



**Escola Politécnica – Universidade de São Paulo**  
**Departamento de Engenharia Elétrica Sistemas Eletrônicos**

## **Projeto Final PSI2672**

**Práticas em Reconhecimento de Padrões,  
Modelagem e Neurocomputação**

**Grupo:**

**Brunno Vodola Martins**

**Franz Sebastián Bedoya LLano**

**Lucas Akira Beppu**

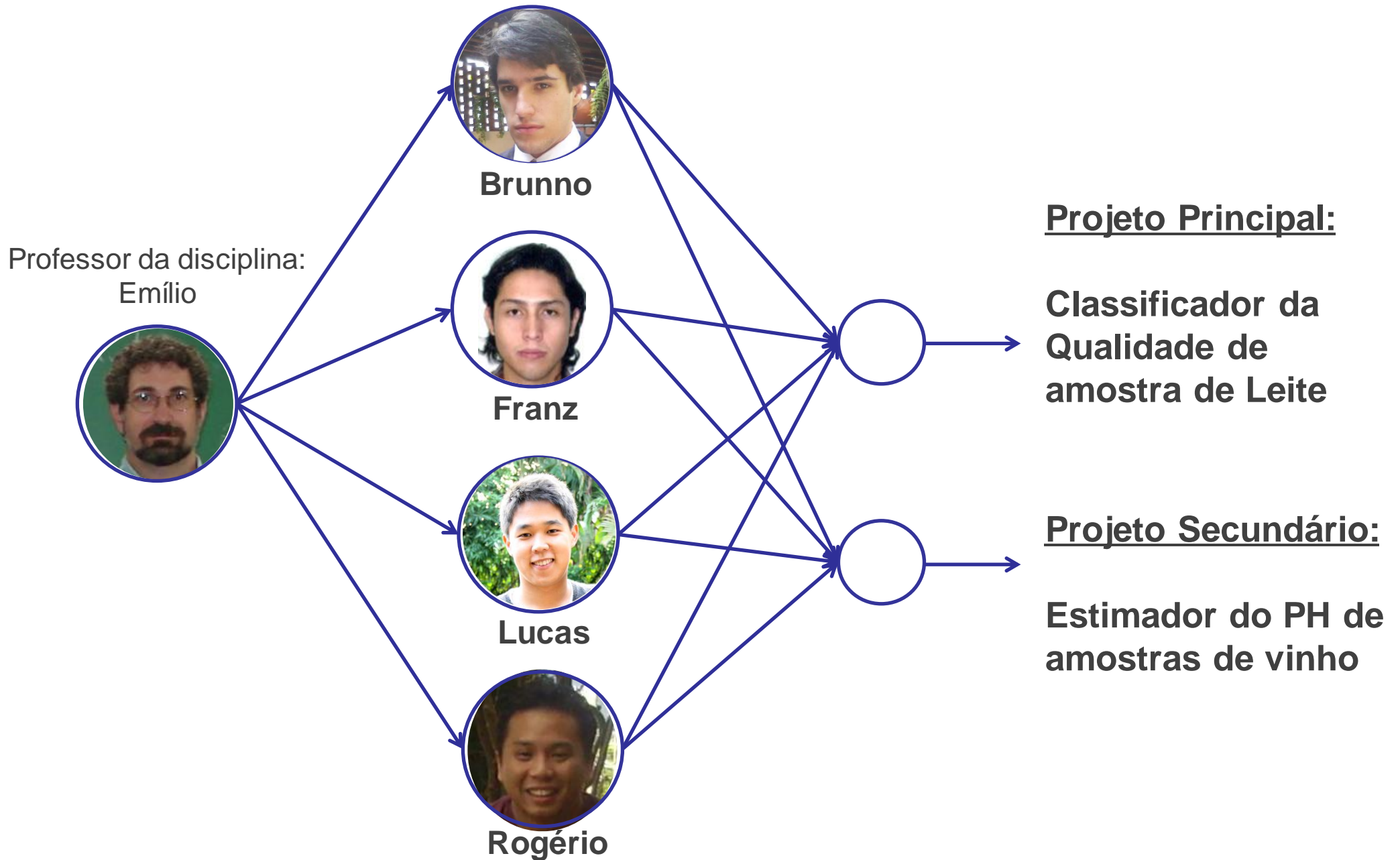
**Rogério Hong Hui Chung**

**Professor:**

**Emílio Del Moral Hernandez**

**Data:** 18 de Junho de 2012

# Apresentação do Grupo e dos Temas dos Projetos



Os **dados** utilizados no **projeto principal** foram retirados do Projeto de Formatura “**Via Láctea**” (2011) dos alunos Daniel Torres, Douglas Veronez e Patrick Santos.

As medidas foram feitas através de uma **Língua Eletrônica**.

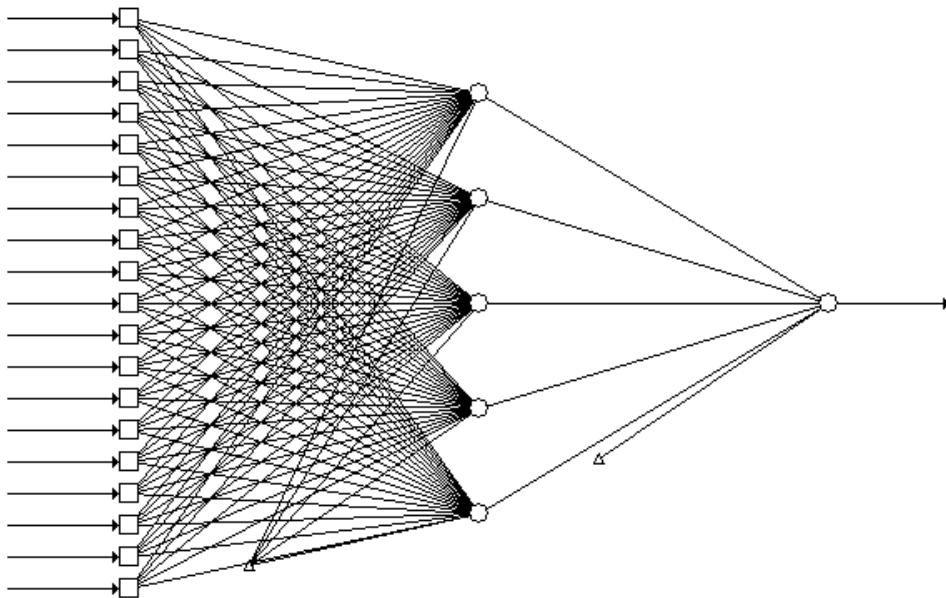


Foto retirada da monografia “Projeto Via Láctea”.

Os exemplares dos conjuntos de treino e teste foram escolhidos  
“cortando” a lista de dados para termos 25 exemplares de leite  
“bom” e 25 de leite “ruim”

10	3,07E-06	3,19E-06	5,73E-06	9,58E-06	9,99E-06	9,57E-06	9,58E-06	1,01E-05	1,46E+06	7,76E+06	6,61E+06	3,97E+06	9,42E+06	1,31E+07	1,53E+07	9,76E+06	9,54E+06	900
11	2,69E-06	2,75E-06	5,01E-06	8,44E-06	9,08E-06	8,80E-06	8,27E-06	8,81E-06	1,46E+06	7,70E+06	6,55E+06	3,95E+06	9,05E+06	1,24E+07	1,43E+07	9,37E+06	9,14E+06	1000
12	1,19E-06	1,12E-06	2,15E-06	3,44E-06	4,36E-06	4,58E-06	3,04E-06	3,41E-06	1,45E+06	7,41E+06	6,34E+06	3,79E+06	7,53E+06	9,05E+06	9,89E+06	7,89E+06	7,53E+06	2000
13	7,93E-07	7,10E-07	1,35E-06	2,02E-06	2,70E-06	2,88E-06	1,70E-06	1,93E-06	1,45E+06	7,27E+06	6,25E+06	3,72E+06	7,03E+06	7,97E+06	8,46E+06	7,48E+06	7,07E+06	3000
14	6,20E-07	5,40E-07	9,88E-07	1,41E-06	1,89E-06	2,02E-06	1,15E-06	1,32E-06	1,44E+06	7,17E+06	6,19E+06	3,68E+06	6,78E+06	7,44E+06	7,78E+06	7,28E+06	6,84E+06	4000
15	5,22E-07	4,48E-07	7,85E-07	1,08E-06	1,44E-06	1,53E-06	8,74E-07	9,96E-07	1,44E+06	7,11E+06	6,15E+06	3,66E+06	6,61E+06	7,13E+06	7,39E+06	7,15E+06	6,70E+06	5000
16	4,63E-07	3,93E-07	6,62E-07	8,83E-07	1,15E-06	1,22E-06	7,08E-07	8,05E-07	1,44E+06	7,05E+06	6,12E+06	3,65E+06	6,50E+06	6,91E+06	7,14E+06	7,07E+06	6,60E+06	6000
17	4,17E-07	3,57E-07	5,76E-07	7,44E-07	9,62E-07	1,01E-06	6,01E-07	6,80E-07	1,44E+06	7,00E+06	6,10E+06	3,63E+06	6,42E+06	6,77E+06	6,96E+06	7,00E+06	6,53E+06	7000
18	3,87E-07	3,32E-07	5,15E-07	6,50E-07	8,25E-07	8,62E-07	5,26E-07	5,94E-07	1,44E+06	6,96E+06	6,09E+06	3,63E+06	6,35E+06	6,66E+06	6,84E+06	6,95E+06	6,47E+06	8000
19	3,64E-07	3,12E-07	4,69E-07	5,84E-07	7,24E-07	7,52E-07	4,76E-07	5,30E-07	1,44E+06	6,92E+06	6,07E+06	3,62E+06	6,30E+06	6,57E+06	6,74E+06	6,91E+06	6,42E+06	9000
20	3,44E-07	2,98E-07	4,34E-07	5,28E-07	6,49E-07	6,69E-07	4,34E-07	4,83E-07	1,43E+06	6,89E+06	6,06E+06	3,62E+06	6,26E+06	6,49E+06	6,66E+06	6,87E+06	6,38E+06	10000
21	3,38E-05	3,48E-05	4,61E-05	4,28E-05	2,93E-05	2,60E-05	4,14E-05	4,12E-05	1,64E+06	1,55E+07	1,23E+07	5,20E+06	4,38E+07	7,37E+07	8,03E+07	6,37E+07	5,18E+07	100
22	1,45E-05	1,49E-05	2,16E-05	2,74E-05	2,13E-05	1,89E-05	3,00E-05	2,89E-05	1,61E+06	1,33E+07	1,08E+07	4,85E+06	2,53E+07	4,22E+07	4,76E+07	3,10E+07	2,91E+07	200
23	8,60E-06	8,85E-06	1,38E-05	1,91E-05	1,64E-05	1,48E-05	2,19E-05	2,16E-05	1,59E+06	1,25E+07	1,02E+07	4,67E+06	1,93E+07	3,08E+07	3,53E+07	2,15E+07	2,12E+07	300
24	5,87E-06	6,07E-06	1,01E-05	1,42E-05	1,32E-05	1,22E-05	1,65E-05	1,68E-05	1,59E+06	1,20E+07	9,84E+06	4,56E+06	1,66E+07	2,53E+07	2,91E+07	1,75E+07	1,74E+07	400
25	4,36E-06	4,52E-06	7,94E-06	1,10E-05	1,09E-05	1,03E-05	1,27E-05	1,34E-05	1,58E+06	1,17E+07	9,58E+06	4,48E+06	1,50E+07	2,21E+07	2,53E+07	1,54E+07	1,53E+07	500

