  
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
DOUTORADO EM MÉTODOS DE APOIO À DECISÃO

**TRABALHO 2: PREVISÃO DE SÉRIES TEMPORAIS**

**Gabriel Baruque**

**2012325**

Rio de Janeiro

2020

GABRIEL BARUQUE

**TRABALHO 1: ANÁLISE DE CRÉDITO BANCÁRIO**

Trabalho apresentado à disciplina ELE2394 – Redes Neurais I, do programa de Pós Graduação em Métodos de Apoio à Decisão da Coordenação de Engenharia Elétrica da PUC-Rio como requisito parcial de avaliação.

Professora: Marley Vellasco

Rio de Janeiro

2020

# Introdução

Neste trabalho, utilizamos uma série temporal de demanda de energia com o intuito de analisar as diferentes configurações possíveis em uma rede neural em termos de janela temporal, codificação dos atributos de entrada, numero de processadores, épocas a serem treinadas e early stopping, além das funções utilizadas na camada escondida e de saída.

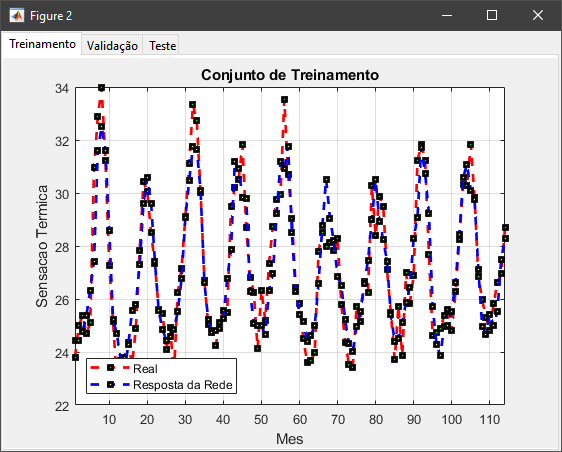
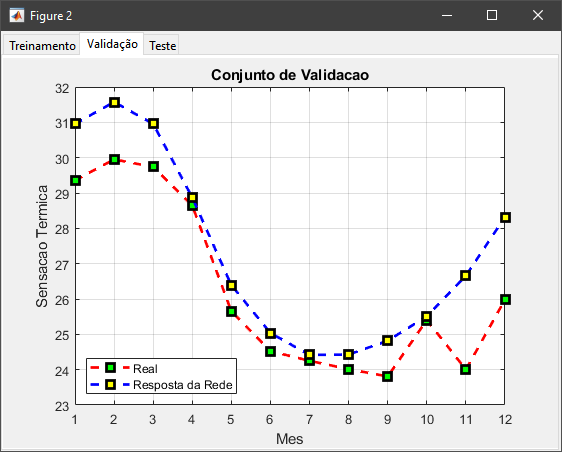
Diversos experimentos foram feitos e comentários sobre cada um, após os resultados obtidos. Por fim, uma conclusão sobre os diversos parâmetros modificados é realizada.

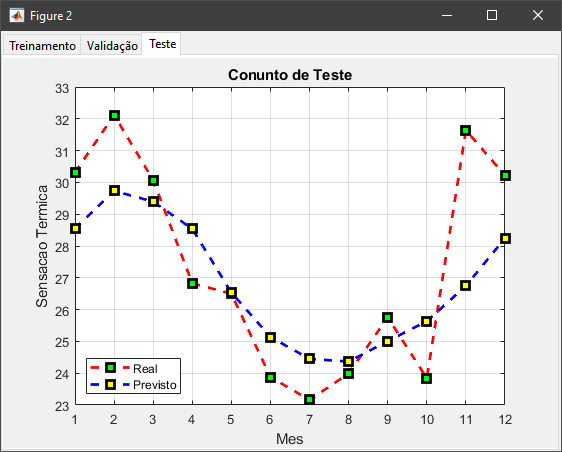
Neste trabalho, entre os 8 microclimas possíveis, foi escolhido o microclima 6 para ser utilizado nos testes. Além disso, para gerar um valor estatístico mais robusto, mas não comprometer muito o tempo de execução dos experimentos, para cada um, 8 redes iguais foram testadas

# Resultados

**TESTE 1 – A**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 1 | Janela = | 6 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 7 | Nepoch = | 100 | Numchkval = | 100 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | trainlm |



Métricas:

MAPE\_v\_f = 3.9873

RMSE\_v\_f = 1.3316

MAPE\_t\_f = 5.5701

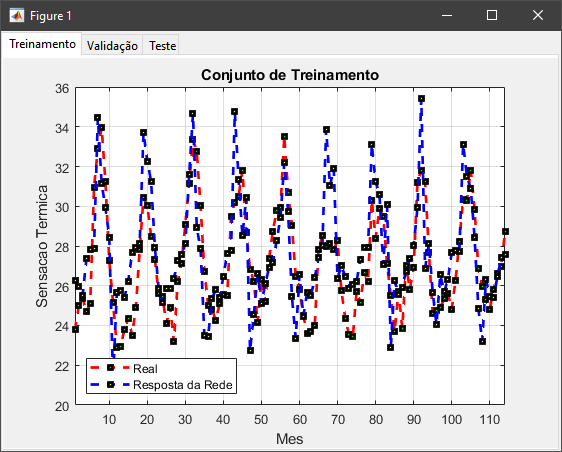
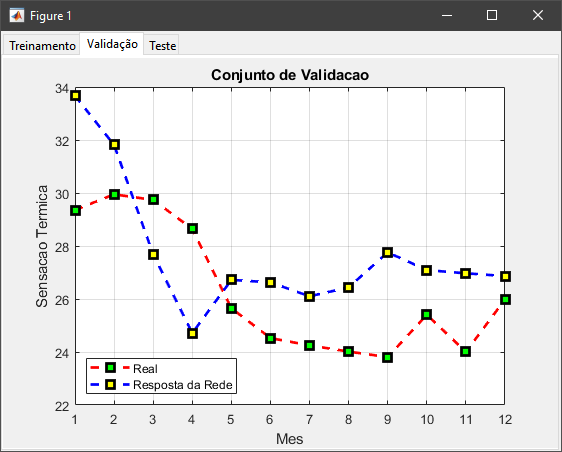
RMSE\_t\_f = 1.9826

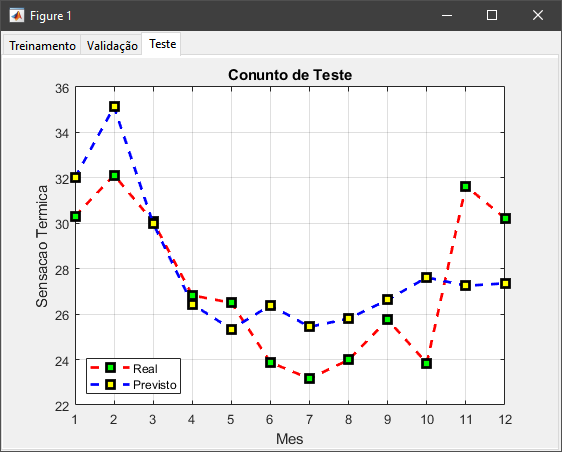
No primeiro teste, as redes foram treinadas com 7 processadores, sem early stopping (100 épocas) e com as funções ‘tansig’ na ativação nas camadas escondidas e ‘purelin’ na saída. O algoritmo de treinamento utilizado foi o ‘trainlm’.

