

Universidade Federal do Pará – UFPA

Instituto de Tecnologia – ITEC

Faculdade de Engenharia Elétrica e Biomédica

Redes Neurais Artificiais

Artigo Redes Neurais Artificiais

Arthur Gutemberg Hora de Miranda – 202007140043

Arthur Felipe dos Santos Fernandes – 201707140065

Lucas Portilho Nunes – 201707140033

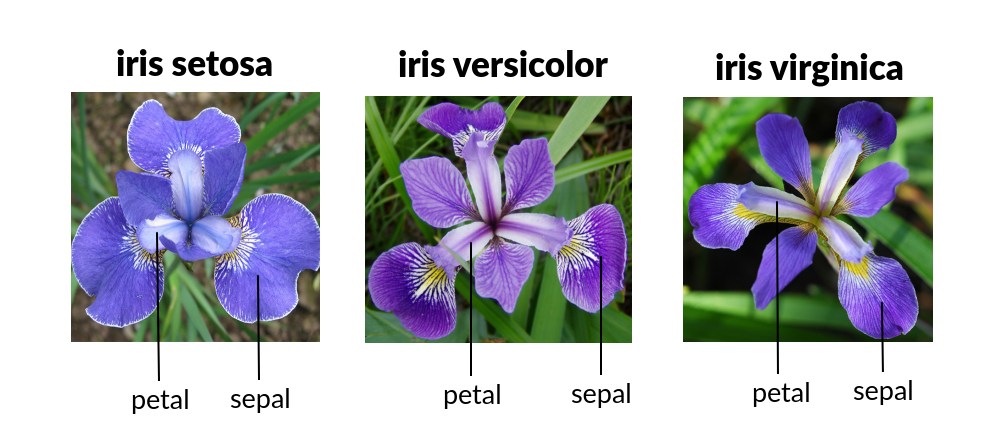
Belém

2021

1. **Introdução**

O banco de dados de Iris, foi criado em 1936 por Ronald A. Fisher, com a premissa de usar medidas para resolver desafios da taxonomia. O objetivo era distinguir espécies da planta Iris, a partir das medições pode-se perceber que as medidas discrepantes permitiam distinguir as espécies (espécies linearmente separáveis), isso permite uma facilidade de identificação para a rede neural, uma vez que as classes e espécies são pré-estabelecidas e linearmente separáveis. O banco consiste em três espécies com 150 amostras de 4 informações: Largura das sépalas; Comprimento das sépalas; Largura das pétalas; Comprimento das pétalas. As espécies analisadas no banco de dados são ilustradas na Figura 1.

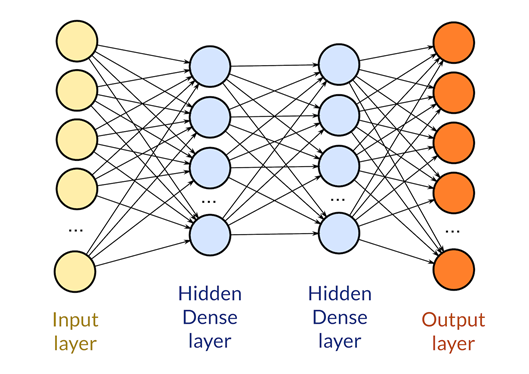
Figura 1 – Espécies de Iris



Fonte: (CHAUHAN, 2021)

Redes neurais são modelos computacionais, inspirados nos neurônios biológicos e na estrutura do cérebro, que a partir de estímulos recebidos, caso a soma destes atinja um certo limiar, o neurônio emite um sinal excitatório para os próximos neurônios, e caso não atinja, o neurônio emite um sinal inibitório. Analogamente aos neurônios biológicos, a rede computacional conta com vários neurônios divididos em três camadas: entrada, oculta ou intermediária, e saída, apresentadas na Figura 2.

Figura 2 – Modelo de rede neural



Fonte:(ALYOUSFI, 2020)

Cada neurônio da rede funciona do mesmo modo, cada entrada possui um peso e a soma de todas as entradas com o *Bias* garante que a contribuição de todos os neurônios afete a resposta, em seguida, a função de ativação é aplicada a saída do somatório, garantindo uma saída entre [0;1]. O processo pode ser observado na figura 3, a equação 1 determina a saída de um dado neurônio.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

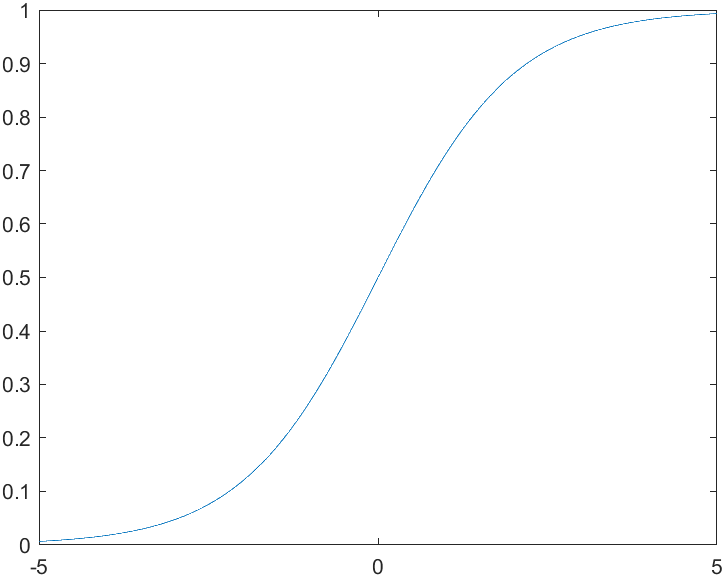
Figura 3 – Modelo de neurônio



Fonte: (HAYKIN, 2001)

Funções de ativação são utilizadas para garantir que a saída do neurônio seja limitada, as funções que normalmente são utilizadas são sigmóide, limiar, limiar por partes, tangente hiperbólica e arco tangente, a função sigmóide é exibido na figura 4.