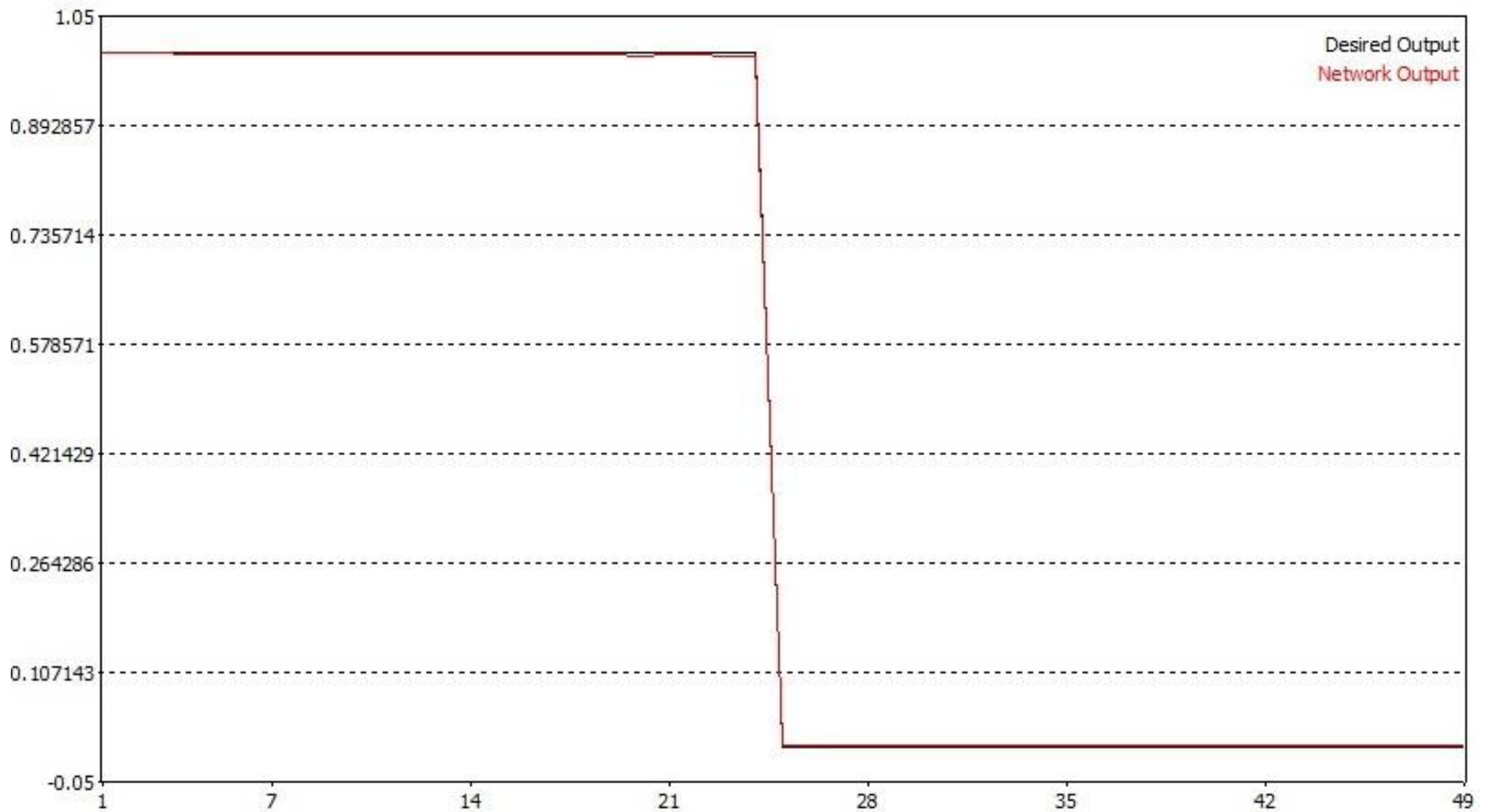
Os ensaios do projeto foram realizados no *software Multiple Back Propagation (MBP)*



Topologia da Rede 19-5-1

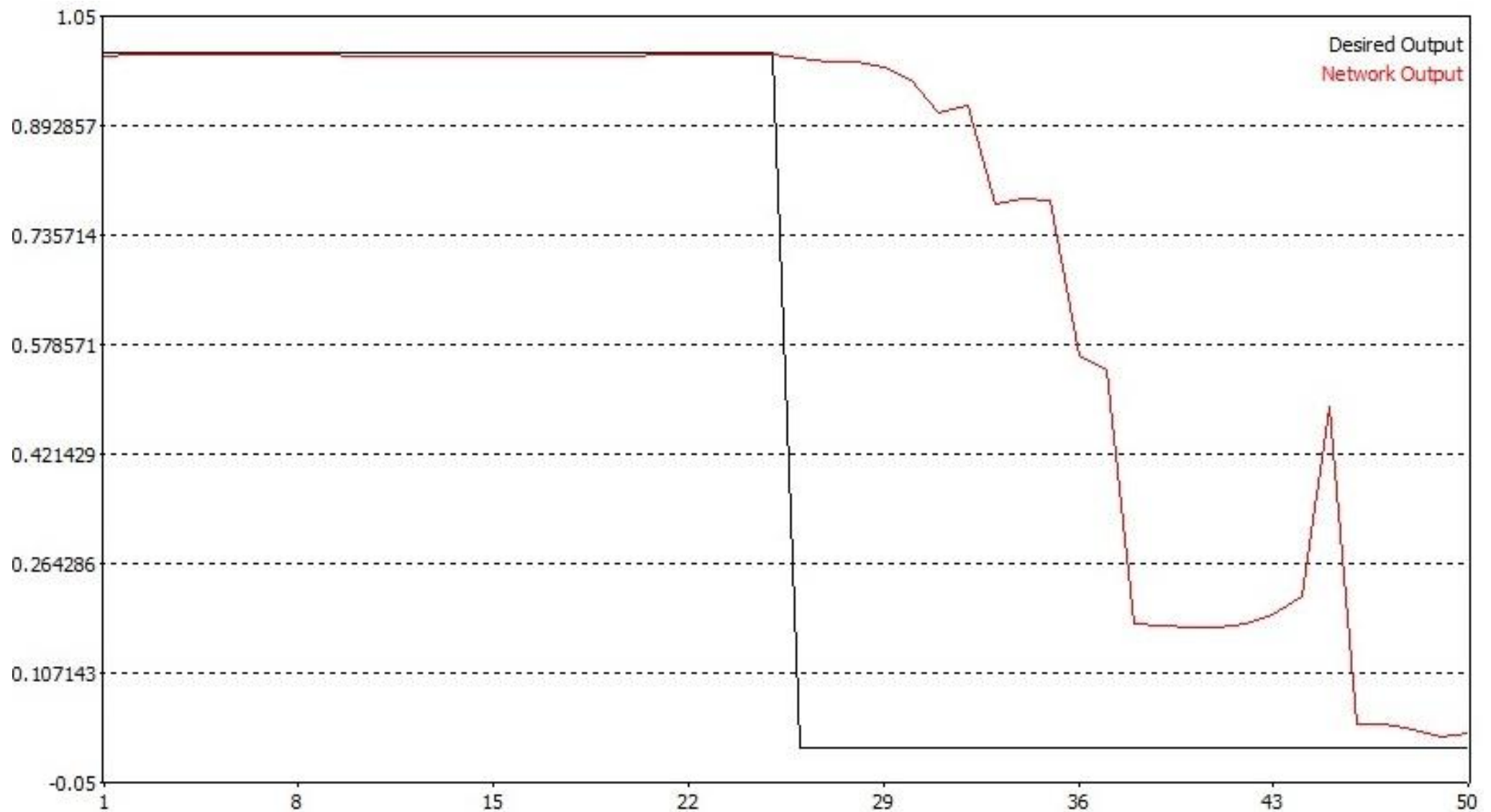
RMS ~ 0.2

# Projeto Principal – Ensaio Inicial - Resultados



Conjunto de treino – taxa de acerto = 100%

# Projeto Principal – Ensaio Inicial - Resultados



Conjunto de teste – taxa de acerto = 76%



Os exemplares do segundo ensaio foram sorteados para o conjunto de treino e de testes. Isso foi uma forma de processar os dados a fim de melhorar o desempenho da rede neural.