Nessa configuração, a rede obteve, a princípio, bons resultados, obtendo valores pequenos de MAPE e RMSE, tanto na validação quanto no teste. Através dos gráficos, podemos imaginar que a rede precisa generalizar menos, ou seja, tentar se adequar mais ao “formato” dos dados, para melhorar seu desempenho na validação e teste.

**TESTE 1 – B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 1 | Janela = | 6 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 10 | Nepoch = | 100 | Numchkval = | 100 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | traingdm |



Métricas:

MAPE\_v\_f = 9.2493

RMSE\_v\_f = 2.6554

MAPE\_t\_f = 7.6337

RMSE\_t\_f = 2.4279

Como mencionado no teste anterior, para conseguir uma generalização menor, mais processadores foram utilizados, nesse caso, 10. Outra mudança feita, foi no algoritmo de treinamento, utilizando agora o ‘traingdm’.

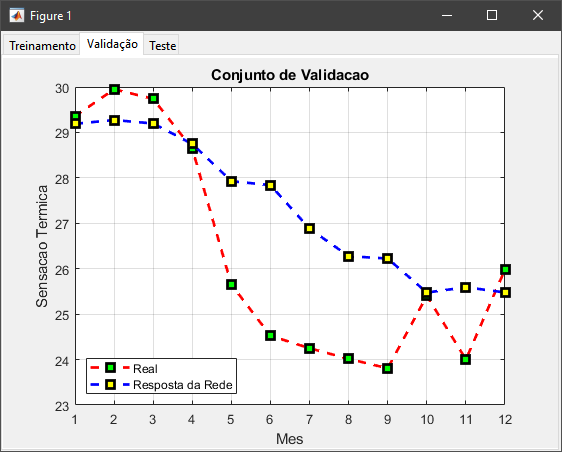
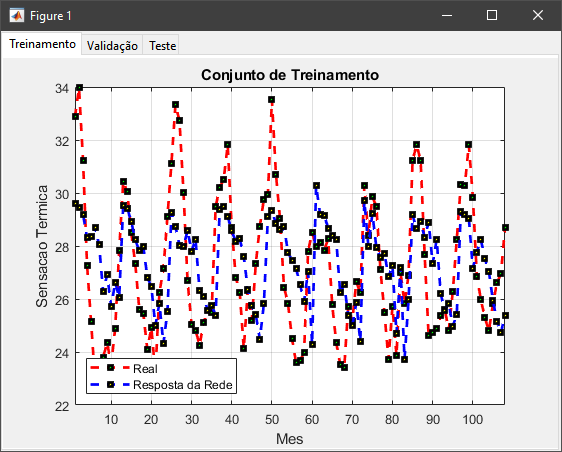
Nessa configuração, percebemos que o resultado obteve uma piora considerável, obtendo padrões bem diferentes do esperado, nos dados de teste e principalmente nos de validação. Mesmo no treino, vemos “picos” muito grandes, onde não era o esperado. É provável que esse comportamento estranho tenha sido obtido pelo algoritmo de treinamento ‘traingdm’, uma vez que apenas este e o número de processadores foram alterados.

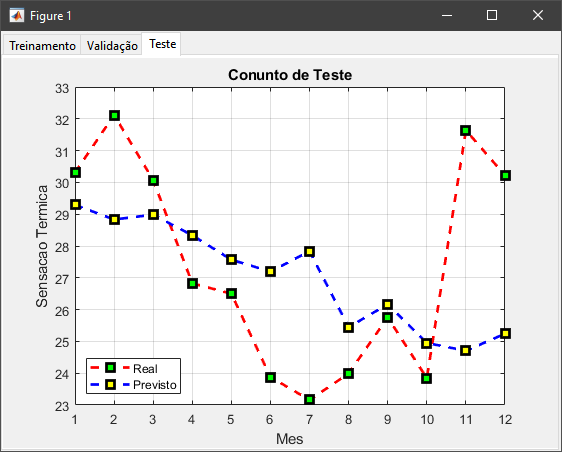
Também, para este teste e o anterior, uma janela de 6 meses foi escolhida como valor de referência para comparação futura.

Podemos ver como os valores de MAPE e RMSE tanto para validação quanto para teste confirmam o pior resultado visualizado nos gráficos. É provável que o treinamento por ‘traingdm’ não seja a melhor opção para esse problema.

**TESTE 2 – A**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 1 | Janela = | 12 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 4 | Nepoch = | 200 | Numchkval = | 30 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | traingdx |



Métricas:

MAPE\_v\_f = 5.5550

RMSE\_v\_f = 1.7706

MAPE\_t\_f = 9.3043

RMSE\_t\_f = 3.2363

Nesta sequência de testes, foi escolhida uma janela de 12 meses e o early stopping foi introduzido, com treinamento em 200 épocas e paciência de 30.

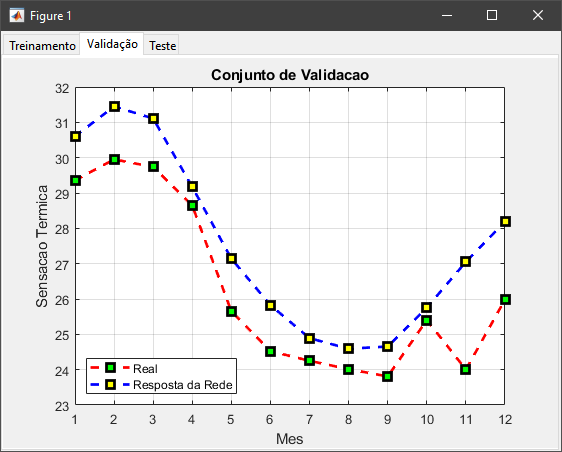
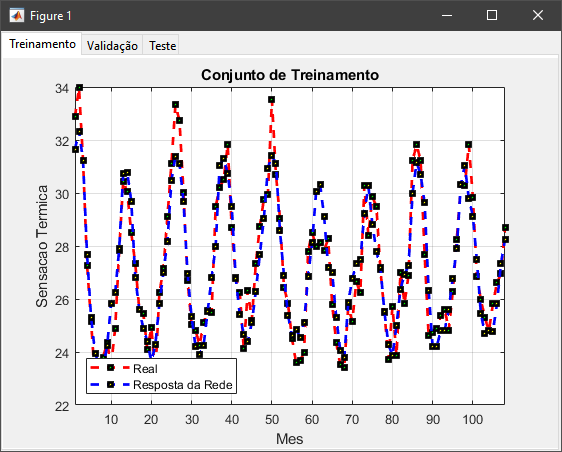
Um número relativamente pequeno de processadores foi escolhido, no teste inicial, 4. Por conta dessa baixa quantidade de processadores, percebemos que a rede não foi capaz de se adequar aos dados no treinamento, e precisaria de mais épocas ou processadores para se aproximar do formato desejado.

