

Construção de um Classificador de Dados de flores Iris utilizando Classificação associativa com Redes Neural

Adão de Souza

7º EGCA

adaosouza19@yahoo.com.br

Davidson Ferreira

7º EGCA

davidson55n@yahoo.com.br

Francisco Jr.

7º EGCA

indiomg@hotmail.co.uk

Philip Rodrigues

7º EGCA

philipmp29@gmail.com

• Flor Iris

Iris é um gênero de plantas com flor, muito apreciado pelas suas diversas espécies, que ostentam flores de cores muito vivas. São, vulgarmente, designadas como lírios, embora tal termo se aplique com mais propriedade a outro tipo de flor. É uma flor muito frequente em jardins.



IRIS SETOSA



IRIS VERSICOLOR



IRIS VIRGINICA

Figura 1 – Classes de flores Iris

• Banco de dados da Flor Iris

Os bancos de dados da flor Iris ou conjunto de dados Iris de Fisher, é um conjunto multivariado de dados introduzido por Ronald Fisher em seu artigo de 1936. O uso de múltiplas medições em problemas taxonômicos como um exemplo de análise discriminante linear. Às vezes é chamado de conjunto de dados Anderson Iris porque Edgar Anderson recolheu os dados para quantificar a variação morfológicas das flores iris de três espécies relacionadas. Duas das três espécies foram coletadas na Península de Gaspé "todos do mesmo pasto, e pegou no mesmo dia e medida na mesmo tempo pela mesma pessoa, com o mesmo aparelho". O conjunto de dados consiste de 50 amostras de cada uma das três espécies de Iris (Iris setosa, Iris versicolor e Iris virginica). Quatro características foram medidas a partir de cada uma das amostras, o comprimento e a largura das sépalas e pétalas, em centímetros. Com base na combinação destas quatro características, Fisher desenvolveu um modelo discriminante linear para distinguir as espécies de cada outro[3].

- **Redes Neurais Artificiais (RNA)**

Redes neurais são uma solução computacional que envolve o desenvolvimento de estruturas matemáticas com a habilidade de aprendizagem. Os métodos são resultados de investigações acadêmicas para modelagem da aprendizagem do sistema nervoso. As redes neurais têm uma notável habilidade de derivar médias de dados complicados ou imprecisos e podem ser utilizadas para extrair padrões e detectar tendências que são muito complexas para serem percebidas tanto por humanos quanto por outras técnicas computacionais. Uma rede neural treinada pode ser considerada como um "expert" na categoria de informação que foi previamente para a análise. Este expert pode então ser utilizado para fornecer projeções, dadas novas situações de interesse[1].

- **Descrição de dados**

Este projeto usa o melhor conhecimento do conjunto de dados da planta Iris do UCI repository of Machine Learning Database, que se refere a 3 classes de 50 instâncias cada. Onde cada classe refere-se a um tipo de planta. A primeira classe é linearmente distinguível das outras duas restantes, com a segunda não sendo linearmente separada das outras. As 150 instâncias, que são igualmente separadas em 3 classes, contém os seguintes quatro atributos: comprimento e largura da sépala, comprimento e largura da pétala. Uma sépala é uma divisão no cálice, que é a camada protetora da flor em broto, e uma pétala é uma divisão da flor. Os valores mínimos para os dados brutos contidos no conjunto de dados são os seguintes (medições em centímetros): comprimento da sépala (4.3), largura da sépala (2.0), comprimento da pétala (1.0), largura da pétala (0.1). Os valores máximos para os dados brutos contidos no conjunto de dados são os seguintes (medições em centímetros): comprimento da sépala (7.9), largura da sépala (4.4), comprimento da pétala (6.9), largura da pétala (2.5). Cada instância vai pertencer a uma das seguintes classes: Iris Setosa, Iris Versicolour ou Iris Virginica[2].

- **Metodologia**

Metodologia para geração do conjunto de Regras da Associação de Classes. Primeiro, encontrando as Regras de Associação de Classes do conjunto de dados pode ser útil em vários contextos. Em geral entendendo as regras que relacionam atributo da classe com os atributos previstos. As regras de extração da associação foram alcançadas usando o algoritmo proposto. As regras são extraídas dependendo da relação entre a previsão e os atributos das classes. O conjunto de dados da fig. 1 é um segmento do conjunto de dados da planta Iris.

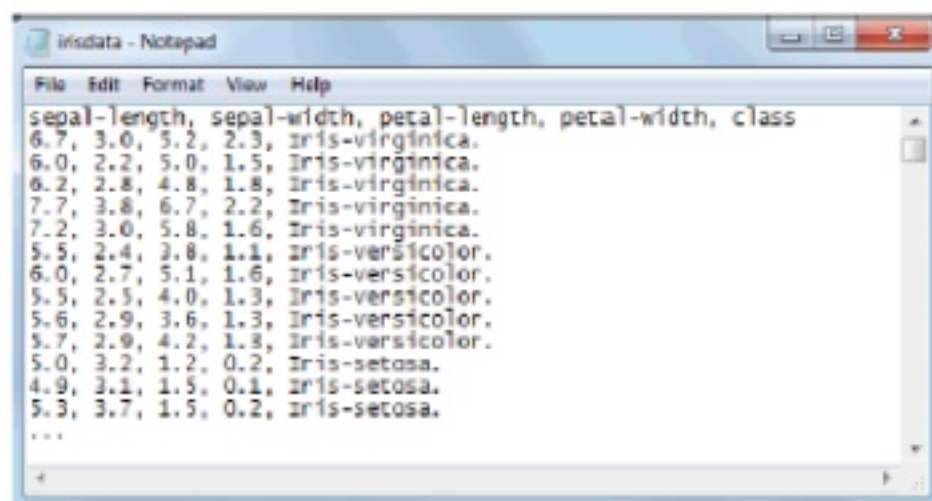


Figura 2 – irisdata

O alvo é usar o algoritmo proposto, para encontrar relações entre as características e as representar nas Regras de Associação de Classes para formar a entrada para a rede neural Feedforward Backpropagation. As três classes da Iris foram alocadas numericamente e representadas como mostra na tabela 1.