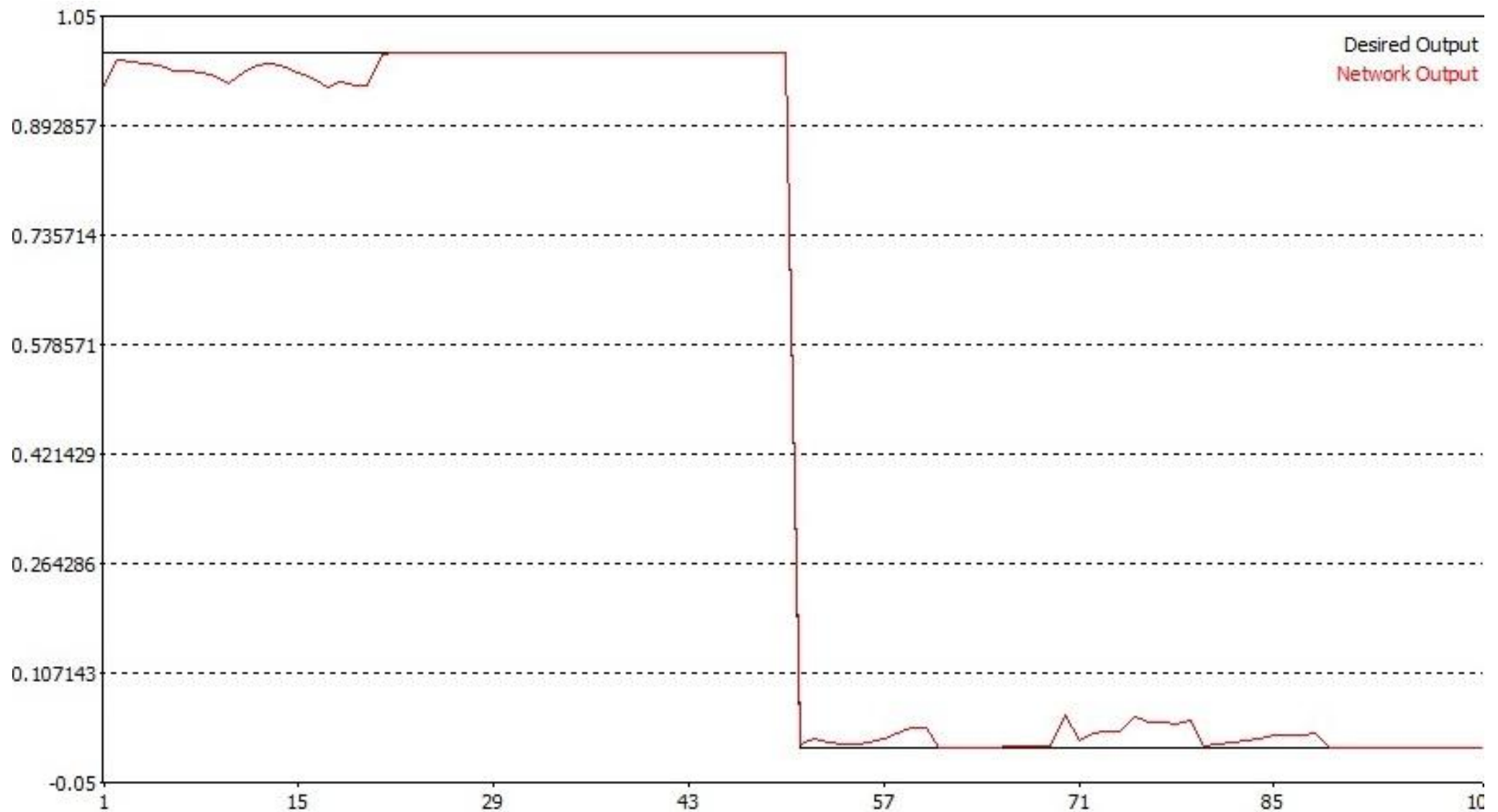
→	7	5,27E-06	5,57E-06	9,52E-06	1,51E-05	1,41E-05	1,28E-05	1,58E-05	1,59E-05	1,47E+06	8,01E+06	6,84E+06	4,09E+06	1,14E+07	1,70E+07	2,01E+07	1,20E+07	1,18E+07	600
→	8	4,28E-06	4,50E-06	7,85E-06	1,28E-05	1,25E-05	1,15E-05	1,32E-05	1,35E-05	1,47E+06	7,91E+06	6,75E+06	4,04E+06	1,05E+07	1,53E+07	1,80E+07	1,10E+07	1,08E+07	700
→	9	3,59E-06	3,74E-06	6,65E-06	1,10E-05	1,11E-05	1,05E-05	1,12E-05	1,16E-05	1,47E+06	7,82E+06	6,67E+06	4,00E+06	9,89E+06	1,41E+07	1,65E+07	1,03E+07	1,01E+07	800
→	10	3,07E-06	3,19E-06	5,73E-06	9,58E-06	9,99E-06	9,57E-06	9,58E-06	1,01E-05	1,46E+06	7,76E+06	6,61E+06	3,97E+06	9,42E+06	1,31E+07	1,53E+07	9,76E+06	9,54E+06	900
→	11	2,69E-06	2,75E-06	5,01E-06	8,44E-06	9,08E-06	8,80E-06	8,27E-06	8,81E-06	1,46E+06	7,70E+06	6,55E+06	3,95E+06	9,05E+06	1,24E+07	1,43E+07	9,37E+06	9,14E+06	1000
→	12	1,19E-06	1,12E-06	2,15E-06	3,44E-06	4,36E-06	4,58E-06	3,04E-06	3,41E-06	1,45E+06	7,41E+06	6,34E+06	3,79E+06	7,53E+06	9,05E+06	9,89E+06	7,89E+06	7,53E+06	2000
→	13	7,93E-07	7,10E-07	1,35E-06	2,02E-06	2,70E-06	2,88E-06	1,70E-06	1,93E-06	1,45E+06	7,27E+06	6,25E+06	3,72E+06	7,03E+06	7,97E+06	8,46E+06	7,48E+06	7,07E+06	3000
→	14	6,20E-07	5,40E-07	9,88E-07	1,41E-06	1,89E-06	2,02E-06	1,15E-06	1,32E-06	1,44E+06	7,17E+06	6,19E+06	3,68E+06	6,78E+06	7,44E+06	7,78E+06	7,28E+06	6,84E+06	4000
→	15	5,22E-07	4,48E-07	7,85E-07	1,08E-06	1,44E-06	1,53E-06	8,74E-07	9,96E-07	1,44E+06	7,11E+06	6,15E+06	3,66E+06	6,61E+06	7,13E+06	7,39E+06	7,15E+06	6,70E+06	5000
→	16	4,63E-07	3,93E-07	6,62E-07	8,83E-07	1,15E-06	1,22E-06	7,08E-07	8,05E-07	1,44E+06	7,05E+06	6,12E+06	3,65E+06	6,50E+06	6,91E+06	7,14E+06	7,07E+06	6,60E+06	6000
→	17	4,17E-07	3,57E-07	5,76E-07	7,44E-07	9,62E-07	1,01E-06	6,01E-07	6,80E-07	1,44E+06	7,00E+06	6,10E+06	3,63E+06	6,42E+06	6,77E+06	6,96E+06	7,00E+06	6,53E+06	7000
→	18	3,87E-07	3,32E-07	5,15E-07	6,50E-07	8,25E-07	8,62E-07	5,26E-07	5,94E-07	1,44E+06	6,96E+06	6,09E+06	3,63E+06	6,35E+06	6,66E+06	6,84E+06	6,95E+06	6,47E+06	8000
→	19	3,64E-07	3,12E-07	4,69E-07	5,84E-07	7,24E-07	7,52E-07	4,76E-07	5,30E-07	1,44E+06	6,92E+06	6,07E+06	3,62E+06	6,30E+06	6,57E+06	6,74E+06	6,91E+06	6,42E+06	9000
→	20	3,44E-07	2,98E-07	4,34E-07	5,28E-07	6,49E-07	6,69E-07	4,34E-07	4,83E-07	1,43E+06	6,89E+06	6,06E+06	3,62E+06	6,26E+06	6,49E+06	6,66E+06	6,87E+06	6,38E+06	10000
→	21	3,38E-05	3,48E-05	4,61E-05	4,28E-05	2,93E-05	2,60E-05	4,14E-05	4,12E-05	1,64E+06	1,55E+07	1,23E+07	5,20E+06	4,38E+07	7,37E+07	8,03E+07	6,37E+07	5,18E+07	100
→	22	1,45E-05	1,49E-05	2,16E-05	2,74E-05	2,13E-05	1,89E-05	3,00E-05	2,89E-05	1,61E+06	1,33E+07	1,08E+07	4,85E+06	2,53E+07	4,22E+07	4,76E+07	3,10E+07	2,91E+07	200
→	23	8,60E-06	8,85E-06	1,38E-05	1,91E-05	1,64E-05	1,48E-05	2,19E-05	2,16E-05	1,59E+06	1,25E+07	1,02E+07	4,67E+06	1,93E+07	3,08E+07	3,53E+07	2,15E+07	2,12E+07	300
→	24	5,87E-06	6,07E-06	1,01E-05	1,42E-05	1,32E-05	1,22E-05	1,65E-05	1,68E-05	1,59E+06	1,20E+07	9,84E+06	4,56E+06	1,66E+07	2,53E+07	2,91E+07	1,75E+07	1,74E+07	400
→	25	4,36E-06	4,52E-06	7,94E-06	1,10E-05	1,09E-05	1,03E-05	1,27E-05	1,34E-05	1,58E+06	1,17E+07	9,58E+06	4,48E+06	1,50E+07	2,21E+07	2,53E+07	1,54E+07	1,53E+07	500
→	26	3,43E-06	3,54E-06	6,51E-06	8,84E-06	9,24E-06	8,89E-06	1,01E-05	1,09E-05	1,58E+06	1,15E+07	9,40E+06	4,41E+06	1,40E+07	1,99E+07	2,27E+07	1,41E+07	1,39E+07	600
→	27	2,79E-06	2,88E-06	5,52E-06	7,30E-06	7,98E-06	7,79E-06	8,23E-06	9,01E-06	1,57E+06	1,13E+07	9,26E+06	4,36E+06	1,33E+07	1,84E+07	2,09E+07	1,33E+07	1,30E+07	700
→	28	2,35E-06	2,43E-06	4,80E-06	6,18E-06	6,97E-06	6,91E-06	6,86E-06	7,62E-06	1,57E+06	1,12E+07	9,13E+06	4,31E+06	1,27E+07	1,73E+07	1,95E+07	1,27E+07	1,23E+07	800

