PROGRAMA DE PÓS-GRADUAÇÃO DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO
ASSOCIAÇÃO UFMA/UFPI
Disciplina: Redes Neurais (DCC40)
Professor: Areolino de Almeida Neto
Aluno: Raimundo Vale

Trabalho sobre Reconhecimento de Fonemas via RNA

Atividade do trabalho

Realizar o treinamento de uma RNA do tipo **MLP** para aprender a reconhecer as sílabas das palavras DIREITA e ESQUERDA. Foram utilizadas 119 amostras gravadas pelos alunos da disciplina. As amostras foram gravadas de diferentes formas, entonações, velocidades e pronúncias para propiciar mais diversidade. As amostras foram separadas em três partes: treino, validação e teste.

Proposta para a solução

Os arquivos foram gerados e agrupados conforme as sílabas das palavras DIREITA e ESQUERDA. Cada arquivo de sílaba foi lido e tratado via FFT obter a representação das senoides. Os dados obtidos após o tratamento com FFT resultava em arquivos muito grandes pelo que optei por agrupar as frequências em tamanhos iguais e iniciar o aprendizado. Ao longo do aprendizado vários cenários (conjunto das variáveis importantes do problema) foram testados para avaliar qual obtinha os melhores resultados. As variáveis que julguei importantes e que foram modificadas nos cenários foram: quantidade de amplitudes agrupadas para formar uma média, quantidade de médias fixa dos arquivos pós-tratramento com FFT, tamanho das bases de validação, treino e teste, taxa de aprendizagem, função de ativação, quantidade de épocas. Os cenários estão descritos detalhadamente no Anexo II deste documento. Segue abaixo a tabela comparativa de resultados.

Tabela comparativa de resultados – **Cenário 5** teve a melhor acurácia combinada na validação e no teste.

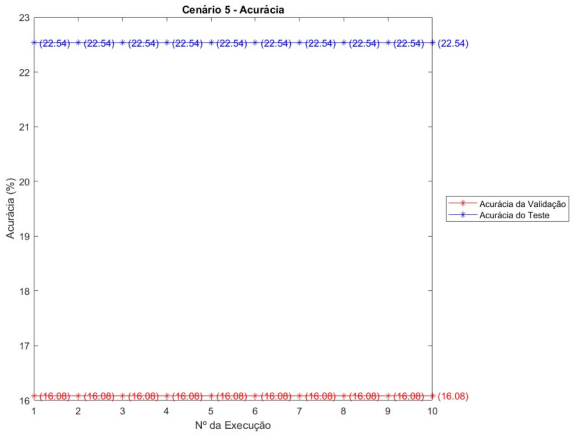
Cenários	Relação T/V/Te	Amplitudes por médias / Tamanho amostra	Acurácia Validação	Acurácia Teste	MSE Validação	MSE Teste	Acertos Validação	Erros Validação	Acertos Teste	Erros Teste
1	60/25/15	600/40	16,76	13,08	3,09	2,99	30	149	14	93
2	70/20/10	300/80	16,08	15,49	3,37	3,12	23	120	11	60
3	85/10/5	300/80	12,68	13,89	3,48	3,13	9	62	5	31
4	60/25/15	600/40	17,88	16,82	2,8	3	32	147	18	89
5	70/20/10	600/40	16,08	22,54	3,86	4	23	120	16	55
6	70/20/10	900/20	9,09	12,68	3,13	3,06	13	130	9	62
7	70/20/10	600/40	13,99	12,68	3,22	3,01	20	123	9	62
8	70/20/10	600/40	13,99	12,68	3,59	3,58	20	123	9	62
9	70/20/10	600/40	13,99	12,68	3,51	3,5	20	123	9	62
10	70/20/10	600/40	13,29	16,9	5,55	6,16	19	124	12	59

Detalhamento do Cenário 5

Cenário 5:
Quantidade de amplitudes agrupadas para formar uma média: 600
Quantidade de médias por arquivo: 40
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 40
Neurônios da camada escondida: 80
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.05
Função de ativação: Tangente hiperbólica
Valor de k = 1 (função linear)

