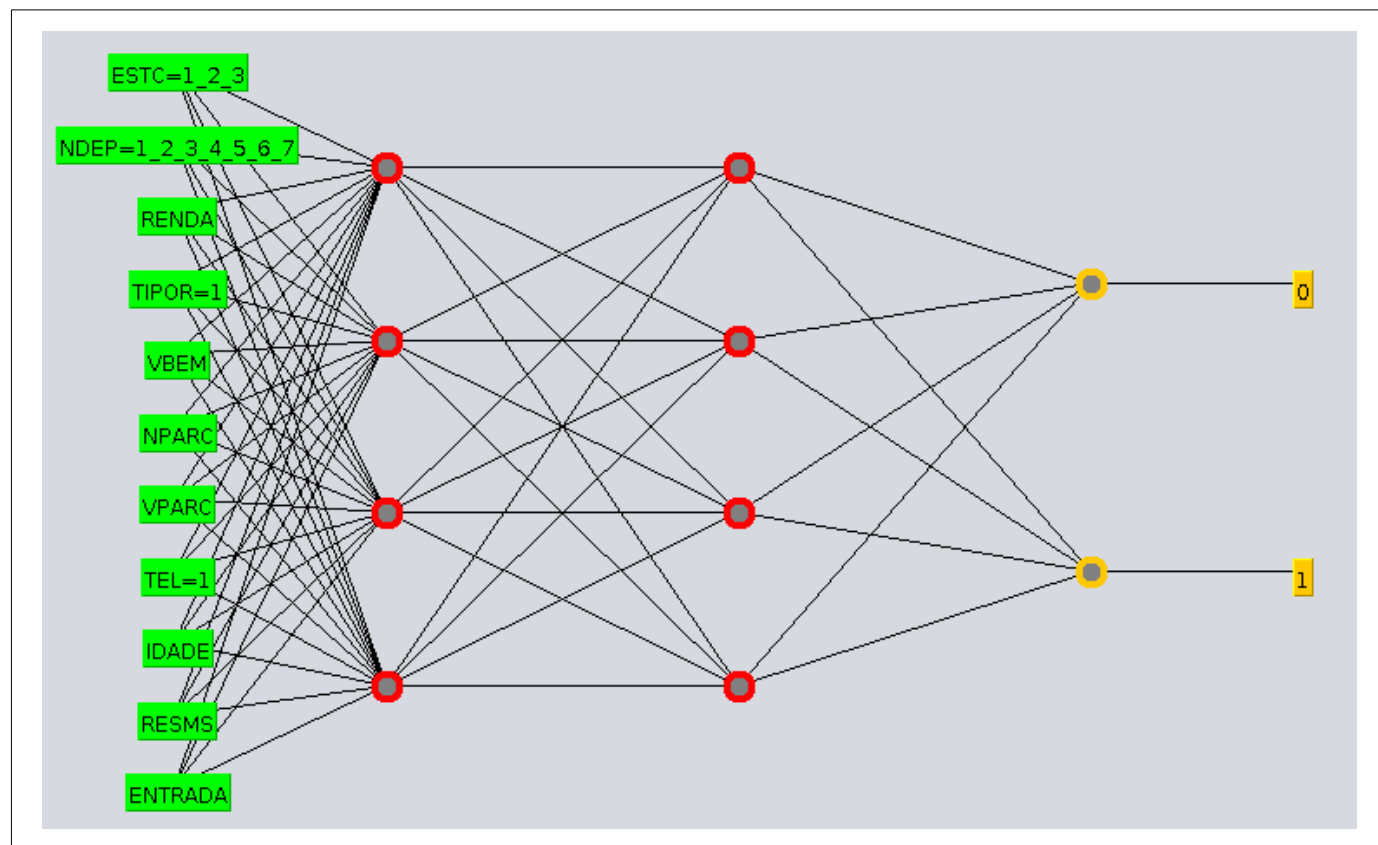


Figura 2. Grafo orientado da rede após se aplicar o filtro merge two values nas nos atributos 1,2,3,4,5,6,7 da classe NDEP e no atributos 1,2,3 da classe ESTC.