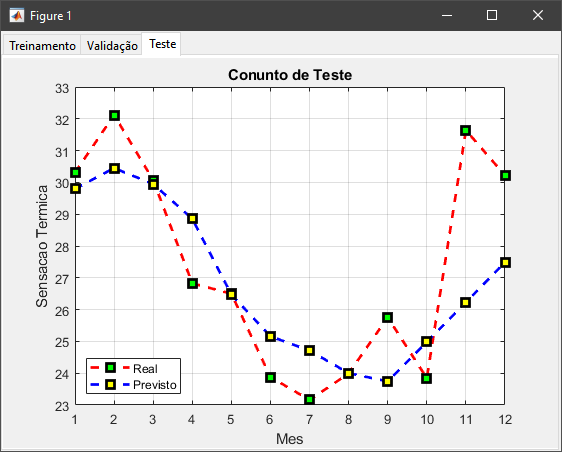
Além disso, o treinamento ‘traingdx’ foi utilizado, podendo influenciar no resultado não esperado da rede.

Apesar do treinamento ter sido aquém do desejado, pela observação gráfica, as métricas, principalmente para os dados de teste, mostram como a rede não foi capaz de gerar resultados promissores.

**TESTE 2 – B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 1 | Janela = | 12 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 9 | Nepoch = | 200 | Numchkval = | 30 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | trainlm |



Métricas:

MAPE\_v\_f = 4.8317

RMSE\_v\_f = 1.4608

MAPE\_t\_f = 5.4861

RMSE\_t\_f = 2.1078

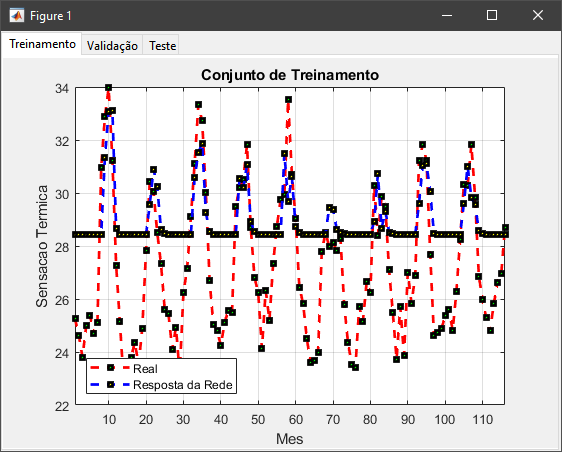
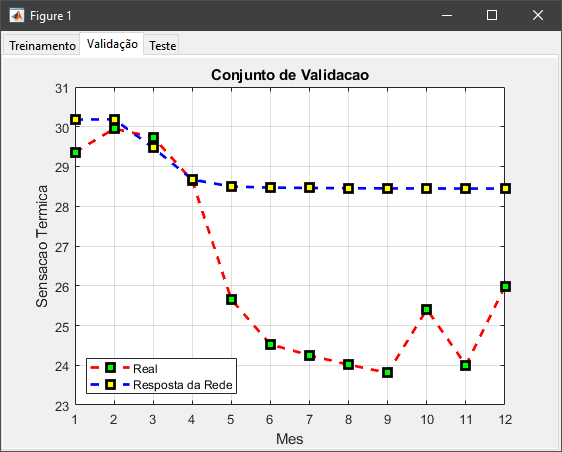
Para melhorar o desempenho da rede, o treinamento foi alterado para ‘trainlm’

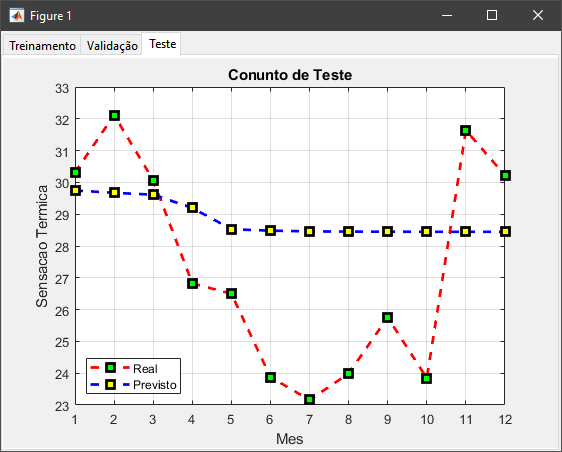
Nesse caso, a rede obteve bons resultados, conseguindo uma boa generalização, vista através dos gráficos, e também refletida nos valores de MAPE e RMSE.

Ainda assim, a configuração com 7 processadores e janela menor obteve melhores resultados, indicando que é possível ter utilizado uma janela muito grande (piorando o treinamento da rede) ou um número de processadores inadequado.

**TESTE 3 - A**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 1 | Janela = | 4 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 13 | Nepoch = | 200 | Numchkval = | 30 |
| Tipofunc = | logsig | Funcsaida = | logsig | Algtrei = | trainlm |





Métricas:

MAPE\_v\_f = 10.5882

RMSE\_v\_f = 3.1422

MAPE\_t\_f = 11.1668

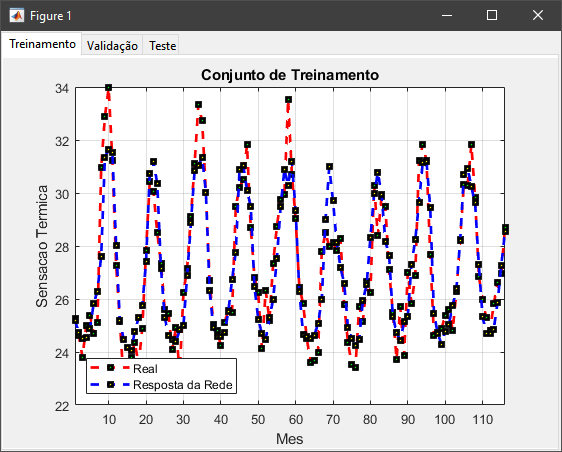
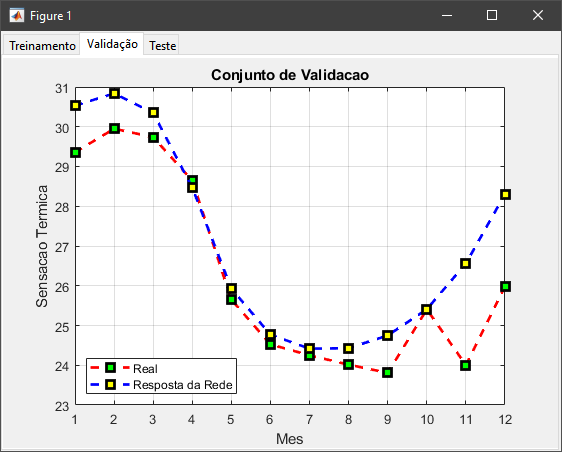
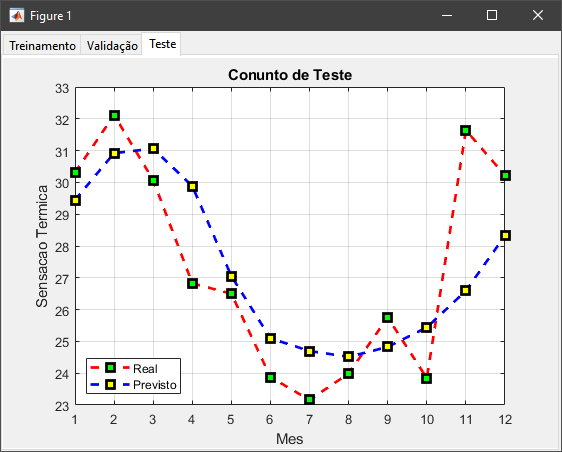
RMSE\_t\_f = 3.2564

