



UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
CURSO DE GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

Luis Felipe Ferreira Soares

**ANOVA como Critério de Parada no Treinamento de Redes Neurais
Artificiais**

São Luís – MA

2024

Luis Felipe Ferreira Soares

**ANOVA como Critério de Parada no Treinamento de Redes Neurais
Artificiais**

Monografia apresentada ao curso de Engenharia da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para a obtenção do grau de Bacharel em Engenharia da Computação.

Graduação em Engenharia da Computação
Universidade Federal do Maranhão

Orientador: Prof. Dr. Paulo Rogerio de Almeida Ribeiro

São Luís – MA

2024

Luis Felipe Ferreira Soares

**ANOVA como Critério de Parada no Treinamento de Redes Neurais
Artificiais**

Monografia apresentada ao curso de Engenharia da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para a obtenção do grau de Bacharel em Engenharia da Computação.

Trabalho aprovado em: São Luís - MA, ____ de _____ de 2024

Prof. Dr. Paulo Rogerio de Almeida Ribeiro

Orientador

Universidade Federal do Maranhão

Prof. Dr. Alex Oliveira Barradas Filho

Examinador

Universidade Federal do Maranhão

Prof. Dr. Bruno Feres de Souza

Examinador

Universidade Federal do Maranhão

São Luís – MA

2024

RESUMO

Este trabalho propõe e avalia o uso da análise de variância com medidas repetidas (ANOVA) como critério de parada no treinamento de redes neurais artificiais (RNA). O objetivo é otimizar o processo de treinamento, reduzindo o tempo de treinamento e possivelmente prevenindo o *overfitting*. Este método aplica o ANOVA ao erro quadrático médio (EQM) durante o treinamento para determinar estatisticamente o momento de que o treinamento deve cessar. Dois problemas são estudados: regressão de polinômios quadráticos com ruído gaussiano e classificação do conjunto de dados Íris. O método proposto é comparado com o treinamento tradicional em termos de tempo e qualidade do modelo. Os resultados mostram uma redução moderada no tempo de treinamento para problemas de regressão e uma redução de baixa à moderada no tempo de treinamento para classificação da Íris, mantendo os desempenhos semelhantes na validação. Conclui-se que o ANOVA com medidas repetidas é uma possível ferramenta para otimizar o treinamento de RNAs, a qual fornece critérios de parada estatisticamente baseados e contribui para a melhoria da eficiência das técnicas de treinamento em aprendizado de máquina.

Palavras-chave: Rede Neural Artificial, Análise de Variância, Critério de Parada, Overfitting, Aprendizado de Máquina, Inteligência Artificial

ABSTRACT

This work proposes and evaluates the use of repeated measures analysis of variance (ANOVA) as a stopping criterion in the training of artificial neural networks (ANN). The objective is to optimize the training process, reducing training time and potentially preventing overfitting. This method applies ANOVA to the mean squared error (MSE) during training to statistically determine the moment when training should cease. Two problems are studied: regression of quadratic polynomials with Gaussian noise and classification of the Iris dataset. The proposed method is compared with traditional training in terms of time and model quality. The results show a moderate reduction in training time for regression problems and a low to moderate reduction in training time for Iris classification, while maintaining similar performance. It is concluded that repeated measures ANOVA is a potential tool for optimizing ANN training, providing statistically based stopping criteria and contributing to the improvement of training efficiency in machine learning techniques.

Keywords: Artificial Neural Network, Analysis of Variance, Stopping Criterion, Overfitting, Machine Learning, Artificial Intelligence.

LISTA DE FIGURAS

Figura 1 - Valor-p ao longo das épocas de treinamento para o polinômio	19
Figura 2 - Evolução do MSE no problema do polinômio.....	19
Figura 3 - Valor-p ao longo das épocas de treinamento para as Íris	23
Figura 4 - Evolução do EQM no problema das Íris.....	24

LISTA DE TABELAS

Tabela 1 - Configuração dos dados a serem passados para o ANOVA	14
Tabela 2 - Dados comparativos para o polinômio	18
Tabela 3 - Progressão do EQM para o polinômio	20
Tabela 4 - Dados comparativos para as Íris	21
Tabela 5 – Progressão do EQM para as Íris.....	24

LISTA DE ABREVIATURAS E SIGLAS

ANOVA	<i>Analysis of Variance</i>
EQM	<i>Erro Quadrático Médio</i>
RNA	<i>Rede Neural Artificial</i>
ReLU	<i>Rectified Linear Units</i>