TABELA 1

Class	Numeric representation
Iris Virginica.	1
Iris Versicolour	2
Iris Setosa	3

- **O sistema proposto:**

Depois de muitas pesquisas e de diversas tentativas sem sucesso, decidimos usar uma Rede Neural de Classificação associativa.

Aplicar uma Rede Neural de Classificação associativa, passa por três etapas :

Pré-processamento, geração de regras de associação de classe e criar rede neural backpropagation de Classificação Associativo.

1. Pré – tratamento:

O principal objetivo desta etapa é preparar os dados para o próximo passo :

Associação de Classe na Mineração de Regras. Este passo inicia-se com o processo de transformação da base de dados original . Depois de que os atributos contínuos são submetidos ao processo de discretização .

2. Transformação:

A extracção de dados com base na rede neural só pode lidar com os dados numéricos por isso é necessária transformar a data personagem em dados numéricos.

3. Discretização:

classificação de conjuntos de dados costumam conter muitos atributos contínuos. Mineirar as regras de associação com atributos contínuos ainda é uma questão de pesquisa . Sendo assim, o nosso sistema envolve a discretizaçãode atributos contínuos na classe alvo (Target) pre-determinada. [2]

	A	B	C	D	E	F	G
41	5.0	3.5	1.3	0.3			iris-setosa
42	4.5	2.3	1.3	0.3			iris-setosa
43	4.4	3.2	1.3	0.2			iris-setosa
44	5.0	3.5	1.6	0.6			iris-setosa
45	5.1	3.8	1.9	0.4			iris-setosa
46	4.8	3.0	1.4	0.3			iris-setosa
47	5.1	3.8	1.6	0.2			iris-setosa
48	4.6	3.2	1.4	0.2			iris-setosa
49	5.3	3.7	1.5	0.2			iris-setosa
50	5.0	3.3	1.4	0.2			iris-setosa
51	7.0	3.2	4.7	1.4			iris-versicolour
52	6.4	3.2	4.5	1.5			iris-versicolour
53	6.9	3.1	4.9	1.5			iris-versicolour
54	5.5	2.3	4.0	1.3			iris-versicolour
55	6.5	2.8	4.6	1.5			iris-versicolour
56	5.7	2.8	4.5	1.3			iris-versicolour
57	6.3	3.3	4.7	1.6			iris-versicolour
58	4.9	2.4	3.3	1.0			iris-versicolour
59	6.6	2.9	4.6	1.3			iris-versicolour
60	5.2	2.7	3.9	1.4			iris-versicolour
61	5.0	2.0	3.5	1.0			iris-versicolour
62	5.9	3.0	4.2	1.5			iris-versicolour
63	6.0	2.2	4.0	1.0			iris-versicolour

Figura 3 – Pré – tratamento.

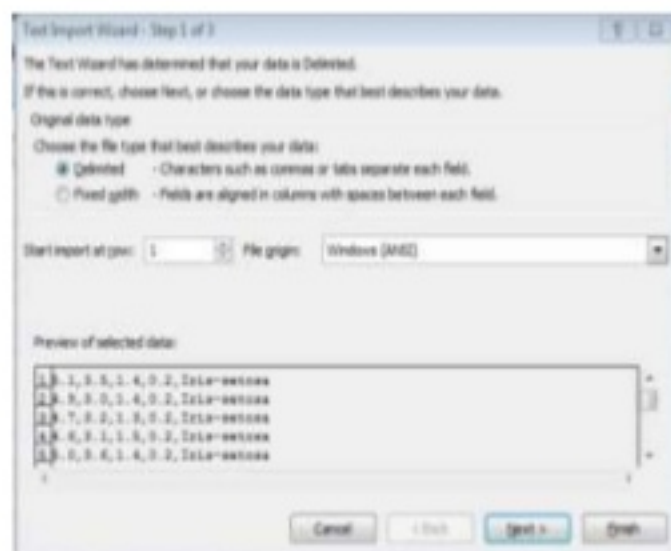


Figura 4 –Transformação.

	A	B	C	D	E	F	G
41	5.1	3.8	1.9	0.4	1		
42	4.8	3.0	1.4	0.3	1		
43	5.1	3.8	1.6	0.2	1		
44	4.6	3.2	1.4	0.2	1		
45	5.3	3.7	1.5	0.2	1		
46	5.0	3.3	1.4	0.2	1		
47	7.0	3.2	4.7	1.4	2		
48	6.4	3.2	4.5	1.5	2		
49	6.9	3.1	4.9	1.5	2		
50	5.5	2.9	4.0	1.9	2		
51	6.5	2.8	4.6	1.5	2		
52	5.7	2.8	4.5	1.3	2		
53	6.3	2.9	4.7	1.8	2		
54	4.9	2.4	3.3	1.0	2		
55	6.6	2.9	4.6	1.3	2		
56	5.2	2.7	3.9	1.4	2		
57	5.0	2.0	3.5	1.0	2		
58	5.3	3.0	4.2	1.5	2		
59	6.0	2.2	4.0	1.0	2		
60	6.1	2.9	4.7	1.4	2		
61	5.6	2.9	3.6	1.3	2		
62	6.7	3.1	4.4	1.4	2		
63	5.6	3.0	4.5	1.5	2		
64	5.8	2.7	4.1	1.0	2		
65	6.2	2.2	4.5	1.5	2		
66	5.6	2.5	3.9	1.1	2		
67	5.9	3.2	4.6	1.0	2		
68	6.1	2.8	4.0	1.3	2		
69	6.3	2.5	4.9	1.5	2		
70	6.1	2.8	4.7	1.2	2		
71	6.4	2.9	4.3	1.3	2		
72	6.6	3.0	4.4	1.4	2		
73	6.8	2.8	4.6	1.4	2		
74	6.7	3.0	5.0	1.7	2		