Figura 4 – Função Sigmóide

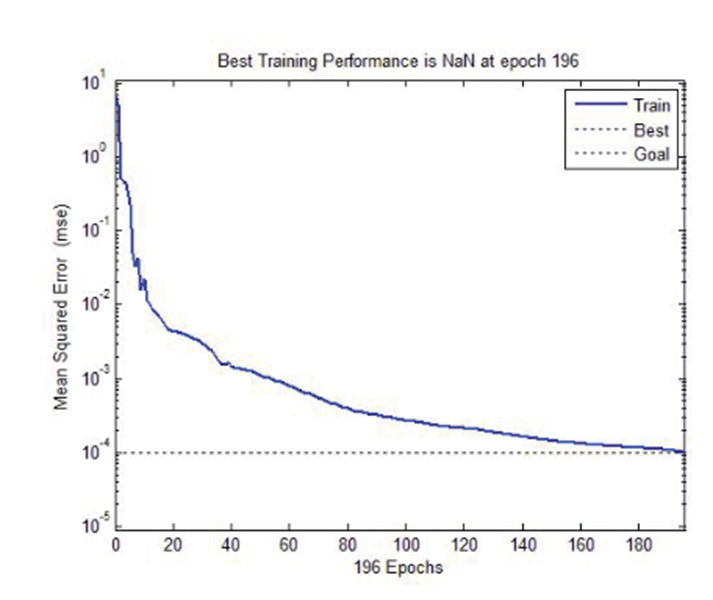


Fonte: Autores

As redes podem ser aplicadas para diversos tipos de problemas, entre eles classificação, categorização e previsão temporal. O banco de dados da Iris representa um problema de classificação, a rede deve ser treinada para ser capaz de distinguir as 3 espécies. No problema abordado, o treinamento é do tipo supervisionado, isto é, a rede vai receber um *feedback* da resposta e dada a diferença entre a saída obtida e saída desejada (performance), junto com o método de treino, os pesos de cada neurônio devem ser atualizados para aumentar o desempenho da rede e se aproximar cada vez mais da resposta correta.

A performance é a representação do aprendizado da rede, a função mais comum é o erro médio quadrático, o gráfico de evolução do erro normalmente é similar a figura 5, quanto menor o erro, mais precisa é a rede; possuindo algumas ressalvas, a minimização da performance é o objetivo de toda fase de treinamento.

Figura 5 – Exemplo de gráfico de performance



Fonte: (MACEDO, 2014)

Um dos algoritmos mais simples para atualização dos pesos é o gradiente descendente que consiste em determinar o gradiente que “aponta” para o mínimo global da função de performance, a cada iteração um novo gradiente é determinado e os pesos são atualizados, à medida que o número de iterações aumenta, a magnitude do gradiente diminui, uma vez que o gradiente esteja abaixo de um certo limiar, o treinamento é finalizado.

Neste artigo foi utilizado o *software* Matlab® que possui funções para o desenvolvimento de redes neurais, permitindo uma facilidade ao usuário para alteração dos parâmetros envolvidos, entre eles, funções de performance e métodos de treino. As funções de performance estão listadas no quadro 1.

Quadro 1 – Funções de performance

|  |  |
| --- | --- |
| **Comando** | **Método** |
| mae | Erro médio absoluto |
| mse | Erro médio quadrático |
| sae | Soma de erro absoluta |
| sse | Soma de erro quadrático |
| crossentropy | Entropia cruzada |
| msesparse | Erro médio quadrático com variáveis acopladas |

Fonte: Adaptado de (MATHWORKS, 2021)

Neste trabalho, foram utilizadas:

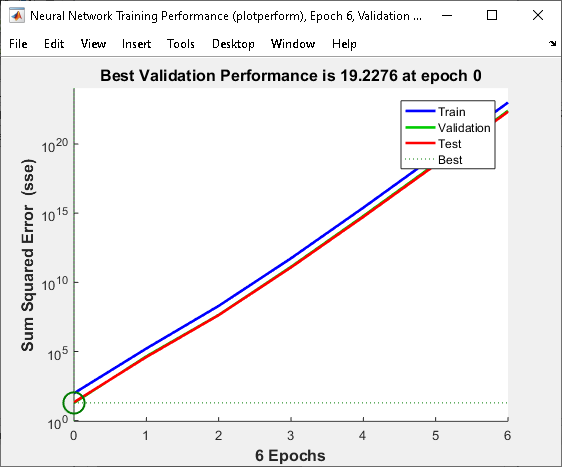
**Erro médio quadrático (MSE)**: Como explicitado anteriormente, este método é comumente utilizado para verificar o processo de evolução do erro em uma MLP, sua expressão é dada pela equação 2. Onde é a saída desejada, é a saída da rede e n é o número de amostras.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

**Soma de erro quadrático (SSE):** O método consiste na aplicação da equação 3. Com base na equação 3, é possível inferir que quanto menor o resultado da soma de erro quadrático, tanto menor é a diferença entre os valores projetados e os obtidos, dessa forma, melhores são os resultados. Neste método são realizadas apenas seis iterações e diferentemente do gráfico do método EQM, seu comportamento é crescente.

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

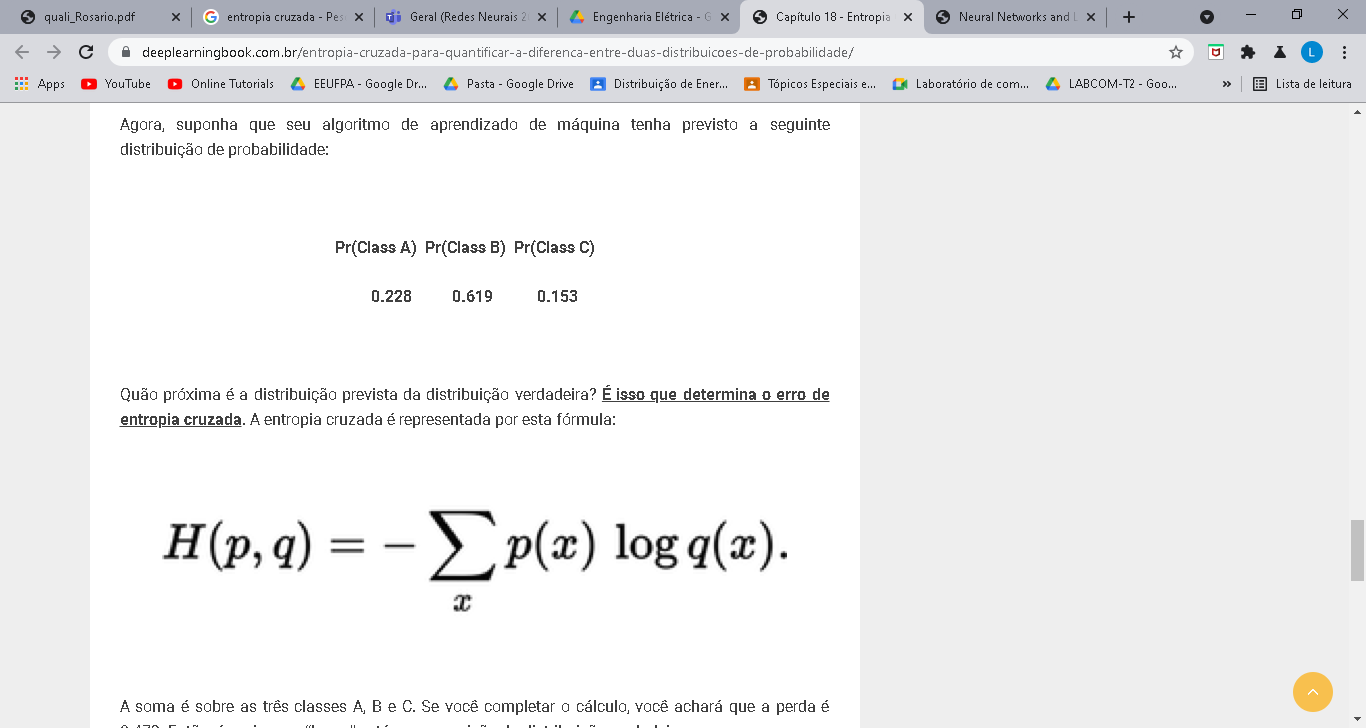
Figura 6: Performance utilizando o método SSE