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 Objetivos.....	11
1.1.1 Objetivos Específicos	11
2 METODOLOGIA.....	12
2.1 Descrição dos Dados	12
2.2 Modelo de Treinamento.....	12
2.3 Aplicação do ANOVA como Critério de Parada o Treinamento	13
2.4 Validação do Método	15
2.5 Considerações sobre a Aplicação do ANOVA com Medidas Repetidas	15
2.6 Ferramentas Utilizadas e Reprodutibilidade.....	16
3 RESULTADOS.....	17
3.1 Polinômio $x^2 + 2x + 1$ com Ruído Gaussiano.....	17
3.2 Conjunto de Dados Íris.....	21
CONCLUSÃO.....	25
REFERÊNCIAS	27

1 INTRODUÇÃO

O treinamento de redes neurais artificiais (RNA) é um processo fundamental no campo do aprendizado de máquina, contudo muitas vezes enfrenta desafios relacionados ao *overfitting*, onde o modelo se sobreajusta aos dados de treinamento, comprometendo assim sua capacidade de generalização (Goodfellow et al., 2016). Para minimizar esse problema, diversas técnicas foram desenvolvidas e aplicadas ao longo dos anos.

Os métodos mais comuns para evitar *overfitting* são a regularização L1 e L2, que adicionam penalidades à função de perda para controlar a complexidade do modelo (Ng, 2004). Outra técnica muito utilizada é o *dropout*, que desativa neurônios aleatoriamente durante o treinamento, reduzindo a dependência de recursos específicos (Srivastava et al., 2014). A validação cruzada também desempenha um papel importante, permitindo uma avaliação mais robusta do desempenho do modelo em dados não vistos (Kohavi, 1995).

A otimização de hiperparâmetros é outro aspecto importante do treinamento de redes neurais artificiais. Métodos como pesquisa em grade, pesquisa aleatória e otimização bayesiana têm sido utilizados para encontrar configurações ideais de hiperparâmetros, incluindo taxa de aprendizagem, tamanho do lote e arquitetura de rede (Bergstra e Bengio, 2012). No entanto, estes métodos são computacionalmente caros e não abordam diretamente o problema de determinar o número ideal de épocas de treinamento.

As técnicas de *early stopping* surgiram como uma solução para interromper o treinamento antes que ocorra o *overfitting*. Prechelt (1998) propôs vários critérios de parada baseados em erros de validação, enquanto Sjöberg e Ljung (1995) introduziram métodos que consideram a estabilidade dos pesos da rede. Embora eficazes, estes métodos nem sempre capturam a complexidade estatística da evolução dos erros durante o treinamento.

Determinar o número ideal de épocas de treinamento continua sendo um ponto chave na modelagem de redes neurais artificiais, pois está intrinsecamente ligado à prevenção do *overfitting* e à otimização do desempenho do modelo. Os métodos tradicionais de *early stopping* frequentemente dependem de heurísticas ou critérios arbitrários, que podem não ser adequados para todos os tipos de problemas.

Neste caso, propõe-se a aplicação da análise de variância de medidas repetidas (ANOVA) como critério de base estatística para determinar o momento ideal para interromper o treinamento. O ANOVA com medidas repetidas é uma técnica estatística

capaz de analisar mudanças nas medidas ao longo do tempo ou em diferentes condições (Field, 2013). Ao aplicar esta técnica ao erro quadrático médio (MSE) durante o treinamento, buscamos desenvolver um método para otimizar de forma genérica a escolha do número de épocas adequado para cada problema.

A motivação para a escolha do ANOVA com medidas repetidas é devido a sua capacidade de detectar alterações significativas na variância do erro ao longo do tempo, potencialmente identificando o ponto em que o modelo começa a sobreajustar ou o momento em que as variações se tornam estatisticamente insignificantes. Essa abordagem não apenas otimiza o tempo de treinamento, mas também aprimora a capacidade de generalização do modelo.

Deste modo, este trabalho tem como objetivo investigar a eficácia do ANOVA com medidas repetidas como critério de parada para treinamento de redes neurais artificiais e comparar o método desenvolvido com o treinamento convencional em problemas de regressão e classificação. Espera-se que esta abordagem contribua para o desenvolvimento de técnicas de treinamento de redes neurais artificiais mais eficientes, com implicações em diversos campos do aprendizado de máquina e da inteligência artificial.