Figura 5 – Discretização.

	A	B	C	D	E	F
1	4.8	3.1	1.6	0.2	1	
2	5.4	3.4	1.5	0.4	1	
3	5.2	4.1	1.5	0.1	1	
4	5.5	4.2	1.4	0.2	1	
5	4.9	3.1	1.5	0.1	1	
6	5.5	2.4	3.8	1.1	2	
7	5.5	2.4	3.7	1.0	2	
8	5.8	2.7	3.9	1.2	2	
9	6.0	2.7	5.1	1.6	2	
10	5.4	3.0	4.5	1.5	2	
11	7.2	3.2	6.0	1.8	3	
12	6.2	2.8	4.8	1.8	3	
13	6.1	3.0	4.9	1.8	3	
14	6.4	2.8	5.6	2.1	3	
15	7.2	3.0	5.8	1.6	3	

Figura 6 – Classe Teste.

A figura 7 mostra a foto principal do nosso projeto. conjunto de dados da Flor Iris na área de trabalho do MATLAB. Como os atributos são contínuos a discretização esta pronta.



Figura 7 – MatLab

- **Metodologia para construir o classificador**

Para treinar a rede neural Backpropagation Regras de Associação de Classes são usadas. Para treinar a rede, tivemos que usar a função `train()` do Matlab. No qual 60% dos dados são usados para treinamento, 20% são usados para testar e 20% para propósito de validação. `traingdm`, `learnqdm` e `tansig` são usados como função de treino, função de aprendizado e função de transferência. Número de nós de camadas escondidos = 9 e o aprendizado vai parar quando o erro for menor que 0.005 ou período(epoca) = 5,000 vezes. Estes valores são constantes em ordem para determinar o momento constante (`net.trainParam.mc`) (mc) e taxa de aprendizado (`net.trainParam.lr`)(lr). A tabela 3 mostra o número de teste padrão não identificado quando momento constante (mc) está variado e guardando taxa de aprendizado para 0.9. O valor do mc selecionado pelo teste padrão não identificado é mínimo. O melhor desempenho é alcançado quando mc=0.7.

Agoras, os valores de taxa de aprendizado variam entre 0.1 e 0.9 e mc é mantido constante em 0.7 e números de teste padrão não identificado são mostrados na tabela 6. Assim, o melhor desempenho pode ser alcançado em 0.

TABELA 2 : Desempenho (mc)

lr	Mc	Ensaios Padrões não id.
0.9	0.1	29
	0.2	24
	0.3	23
	0.4	33
	0.5	26
	0.6	24
	0.7	7
	0.8	8
	0.9	19

TABELA 3 : Desempenho de taxa de aprendizado

lr	Mc	Ensaio Padrões não id.
0.7	0.1	8
	0.2	2
	0.3	4
	0.4	3
	0.5	4
	0.6	3
	0.7	5
	0.8	9
	0.9	7

Agora, para determinar o número de períodos apropriados. Mantém os valores de todas as outras propriedades constantes. Tabela 4 mostra o número de padrão não identificado com períodos variados. O melhor desempenho é alcançado com período = 2000 e tabela 5 mostra a melhor propriedade para a rede neural backpropagation em ordem para construir um classificador para planta Iris.

TABELA 4 : Desempenho de período

Epoca	Ensaio Padrões não id.
500	11
1000	6
1500	3
2000	2
2500	2
5000	2

TABELA 5 : Propriedades para construir a RN Backpropagation para Flores Iris.

Neural network Properties	Value
Number of hidden neurons	9
Transfer function	tansig
Learning rate	0.2
Momentum constant	0.7
Training technique	Gradient descent with momentum Backpropagation
Epochs	2000
Misidentified patterns	2
Training patterns	181
Validation patterns	60
Testing patterns	60

Criando a rede usando a ferramenta nntool do MatLab.