Fonte: Autores

**Crossentropy**: Este método é particularmente válido quando temos redes com saídas binárias, como normalmente é o caso das redes neurais. Dessa forma imagine que, para um determinado treino, foi proposta a distribuição de probabilidade para a saída de uma rede conforme abaixo.

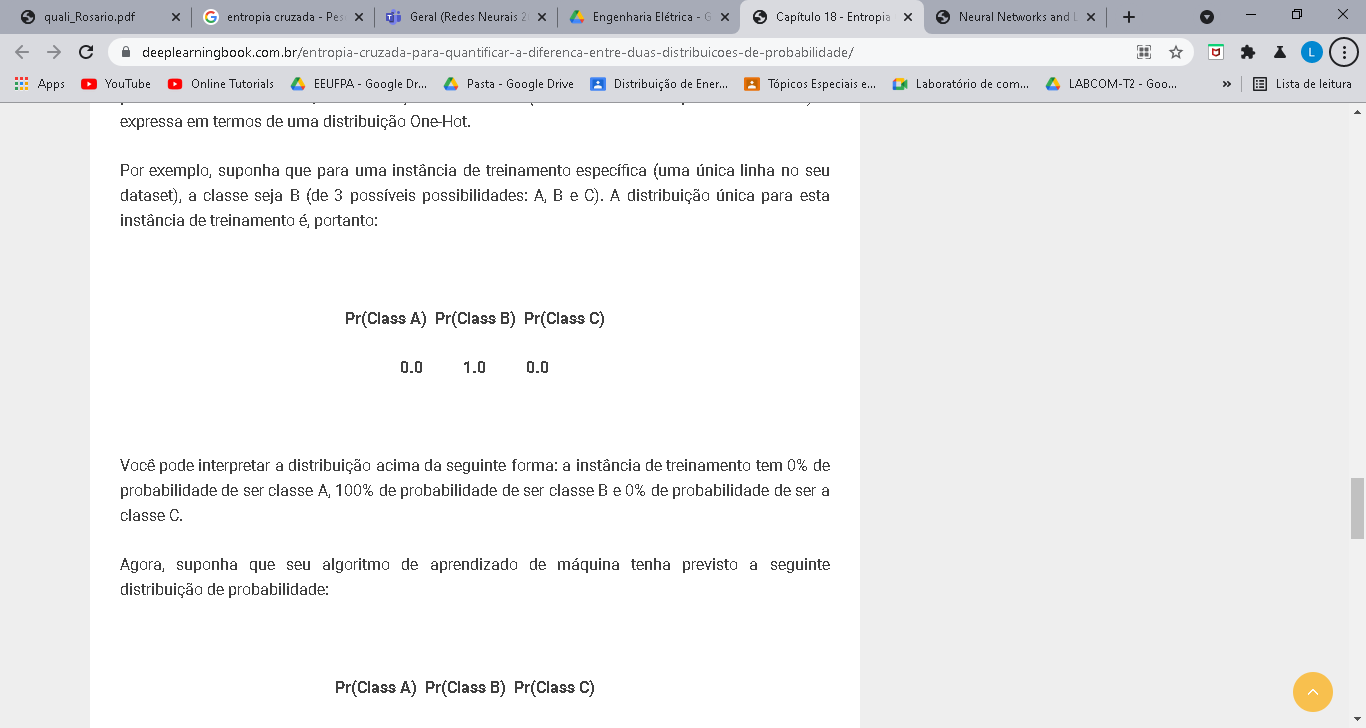
Figura 6 – Distribuição de probabilidade



Fonte: (“Capítulo 18 - Entropia Cruzada Para Quantificar a Diferença Entre Duas Distribuições de Probabilidade - Deep Learning Book”, 2021)

Existem três possíveis classes que podem ser saídas da rede, mas apenas uma é a verdadeira, isto é, deveria possuir 100% de chance de ocorrer:

Figura 7 – Probabilidade por evento



Fonte: (“Capítulo 18 - Entropia Cruzada Para Quantificar a Diferença Entre Duas Distribuições de Probabilidade - Deep Learning Book”, 2021)

Diante disso, dada a distribuição de probabilidade proposta pela máquina e a distribuição de probabilidade real, busca-se analisar, por meio da *crossentropy*, o quão distante elas estão. Sua definição é dada pela equação 4. Perceba que o somatório é dado em função do número de classes na saída.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

**Erro médio absoluto (MAE):** o método consiste na obtenção da média do módulo da diferença entre os valores desejados e os valores obtidos, conforme consta na equação 5.