Gráficos Resultantes

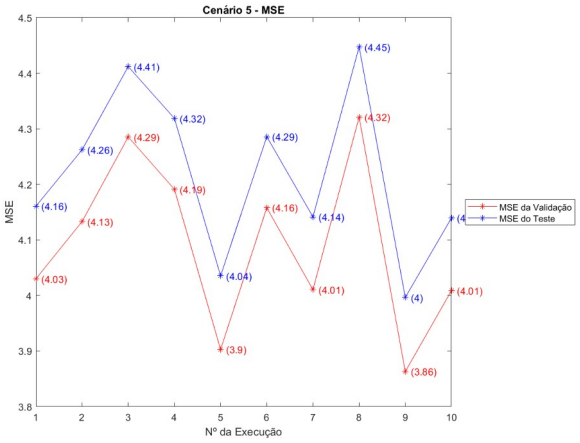
Acurácia



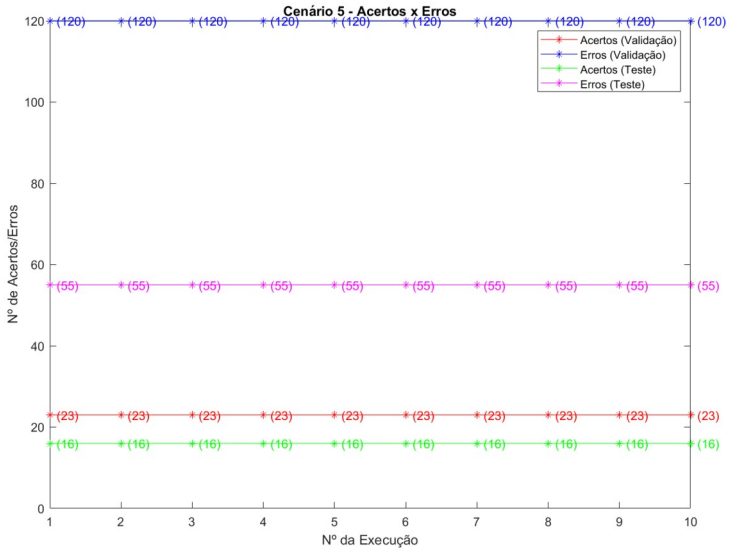
Este cenário proveu a melhor acurácia na fase de teste 22,54% que é um valor baixo para uma aplicação para uso, porém dada pequena base de treinamento, validação e teste foi o melhor resultado dentre todos as tentativas feitas. O resultado da validação foi quarto melhor dentre todos os cenários. E os resultados combinado de ambas as fases foi o melhor de todos pelo que concluí sendo este o melhor cenário do aprendizado.

MSE

Apesar de ter havido outros cenários com MSEs menores, no Cenário 5 os menores MSEs foram 3,8 e 4, havendo ainda picos de 4,45 e 4,32. Mesmo com MSE tão alto a acurácia do cenário foi determinante para a escolha. Os MSEs altos necessariamente não implicaram baixa acurácia ou poucos de acertos.



Acertos vs Erros



Apesar de resultados bem lineares e não tendo sido os melhores em valores absolutos – O Cenário 1 foi o melhor devido a base de validação e teste maior. O Cenário 5 em termos percentuais teve melhor retorno.

Conclusão:

1. Em não havendo uma base de comparação entre os resultados que obtive, conclui que os resultados foram baixos, porém dada o tamanho da base que consegui talvez os resultados estejam dentro do razoável.
2. O tamanho das bases foram determinantes para a variação dos valores. A maior base de treinamento (85%) não proporcionou os melhores resultados. A menor base de treinamento (60%) por sua vez proporcionou até bons resultados comparativamente aos demais cenários.
3. Uma opção que poderia ser testada porém pela exiguidade do tempo após conseguir um algorimo estável seria avaliar agrupar uma quantidade menor de amplitude por média (150) e comparar resultados já que optei por agrupo a partir de 300 amplitudes.