→ Conjunto de treino

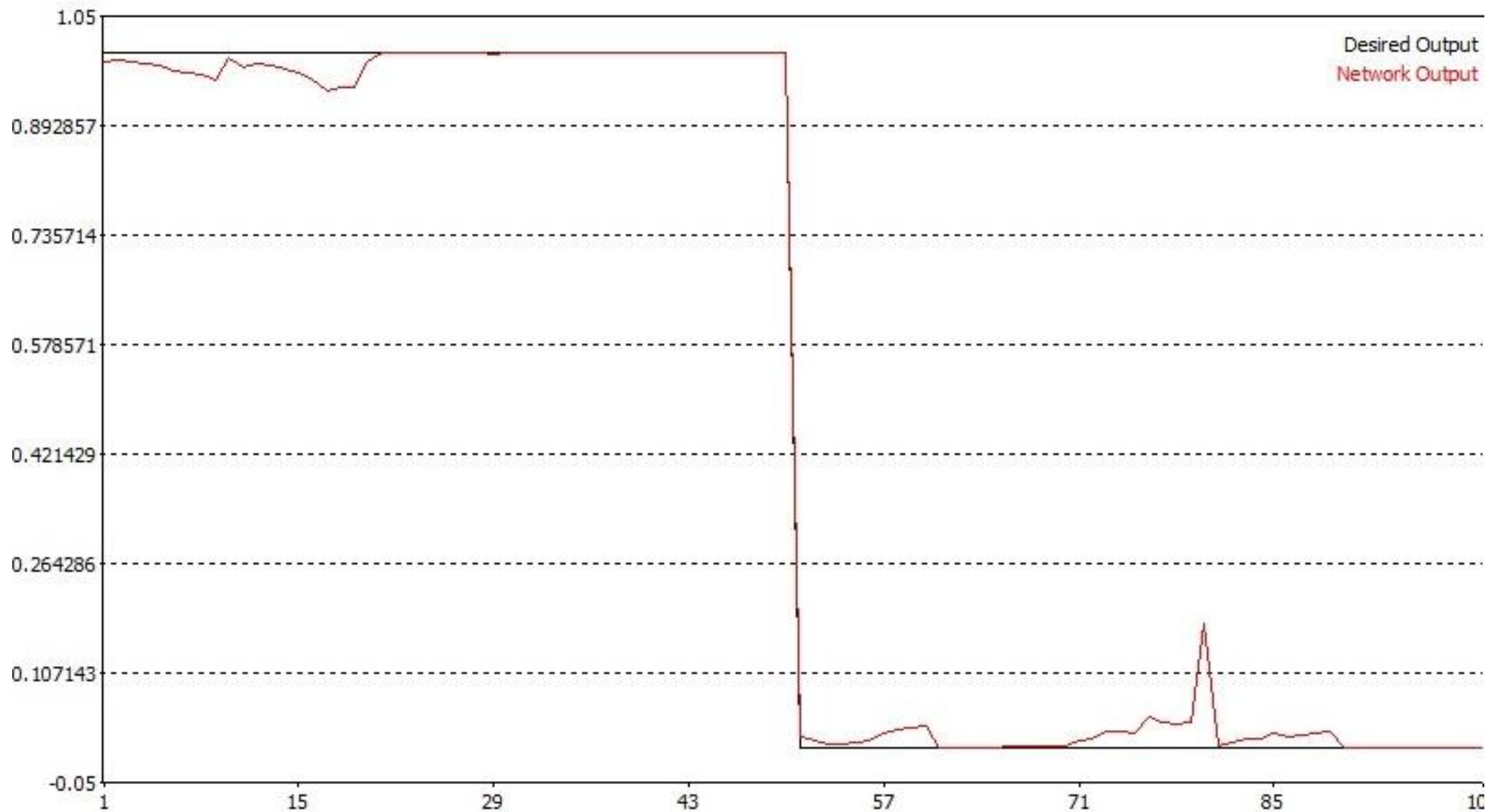
→ Conjunto de teste



# Projeto Principal – Resultados

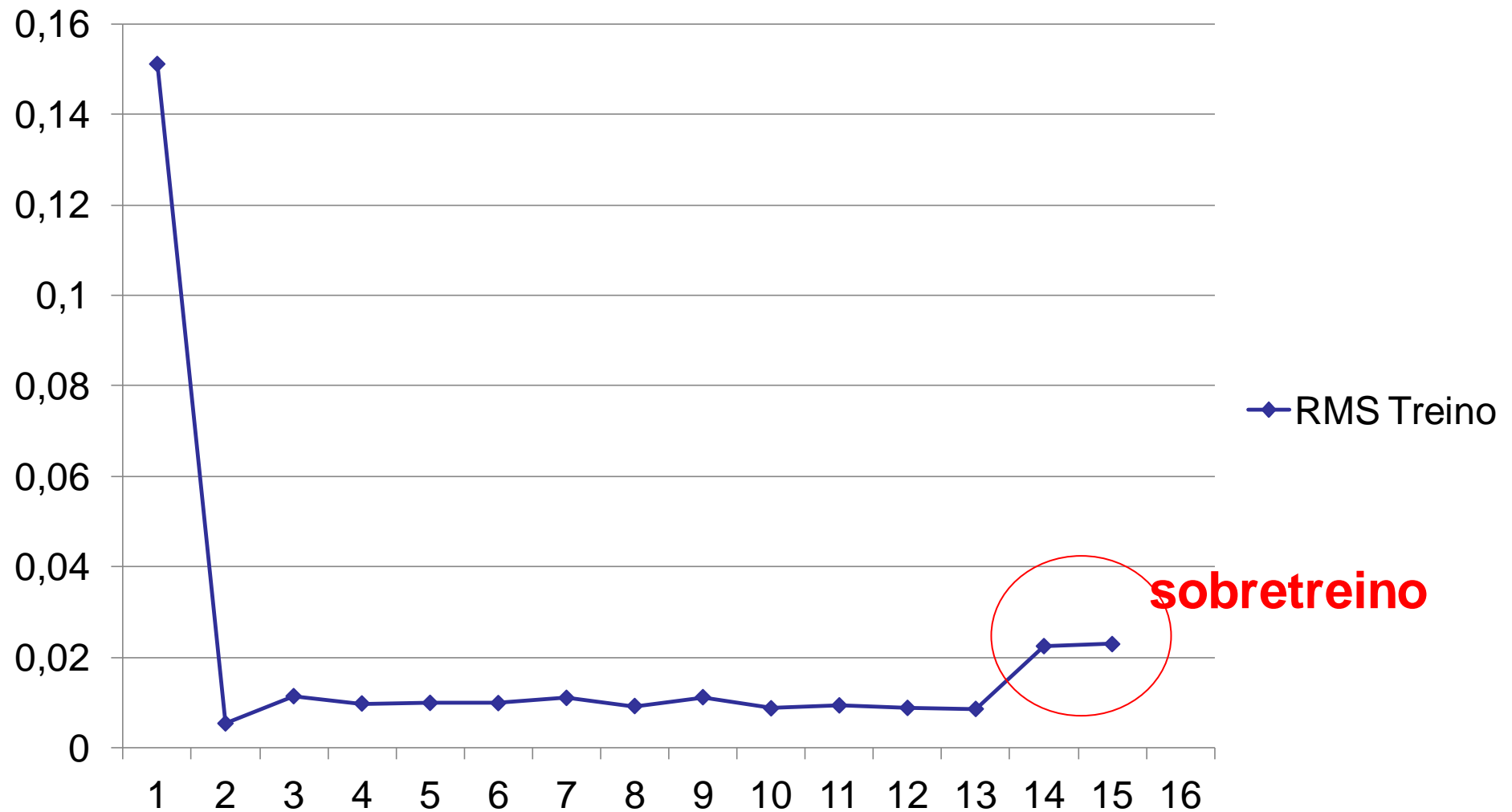


Conjunto de treino – taxa de acerto = 100%



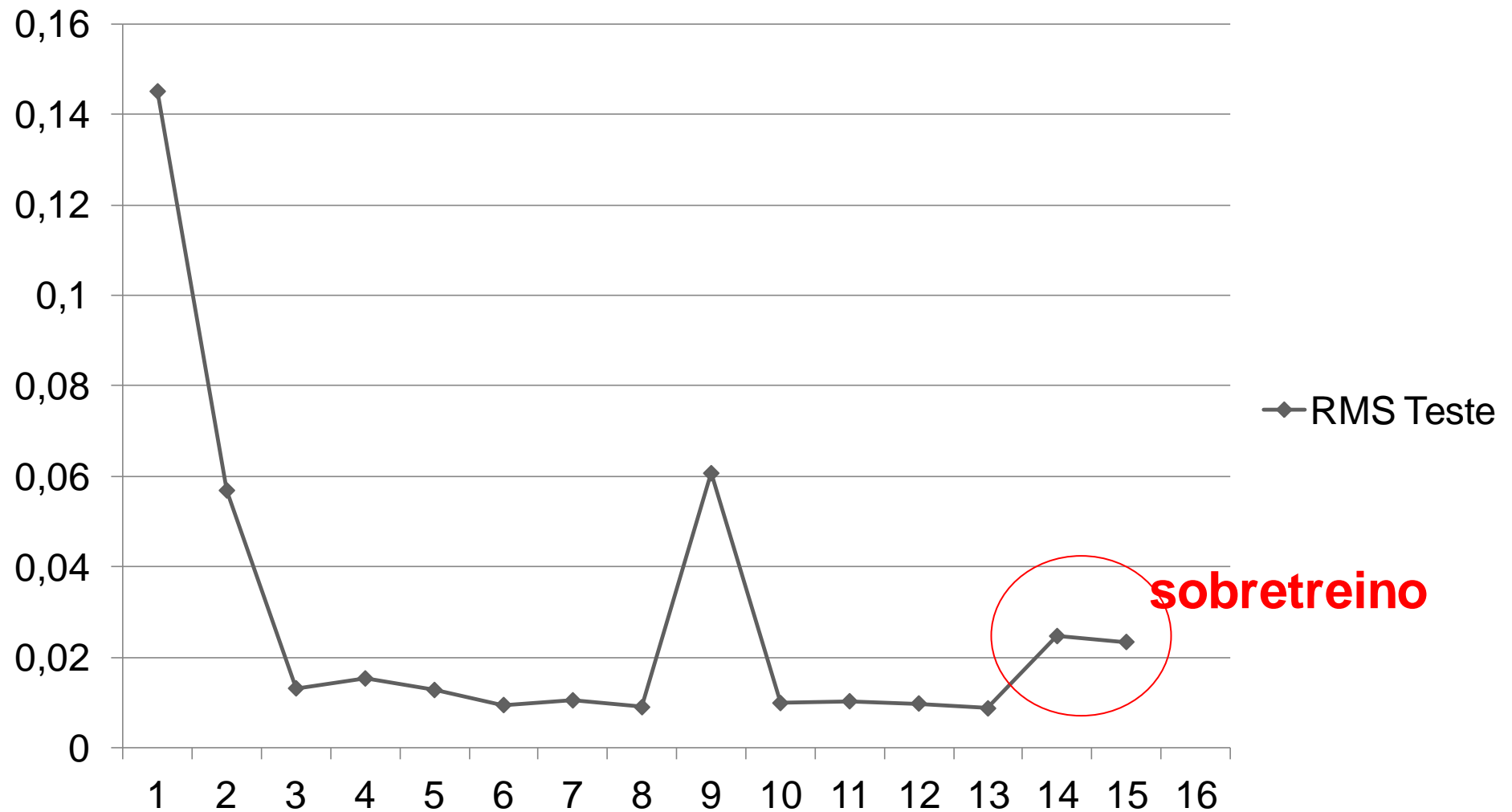
Conjunto de teste – taxa de acerto = 100%

## RMS Treino



Menor RMS de treino = 2 nós na camada escondida

## RMS Teste



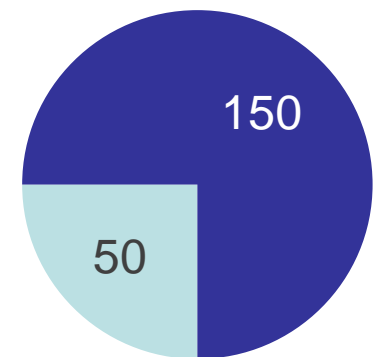
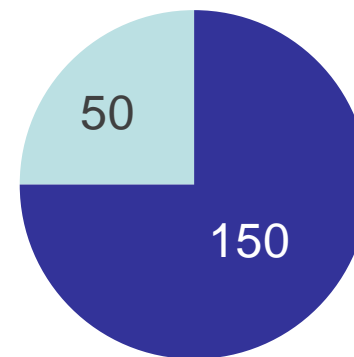
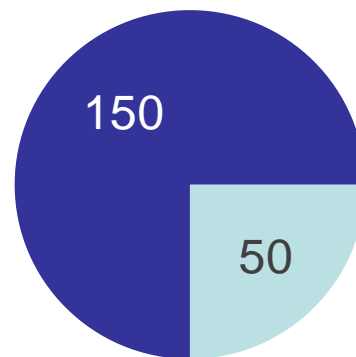
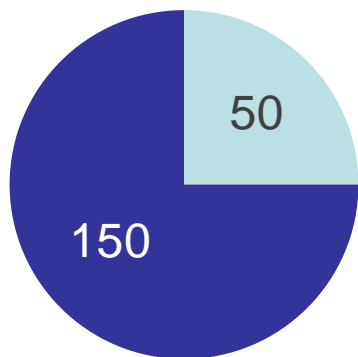
Menor RMS de teste = 13 nós na camada escondida

Para verificação do “poder de extrapolação” da rede foi feito o 25-fold, ou seja, 25% do conjunto de dados foi separado para conjunto de treino e fizemos uma validação cruzada.