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Os métodos de treino estão dispostos no quadro 2, neste trabalho foram utilizados os métodos: Levenberg-Marquardt, gradiente conjugado escalonado e o gradiente descendente;

Quadro 2 – Métodos de treino

|  |  |
| --- | --- |
| **Comando** | **Algoritmo** |
| trainlm | Levenberg-Marquardt |
| trainbr | Regularização Bayesiana |
| trainbfg | BFGS Quasi-Newton |
| trainrp | Resilient Backpropagation |
| trainscg | Gradiente Conjugado Escalonado |
| traincgb | Gradiente Conjugado de Powell/Beale |
| traincgf | Gradiente Conjugado de Fletcher-Powell |
| traincgp | Gradiente Conjugado de Polak-Ribiére |
| trainoss | One Step Secant |
| traingdx | Gradiente descendente com taxa de aprendizado variável |
| traingdm | Gradiente descendente com momento |
| traingd | Gradiente descendente |

Fonte: Adaptado de (MATHWORKS, 2021)

1. **Metodologia**

Os experimentos consistiam primeiro na separação dos dados, para garantir que possuiríamos o suficiente para a realização de todas as etapas. Juntamente nessa fase, declaramos o número de neurônios que cada camada teria, para em seguida executarmos os comandos que seriam responsáveis por dar início a etapa de treinamento.

Dividimos as informações obtidas pelo banco de dados de forma que 70% fosse destinado ao treinamento da nossa rede, 15% fosse utilizado para a validação dos resultados, e os últimos 15% reservados para a etapa de teste. A rede possuía apenas 1 camada escondida, e cada experimento variava a sua quantidade entre 1 e 10 neurônios, além de possuir um par único de método de treinamento e avaliador de performance. Cada experimento foi realizado com 3 ensaios consecutivo, com exceção do método de Levenberg-Marquardt que será melhor abordado nas considerações finais. O memorial dos experimentos está apresentado no quadro 3.

Quadro 3 – Memorial de experimento

|  |  |  |
| --- | --- | --- |
| **Teste** | **Método de treino** | **Função de performance** |
| 1 | trainscg | crossentropy |
| 2 | trainscg | mae |
| 3 | trainscg | mse |
| 4 | traingd | sse |
| 5 | traingd | mae |
| 6 | traingd | mse |
| 8 | trainlm | sse |
| 9 | trainlm | mse |

Fonte: Autores

1. **Resultados e discussões**

Tabela 1 – Experimento 1 (trainscg e crossentropy)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 24 | 0,0203 | 96,2 | 91,3 | 95,7 | 95,3 |
| 1 | 2 | 51 | 0,0308 | 96,2 | 100 | 100 | 97,3 |
| 1 | 3 | 29 | 0,027 | 98,1 | 91,3 | 95,7 | 96,7 |
| 2 | 1 | 25 | 0,00707 | 98,1 | 95,7 | 95,7 | 97,3 |
| 2 | 2 | 36 | 0,000495 | 100 | 100 | 91,3 | 98,7 |
| 2 | 3 | 58 | 0,0178 | 98,1 | 100 | 100 | 98,7 |
| 3 | 1 | 14 | 0,00526 | 93,3 | 91,3 | 87 | 92 |
| 3 | 2 | 20 | 0,011 | 98,1 | 100 | 95,7 | 98 |
| 3 | 3 | 42 | 0,000269 | 100 | 100 | 82,6 | 97,3 |
| 4 | 1 | 20 | 0,0156 | 97,1 | 100 | 95,7 | 97,3 |
| 4 | 2 | 14 | 0,00948 | 97,1 | 100 | 95,7 | 97,3 |
| 4 | 3 | 18 | 0,0145 | 97,1 | 91,3 | 100 | 96,7 |
| 5 | 1 | 7 | 0,316 | 41,3 | 56,5 | 26,1 | 41,3 |
| 5 | 2 | 57 | 0,00684 | 99 | 100 | 95,7 | 98,7 |
| 5 | 3 | 28 | 0,00518 | 100 | 95,7 | 95,7 | 98,7 |
| 6 | 1 | 13 | 0,0102 | 96,2 | 95,7 | 95,7 | 96 |
| 6 | 2 | 20 | 0,0131 | 97,1 | 95,7 | 100 | 97,3 |
| 6 | 3 | 15 | 0,0179 | 96,2 | 91,3 | 100 | 96 |
| 7 | 1 | 24 | 0,0144 | 98,1 | 100 | 95,7 | 98 |
| 7 | 2 | 34 | 0,00263 | 100 | 100 | 91,3 | 98,7 |
| 7 | 3 | 11 | 0,00941 | 98,1 | 95,7 | 95,7 | 97,3 |
| 8 | 1 | 11 | 0,0121 | 93,3 | 95,7 | 100 | 94,7 |
| 8 | 2 | 22 | 0,0173 | 98,1 | 100 | 100 | 98,7 |
| 8 | 3 | 26 | 0,0123 | 97,1 | 100 | 95,7 | 97,3 |
| 9 | 1 | 17 | 0,0125 | 98,1 | 95,7 | 100 | 98 |
| 9 | 2 | 46 | 0,00248 | 100 | 100 | 95,7 | 99,3 |
| 9 | 3 | 21 | 0,0154 | 98,1 | 100 | 95,7 | 98 |
| 10 | 1 | 20 | 0,0163 | 97,1 | 100 | 100 | 98 |
| 10 | 2 | 36 | 0,0136 | 98,1 | 100 | 95,7 | 98 |
| 10 | 3 | 29 | 0,0152 | 98,1 | 100 | 95,7 | 98 |