Anexo I - Detalhamento dos algoritmos

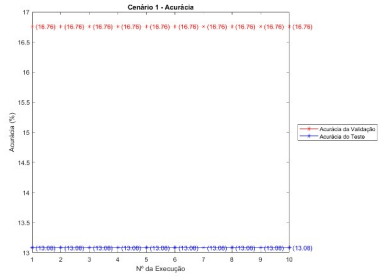
Para atacar problema utilizei tempo e esforço excessivo, com pouco retorno para desenvolver um algoritmo próprio funcional, em não conseguindo optei por utilizar os algoritmos de MLP e RBF disponibilizados pelos alunos da disciplina. Depois de avaliar bastante o resultado e andamento da atividade, escolhi utilizar o MLP para a RNA proposta. Somente na segunda tentativa de modificação do algoritmo é que consegui executar a RNA. A solução consiste em:

Módulo	O que faz	Saídas
preProcessamento_Fonemas_v0.m	1. Lê os arquivos *.wav dentro da pasta fonemas; 2. Define as variáveis: quantidade de amplitudes agrupadas por média (indiceCompac); quantidade fixa de médias de amplitudes por amostra (tamSaida); 3. Aplica FFT nas amostras; 4. Divide as bases para serem processadas.	arquivo: preProc.mat
training_Fonemas_v0.m	1. Define variáveis: eta, função de ativação, épocas, quantidade fixa de médias de amplitudes por amostra (tamSaida); 2. Chama em um laço na quantidade dos treinamentos as funções: rnafit (treinamento da rede); rnapredict (validação) e para teste. 3. Gera os gráficos resultantes	arquivo: MelhorRede.mat
rnafit.m	Na quantidade épocas definidas aplica o cálculo dos pesos, dos erros, utiliza a função de ativação e a taxa de aprendizagem, faz propagação dos pesos nas camadas sobre a base de treinamento e acumula o erro quadrático.	Obs.: desenvolvido por Alisson Mendonça.
rnapredict.m	Na base de validação e de teste compara os resultados de saída com os de referência e assim acumula os resultados de acurácia, acertos e erros.	Obs.: desenvolvido por Alisson Mendonça.
grafico.m	Gera os gráficos dos dados acumulados nos processamentos anteriores.	Obs.: desenvolvido por Alisson Mendonça.

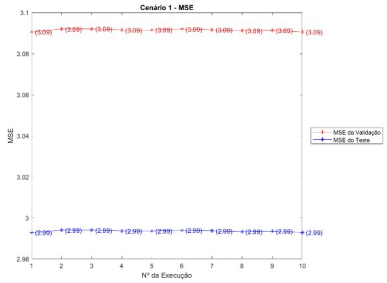
Anexo II – Descrição dos cenários

Cenário 1:
Quantidade de amplitudes agrupadas para formar uma média: 300
Quantidade de médias por arquivo: 80
Tamanho dos grupos de treino, validação e teste: 60%-25%-15%
Tamanho da base de treino: 428
Neurônios da camada de entrada: 80
Neurônios da camada escondida: 160
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.07
Função de ativação: Sigmoide
Valor de k = 1 (função linear)
Gráficos Resultantes

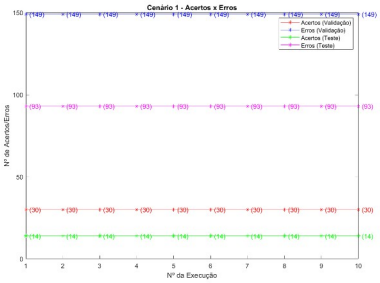
Acurácia



MSE



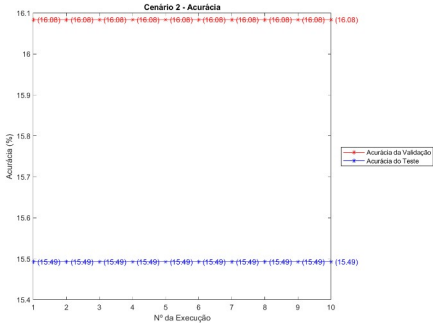
Acertos vs Erros



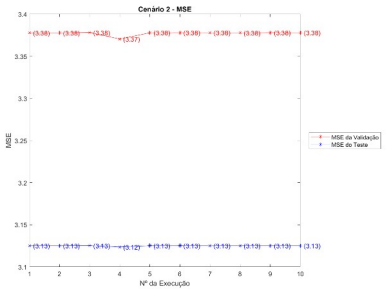
Conclusão do cenário
Apesar de o MSE ter se estabilizado ao longo dos treinamentos a números iguais e a base de treino menor - somente 428 exemplos – a quantidade de erros se manteve alta.