Tabela III
GRUPO 02

Configuração	I	II	III	IV	V	VI
Neuronios na CO1	4	4	4	-	4	4
Neuronios na CO2	4	4	8	-	4	4
Cl. Correta (%)	53.30	91.47	91.67	-	53.27	91.27
Cl. Incorreta (%)	46.7	8.53	8.33	-	46.73	8.73
MAE (%)	0.49	0.12	0.12	-	0.5	0.13
RMSE (%)	0.50	0.26	0.26	-	0.5	0.27
RAE (%)	98.96	25.05	24.36	-	100.4	24.66
RRSE(%)	101.26	52.85	53.24	-	100.2	54.12
Dados Totais	1500	1500	1500	-	1500	1500

Tabela V
GRUPO 03

Configuração	I	II	III	IV	V	VI
Neuronios na CO1	4	4	1	-	4	4
Neuronios na CO2	4	4	12	-	4	4
Cl. Correta (%)	49	91.4	90.9	-	51	90
Cl. Incorreta (%)	51	8.6	9.1	-	49	10
MAE (%)	0.5	0.1	0.1	-	0.5	0.1399
RMSE (%)	0.5	0.2	0.3	-	0.5	0.2738
RAE (%)	100.1	24.5	32.7	-	100	27.986
RRSE(%)	100.3	53.1	55.6	-	100	54.7797
Dados Totais	1500	1500	1500	-	1500	1500

Tabela IV
GRUPO 02: VARIANDO AS ÉPOCAS NO TESTE IV

Configuração	IV ₁	IV ₁₀₀	IV ₁₀₀₀
Neuronios na CO1	4	4	4
Neuronios na CO2	4	4	4
Cl. Correta (%)	53.2667	91.4	91.8667
Cl. Incorreta (%)	46.7333	8.6	8.1333
MAE (%)	0.4913	0.1384	0.1165
RMSE (%)	0.5083	0.2669	0.2584
RAE (%)	98.6867	27.807	23.3909
RRSE(%)	101.8876	53.4966	51.7846
Dados Totais	1500	1500	1500

Tabela VI
GRUPO 03: VARIANDO AS ÉPOCAS NO TESTE IV

Configuração	IV ₁	IV ₁₀₀	IV ₁₀₀₀
Neuronios na CO1	4	4	4
Neuronios na CO2	4	4	4
Cl. Correta (%)	49	90.3333	91
Cl. Incorreta (%)	51	9.6667	9
MAE (%)	0.5002	0.1321	0.1318
RMSE (%)	0.5004	0.2748	0.2653
RAE (%)	100.0891	26.4304	26.3714
RRSE(%)	100.099	54.9758	53.0768
Dados Totais	1500	1500	1500

A tabela VI mostra os testes de variação das épocas do grupo 3.

Os parâmetros do modelo que permaneceram inalterados foram a taxa de aprendizado com um valor de 0.3, momentum 0.2, limiar de validação 20.

V. CONCLUSÕES

Ao analisar previamente os dados é possível que o número de dependentes NDEP é majoritariamente composto por uma pessoa.

Para agrupamento 1, tabela I ,a classificação correta, quando

não se utiliza a normalização é insatisfatória com uma taxa de acerto de 52.3%, quase muito semelhante a classificação incorreta que foi de 47.7%. Assim como o erro médio absoluto apresenta 49% e o erro médio quadrático 50%, ou seja o erro está na mesma escala das medidas corretas. Esse desempenho se deve ao fato de que a rede não consegue identificar com clareza os atributos das classes que tem escalas de valores absolutos muito discrepantes. Esse tipo de comportamento é verificado para todos os grupos 1,2 e 3, no teste I.

Ainda no agrupamento 1 teste II, ao se normalizar os dados sem codificação binária dos atributos categóricos nota-se uma melhora no desempenho da rede aonde a classificação correta atinge o valor de 91.3% e a incorreta 8.7%. O erro quadrado médio e o erro absoluto atingem respectivamente as marcas de 0.26% e 0.13%. Esta melhora no desempenho se deve ao fato da normalização trazer para o mesmo patamar de valores os atributos de entrada facilitando o processo de classificação. Este mesmo comportamento se verificou em ambos os grupos de teste.