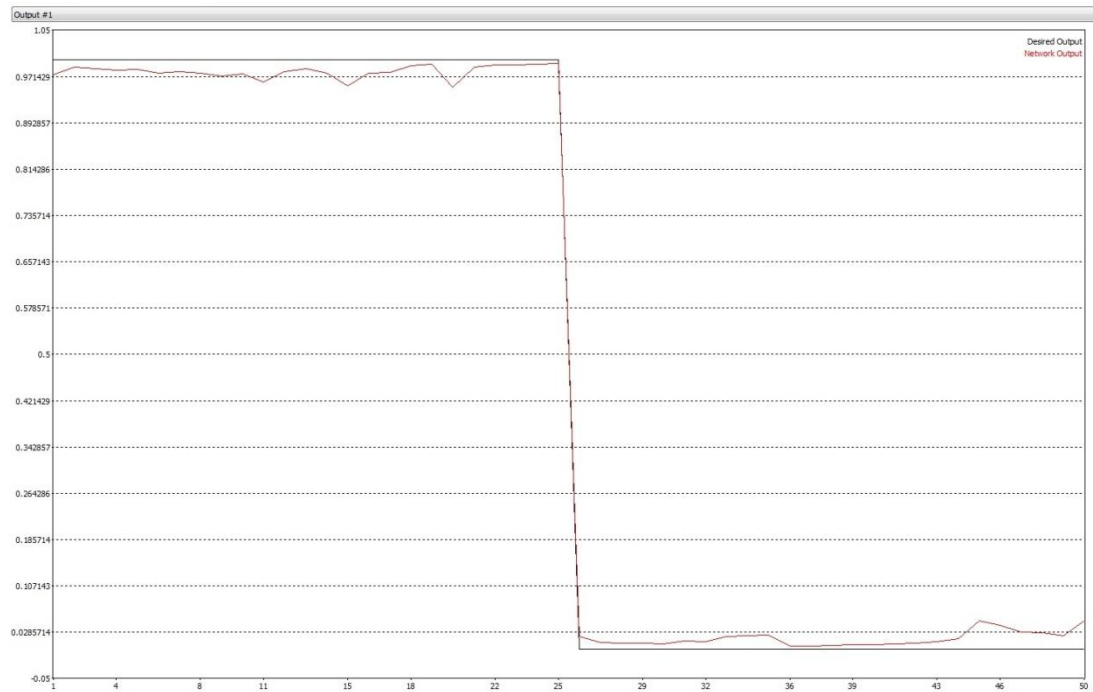
Número total de amostras = 200 exemplares

Amostras de treino = 50 exemplares

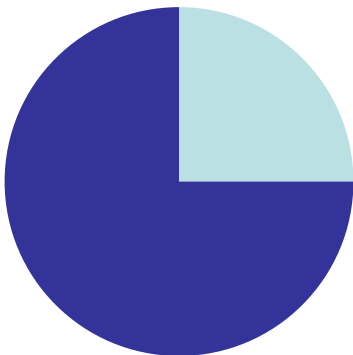
Amostras de teste = 150 exemplares



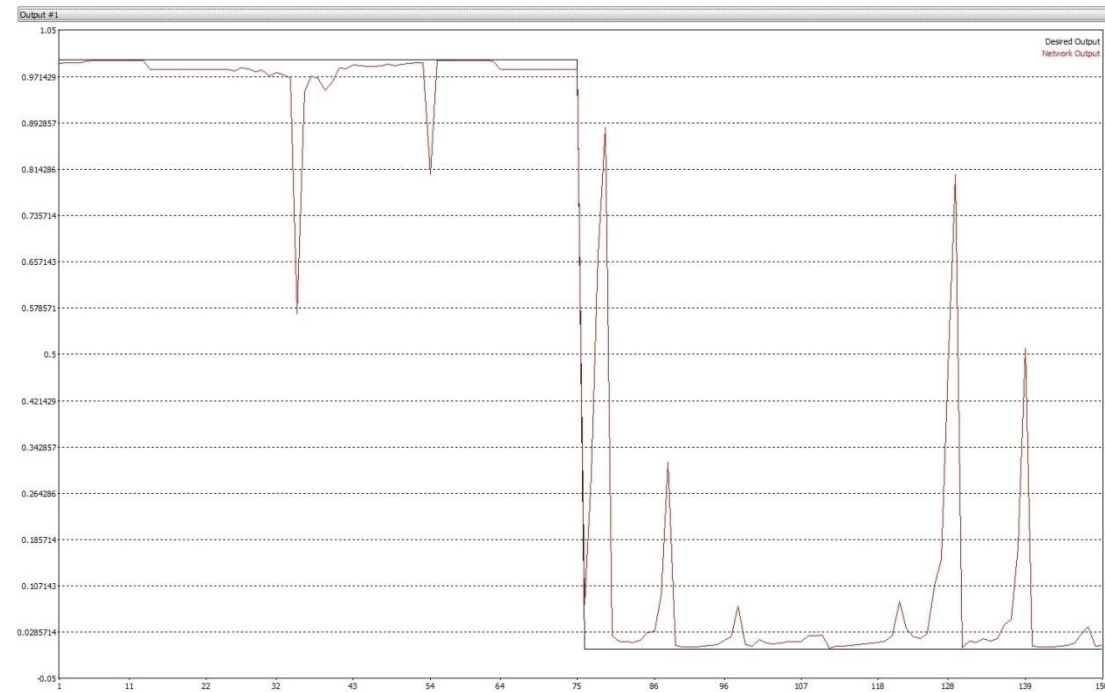
## Conjunto de treino



Taxa de acerto = 100%



## Conjunto de teste



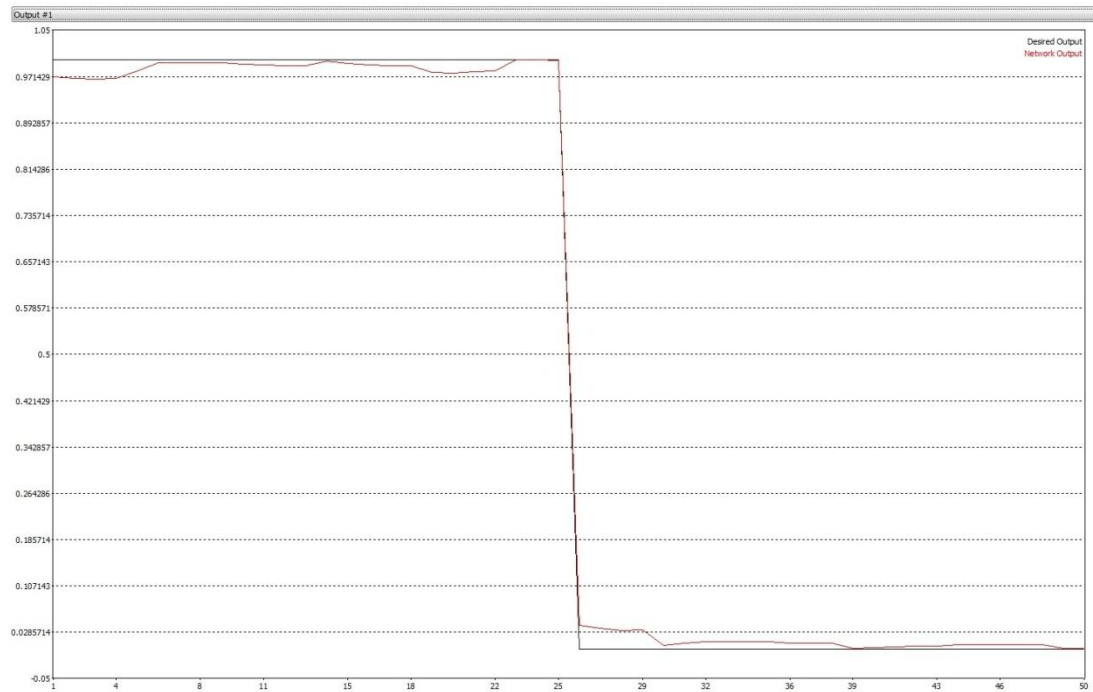
Taxa de acerto = 97%

Matriz de confusão  
do teste

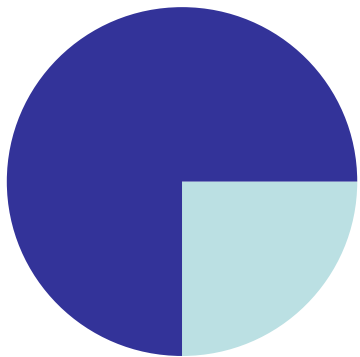
		SAIDA CLASSIFICADOR	
		BOM	RUIM
AMOSTRAS	BOM	75	0
	RUIM	4	71