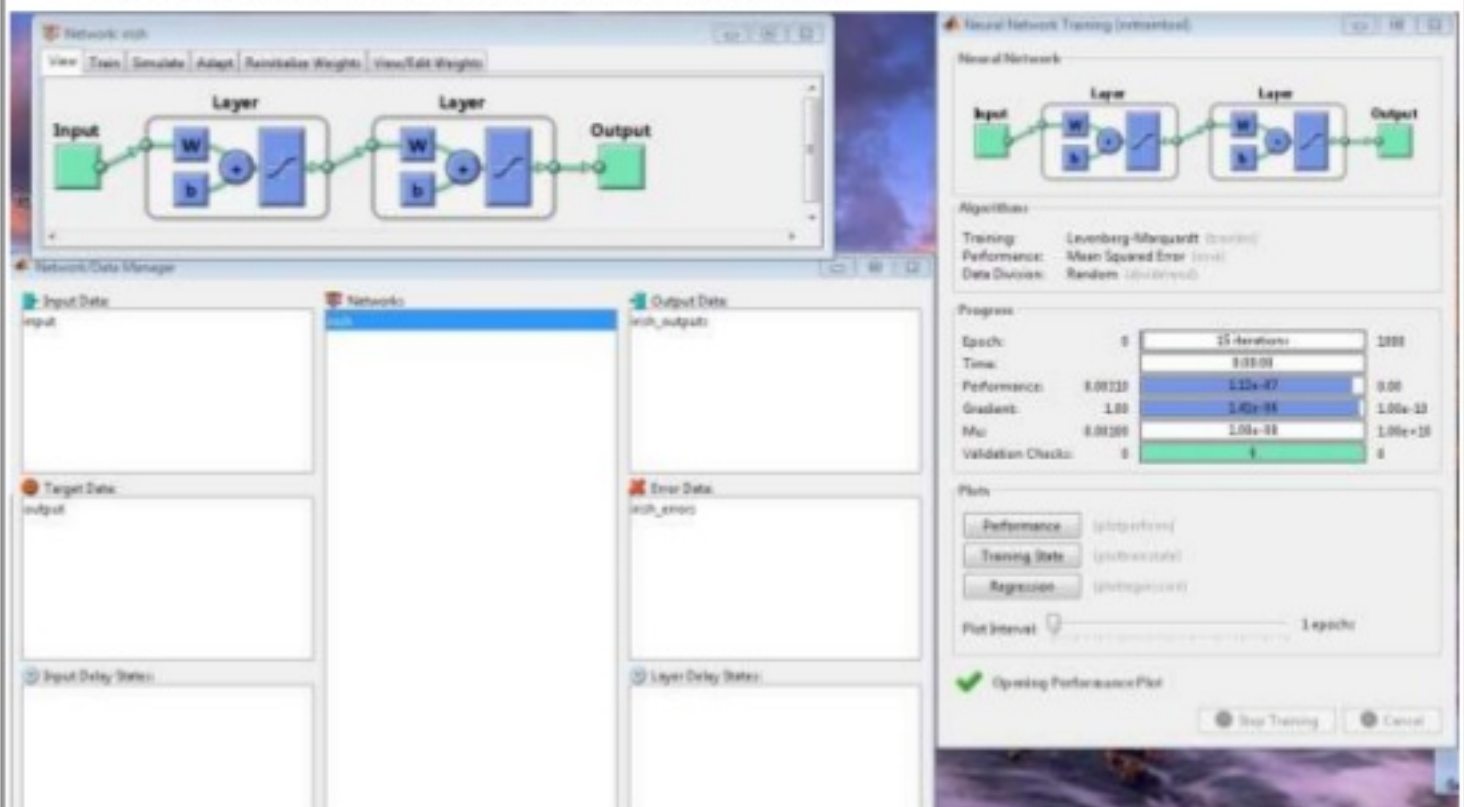


Figura 8 – Ferramenta nntool.

Proxima figura mostra o treinamento da Rede.