Neste teste, com uma janela de 4 meses e utilizando 13 processadores, foi utilizada a função ‘logsig’ para a ativação das camadas escondidas e para a camada de saída. Pode-se observar nos gráficos que alguma das funções fez com que o treinamento ficasse prejudicado

É possível que a rede tenha sofrido paralisia ou a função de ativação na camada de saída não seja adequada para o problema enfrentado no trabalho. Posteriormente, vemos que o mais provável é que a segunda opção tenha acontecido.

**TESTE 3 - B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 1 | Janela = | 4 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 7 | Nepoch = | 400 | Numchkval = | 50 |
| Tipofunc = | logsig | Funcsaida = | purelin | Algtrei = | trainlm |



Métricas:

MAPE\_v\_f = 3.1387

RMSE\_v\_f = 1.1447

MAPE\_t\_f = 5.8170

RMSE\_t\_f = 2.0289

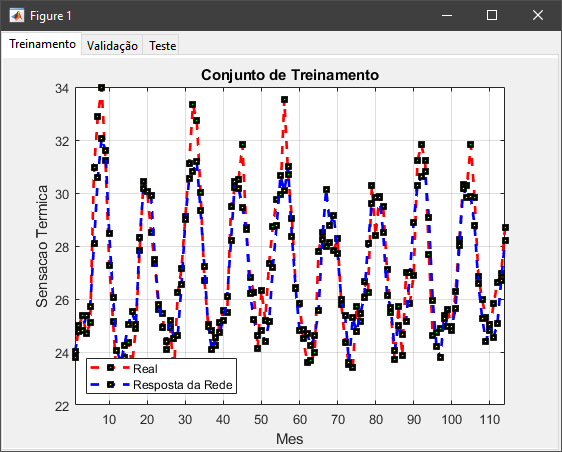
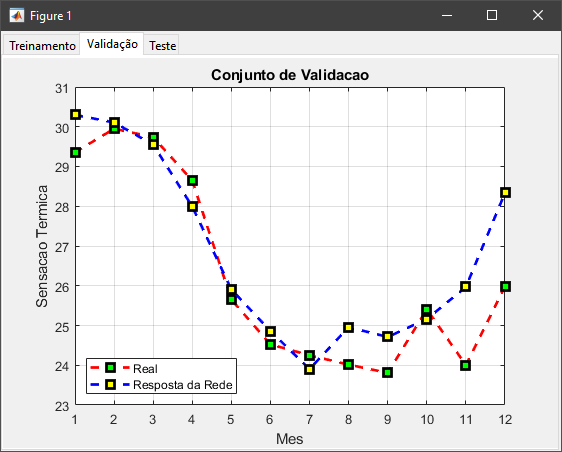
Nesse teste, o número de processadores utilizados foi de 7, e a rede foi treinada por 400 épocas com early stopping de 50.

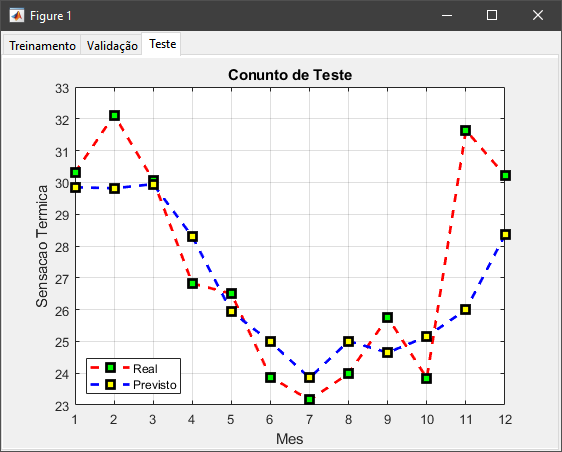
Este teste foi interessante pois mostrou que o problema encontrado anteriormente não foi de paralisia, uma vez que a rede conseguiu bom desempenho utilizando a função de ativação ‘logsig’ nas camadas escondidas.

Observamos pelas métricas que a rede obteve um dos melhores resultados até o momento, tanto em Mean Absolute Percentage Errors quanto em Root Means Squared Errors.

**TESTE 4 - A**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 2 | Janela = | 6 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 10 | Nepoch = | 400 | Numchkval = | 50 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | trainlm |





Métricas:

MAPE\_v\_f = 3.0313

RMSE\_v\_f = 1.0346

MAPE\_t\_f = 5.2010

RMSE\_t\_f = 2.0213

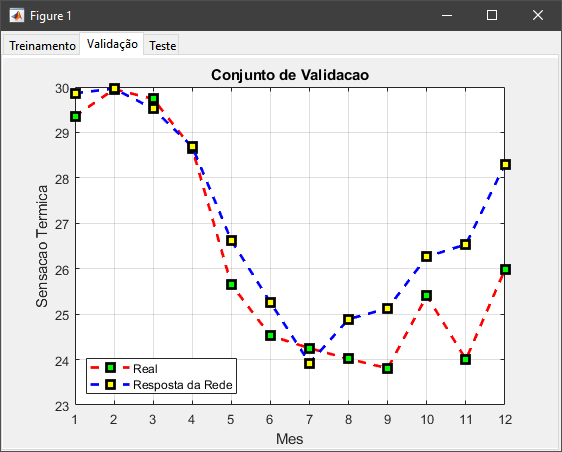
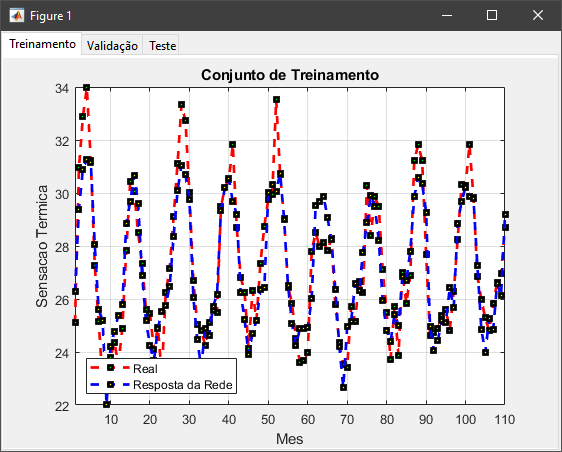
Os testes 1 a 3 foram feitos utilizando a codificação 1, onde os meses eram representadas por valores reais. Nos próximos 3testes a codificação utilizada foi a binária, com 4 bits.