1.1 Objetivos

Avaliar a eficácia da utilização da análise de variância com medidas repetidas (ANOVA) como critério de parada no treinamento de redes neurais artificiais (RNA), visando otimizar o processo de treinamento e possivelmente atenuar problemas de *overfitting*.

1.1.1 Objetivos Específicos

- Implementar um método de parada de treinamento de uma rede neural artificial baseado no ANOVA com medidas repetidas como um fator condicional.
- Comparar o desempenho do método proposto (condicional) com o treinamento convencional em termos de: a) Tempo de treinamento (número de épocas). b) Qualidade do modelo final (erro quadrático médio).
- Avaliar a capacidade do método proposto em prevenir o *overfitting* e comparar o erro de validação entre o método condicional e o método tradicional.
- Identificar potenciais limitações e desafios na aplicação da abordagem proposta em diferentes cenários do treinamento de redes neurais artificiais.

2 METODOLOGIA

Este trabalho realizou experimentos computacionais com a finalidade de avaliar a eficácia da utilização da Análise de Variância (ANOVA) com medidas repetidas como condição de parada do treinamento de uma Rede Neural Artificial (RNA), utilizando o erro quadrático médio como métrica a ser passada para o ANOVA.

2.1 Descrição dos Dados

Para tanto, foram escolhidos dois problemas como alvo deste trabalho: o treinamento de um polinômio de 2º grau ($x^2 + 2x + 1$) e o conjunto de dados das espécies de flores do gênero *Íris*, comumente utilizado em aprendizado de máquina. Ambos foram escolhidos por sua simplicidade.

Inicialmente, os pares de valores (x, y) do polinômio foram calculados avaliando a função $f(x) = x^2 + 2x + 1$ em um intervalo fixo e simétrico de $[-5, 5]$, com um incremento de 0,001 unidades entre cada ponto do eixo x , configurações que produziram 10.000 pares de valores (x, y) . Estes pontos foram então somados a um ruído gaussiano para simular uma amostragem real dos valores. Os dados gerados foram utilizados de forma bruta, sem nenhum pré-processamento.

O conjunto de dados das espécies de flores do gênero *Íris* foi obtido da base de dados *UCI Machine Learning Repository*, que é uma versão ligeiramente diferente da apresentada por Fisher (1936). Este conjunto de dados apresenta 150 instâncias, que possuem como vetor de características o comprimento da sépala, a largura da sépala, comprimento da pétala e largura da pétala, e como rótulo uma das possíveis espécies (*Iris setosa*, *Iris versicolor* ou *Iris virginica*). Estes dados também foram utilizados sem nenhum pré-processamento.

Complementarmente, cada conjunto de dados foi dividido em conjunto de treinamento e conjunto de validação, em uma razão de 80% e 20% respectivamente, a fim de avaliar a ocorrência de *overfitting* em ambos os problemas.

2.2 Modelo de Treinamento

Para o treinamento do polinômio, foi utilizado um modelo sequencial, composto de uma camada de entrada unidimensional, duas camadas densas ocultas com 10 neurônios e função de ativação *ReLU*, e uma camada de saída unidimensional. Estas configurações são adequadas para previsão de valores contínuos, isto é, regressões lineares. Durante o treinamento, o modelo utilizou o Erro Quadrático Médio (EQM) como

função de perda e o otimizador Adam, um método de descida de gradiente, trabalhou em direção à minimização do erro entre as previsões de treinamento e os valores reais. Para o propósito deste trabalho, o único hiperparâmetro de interesse é o número de épocas, que para este problema foi estabelecido em 150 para o modelo que não utilizou a condição de parada (modelo convencional), enquanto outros hiperparâmetros, como taxa de aprendizagem, permaneceram com os padrões atribuídos pela biblioteca utilizada, no caso o *TensorFlow*.

Para o treinamento do conjunto de dados Iris, utilizou-se um modelo sequencial com uma camada de entrada de dimensão 4, duas camadas densas ocultas com 10 neurônios e função de ativação *ReLU*, e uma camada de saída com 3 neurônios e função de ativação *softmax*. O otimizador também foi o Adam, mas utilizou a função de perda categórica cruzada (*categorical crossentropy*) na minimização do erro de treinamento, que é mais adequada para problemas de classificação. Ainda assim, o erro quadrático médio foi uma métrica monitorada nesse modelo durante sua execução e foi a métrica utilizada no ANOVA com medidas repetidas para condicionar a parada de execução, assim como no problema anterior. Novamente, o único hiperparâmetro de interesse é o número de épocas, que foi estabelecido em 500 épocas para este problema no modelo convencional.