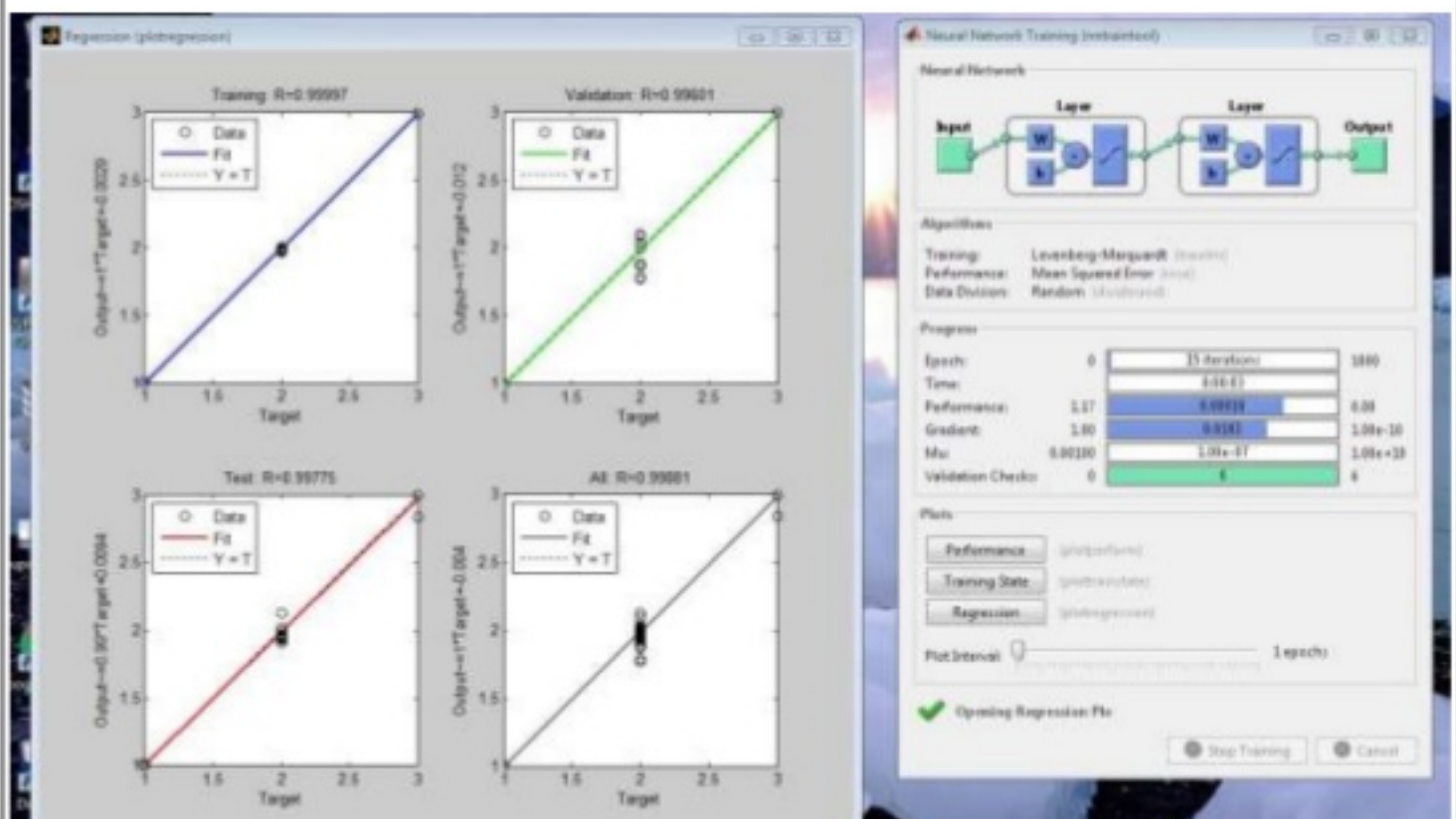


Figura 8 – Treinamento e Gráfico plotado.

A próxima figura mostra o Gráfico de melhor Validação, com valores que superaram o esperado.