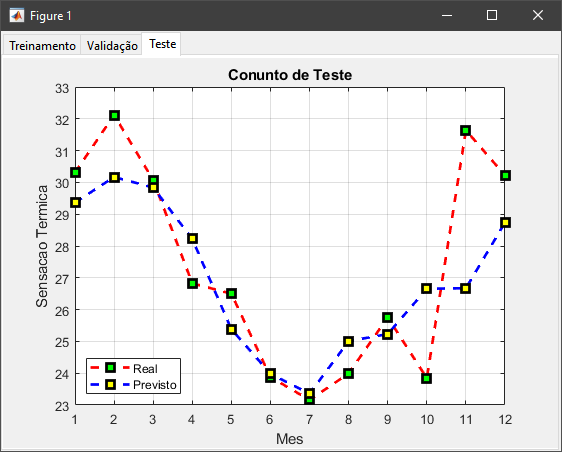
Inicialmente, foram utilizados 10 processadores e uma janela de 6 meses. As funções utilizadas nas camadas escondidas e de saída foram, respectivamente, ‘tansig’ e ‘purelin’. Esse teste foi uma tentativa de se fazer um paralelo com o primeiro teste realizado, ajustando o número de processadores em função da maior quantidade de entradas.

Os resultados foram bem próximos, com uma melhora, com uma melhora pequena nos valores de MAPE. Já para os valores de RMSE, houve uma melhora na validação, e um desempenho ligeiramente inferior (mas muito semelhante) no teste.

**TESTE 4 - B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 2 | Janela = | 10 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 10 | Nepoch = | 400 | Numchkval = | 400 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | trainlm |





Métricas:

MAPE\_v\_f = 3.5680

RMSE\_v\_f = 1.1876

MAPE\_t\_f = 4.9448

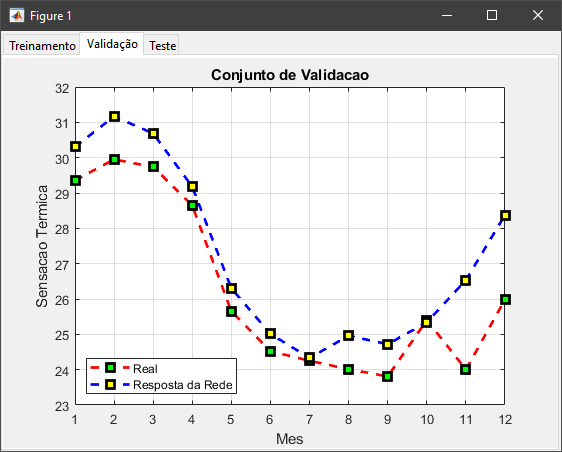
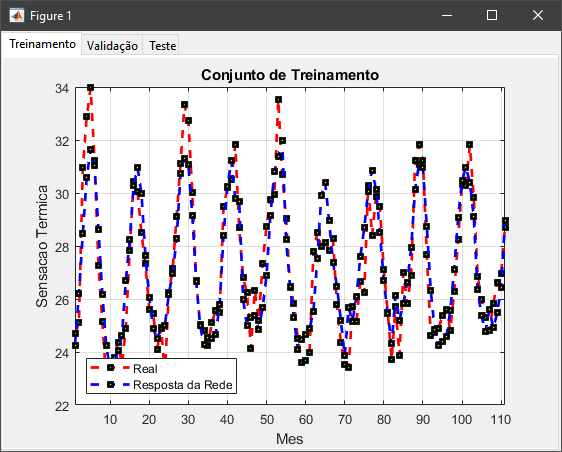
RMSE\_t\_f = 1.9164

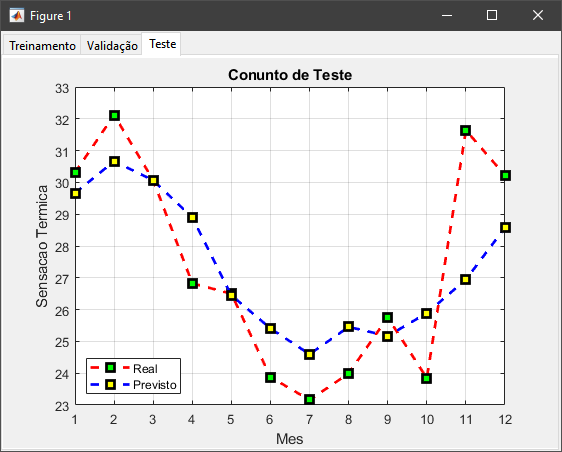
Neste teste, a codificação, o número de processadores e as funções utilizadas foram mantidas inalteradas. As mudanças foram feitas no tamanho da janela, que para este teste foi configurada como 10 meses, e no early stopping, que fo desativado.

Os resultados foram, novamente, bons. Na realidade, foi o melhor resultado obtido para os dados de teste. Estes resultados podem indicar que é necessário utilizar mais épocas ou uma “paciência” maior.

**TESTE 5 - A**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 2 | Janela = | 9 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 15 | Nepoch = | 600 | Numchkval = | 200 |
| Tipofunc = | logsig | Funcsaida = | purelin | Algtrei = | trainlm |





Métricas:

MAPE\_v\_f = 3.7480

RMSE\_v\_f = 1.2280

MAPE\_t\_f = 5.3708

RMSE\_t\_f = 1.8848

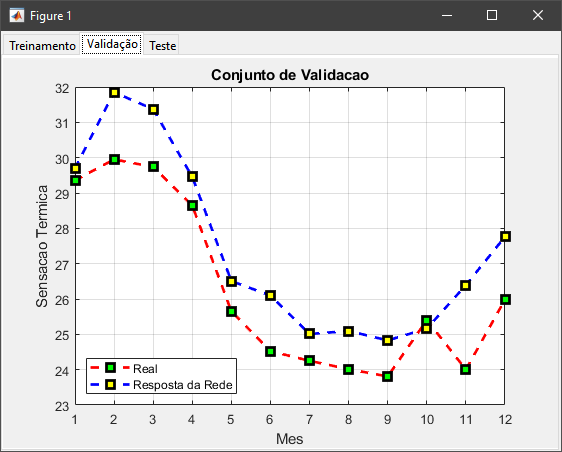
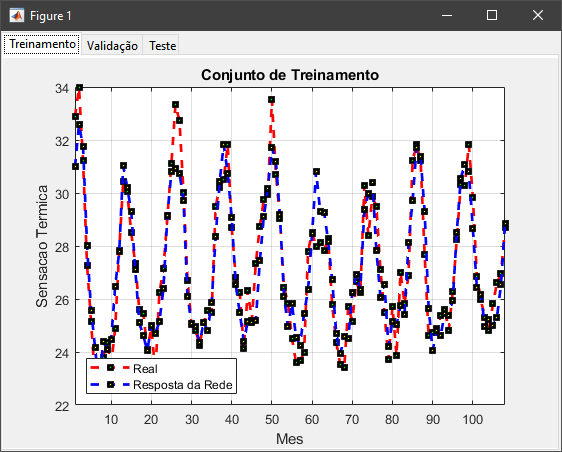
Como, em testes anteriores, a função ‘logsig’ na camada escondida obteve resultados bons, novos testes foram feitos com essa mesma função, utilizando a codificação com 4 bits.

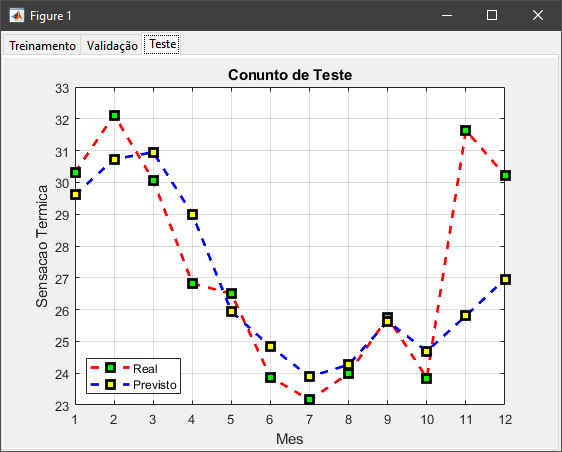
É interessante ressaltar que, como a codificação muda a quantidade de atributos de entrada, a quantidade de neurônios para essas codificações irá também aumentar, para que a estimativa da rede não seja muito afetada (por ter poucos neurônios para o número de entradas mais elevado). Sabe-se que não existe uma heurística ou “regra” definitiva, e por esse motivo, várias experimentações estão sendo feitas neste trabalho.

Os resultados para o teste, foram também bons, principalmente para os dados de teste, em específico para a métrica RMSE, que obteve o valor mais baixo até agora.

**TESTE 5 - B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 2 | Janela = | 12 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 20 | Nepoch = | 800 | Numchkval = | 200 |
| Tipofunc = | logsig | Funcsaida = | purelin | Algtrei = | trainlm |





Métricas:

MAPE\_v\_f = 4.6061

RMSE\_v\_f = 1.3538

MAPE\_t\_f = 5.1155

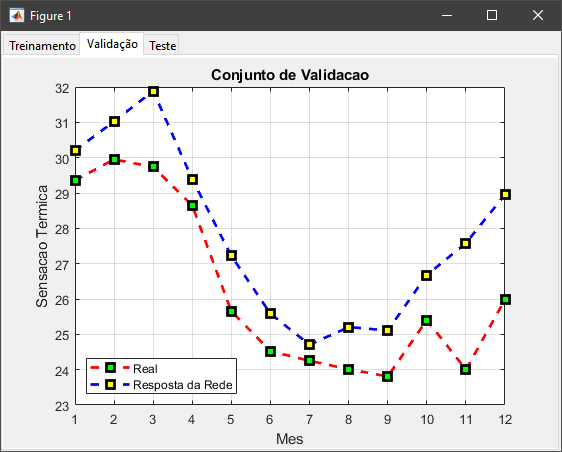
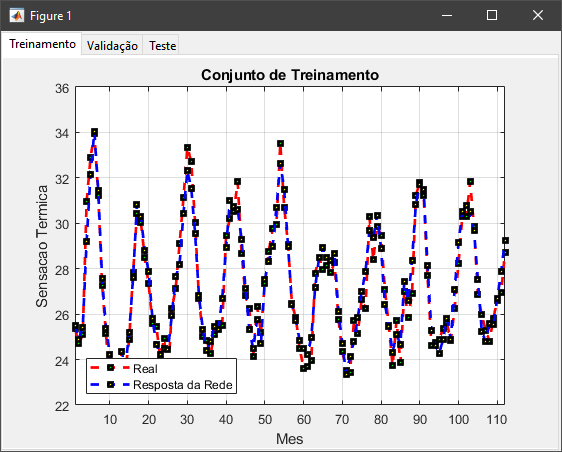
RMSE\_t\_f = 2.1437

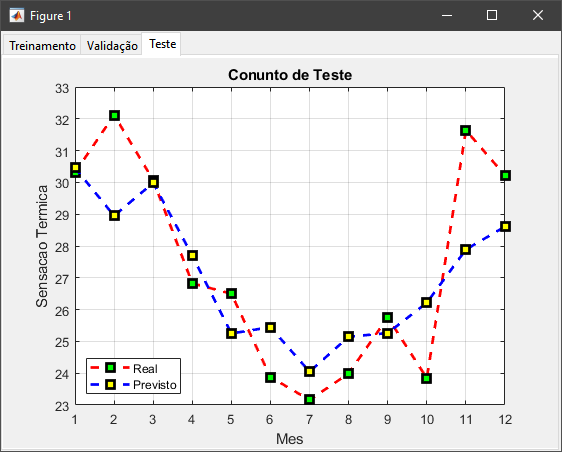
Neste teste, o modelo foi treinado por mais épocas, ainda com um early stopping de 200 épocas.

Pode-se observar que os as métricas voltaram a possuir valores ligeiramente maiores, o que não é desejado. É provável então que 12 meses, seja uma janela muito grande, e prejudica a predição do modelo, assim como suposto no Teste 2 – B.