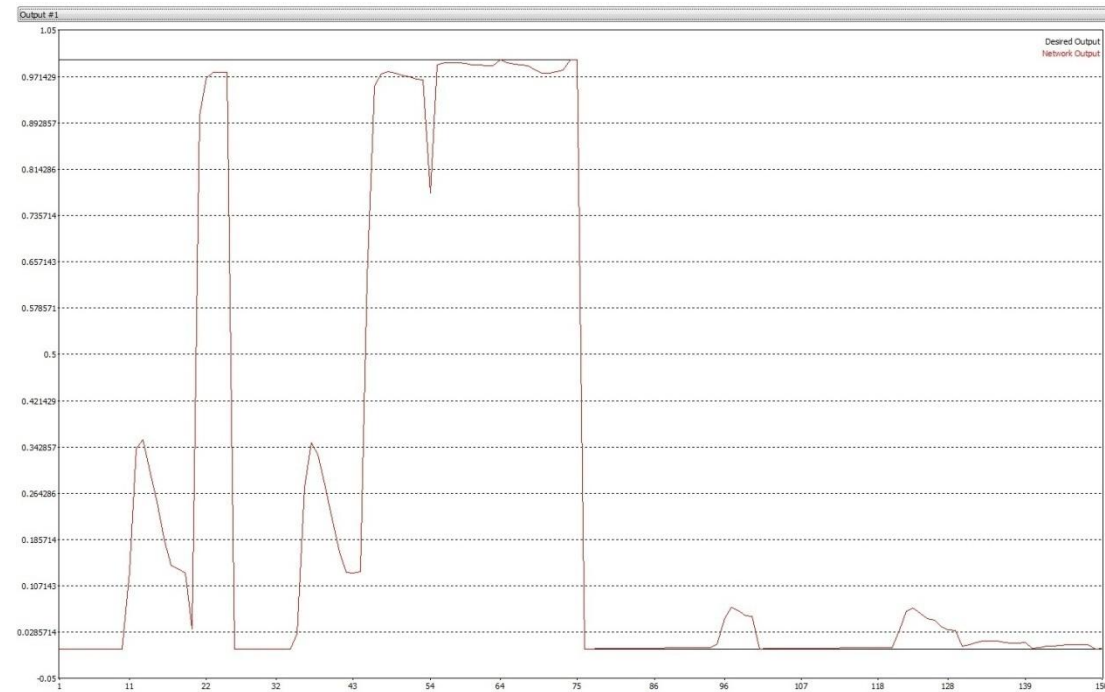
## Conjunto de treino



Taxa de acerto = 100%



## Conjunto de teste

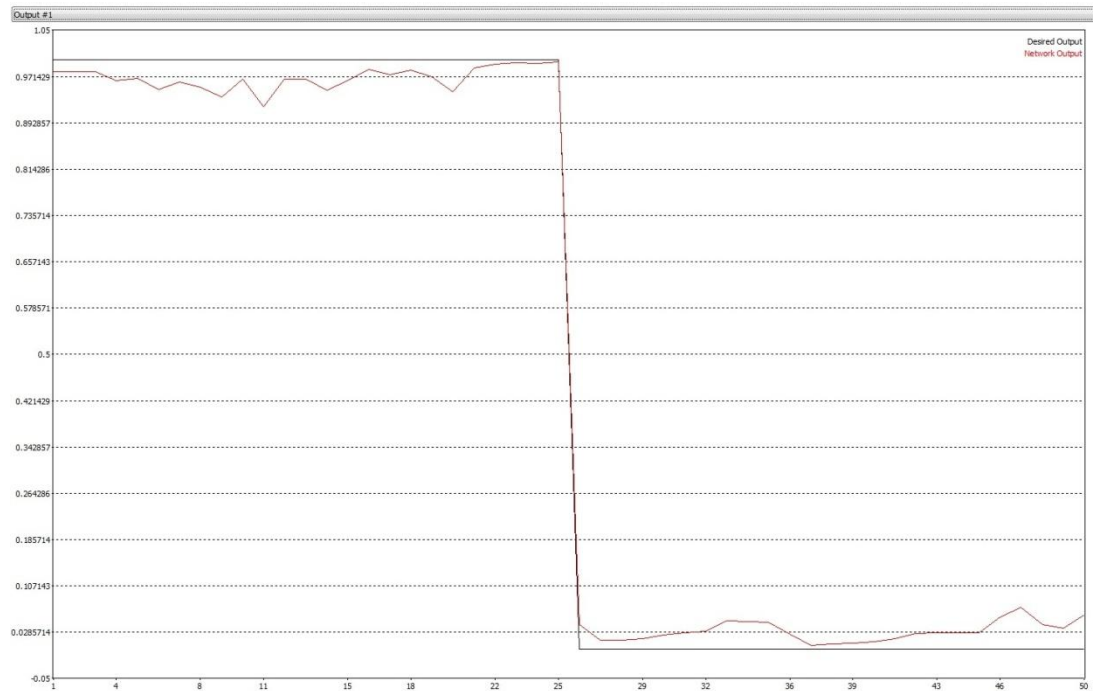


Taxa de acerto = 74%

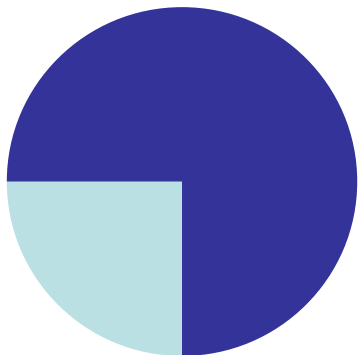
Matriz de confusão  
do teste

		SAIDA CLASSIFICADOR	
		BOM	RUIM
AMOSTRAS	BOM	36	39
	RUIM	0	75

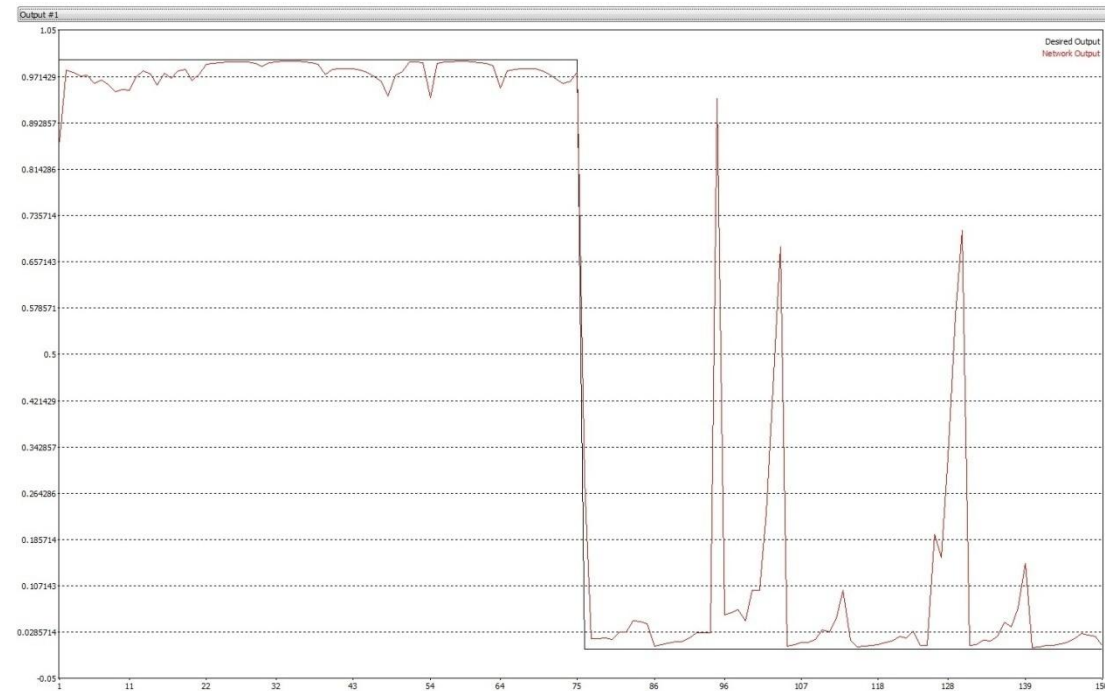
## Conjunto de treino



Taxa de acerto = 100%



## Conjunto de teste

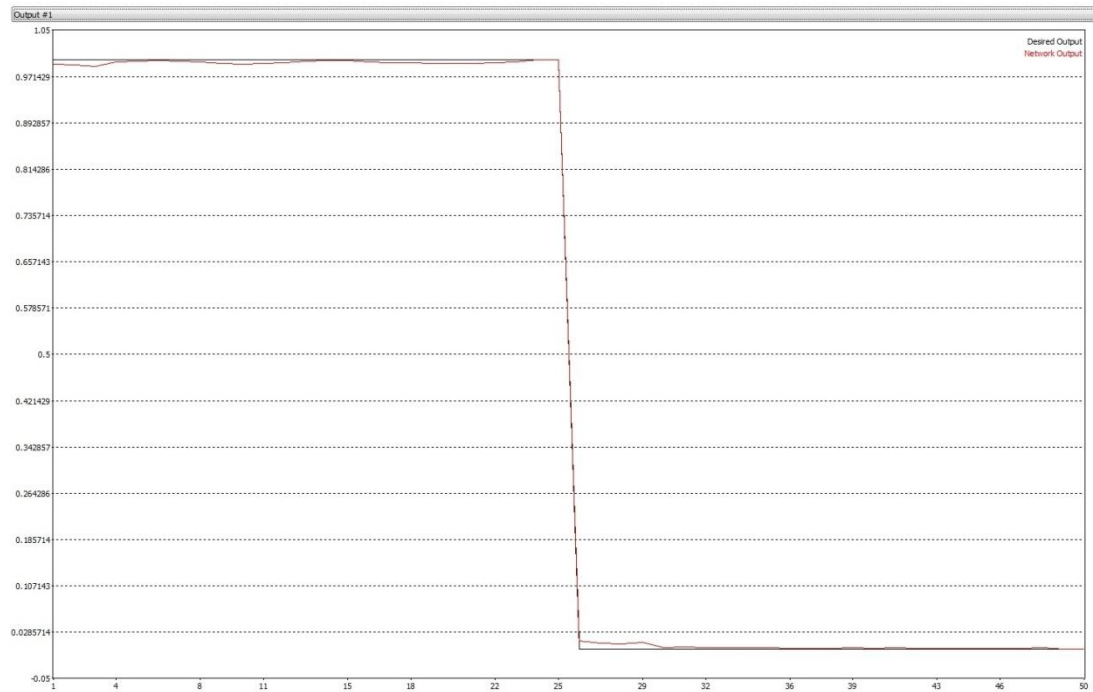


Taxa de acerto = 97%

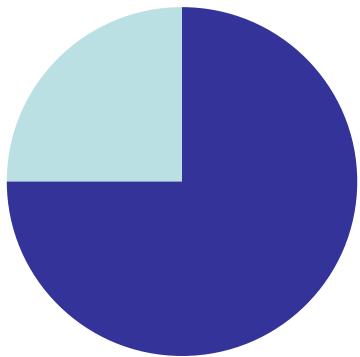
Matriz de confusão  
do teste

		SAIDA CLASSIFICADOR	
		BOM	RUIM
AMOSTRAS	BOM	75	0
	RUIM	4	71

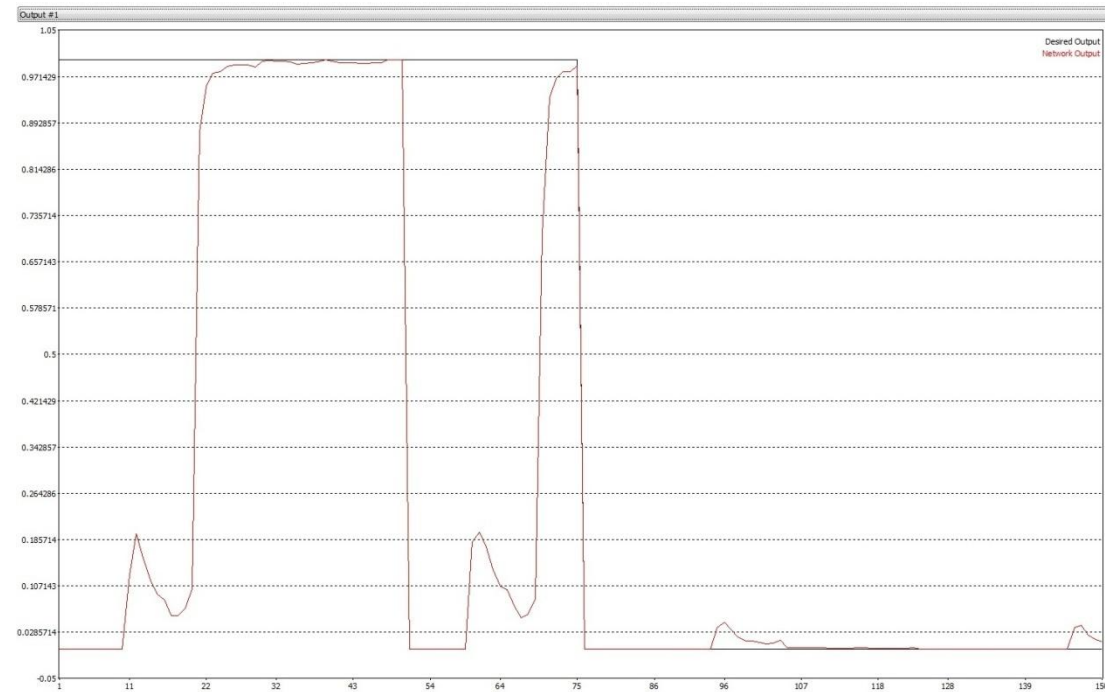
## Conjunto de treino



Taxa de acerto = 100%



## Conjunto de teste



Taxa de acerto = 74%

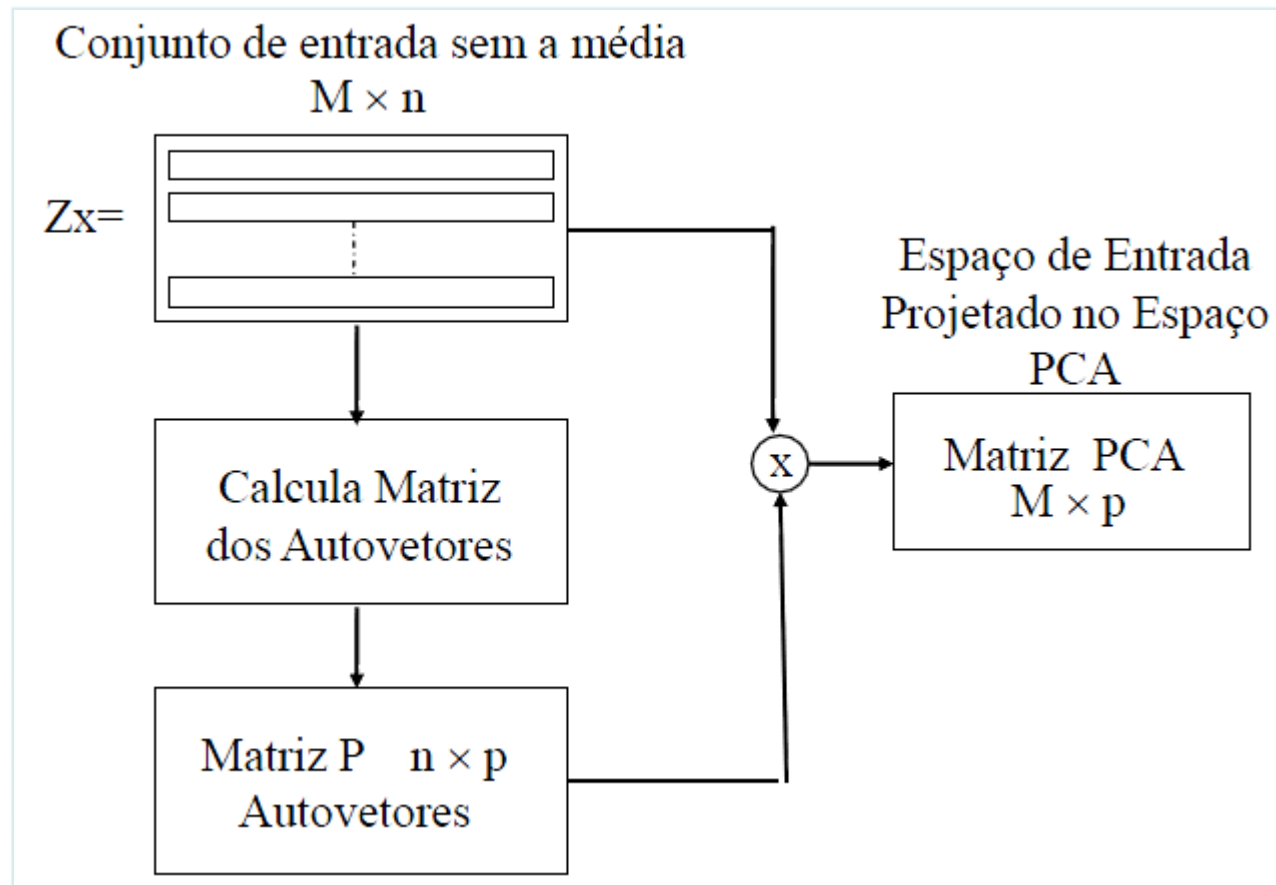
Matriz de confusão  
do teste

		SAIDA CLASSIFICADOR	
		BOM	RUIM
AMOSTRAS	BOM	36	39
	RUIM	0	75

Os dados coletados da base de dados, apresentam um alto nível de otimização por causa de uma boa implementação em hardware dos sensores de capacitância e resistência. Porém todos juntos geram uma alta densidade de dados para a rede.

Para degenerar um pouco os dados e obter menor densidade na entrada da rede neural artificial foram utilizados dois diferentes técnicas de pré-processamento dos dados. PCA e media estatística dos dados.

- Algoritmo PCA

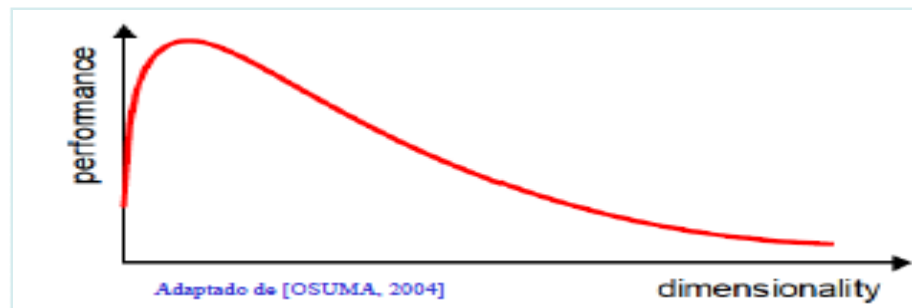


Esquema geral da transformação pro espaço PCA.

Em termos gerais procura-se os componentes com mais distancia estatística ou componentes principais a partir dos autovalores e auto vetores da matriz de covariância do conjunto original de dados menos sua media.

A media representa a repetitividade dos dados, por tanto queremos a principio subtrair dos dados originais aquela repetitividade que acrescenta a dimensionalidade dos mesmos.

O alvo da redução da dimensionalidade dos dados mediante o analise de componentes principais, tem a ver com o aumento do desempenho e a redução de carga de processamento.