Cenário 2: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 300
Quantidade de médias por arquivo: 80
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 80
Neurônios da camada escondida: 160
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.06
Função de ativação: Sigmoide
Valor de k = 1 (função linear)
Gráficos Resultantes

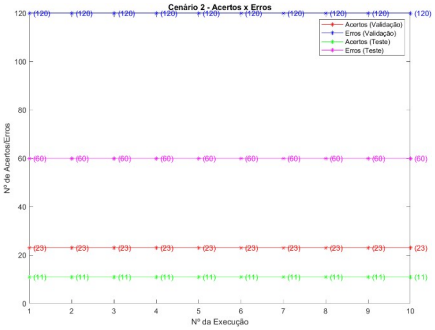
Acurácia



MSE



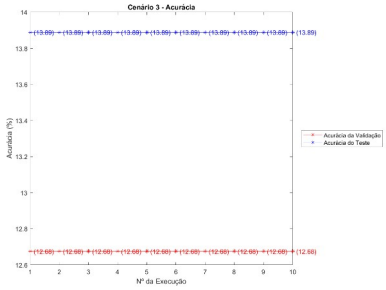
Acertos vs Erros



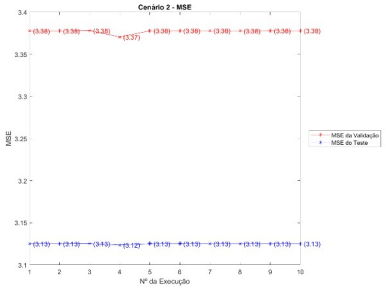
Conclusão do cenário
Em relação ao cenário 1 – a base de treino aumentou e ocorreu uma melhoria da acurácia (de 13,08 para 15,49). O MSE foi aumentou tanto na validação e teste. Já a quantidade de acertos melhorou em relação ao cenário 1.

Cenário 3: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 300
Quantidade de médias por arquivo: 80
Tamanho dos grupos de treino, validação e teste: 85%-10%-5%
Tamanho da base de treino: 607
Neurônios da camada de entrada: 80
Neurônios da camada escondida: 160
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.06
Função de ativação: Sigmoide
Valor de k = 1 (função linear)
Gráficos Resultantes

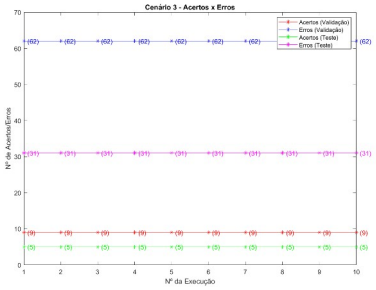
Acurácia



MSE



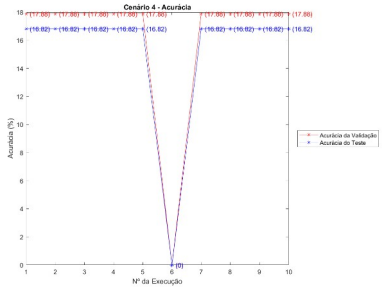
Acertos vs Erros



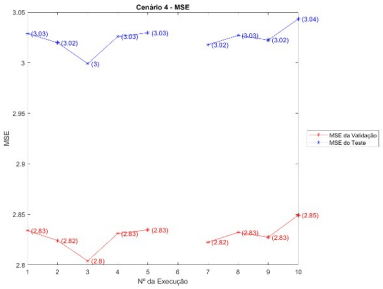
Conclusão do cenário
Em relação ao cenário 2 – as bases de validação e treino diminuíram e tiveram pior desempenho. Já o MSE da validação aumentou. Por fim a relação de acertos e erros piorou. Entendemos pelos resultados obtidos que a melhor amostra de treino, validação e teste está entre a relação 60-25-15 e 70-20-10.