Fonte: Autores

Tabela 2 – Experimento 2 (trainscg e mae)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 81 | 0,253 | 63,5 | 73,9 | 73,9 | 55,7 |
| 1 | 2 | 44 | 0,227 | 67,3 | 60,9 | 69,6 | 66,7 |
| 1 | 3 | 18 | 0,297 | 68,3 | 56,5 | 65,2 | 66 |
| 2 | 1 | 159 | 0,236 | 66,3 | 65,2 | 69,6 | 66,7 |
| 2 | 2 | 43 | 0,285 | 67,3 | 56,5 | 73,9 | 66,7 |
| 2 | 3 | 68 | 0,338 | 48,1 | 52,2 | 43,5 | 48 |
| 3 | 1 | 63 | 0,228 | 67,3 | 73,9 | 56,5 | 66,7 |
| 3 | 2 | 102 | 0,201 | 69,2 | 56,5 | 65,2 | 66,7 |
| 3 | 3 | 15 | 0,345 | 48,1 | 52,2 | 52,2 | 49,3 |
| 4 | 1 | 194 | 0,0918 | 93,3 | 100 | 95,7 | 94,7 |
| 4 | 2 | 274 | 0,0731 | 95,2 | 100 | 95,7 | 96 |
| 4 | 3 | 19 | 0,243 | 73,1 | 65,2 | 78,3 | 72,7 |
| 5 | 1 | 143 | 0,0702 | 95,2 | 87 | 87 | 92,7 |
| 5 | 2 | 49 | 0,246 | 71,2 | 69,6 | 82,6 | 72,7 |
| 5 | 3 | 41 | 0,223 | 67,3 | 73,9 | 56,5 | 66,7 |
| 6 | 1 | 179 | 0,0777 | 94,2 | 95,7 | 100 | 95,3 |
| 6 | 2 | 67 | 0,131 | 95,2 | 100 | 91,3 | 95,3 |
| 6 | 3 | 46 | 0,106 | 95,2 | 95,7 | 96,7 | 95,3 |
| 7 | 1 | 86 | 0,123 | 94,2 | 95,7 | 91,3 | 94 |
| 7 | 2 | 50 | 0,146 | 92,3 | 100 | 100 | 94,7 |
| 7 | 3 | 306 | 0,0566 | 98,1 | 100 | 100 | 98,7 |
| 8 | 1 | 89 | 0,118 | 96,2 | 91,3 | 100 | 96 |
| 8 | 2 | 246 | 0,115 | 91,3 | 82,6 | 95,7 | 90,7 |
| 8 | 3 | 31 | 0,176 | 88,5 | 91,3 | 87 | 88,7 |
| 9 | 1 | 19 | 0,301 | 63,5 | 78,3 | 56,5 | 64,7 |
| 9 | 2 | 36 | 0,225 | 74 | 82,6 | 69,6 | 74,7 |
| 9 | 3 | 23 | 0,217 | 82,7 | 87 | 78,3 | 82,7 |
| 10 | 1 | 87 | 0,102 | 96,2 | 100 | 91,3 | 96 |
| 10 | 2 | 114 | 0,0862 | 96,2 | 95,7 | 91,3 | 95,3 |
| 10 | 3 | 25 | 0,229 | 71,2 | 73,9 | 73,9 | 72 |

Fonte: Autores

Tabela 3 – Experimento 3 (trainscg e mse)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 52 | 0,112 | 67,3 | 56,5 | 73,9 | 66,7 |
| 1 | 2 | 17 | 0,106 | 71,2 | 52,2 | 60,9 | 66,7 |
| 1 | 3 | 41 | 0,109 | 67,3 | 56,5 | 73,9 | 66,7 |
| 2 | 1 | 35 | 0,00716 | 99 | 95,7 | 95,7 | 98 |
| 2 | 2 | 18 | 0,12 | 72,1 | 73,9 | 78,3 | 73,3 |
| 2 | 3 | 42 | 0,0092 | 99 | 91,3 | 100 | 98 |
| 3 | 1 | 34 | 0,0104 | 98,1 | 95,7 | 100 | 98 |
| 3 | 2 | 46 | 0,0113 | 98,1 | 100 | 100 | 98,7 |
| 3 | 3 | 48 | 0,00687 | 98,1 | 100 | 95,7 | 98 |
| 4 | 1 | 20 | 0,0205 | 94,2 | 100 | 100 | 96 |
| 4 | 2 | 26 | 0,0146 | 98,1 | 95,7 | 100 | 98 |
| 4 | 3 | 65 | 0,00834 | 98,1 | 100 | 91,3 | 97,3 |
| 5 | 1 | 20 | 0,0159 | 97,1 | 95,7 | 100 | 97,3 |
| 5 | 2 | 38 | 0,00906 | 99 | 100 | 95,7 | 98,7 |
| 5 | 3 | 50 | 0,0113 | 98,1 | 95,7 | 95,7 | 97,3 |
| 6 | 1 | 36 | 0,0151 | 97,1 | 100 | 95,7 | 97,3 |
| 6 | 2 | 27 | 0,0134 | 98,1 | 100 | 95,7 | 98 |
| 6 | 3 | 40 | 0,0129 | 97,1 | 100 | 100 | 98 |
| 7 | 1 | 11 | 0,0195 | 93,3 | 91,3 | 100 | 94 |
| 7 | 2 | 36 | 0,0118 | 98,1 | 95,7 | 91,3 | 96,3 |
| 7 | 3 | 31 | 0,0139 | 97,1 | 100 | 100 | 98 |
| 8 | 1 | 57 | 0,0115 | 96,2 | 100 | 95,7 | 96,7 |
| 8 | 2 | 37 | 0,00932 | 99 | 95,7 | 91,3 | 97,3 |
| 8 | 3 | 34 | 0,0175 | 97,1 | 100 | 100 | 98 |
| 9 | 1 | 31 | 0,015 | 97,1 | 100 | 100 | 98 |
| 9 | 2 | 48 | 0,0102 | 99 | 100 | 95,7 | 98,7 |
| 9 | 3 | 33 | 0,0167 | 97,1 | 100 | 100 | 98 |
| 10 | 1 | 33 | 0,0139 | 98,1 | 95,7 | 91,3 | 96,7 |
| 10 | 2 | 33 | 0,00953 | 98,1 | 95,7 | 95,7 | 97,3 |
| 10 | 3 | 17 | 0,011 | 98,1 | 91,3 | 91,3 | 96 |