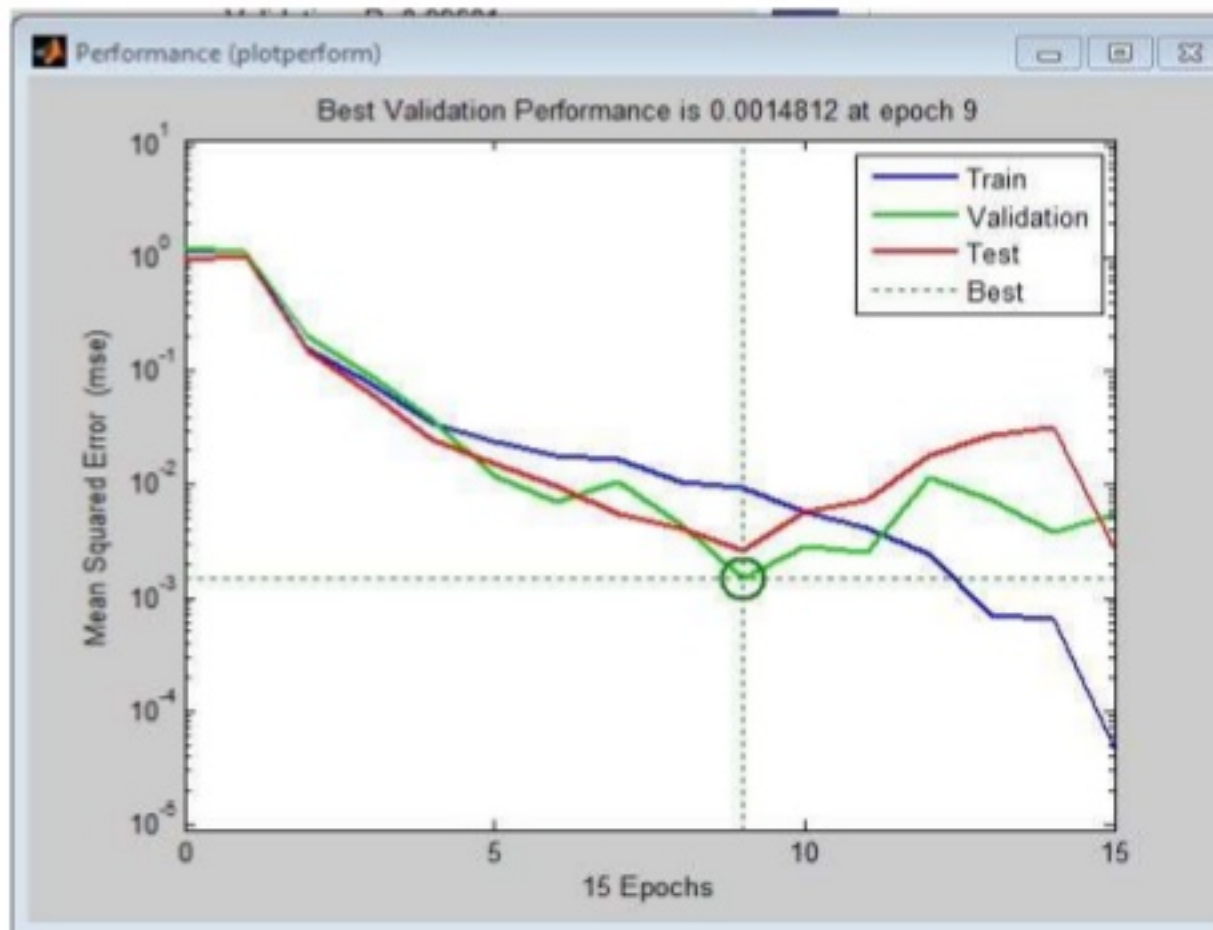


Figura 9 –Performa

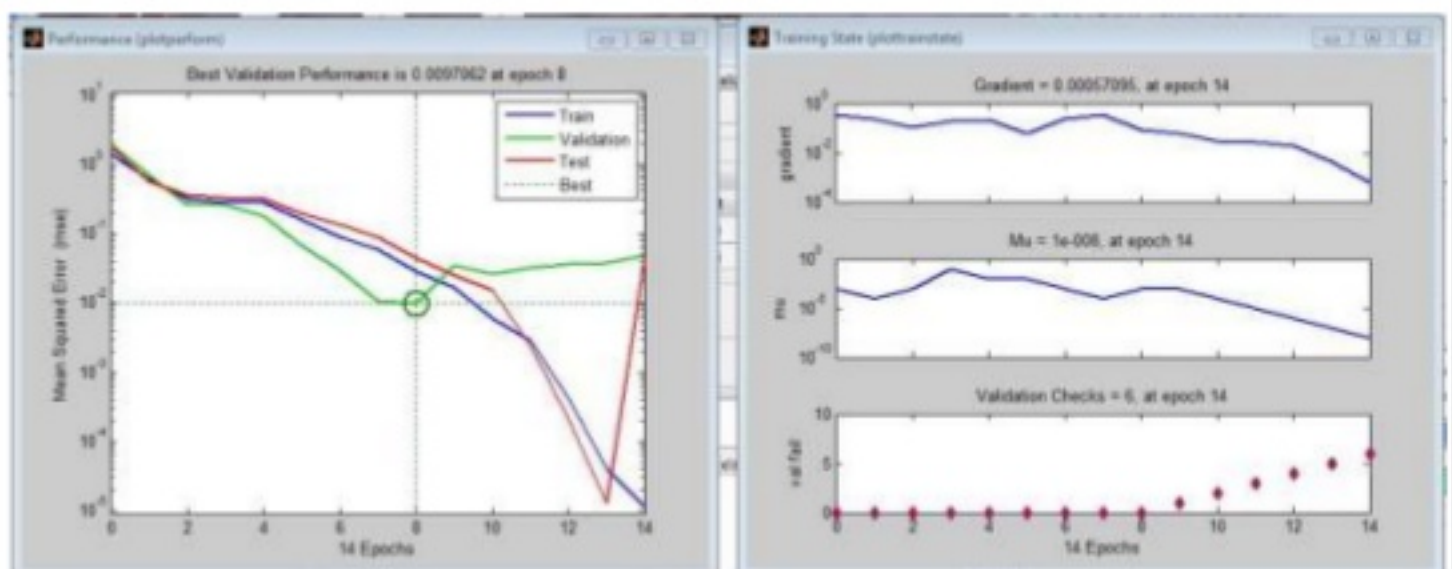


Figura 10 –Amostra de treinamentos anteriores.

Imagem dos parâmetros de Rede.