2.3 Aplicação do ANOVA como Critério de Parada o Treinamento

A aplicação do ANOVA com medidas repetidas entrou como uma etapa adicional no processo de treinamento. Para tanto, primeiro reestruturou-se o agrupamento dos erros quadráticos médios coletados para satisfazer as condições de aplicação do ANOVA com medidas repetidas. Definiu-se o conceito de "janela de épocas", que consiste em um agrupamento de erros quadráticos médios ao longo de um número arbitrário de épocas e atua como um sujeito, bem como cada época dentro de uma janela atua como um diferente nível a ser avaliado. Esta convenção permite que os dados sejam estruturados de acordo com o formato requerido pelas bibliotecas utilizadas e pelas suposições do ANOVA com medidas repetidas. No caso, foram escolhidos o valor de cinco épocas para constituir uma janela e cinco janelas foram submetidas a uma etapa de avaliação pelo ANOVA com medidas repetidas, constituindo uma tabela de formato 5x5, assim como na Tabela 1, onde cada célula representa o número da época de onde o erro quadrático médio foi coletado.

Tabela 1 - Configuração dos dados a serem passados para o ANOVA

Janela	1ª Época	2ª Época	3ª Época	4ª Época	5ª Época
1	1	2	3	4	5
2	6	7	8	9	10
3	11	12	13	14	15
4	16	17	18	19	20
5	21	22	23	24	25

Inicialmente, os erros quadráticos médios foram coletados até preencher completamente todas as cinco janelas de épocas. Esses dados foram então submetidos ao ANOVA com medidas repetidas, o qual analisou como o histórico de erros quadráticos médios coletados variou entre as diferentes épocas (da 1ª à 5ª) para cada janela. Esta análise produziu diversos resultados estatísticos, porém apenas avaliou-se o valor-p. A hipótese nula constituída foi H_0 : "Não há diferença significativa entre as médias das janelas" e um nível de significância $\alpha = 0,05$ foi estabelecido.

A condição de parada baseia-se no fato de que um valor-p $\leq \alpha$ indica que se deve rejeitar a hipótese nula, isto é, a média de pelo menos uma janela é significativamente diferente das outras e, portanto, deve-se continuar o treinamento. Caso o valor-p $> \alpha$, conclui-se que não há evidências suficientes para rejeitar a hipótese nula e o treinamento deve parar. Se a análise indicar que o treinamento deve continuar, descarta-se a janela mais antiga e prossegue-se com a coleta de uma nova janela. O processo de avaliar o valor-p para cada novo conjunto de janelas de épocas continuará até que a condição de parada ocorra ou o limite máximo definido de épocas seja atingido (150 para o polinômio e 500 para as Íris).

Devido às características inerentes observadas durante iterações iniciais do problema do polinômio, onde houve valores altos para o erro quadrático médio, optou-se por realizar uma segunda verificação para a condição de parada, após a avaliação do ANOVA com medidas repetidas: a pontuação padrão ou *Z-Score*, utilizada para identificar a presença de outliers no conjunto de dados analisado. Definiu-se que a presença de qualquer conjunto de janelas de épocas que possuíssem valores com pontuação padrão acima de 3 unidades seria recusada como válida na parada do treinamento, prosseguindo normalmente para a coleta da próxima janela. O problema da Íris não passou por essa segunda condição, pois não se observou a anomalia mencionada para o problema do polinômio.

2.4 Validação do Método

Para efeito de comparação e controle, foram treinados ambos os modelos sem aplicação do ANOVA com medidas repetidas como condição de parada, definido como "método convencional", e com a sua aplicação, definido como "método condicional". Ambos os modelos foram treinados com o mesmo conjunto de dados de treinamento e validação. A fim de investigar o impacto da condição de parada na prevenção de *overfitting*, após a avaliação do conjunto de validação, os erros quadráticos médios de cada modelo foram comparados por meio do teste t de *Student*, onde a hipótese nula é H_0 : "*Não há diferença significativa entre os erros quadráticos médios*" e o nível de significância foi definido como $\alpha = 0,05$. Esta comparação foi utilizada para identificar se a diferença entre os erros quadráticos médios dos métodos convencional e condicional foi de fato significativa, isto é, se os achados indicam que um modelo foi superior ao outro na ocasião ou se as diferenças estavam estatisticamente dentro do esperado.