**TESTE 6 - A**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 3 | Janela = | 8 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 30 | Nepoch = | 1000 | Numchkval = | 200 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | trainlm |





Métricas:

MAPE\_v\_f = 5.8655

RMSE\_v\_f = 1.7608

MAPE\_t\_f = 5.2404

RMSE\_t\_f = 1.8143

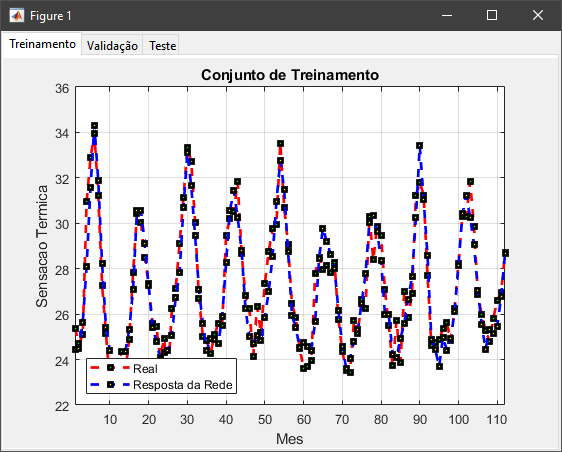
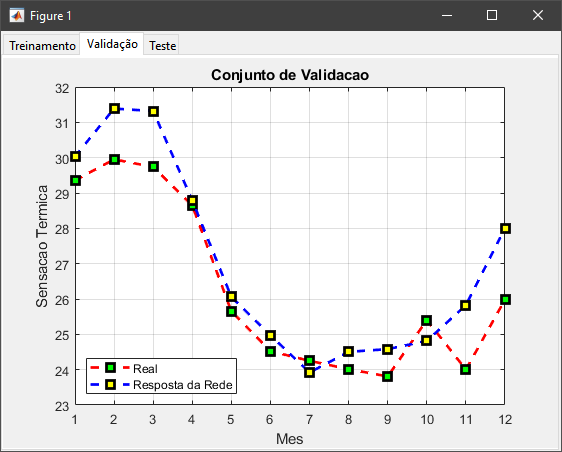
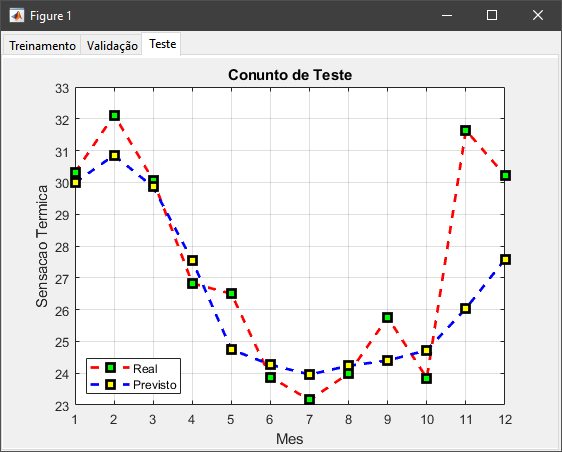
Neste teste, introduziu-se a codificação binária com 12 bits. Com isso, a quantidade de atributos de entrada é maior, necessitando, em teoria, de mais processadores para que a rede possa ter um desempenho adequado.

Nesse caso, 30 processadores foram utilizados em uma camada escondida, e a quantidade de épocas foi aumentada para 1000, mantendo-se a paciência em 200 épocas. A janela utilizada foi de 8 meses, não tão grande a ponto de diminuir o desempenho, como no último teste.

Podemos observar que os resultados das métricas foram piores que outros experimentos, apesar do RMSE no teste ter tido o melhor resultado até então. Mesmo observando os gráficos, podemos observar alguns pontos que poderiam ter sido preditos de uma forma mais acurada.

**TESTE 6 - B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 3 | Janela = | 8 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 13 | Nepoch = | 1000 | Numchkval = | 500 |
| Tipofunc = | tansig | Funcsaida = | purelin | Algtrei = | trainlm |



Métricas:

MAPE\_v\_f = 3.3892

RMSE\_v\_f = 1.0845

MAPE\_t\_f = 4.7067

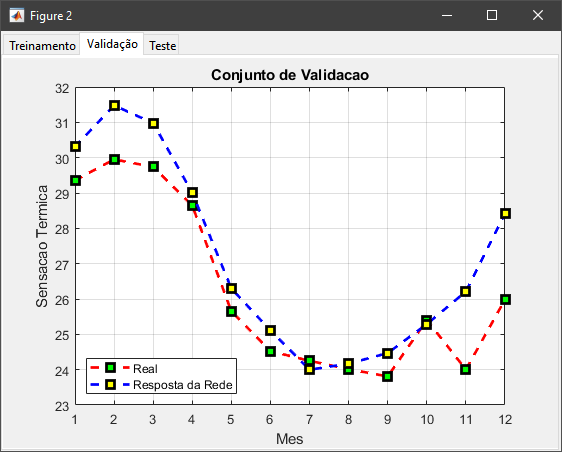
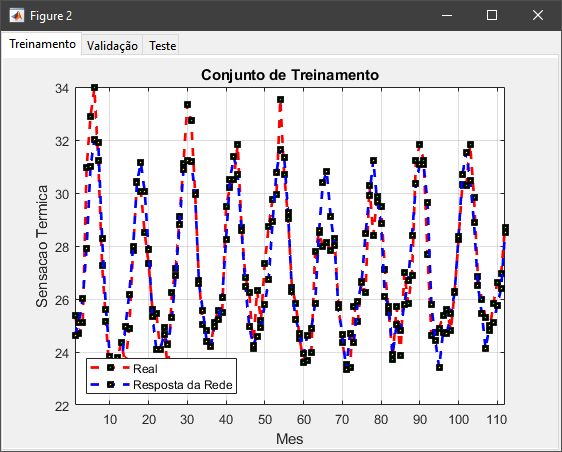
RMSE\_t\_f = 1.9824

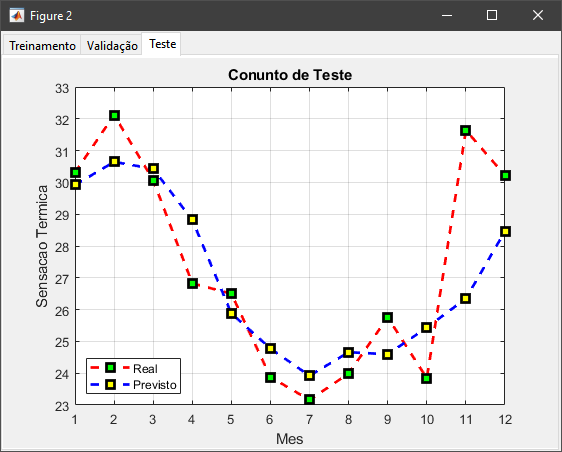
Para alcançar uma generalização maior, evitando pontos muito “fora da curva” desejada, diminuiu-se a quantidade de processadores utilizados para o teste atual. Além disso, a paciência do modelo foi configurada para 500 épocas. Novamente, para este teste e posteriores, a codificação com 12 bits foi selecionada.

Com as alterações feitas, o modelo obteve uma melhor generalização. As métricas foram, em gral, melhores que o experimento anterior. A menor quantidade de processadores foi, provavelmente, fundamental para que a generalização pudesse acontecer melhor.