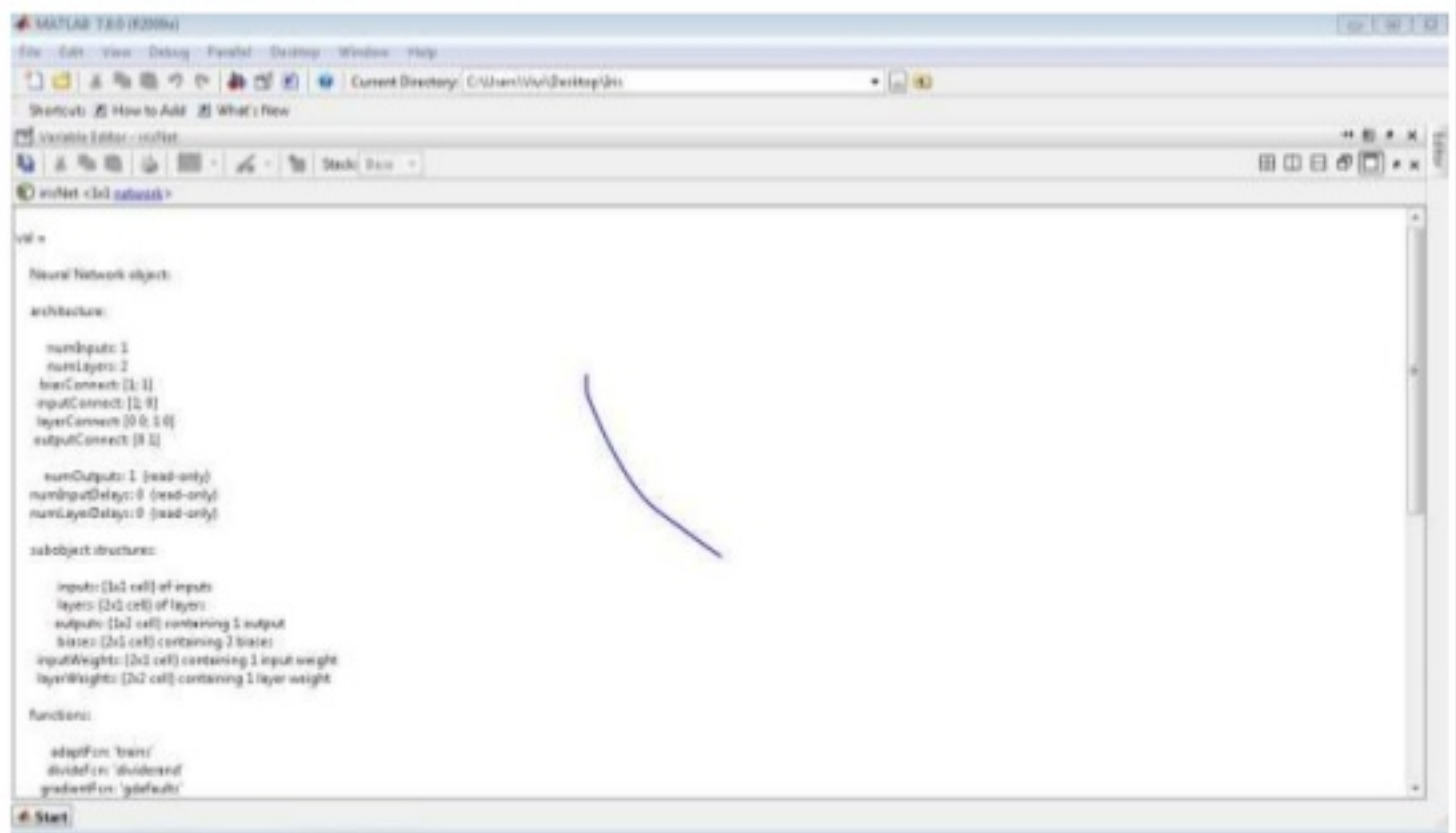


Figura 11 – Parametros da RN Backpropagation.

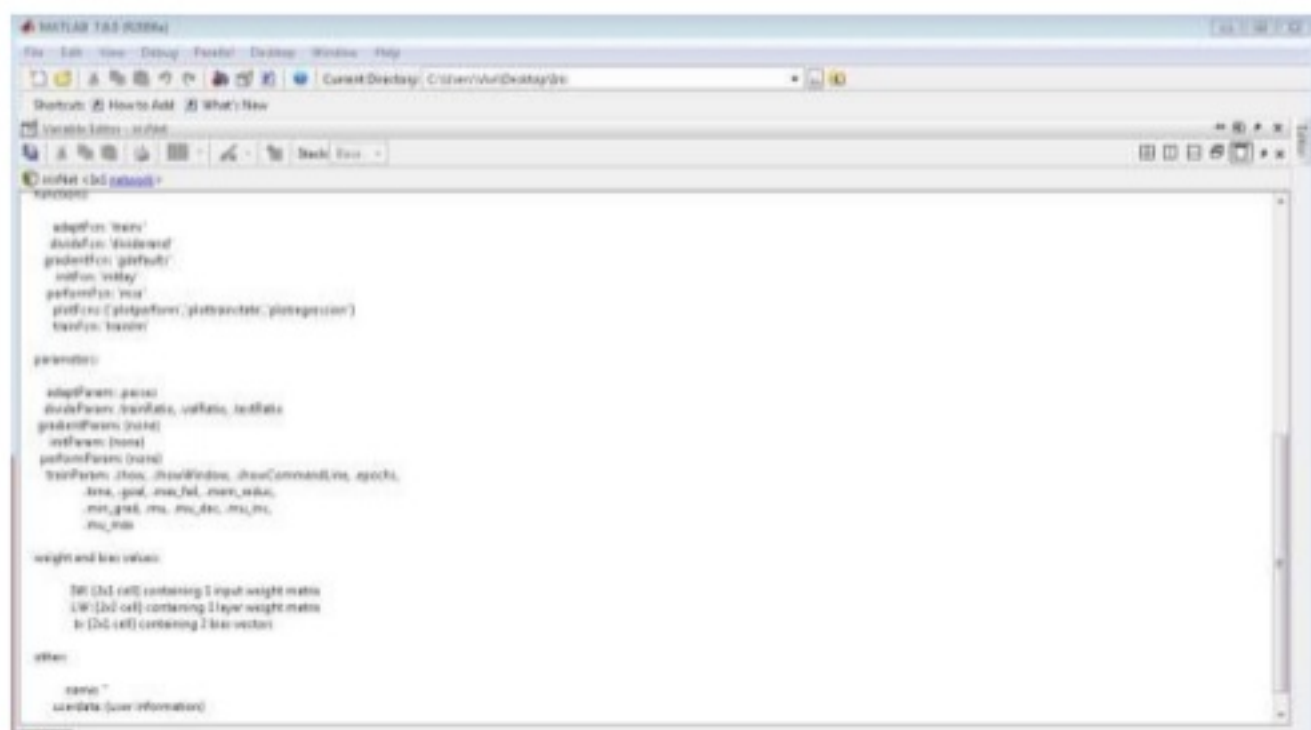


Figura 12 – Parametros da RN Backpropagation.

Agora já treinamos a rede que pode prever mais precisamente a classe baseado nos atributos previstos.

Proxima figura mostra os Parametros de saída da Rede.