Ambos os métodos, convencional e condicional, foram executados trinta vezes utilizando as diretrizes descritas anteriormente, a fim de determinar o impacto da estocasticidade inerente à aprendizagem de máquina, também permitindo avaliar a frequência com que certos resultados foram produzidos. O erro quadrático médio dos métodos convencional e condicional, o valor-p do teste t de *Student* realizado entre estes e a época de parada de todas as execuções foram salvos em disco para posterior análise de resultados.

2.5 Considerações sobre a Aplicação do ANOVA com Medidas Repetidas

O ANOVA com medidas repetidas foi utilizada como teste estatístico na condição de parada no treinamento dos modelos dos problemas selecionados devido à sua capacidade de detectar diferenças significativas entre as médias de múltiplas condições, no caso, das janelas de épocas. Assim, foi aplicada para monitorar a evolução do erro quadrático médio durante o treinamento dos conjuntos de dados de cada problema.

Contudo, é importante ressaltar que existem certas limitações quanto à forma de aplicação do ANOVA com medidas repetidas na condição de parada do treinamento das redes neurais artificiais.

Primeiramente, deve-se destacar alguns aspectos sobre a utilização da hipótese nula H_0 : "*Não há diferença significativa entre as médias das janelas*". Convencionou-se que um valor-p maior que o nível de significância indica que o treinamento deve parar; porém, a aceitação da hipótese nula não necessariamente indica que as médias entre as

janelas não possuem diferenças significativas, apenas que não há evidências suficientes para rejeitá-la. Quando isso ocorre, uma possível ação é aumentar o tamanho da amostra ou, no caso, o número de janelas. Contudo, este trabalho utilizou um conjunto de janelas fixas. Ainda a respeito da aceitação da hipótese, deve-se também considerar a ocorrência do erro do tipo II; porém, a execução de múltiplos treinamentos, no caso trinta, tem justamente o intuito de avaliar e minimizar o efeito de sua ocorrência.

Em segundo lugar, deve-se mencionar que nem todo conjunto de treinamento necessariamente obedecerá às suposições do ANOVA com medidas repetidas, sendo elas: normalidade, esfericidade e independência. Na execução do ANOVA com medidas repetidas pelas bibliotecas utilizadas, essas suposições são verificadas e, quando violadas, encerram a execução do algoritmo.

2.6 Ferramentas Utilizadas e Reprodutibilidade

Todos os algoritmos de treinamento foram implementados na linguagem de programação Python, com o auxílio das bibliotecas *TensorFlow* e *scikit-learn*, destinadas a aplicações de aprendizagem de máquina e inteligência artificial. Também foram utilizadas diversas bibliotecas de computação numérica e estatística, como *NumPy*, *SciPy* e *Pandas*, para realizar os testes estatísticos, como a Análise de Variância (ANOVA) com medidas repetidas. Os gráficos foram produzidos com o auxílio da biblioteca *Matplotlib*.

3 RESULTADOS

A seguir, estão descritos individualmente os resultados para cada um dos problemas selecionados.

3.1 Polinômio $x^2 + 2x + 1$ com Ruído Gaussiano

Inicialmente, foi treinada a rede neural artificial (RNA) para a aprendizagem dos dados do polinômio $x^2 + 2x + 1$ com o ruído gaussiano, com e sem a aplicação da condição de parada. Esses modelos foram então submetidos à avaliação com o conjunto de dados de validação, produzindo como métrica o erro quadrático médio (EQM). Os resultados foram submetidos ao teste t de *Student* para verificar se há diferenças significativas entre utilizar ou não a condição de parada. Este procedimento foi repetido trinta vezes, e os resultados obtidos são apresentados na Tabela 2.

Considerando a hipótese nula H_0 : “*Não há diferença significativa entre os EQMs*” e um nível de significância $\alpha = 0,05$, rejeitamos a hipótese nula em um total de 5 execuções, destacadas em vermelho na Tabela 2. A média de épocas executadas pelo método condicionado foi de 59 ± 21 , o que representa uma redução do tempo de treinamento de 47% a 75% em relação ao método convencional, que utilizou 150 épocas.