Fonte: Autores

Tabela 4 – Experimento 4 (traingd e sse)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 6 | 99,1 | 60,6 | 65,2 | 73,9 | 63,3 |
| 1 | 2 | 6 | 110 | 64,4 | 69,6 | 73,9 | 66,7 |
| 1 | 3 | 6 | 72,4 | 67,3 | 56,5 | 69,6 | 66 |
| 2 | 1 | 6 | 160 | 64,4 | 69,6 | 69,6 | 66 |
| 2 | 2 | 6 | 90,3 | 54,8 | 56,5 | 60,9 | 56 |
| 2 | 3 | 6 | 84,5 | 66,3 | 65,2 | 69,6 | 66,7 |
| 3 | 1 | 6 | 96,7 | 55,8 | 43,5 | 52,2 | 53,3 |
| 3 | 2 | 6 | 121 | 66,3 | 73,9 | 60,9 | 66,7 |
| 3 | 3 | 6 | 89,1 | 56,7 | 47,8 | 73,9 | 58 |
| 4 | 1 | 6 | 96,3 | 69,2 | 47,8 | 69,6 | 66 |
| 4 | 2 | 6 | 144 | 56,7 | 69,6 | 60,9 | 59,3 |
| 4 | 3 | 6 | 62,5 | 63,5 | 60,9 | 65,2 | 63,3 |
| 5 | 1 | 6 | 145 | 62,5 | 78,3 | 73,9 | 66,7 |
| 5 | 2 | 6 | 263 | 61,5 | 78,3 | 65,2 | 64,7 |
| 5 | 3 | 6 | 66,8 | 63,5 | 56,5 | 56,5 | 61,3 |
| 6 | 1 | 6 | 247 | 60,6 | 69,6 | 69,6 | 63,3 |
| 6 | 2 | 6 | 252 | 57,7 | 60,9 | 60,9 | 58,7 |
| 6 | 3 | 6 | 131 | 57,7 | 73,9 | 60,9 | 60,7 |
| 7 | 1 | 6 | 148 | 68,3 | 73,9 | 56,5 | 67,3 |
| 7 | 2 | 6 | 84 | 67,3 | 60,9 | 65,2 | 66 |
| 7 | 3 | 6 | 193 | 64,4 | 65,2 | 56,5 | 63,3 |
| 8 | 1 | 6 | 238 | 55,8 | 60,9 | 52,2 | 56 |
| 8 | 2 | 6 | 277 | 50 | 56,5 | 52,2 | 51,3 |
| 8 | 3 | 6 | 128 | 64,4 | 78,3 | 65,2 | 66,7 |
| 9 | 1 | 6 | 145 | 68,3 | 65,2 | 60,9 | 66,7 |
| 9 | 2 | 6 | 131 | 73,1 | 56,5 | 56,5 | 68 |
| 9 | 3 | 6 | 152 | 55,8 | 65,2 | 56,5 | 57,3 |
| 10 | 1 | 6 | 133 | 56,7 | 60,9 | 52,2 | 56,7 |
| 10 | 2 | 6 | 323 | 65,4 | 73,9 | 56,5 | 65,3 |
| 10 | 3 | 6 | 117 | 65,4 | 69,6 | 73,9 | 67,3 |

Fonte: Autores

Tabela 5 – Experimento 5 (traingd e mae)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 52 | 0,208 | 71,2 | 52,2 | 60,9 | 66,7 |
| 1 | 2 | 57 | 0,233 | 66,3 | 60,9 | 73,9 | 66,7 |
| 1 | 3 | 34 | 0,257 | 62,5 | 78,3 | 73,9 | 66,7 |
| 2 | 1 | 15 | 0,34 | 50 | 47,8 | 43,5 | 48,7 |
| 2 | 2 | 37 | 0,259 | 68,3 | 60,9 | 60,9 | 66 |
| 2 | 3 | 29 | 0,34 | 40,4 | 43,5 | 26,1 | 38,7 |
| 3 | 1 | 55 | 0,134 | 90,4 | 82,6 | 100 | 90,7 |
| 3 | 2 | 33 | 0,252 | 64,4 | 73,9 | 69,6 | 66,7 |
| 3 | 3 | 57 | 0,245 | 64,4 | 73,9 | 69,6 | 66,7 |
| 4 | 1 | 149 | 0,101 | 94,2 | 100 | 100 | 96 |
| 4 | 2 | 21 | 0,252 | 68,3 | 56,5 | 69,6 | 66,7 |
| 4 | 3 | 77 | 0,121 | 94,2 | 87 | 95,7 | 93,3 |
| 5 | 1 | 115 | 0,119 | 96,2 | 95,7 | 100 | 96,7 |
| 5 | 2 | 40 | 0,185 | 77,9 | 82,6 | 73,9 | 78 |
| 5 | 3 | 123 | 0,111 | 96,2 | 87 | 91,3 | 94 |
| 6 | 1 | 35 | 0,172 | 93,3 | 100 | 78,3 | 92 |
| 6 | 2 | 67 | 0,139 | 95,2 | 95,7 | 100 | 96 |
| 6 | 3 | 100 | 0,105 | 99 | 95,7 | 100 | 98,7 |
| 7 | 1 | 55 | 0,18 | 88,5 | 73,9 | 91,3 | 86,7 |
| 7 | 2 | 55 | 0,155 | 97,1 | 95,7 | 95,7 | 96,7 |
| 7 | 3 | 112 | 0,129 | 95,2 | 100 | 95,7 | 96 |
| 8 | 1 | 120 | 0,145 | 91,3 | 95,7 | 100 | 93,3 |
| 8 | 2 | 33 | 0,189 | 89,4 | 95,7 | 87 | 90 |
| 8 | 3 | 23 | 0,224 | 80,8 | 87 | 91,3 | 83,3 |
| 9 | 1 | 69 | 0,17 | 90,4 | 69,6 | 87 | 86,7 |
| 9 | 2 | 66 | 0,17 | 96,2 | 95,7 | 95,7 | 96 |
| 9 | 3 | 24 | 0,208 | 83,7 | 100 | 78,3 | 85,3 |
| 10 | 1 | 51 | 0,155 | 98,1 | 87 | 95,7 | 96 |
| 10 | 2 | 32 | 0,217 | 86,5 | 91,3 | 91,3 | 88 |
| 10 | 3 | 7 | 0,308 | 62,5 | 78,3 | 65,2 | 65,3 |