Cenário 4: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 600
Quantidade de médias por arquivo: 40
Tamanho dos grupos de treino, validação e teste: 60%-25%-15%
Tamanho da base de treino: 428
Neurônios da camada de entrada: 40
Neurônios da camada escondida: 80
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.05
Função de ativação: Tangente hiperbólica
Valor de k = 1 (função linear)
Gráficos Resultantes

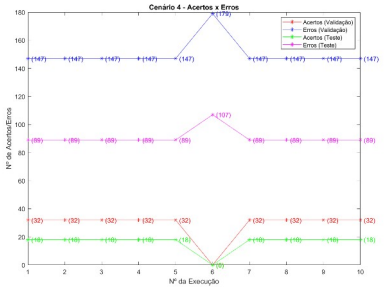
Acurácia



MSE



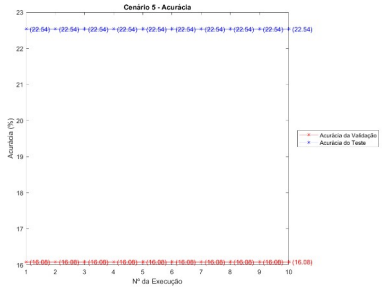
Acertos vs Erros



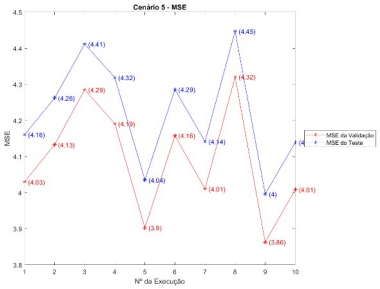
Conclusão do cenário
Neste cenário no treinamento 6 apresentou ocorrência de NaN acarretando resultados impróprios para avaliação e por isso entendemos que este não é o melhor cenário para continuação do trabalho.

Cenário 5: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 600
Quantidade de médias por arquivo: 40
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 40
Neurônios da camada escondida: 80
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.05
Função de ativação: Tangente hiperbólica
Valor de k = 1 (função linear)
Gráficos Resultantes

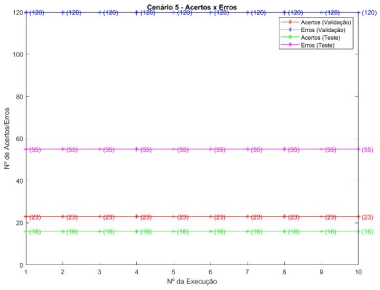
Acurácia



MSE



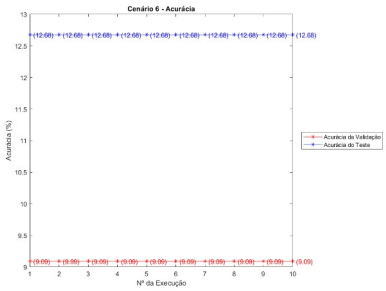
Acertos vs Erros



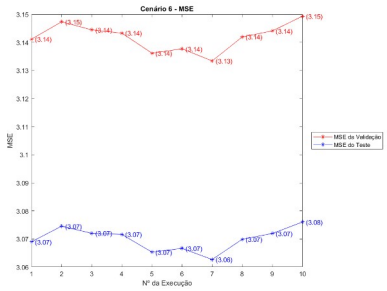
Conclusão do cenário
Os resultados deste cenário apresentaram os melhores índices pelo que adotarei a relação de treinamento de 70%-20%-10% para as bases de treinamento, validação e teste. Nos cenários seguintes variarei outros elementos como as médias de amplitude. Sendo até o momento o melhor cenário.