Das 25 execuções em que a hipótese nula foi aceita, de um total de trinta execuções, 40% produziram EQMs menores que o método convencional, estando destacadas em verde na Tabela 2. Nas execuções onde a hipótese nula foi rejeitada, essa proporção diminui para 33%. O método convencional apresentou um EQM médio de 0,0162, com desvio padrão de 0,0035, enquanto o método condicionado resultou em um EQM médio de 0,02, com desvio padrão de 0,0079. Embora o EQM do método condicionado seja ligeiramente maior, pode-se considerá-lo aceitável, levando em conta a significativa redução no número de épocas de treinamento.

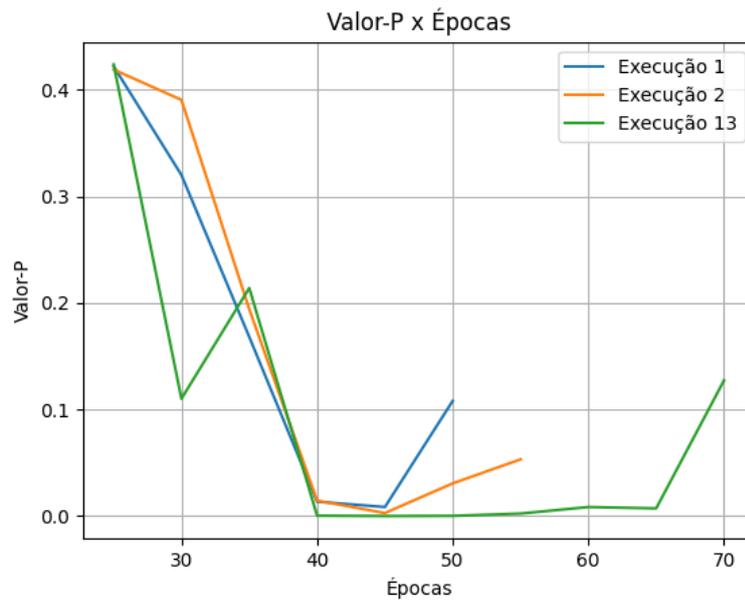
É possível observar um padrão no comportamento dos valores-p do ANOVA com medidas repetidas para estas execuções. Inicialmente, os valores-p são altos, o que validaria a hipótese nula. Em seguida, reduzem-se abruptamente para valores mais baixos, permanecendo assim por algumas épocas, até finalmente aumentarem acima do nível de significância estabelecido ($\alpha = 0,05$).

Tabela 2 - Dados comparativos para o polinômio

Execução	EQM no método condicionado	EQM no método convencional	Valor-P do teste t de Student	Época de parada
1	0,01658511	0,02193924	0,22256362	50
2	0,02101158	0,01685170	0,33909889	55
3	0,02067068	0,01727260	0,43534576	30
4	0,01647313	0,01568525	0,84424027	55
5	0,01393328	0,01559376	0,66565305	65
6	0,01311747	0,01435749	0,73797362	70
7	0,02292446	0,02724322	0,38857110	55
8	0,01581524	0,01763083	0,65708395	55
9	0,05138038	0,01267878	0,00000000	30
10	0,01606524	0,01285624	0,39879405	65
11	0,02014745	0,01773472	0,57935065	70
12	0,01942299	0,01802837	0,74725502	70
13	0,02884174	0,01422026	0,00164010	40
14	0,02285295	0,01323391	0,02360065	70
15	0,01662388	0,01385853	0,47877665	35
16	0,01420467	0,01777071	0,37252576	95
17	0,01713845	0,01689138	0,95224018	65
18	0,03726879	0,01447226	0,00000764	30
19	0,02151847	0,01416203	0,08164512	115
20	0,01494190	0,01689677	0,62419283	70
21	0,01239486	0,01127829	0,74554419	45
22	0,01837562	0,02139857	0,49789463	30
23	0,01319502	0,01295629	0,94736367	55
24	0,01519406	0,01821272	0,46019044	75
25	0,02051128	0,01438127	0,14229255	90
26	0,01991674	0,02206269	0,63952520	65
27	0,01950785	0,01328805	0,12462872	55
28	0,02809808	0,01310868	0,00096854	30
29	0,01598333	0,01541991	0,88693862	35
30	0,01666883	0,01411872	0,51575637	85

A seguir, a Figura 1 mostra o comportamento do valor-p do ANOVA com medidas repetidas durante o treinamento condicionado. Foram escolhidas três execuções distintas: a execução 1, em que a hipótese nula foi aceita e o EQM foi menor que o do método convencional; a execução 2, em que a hipótese nula foi aceita, mas o EQM foi maior que o do método convencional; e a execução 14, em que a hipótese nula foi rejeitada.