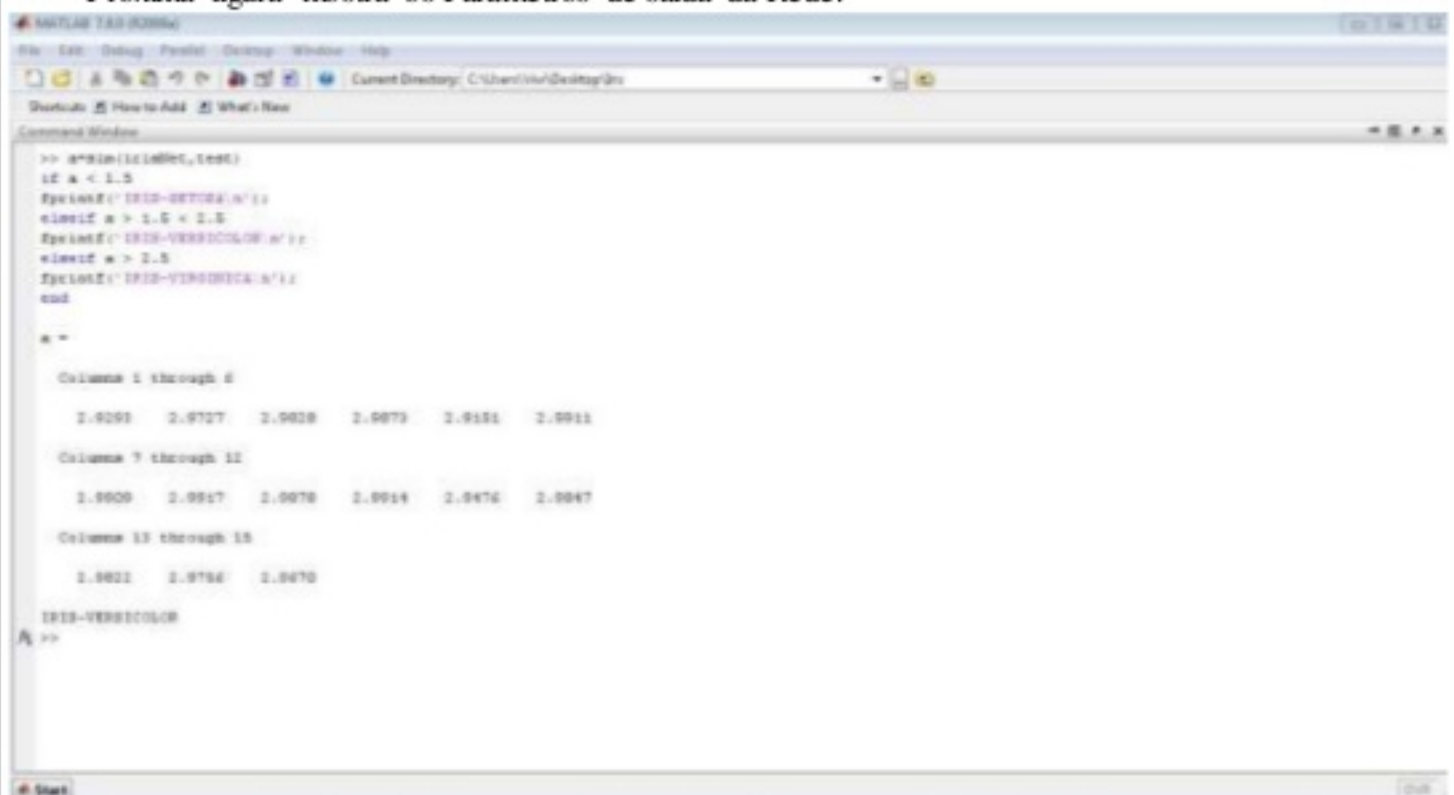


Figura 12 –Parametros de saída da RN.

Foi implementado uma função para a resposta ser impressa como palavra, conforme dados originais.

- **Sumario :**

A rede neural de sistema associativo de classificação é usado para a construção precisa e eficientes de classificadores de flores íris, a fim de melhorar a sua precisão. A rede de três camadas feed-forward usando o Matlab. Os valores ótimos para taxa de aprendizagem $m(Ir)$, Momento constante (mc) e épocas foram encontrados para ser de 0,7, 0,2 e 2000 respectivamente. A estrutura das redes reflete o conhecimento descoberto na fase anterior de descoberta. A rede treinada é então utilizado para classificar os dados invisíveis da Flor Iris.

O Classificador da Flor Iris usando o sistema associativo de classificação é mais preciso do que construir um outro modelo classificador conforme mostrado em pesquisas online, porque os pesos são ajustados de acordo com as regras de associação de classe e melhor é usado para classificar os dados.

• **Referencias :**

- [1] Data Mining-(Mineração de Dados). Joel De Bortoli-Graduação em Gestão de TI, 08/2012, disponível no Link: <http://www.joeldebortoli.com/2012/08/data-mining-mineracao-de-dados.html>
- [2] International Journal of Advancements in Technology, Prof. Sheetal S. Dhande, 04/10/2011, disponível no Link: <http://www.omicsonline.org/open-access/building-an-iris-plant-data-classifier-using-neural-network-associative-classification-0976-4860-2-491-506.pdf>
- [3] **Iris Data Set, Abstract:** Famous database; from Fisher, 1936, disponível no Link: <https://archive.ics.uci.edu/ml/datasets/Iris>
- [4] Treinamento de Redes Neurais Feedforward, Hegler C. Tissot, Luiz C. Camargo, Aurora T. R. Pozo, ArtigoENIAr0409, disponível no Link: http://www.ppgia.pucpr.br/~enia/anais/enia/artigos/105243_2.pdf
- [5] Iris (gênero), Wikipédia, a enciclopédia livre, 07h09min de 23 de dezembro de 2015. Disponível no Link: [https://pt.wikipedia.org/wiki/Iris_\(g%C3%A9nero\)](https://pt.wikipedia.org/wiki/Iris_(g%C3%A9nero))

• **Currículos :**

Adão de souza

Estudante do 7º Período de Engenharia de Controle e Automação, funcionário Responsável pelo setor de Transporte da prefeitura da Cidade de Bela Vista de Minas – MG.

Davidson Ferreira

Estudante do 7º Período de Engenharia de Controle e Automação, Inspetor Técnico de Instrumentação na Usiminas S/A na cidade de Ipatinga - MG.

Francisco Jr.

Estudante do 7º Período de Engenharia de Controle e Automação, Analista Técnico Eletricista na Aperam south america S/A na cidade de Timoteo – MG.

Philip Rodrigues

Estudante do 7º Período de Engenharia de Controle e Automação, Head Chef no Hotel Don Henrique na cidade de Timoteo – MG.