A matriz de covariância não considerará todas as características do espaço de entrada . Consequentemente, haverá um espaço de características nulo na matriz de covariância, e essa matriz será considerada degenerada, tendo no máximo ordem  $N-1$  com  $N$  para o nosso caso a quantidade de amostras pegadas pelo conjunto de 18 sensores ( $N=100$ ). Esse problema é conhecido também com problema de SSS (Small Sample Size) e faz com que a base vetorial formada pelos auto vetores fique limitada a  $N-1$  vetores linearmente independentes, limitando o grau de liberdade das projeções.

## ROTINA EM MATLAB

```
clear all % limpa todas as variaveis
clc      % limpa a tela
arquivo='Treino e Teste v3.xlsx'; % carrega os dados de entrada
[fid, message] = fopen(arquivo,'r'); % abre arquivo para leitura
header_a01 = fgetl(fid); % Leitura do cabeçalho
d=importdata(arquivo); % Leitura dos dados
dados=d.data;
t=dados.TreinoV3;
status = fclose(fid); % Fecha o arquivo
for i=2:20;
    Zx(:,i-1)=t(:,i); % Criação da matriz Zx
End
%%%PRINCIPAIS COMPONENTES DE ANALISE%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
MZx=ones(size(Zx,1),1)*mean(Zx); % Tira a media dos dados
Zax=Zx-MZx; % Substrae a media ou repetitividade dos dados
```

```
V=Zax*Zax'; % Definição a matriz de covariancia
```

```
[Qv,Qa]=eig(V,'nobalance'); autovetores e autovalores da matriz de covariancia
```

```
Pt = (Zax'*Qv*(Qa^(-0.5)))./(sqrt(90-1));
```

```
[Ks, Ki] = sort(diag(Qa));
```

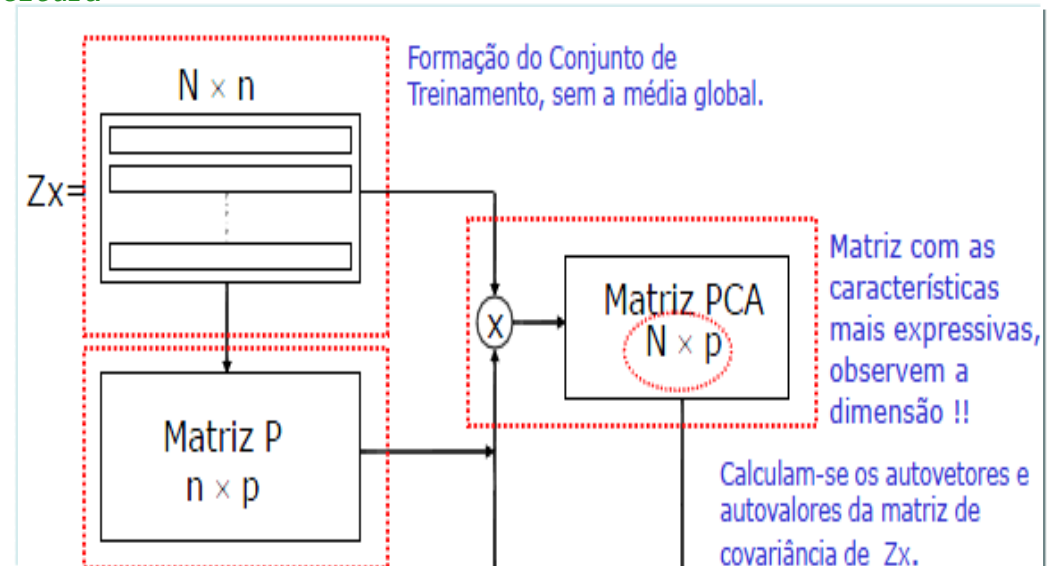
```
Ki = flipud(Ki);
```

```
P=Pt(:,Ki(1:10));
```

```
K = flipud(Ks);
```

```
K=K(1:10);
```

```
PCA=Zx*P; % Matriz PCA com as componentes principais da matriz de dados de entrada
```



## MEDIA ESTATÍSTICA

Mas um médio utilizado para redução de custo computacional foi a média estatística. A partir dos dados de entrada de resistência e capacitância dos 18 sensores foi calculada uma media por amostra ou por frequência.