Na tabela I, variou-se o número de neurônios ocultos da primeira camada para a última camada de 4 para 2. Neste teste pouca diferença foi notada e os valores de erros e acertos permaneceram constantes. Na tabela III, ao invés de diminuir o número de neurônios na camada oculta da primeira para a segunda camadas, optou-se por aumentar o número de neurônios de 4 para 8. Neste segundo caso de estudo os valores permaneceram praticamente inalterados. Na tabela V, aumentou-se o número de neurônios ocultos de 4 para 12, na segunda camada e diminuiu-se o número de neurônios ocultos de 4 para 1 na primeira camada. Neste caso de estudo notou-se uma piora dos índices da rede. A classificação correta caiu para 90.9% e a incorreta aumentou para 9.1% e os erros aumentaram. Esse aumento do erro se deve ao fato de que o arranjo topológico escolhido aglutina em um único nó oculto todos os parâmetros da rede tornando-a um mais ineficiente. Esse comportamento não se verificou nos agrupamentos 1 e 2.

No teste IV, tabela II, ao se variar o número de épocas notou-se melhora nos índices da rede, subindo, por exemplo, o número de acertos de 52.3% para 91.53%. Esse comportamento se verificou para os grupos 2 e 3. Isso se deve ao fato de que na primeira iteração a rede ainda não aprendeu o suficiente do banco de dados ao ponto de realizar uma classificação eficiente. Contudo, a medida que se aumentam o número de épocas para 100 nota-se um rápido aumento de desempenho da rede. O melhor desempenho verificou-se na época 1000, aonde os índices de classificação correta e incorreta obtiveram os melhores resultados atingindo valores de 91.53% e 8.47%. Neste ponto a rede está totalmente treinada o que leva a um aumento expressivo no número de acertos.

Com normalização dos atributos de entrada e utilizando um conjunto de validação, nota-se uma piora nos índices da rede, o que é verificado em todos os grupos de treinamento.

Os melhores resultados foram obtidos agrupando algumas categorias das variáveis ESTC e NDEP. Para isto, utilizou-se o filtro não supervisionado *Merge Two Values*. A Fig. 2, ilustra que o número de atributos utilizados na rede é reduzido a medida que, por exemplo, pessoas solteiras, divorciadas e

viúvas, podem ser consideradas dentro da mesma categoria, a de solteiro, para uma análise bancária. Bem como o atributo número de dependentes pode ser reduzido para um filho ou nenhum filho. Estas considerações fazem com que o desempenho da rede se torne melhor impactando no número de classificações corretas atingindo o seu melhor valor de 92% de acerto na tabela I.

REFERÊNCIAS

- Feldman, J. A., Fanty, M. A., and Goddard, N. H. (1988). Computing With Structured Neural Networks. *Computer*, 21(3):91–103.
- Hagan, M. T., Demuth, H. B., Beale, M. H., and De Jesús, O. (1996). *Neural Network Design*. page 1012.
- Haykin, S. (1999). *Neural Networks: a comprehensive foundation*. Prentice Hall Inc., Porto Alegre, 2 edition.
- Kanal, L. N. (2001). Perceptrons. *Encyclopedia of Computer Science*, pages 11215–11218.
- Krogh, A. (2008). What are artificial neural networks? *Nature biotechnology*, 26(2):195–197.
- Levy, S. (1997). The Computer. *Newsweek*, 130(22):28.
- Mao, J. (1996). Why artificial neural networks? *Communications*, 29:31–44.
- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.
- Minsky, M. and Papert, S. (1969). *Perceptrons*. 2 edition.
- Misra, J. and Saha, I. (2010). Artificial neural networks in hardware: A survey of two decades of progress. *Neurocomputing*, 74(1-3):239–255.
- Nedjah, N., da Silva, F. P., de Sá, A. O., Mourelle, L. M., and Bonilla, D. A. (2016). A massively parallel pipelined reconfigurable design for M-PLN based neural networks for efficient image classification. *Neurocomputing*, 183:39–55.
- Poulton, M. M. (2002). Neural networks as an intelligence amplification tool: A review of applications. *Geophysics*, 67(3):979.
- Rosenblatt, F. (1962). Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms. *Archives of General Psychiatry*, 7:218–219.