Fonte: Autores

Tabela 6 – Experimento 6 (traingd e mse)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 112 | 0,11 | 74 | 82,6 | 91,3 | 78 |
| 1 | 2 | 43 | 0,12 | 96,2 | 95,7 | 100 | 96,7 |
| 1 | 3 | 115 | 0,12 | 80,8 | 78,3 | 82,6 | 80,7 |
| 2 | 1 | 541 | 0,02 | 94,2 | 100 | 95,7 | 95,3 |
| 2 | 2 | 745 | 0,01 | 98,1 | 95,7 | 91,3 | 96,7 |
| 2 | 3 | 407 | 0,02 | 97,1 | 95,7 | 100 | 97,3 |
| 3 | 1 | 36 | 0,11 | 70,2 | 60,9 | 78,3 | 70 |
| 3 | 2 | 167 | 0,02 | 96,2 | 95,7 | 100 | 96,7 |
| 3 | 3 | 47 | 0,11 | 71,2 | 69,6 | 56,5 | 68,7 |
| 4 | 1 | 191 | 0,01 | 99 | 95,7 | 91,3 | 97,3 |
| 4 | 2 | 355 | 0,02 | 99 | 100 | 95,7 | 98,7 |
| 4 | 3 | 115 | 0,02 | 97,1 | 100 | 95,7 | 97,3 |
| 5 | 1 | 112 | 0,02 | 96,2 | 100 | 91,3 | 96 |
| 5 | 2 | 271 | 0,01 | 98,1 | 95,7 | 95,7 | 97,3 |
| 5 | 3 | 215 | 0,02 | 97,1 | 100 | 91,3 | 96,7 |
| 6 | 1 | 187 | 0,03 | 96,2 | 100 | 100 | 97,3 |
| 6 | 2 | 423 | 0,01 | 98,1 | 91,3 | 100 | 97,3 |
| 6 | 3 | 221 | 0,02 | 97,1 | 100 | 95,7 | 97,3 |
| 7 | 1 | 77 | 0,02 | 96,2 | 95,7 | 95,7 | 96 |
| 7 | 2 | 161 | 0,02 | 98,1 | 91,3 | 95,7 | 96,7 |
| 7 | 3 | 231 | 0,01 | 99 | 91,3 | 95,7 | 97,3 |
| 8 | 1 | 125 | 0,02 | 95,2 | 100 | 95,7 | 96 |
| 8 | 2 | 280 | 0,02 | 98,1 | 95,7 | 95,7 | 97,3 |
| 8 | 3 | 103 | 0,02 | 99 | 100 | 87 | 97,3 |
| 9 | 1 | 315 | 0,02 | 97,1 | 100 | 91,3 | 96,7 |
| 9 | 2 | 196 | 0,01 | 99 | 95,7 | 95,7 | 98 |
| 9 | 3 | 91 | 0,02 | 100 | 87 | 91,3 | 96,7 |
| 10 | 1 | 329 | 0,01 | 98,1 | 100 | 95,7 | 98 |
| 10 | 2 | 188 | 0,03 | 97,1 | 100 | 95,7 | 97,3 |
| 10 | 3 | 188 | 0,02 | 98,1 | 95,7 | 100 | 98 |

Fonte: Autores

Tabela 7 – Experimento 7 (trainlm e sse)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 184 | 34,7 | 69,2 | 60,9 | 60,9 | 66,7 |
| 1 | 2 | 27 | 36 | 65,4 | 78,3 | 60,9 | 66,7 |
| 1 | 3 | 256 | 34,2 | 69,2 | 60,9 | 60,9 | 66,7 |
| 2 | 1 | 20 | 1,94 | 98,1 | 100 | 100 | 98,7 |
| 2 | 2 | 26 | 3,46 | 98,1 | 100 | 100 | 98,7 |
| 2 | 3 | 25 | 8,74E-12 | 100 | 100 | 95,7 | 99,3 |
| 3 | 1 | 15 | 0,000228 | 99 | 95,7 | 95,7 | 98 |
| 3 | 2 | 13 | 0,000323 | 98,1 | 95,7 | 95,7 | 97,3 |
| 3 | 3 | 12 | 2 | 99 | 91,3 | 100 | 98 |
| 4 | 1 | 15 | 0,0000115 | 100 | 100 | 95,7 | 99,3 |
| 4 | 2 | 9 | 0,00000196 | 100 | 91,3 | 91,3 | 97,3 |
| 4 | 3 | 100 | 0,0113 | 100 | 91,3 | 100 | 98,7 |
| 5 | 1 | 12 | 1,97 | 99 | 100 | 91,3 | 98 |
| 5 | 2 | 28 | 1,7E-10 | 100 | 100 | 95,7 | 99,3 |
| 5 | 3 | 11 | 2,72 | 98,1 | 100 | 95,7 | 98 |
| 6 | 1 | 12 | 0,0000088 | 100 | 91,3 | 100 | 98,7 |
| 6 | 2 | 13 | 3,12 | 96,2 | 100 | 100 | 97,3 |
| 6 | 3 | 16 | 1,99 | 98,1 | 100 | 95,7 | 98 |
| 7 | 1 | 10 | 1 | 97,1 | 91,3 | 100 | 96,7 |
| 7 | 2 | 16 | 1,21 | 99 | 95,7 | 100 | 98,7 |
| 7 | 3 | 13 | 0,763 | 99 | 95,7 | 95,7 | 98 |
| 8 | 1 | 61 | 5,83E-13 | 100 | 100 | 87 | 98 |
| 8 | 2 | 13 | 3,21 | 97,1 | 100 | 100 | 98 |
| 8 | 3 | 15 | 1,73 | 97,1 | 100 | 100 | 98 |
| 9 | 1 | 210 | 6,43E-08 | 100 | 100 | 91,3 | 98,3 |
| 9 | 2 | 12 | 0,605 | 99 | 91,3 | 95,7 | 97,3 |
| 9 | 3 | 15 | 1,8 | 99 | 95,7 | 91,3 | 97,3 |
| 10 | 1 | 15 | 2,79 | 97,1 | 100 | 95,7 | 97,3 |
| 10 | 2 | 17 | 1,79 | 98,1 | 100 | 91,3 | 97,3 |
| 10 | 3 | 11 | 0,975 | 100 | 95,7 | 95,7 | 98,7 |