**TESTE 7 - A**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 3 | Janela = | 8 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 13 | Nepoch = | 1000 | Numchkval = | 500 |
| Tipofunc = | logsig | Funcsaida = | purelin | Algtrei = | trainlm |





Métricas:

MAPE\_v\_f = 3.5131

RMSE\_v\_f = 1.1907

MAPE\_t\_f = 5.0301

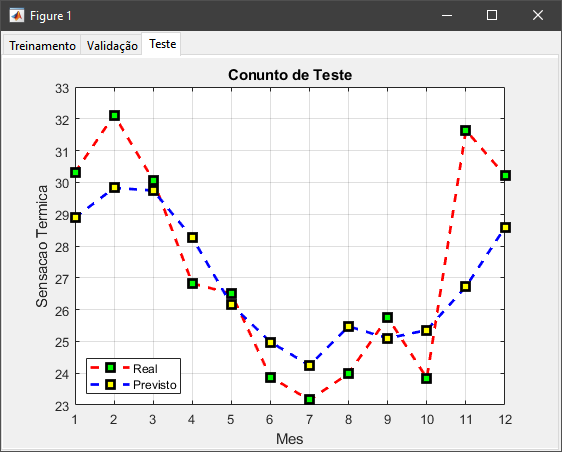
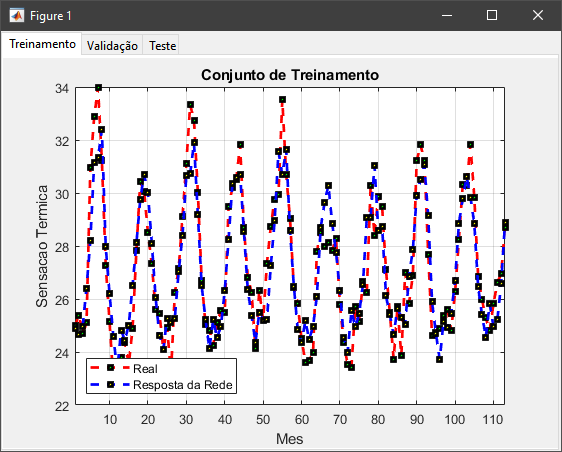
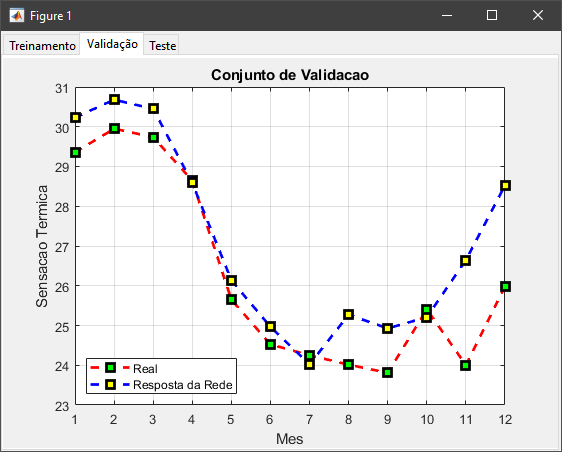
RMSE\_t\_f = 1.9054

Neste experimento, foi alterada apenas a função na camada escondida para a ‘logsig’ a fim de se observar as diferenças obtidas nos resultados.

Pelas métricas, observamos que o desempenho da rede foi ligeiramente pior na validação, porém obteve um valor RMSE melhor no teste.

**TESTE 7 - B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parâmetros gerais** | | | | | |
| Codificação = | 3 | Janela = | 7 |  |  |
| **Parâmetros da rede:** | | | | | |
| Numproc = | 10 | Nepoch = | 2000 | Numchkval = | 1000 |
| Tipofunc = | logsig | Funcsaida = | purelin | Algtrei = | trainlm |



Métricas:

MAPE\_v\_f = 3.6780

RMSE\_v\_f = 1.2449

MAPE\_t\_f = 5.4076

RMSE\_t\_f = 1.9085

Por fim, ainda utilizando a função ‘logsig’ na camada escondida, aumentou-se o número de épocas e a paciência do modelo para, respectivamente, 2000 e 1000. Além disso, o número de processadores utilizados foi de 10, e uma janela de 7 meses foi utilizada.

Pelas métricas fornecidas, vemos que o desempenho da rede piorou. Nesse caso é possível que mais neurônios sejam necessários para alcançar um resultado melhor.

# Conclusão

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resultados obtidos** | | | | | | |
| Testes: | **1 - A** | **1 - B** | **2 - A** | **2 - B** | **3 - A** | **3 - B** |
| MAPE\_v\_f: | 3.9873 | 9.2493 | 5.5550 | 4.8317 | 10.5882 | **3.1387** |
| RMSE\_v\_f: | 1.3316 | 2.6554 | 1.7706 | 1.4608 | 3.1422 | 1.1447 |
| MAPE\_t\_f: | 5.5701 | 7.6337 | 9.3043 | 5.4861 | 11.1668 | 5.8170 |
| RMSE\_t\_f: | 1.9826 | 2.4279 | 3.2363 | 2.1078 | 3.2564 | 2.0289 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Resultados obtidos** | | | | | | | |
| **4 - A** | **4 - B** | **5 - A** | **5 - B** | **6 - A** | **6 - B** | **7 - A** | **7 - B** |
| 3.0313 | 3.5680 | 3.7480 | 4.6061 | 5.8655 | 3.3892 | 3.5131 | 3.6780 |
| 1.0346 | 1.1876 | 1.2280 | 1.3538 | 1.7608 | **1.0845** | 1.1907 | 1.2449 |
| 5.2010 | 4.9448 | 5.3708 | 5.1155 | 5.2404 | **4.7067** | 5.0301 | 5.4076 |
| 2.0213 | 1.9164 | 1.8848 | 2.1437 | **1.8143** | 1.9824 | 1.9054 | 1.9085 |

Observando nas tabelas os resultados obtidos, pode-se notar que são, em muitos casos, bem semelhantes. Os testes 1 – B, 2 – A e 3 – A foram os que obtiveram resultados piores. Isso ocorreu principalmente pelo método de treinamento utilizado em 1 – B e 3 – A, e a função de saída ‘logsig’ em 2 – A.

Os melhores resultados obtidos foram, em geral, com codificações dos atributos de entrada com 12 bits, embora tenham sido resultados próximos à codificação com 4 bits.

Durante os testes, pudemos observar que a quantidade de processadores muito maior que a quantidade de entradas, ou ainda uma quantidade muito menor, faz com que, em geral, a rede tenha uma perda de desempenho, mas em alguns casos, o comportamento se mostra um pouco diferente (como por exemplo no teste 6 – A, que obteve bons resultados, mesmo utilizando uma quantidade de processadores maior que a de entradas).

O mesmo pode ser mencionado sobre as janelas. Uma janela muito grande, em geral, piora ligeiramente o resultado da rede, assim como uma janela pequena. O ideal é ajustar o tamanho da janela de acordo com a quantidade de dados que a base possui.

Com relação ao número de épocas utilizada, verificou-se que, em geral, mais épocas produzem um resultado melhor, contanto que seja utilizado o early stopping para evitar o overfitting da rede.