Cenário 6: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 900
Quantidade de médias por arquivo: 20
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 20
Neurônios da camada escondida: 40
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.05
Função de ativação: Tangente hiperbólica
Valor de k = 1 (função linear)
Gráficos Resultantes

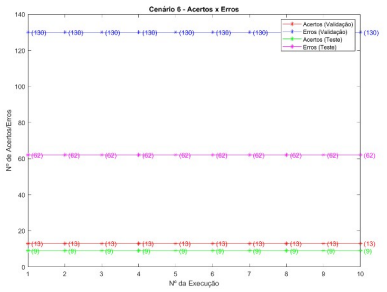
Acurácia



MSE



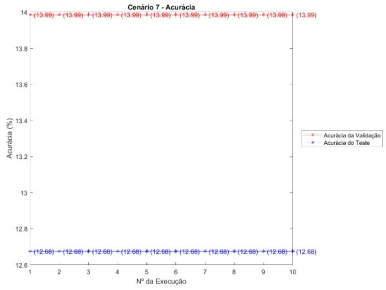
Acertos vs Erros



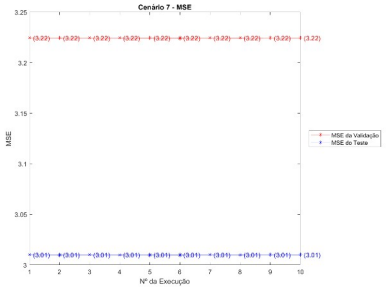
Conclusão do cenário
O cenário 6 melhorou em relação ao cenário 5 somente no MSE menor tanto na validação quanto no teste, porém obteve resultados piores na Acurácia e na relação de Acertos vs Erros.

Cenário 7: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 600
Quantidade de médias por arquivo: 40
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 40
Neurônios da camada escondida: 80
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.06
Função de ativação: Sigmoide
Valor de k = 1 (função linear)
Gráficos Resultantes

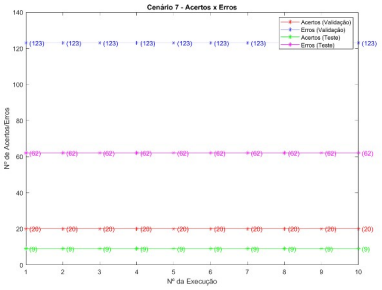
Acurácia



MSE



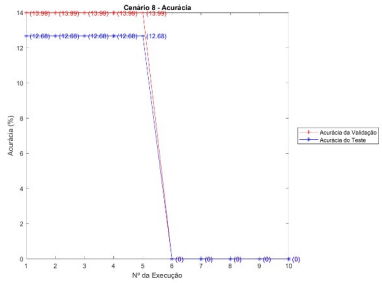
Acertos vs Erros



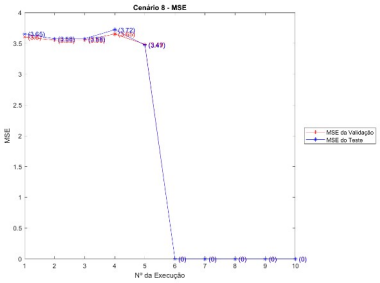
Conclusão do cenário
Os resultados deste cenário apresentaram os melhores índices pelo que adotarei a relação de treinamento de 70%-20%-10% para as bases de treinamento, validação e teste. Em relação ao cenário 5 que obteve os melhores resultados até agora variei a taxa de aprendizagem e a função de ativação. Ainda assim o Cenário 7 não melhorou ao ponto de superar o Cenário 5.

Cenário 8: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 600
Quantidade de médias por arquivo: 40
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 40
Neurônios da camada escondida: 80
Neurônios da camada de saída: 1
Total de épocas por treinamento: 50 por treinamento
Número de treinamento: 5
Valor do eta: 0.06
Função de ativação: Sigmoide
Valor de k = 1 (função linear)
Gráficos Resultantes