### ROTINA IMPLEMENTADA EM MATLAB

$$\bar{\mathbf{x}} = \left[ \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \right]$$

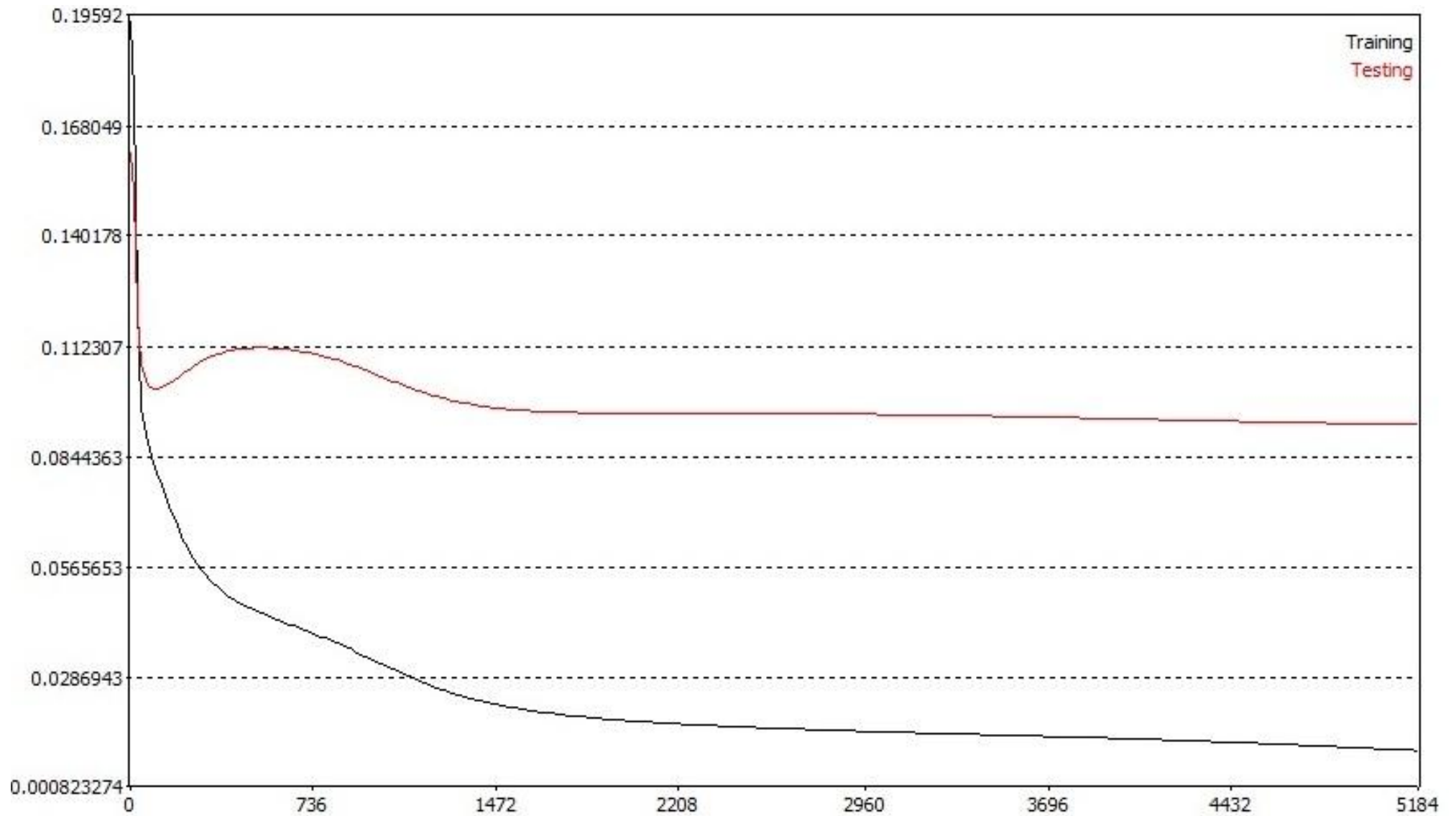
- `for j=1:length(Zx);`
- `medC(j,:)=mean(Zx(j,1:9));`     %media das capacitancias capturadas por frequencia
- `medR(j,:)=mean(Zx(j,10:18));` %media das resistencias capturadas por frequencia
- `freq(j,:)=Zx(j,19);`             %flaquencia
- `end`

Os **dados** utilizados no **projeto secundário** foram retirados do repositório de aprendizado de máquinas da Universidade da Califórnia em Irvine. Os trabalhos foram desenvolvidos por P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.

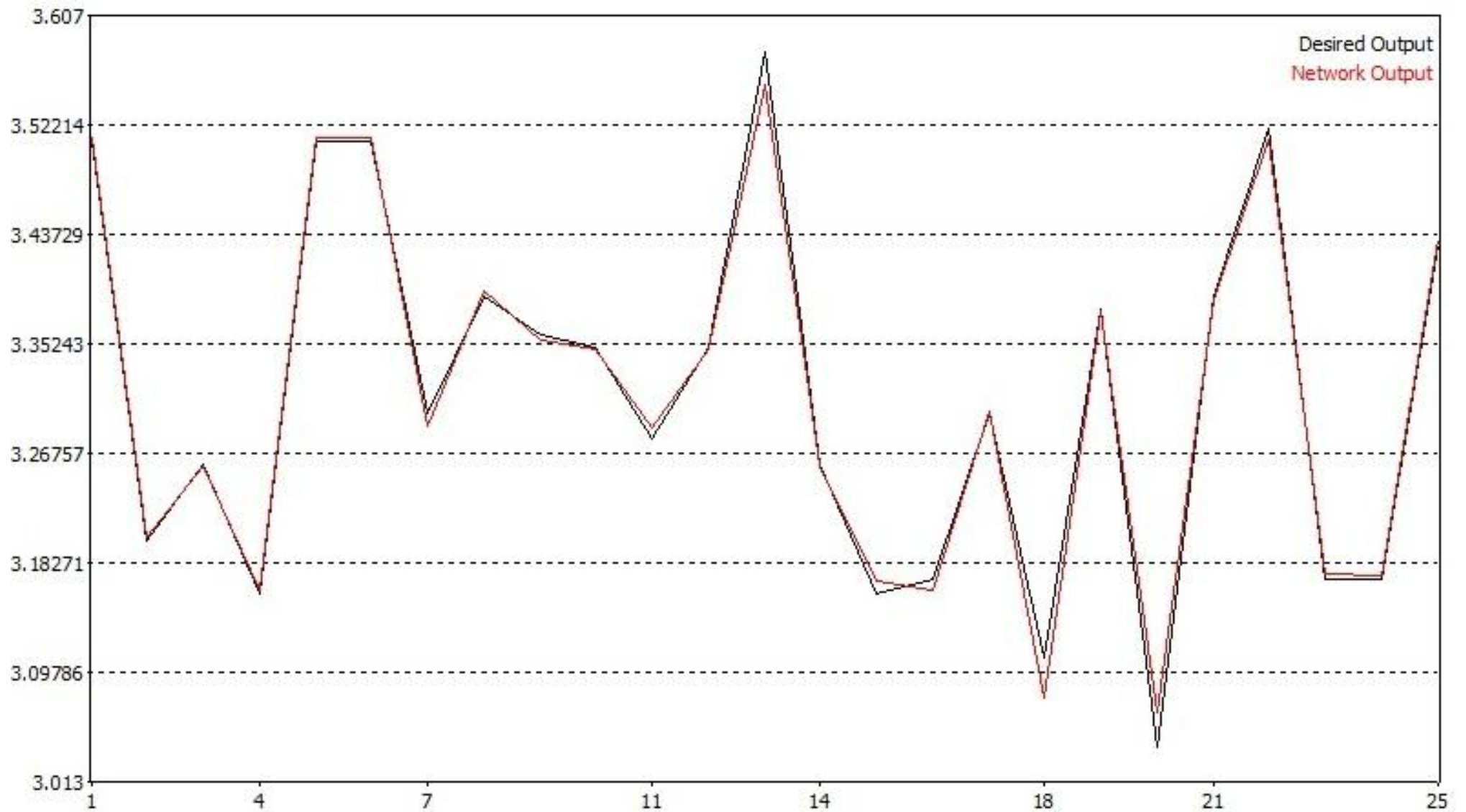
Dados disponíveis em  
<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>



## Projeto Secundário - Resultados

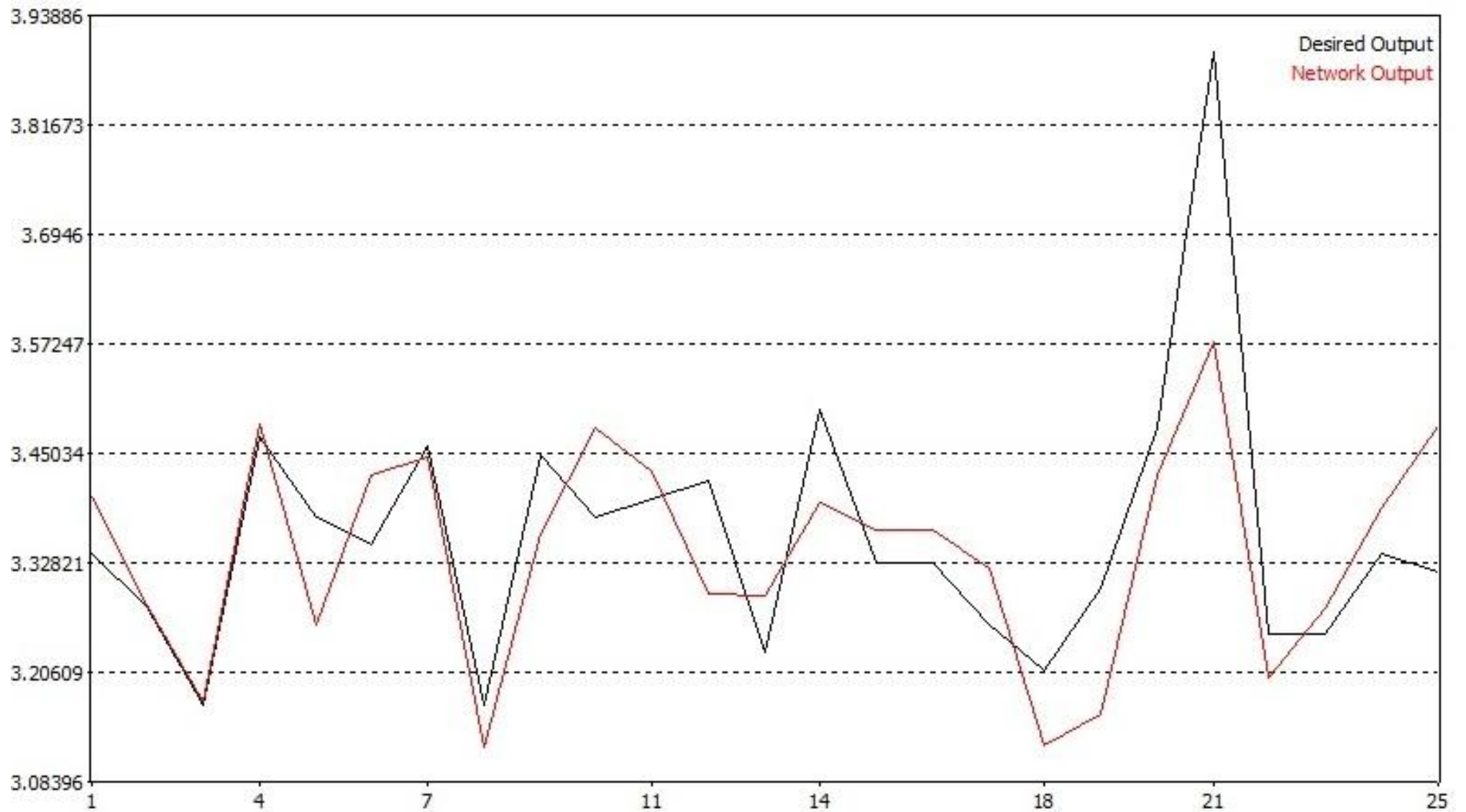


## Projeto Secundário - Resultados





## Projeto Secundário - Resultados



---

# OBRIGADO!