Fonte: Autores

Tabela 8 – Experimento 8 (trainlm e mse)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neurônios | Ensaio | Épocas | Performance | Treinamento (%) | Validação (%) | Teste (%) | Total (%) |
| 1 | 1 | 9 | 0,121 | 63,5 | 78,3 | 69,6 | 66,7 |
| 1 | 2 | 16 | 0,119 | 65,4 | 78,3 | 60,9 | 66,7 |
| 1 | 3 | 49 | 0,114 | 66,3 | 65,2 | 69,6 | 66,7 |
| 2 | 1 | 12 | 0,00000279 | 99 | 91,3 | 100 | 98 |
| 2 | 2 | 12 | 0,00000439 | 98,1 | 100 | 95,7 | 98 |
| 2 | 3 | 22 | 0,00623 | 99 | 100 | 100 | 99,3 |
| 3 | 1 | 13 | 0,00000245 | 99 | 95,7 | 100 | 98,7 |
| 3 | 2 | 12 | 0,00624 | 99 | 100 | 95,7 | 98,7 |
| 3 | 3 | 17 | 1,74E-08 | 100 | 95,7 | 91,3 | 98 |
| 4 | 1 | 10 | 0,00675 | 98,1 | 95,7 | 100 | 98 |
| 4 | 2 | 22 | 0,00624 | 99 | 100 | 95,7 | 98,7 |
| 4 | 3 | 12 | 0,000000368 | 100 | 95,7 | 95,7 | 98,7 |
| 5 | 1 | 15 | 0,0108 | 98,1 | 100 | 100 | 98,7 |
| 5 | 2 | 38 | 0,00558 | 99 | 100 | 95,7 | 98,7 |
| 5 | 3 | 11 | 0,11 | 98,1 | 100 | 100 | 98,7 |
| 6 | 1 | 12 | 0,000272 | 99 | 95,7 | 100 | 98,7 |
| 6 | 2 | 15 | 0,00000119 | 99 | 100 | 95,7 | 98,7 |
| 6 | 3 | 13 | 0,0000826 | 100 | 100 | 91,3 | 98,7 |
| 7 | 1 | 11 | 0,00596 | 99 | 95,7 | 95,7 | 98 |
| 7 | 2 | 10 | 0,00831 | 97,1 | 95,7 | 95,7 | 96,7 |
| 7 | 3 | 11 | 0,00000271 | 99 | 91,3 | 95,7 | 97,3 |
| 8 | 1 | 18 | 7,08E-13 | 100 | 91,3 | 95,7 | 98 |
| 8 | 2 | 10 | 0,00606 | 98,1 | 95,7 | 91,3 | 96,7 |
| 8 | 3 | 8 | 0,00000255 | 98,1 | 95,7 | 100 | 98 |
| 9 | 1 | 13 | 0,0011 | 100 | 91,3 | 100 | 98,7 |
| 9 | 2 | 10 | 0,00169 | 99 | 87 | 100 | 97,3 |
| 9 | 3 | 10 | 0,00617 | 98,1 | 91,3 | 95,7 | 96,7 |
| 10 | 1 | 9 | 0,00161 | 97,1 | 100 | 87 | 96 |
| 10 | 2 | 10 | 0,00344 | 99 | 100 | 95,7 | 98,7 |
| 10 | 3 | 10 | 0,00243 | 99 | 95,7 | 95,7 | 98 |

Fonte: Autores

Embora em diversos experimentos a taxa de acerto dos dados de treinamento e teste tenham sido superiores a 95%, a melhor topologia é justamente a que apresenta melhor aproveitamento com menor número de neurônios na camada oculta. No quadro 4 são apresentados os melhores resultados de cada experimento. Foi levado em consideração os valores das fases de validação e teste para uma melhor observação do desenvolvimento da rede.

Quadro 4 – Melhores resultados

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Neurônios** | **Ensaio** | **Método de treino** | **Função de performance** | **Validação (%)** | **Teste (%)** |
| 1 | 2 | trainscg | crossentropy | 100 | 100 |
| 7 | 2 | trainscg | mae | 100 | 100 |
| 3 | 2 | trainscg | mse | 100 | 100 |
| 5 | 1 | traingd | sse | 78,3 | 73,9 |
| 4 | 1 | traingd | mae | 100 | 100 |
| 6 | 1 | traingd | mse | 100 | 100 |
| 2 | 1 | trainlm | sse | 100 | 100 |
| 2 | 3 | trainlm | mse | 100 | 100 |

Fonte: Autores

Por fim, foi possível analisar que a rede apresentou um bom desenvolvimento com quase todos os métodos, entretanto, como mencionado anteriormente, a melhor rede é a que apresenta o melhor resultado enquanto trabalha com o menor número de neurônios na camada escondida, com isso, a melhor rede é a que apresenta 2 (dois) neurônios na camada oculta.

O método de treino Levenberg-Marquardt se demonstrou incapaz de funcionar com as funções *crossentropy*, MAE e SAE. O método de treino gradiente descendente apresentou baixo aprendizado com a função *crossentropy*.

1. **Conclusão**

A partir dos experimentos foi possível perceber que o método do gradiente conjugado escalonado apresentou as maiores taxas de validação e teste, enquanto que o método de Levenberg-Marquardt obtinha os menores valores de performance. Em contrapartida, a função de performance MSE apresentou a maior consistência, pecando apenas no número de épocas.

Também foi possível perceber que a rede apresenta dificuldades para diferenciar as 3 classes de saída com apenas 1 neurônio na camada intermediária, como nos experimentos 3, 4, 5, 7 e 8.

Apesar das diversas inicializações, durante a elaboração deste trabalho ficou evidente a necessidade de se assumir uma quantidade superior de ensaios, uma vez que, a inicialização randômica dos pesos aliada a amostragem aleatória do banco de dados, acabam por vezes apresentando redes com alto desvio padrão na taxa de acerto nas camadas de treino, validação e teste, como é o caso da rede 5-1 no experimento 1 e das redes 2-1 e 2-3 no experimento 5. Também ficou evidente como a utilização de diversos métodos de treinamento pode enriquecer um trabalho.

Referências:

ALYOUSFI, A. **Deep Learning in Keras - Building a Deep Learning Model**. Disponível em: <https://stackabuse.com/deep-learning-in-keras-building-a-deep-learning-model>. Acesso em: 15 jun. 2021.

**Capítulo 18 - Entropia Cruzada Para Quantificar a Diferença Entre Duas Distribuições de Probabilidade - Deep Learning Book**. Disponível em: <https://www.deeplearningbook.com.br/entropia-cruzada-para-quantificar-a-diferenca-entre-duas-distribuicoes-de-probabilidade/>. Acesso em: 17 jun. 2021.

CHAUHAN, G. **Iris Dataset Project from UCI Machine Learning Repository - Machine Learning HD**. Disponível em: <http://machinelearninghd.com/iris-dataset-uci-machine-learning-repository-project/>. Acesso em: 15 jun. 2021.

HAYKIN, S. **Redes Neurais - Princípios e Prática**, 2001.

MACEDO, M. DAS N. Q. DE. **Classificação automática de curvas de carga usando RNA no ambiente Smart Grid**. [s.l.] Universidade Federal da Bahia, 2014.

MATHWORKS. **Matlab documentationMatlab**, 2021. Disponível em: <https://www.mathworks.com/help/matlab/>. Acesso em: 15 jun. 2021