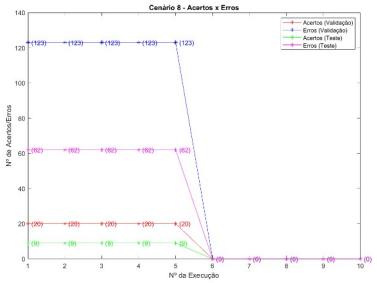
Acurácia



MSE



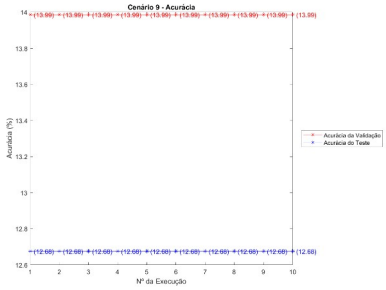
Acertos vs Erros



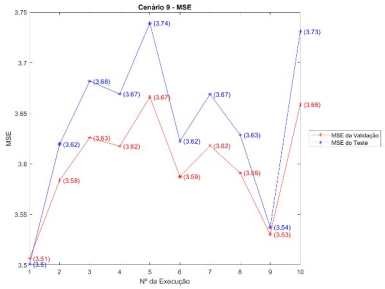
Conclusão do cenário
O Cenário 8 aumentou em 10 vezes a quantidade de épocas (mesmo diminuindo o número de treinamento) para avaliar se o MSE seria menos. Diminui pouco em relação ao Cenário 5 que continuou com melhor acurácia entre os cenários testados.

Cenário 9: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 600
Quantidade de médias por arquivo: 40
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 40
Neurônios da camada escondida: 80
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.05 → na metade da base o eta recebe o valor 0.09.
Função de ativação: Tangente hiperbólica
Valor de k = 1 (função linear)
Gráficos Resultantes

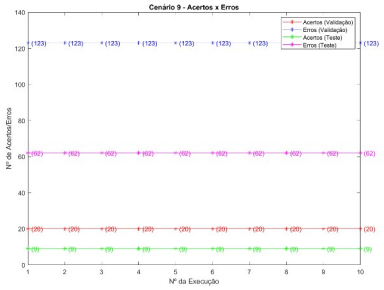
Acurácia



MSE



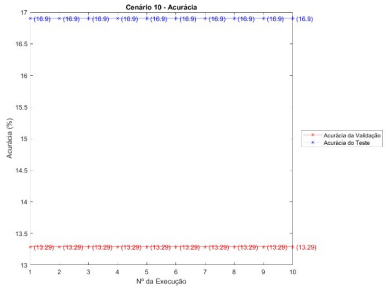
Acertos vs Erros



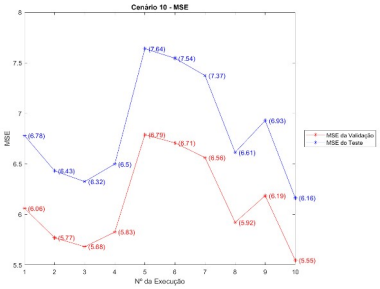
Conclusão do cenário
O Cenário 9 variou (de 0.05 para 0.09) a taxa de aprendizagem (eta) uma durante cada época. Houve uma melhora pequena no MSE, porém o Cenário 5 continuou com melhor acurácia entre os cenários testados.

Cenário 10: Variáveis que mudaram em relação ao cenário anterior estão destacadas em vermelho
Quantidade de amplitudes agrupadas para formar uma média: 600
Quantidade de médias por arquivo: 40
Tamanho dos grupos de treino, validação e teste: 70%-20%-10%
Tamanho da base de treino: 500
Neurônios da camada de entrada: 40
Neurônios da camada escondida: 80
Neurônios da camada de saída: 1
Total de épocas por treinamento: 5 por treinamento
Número de treinamento: 10
Valor do eta: 0.05 → a cada terço da base o eta recebe o valor 0.09.
Função de ativação: Tangente hiperbólica
Valor de k = 1 (função linear)
Gráficos Resultantes

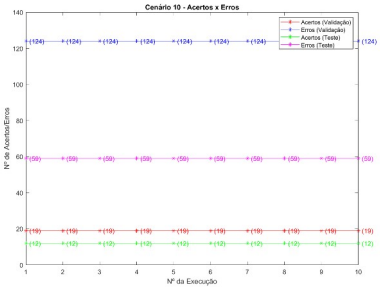
Acurácia



MSE



Acertos vs Erros



Conclusão do cenário
O Cenário 10 variou (de 0.05 para 0.09) a taxa de aprendizagem (eta) por duas vezes em cada época. Houve uma melhora pequena no MSE aumentou bastante e ocorreu uma melhora na acurácia mas não superou o Cenário 5.