Figura 1 - Valor-p ao longo das épocas de treinamento para o polinômio



Contudo, sabe-se que os erros quadráticos médios iniciais têm escalas completamente diferentes dos finais, como pode ser observado na Tabela 3, cujos dados correspondem a primeira execução. Isso justifica o uso da pontuação padrão (*Z-Score*) como segunda condição após a comparação entre o valor-p e a significância, permitindo ignorar estes valores discrepantes e possibilitando que a execução continue até que os erros se estabilizem, como ilustrado na Figura 2, onde se excluem os cinco valores discrepantes iniciais.

Figura 2 - Evolução do MSE no problema do polinômio

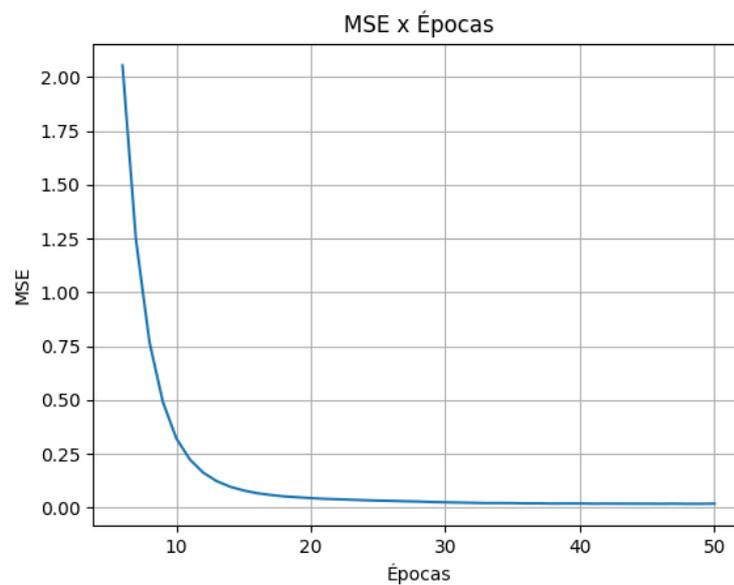


Tabela 3 - Progressão do EQM para o polinômio

Época	EQM	Época	EQM	Época	EQM
1	139,29652405	18	0,05212945	35	0,02075770
2	18,30364037	19	0,04812206	36	0,01952258
3	7,98627472	20	0,04411111	37	0,01965254
4	5,27517891	21	0,04071367	38	0,01866022
5	3,33999705	22	0,03867115	39	0,01914839
6	2,05441260	23	0,03654680	40	0,01906470
7	1,24468851	24	0,03433970	41	0,01818444
8	0,77050471	25	0,03246681	42	0,01841763
9	0,49043354	26	0,03133863	43	0,01824439
10	0,32151699	27	0,02933855	44	0,01807125
11	0,22391310	28	0,02804780	45	0,01798846
12	0,16220497	29	0,02575952	46	0,01762711
13	0,12292250	30	0,02450363	47	0,01813818
14	0,09693938	31	0,02342085	48	0,01750056
15	0,07961185	32	0,02208819	49	0,01760569
16	0,06698086	33	0,02103030	50	0,01832074
17	0,05868140	34	0,02102833		

O ANOVA com medidas repetidas apenas compara as médias entre os grupos, sem necessariamente considerar as escalas entre eles. Devido às características específicas dos dados em questão, torna-se necessário aplicar este segundo procedimento.

Os dados apresentados corroboram a hipótese de que o ANOVA com medidas repetidas pode ser aplicada como uma ferramenta válida e eficaz no critério de parada do treinamento de redes neurais artificiais. Isso ocorre sem perdas significativas na qualidade do modelo, apesar das características do problema, que seria mais bem modelado por uma regressão linear.

Ajustes nos parâmetros, como o tamanho da janela de épocas, a normalização dos erros quadráticos médios e técnicas secundárias após a aplicação do ANOVA com medidas repetidas, podem potencialmente promover uma melhora na qualidade do treinamento em relação ao método convencional.

Apesar disso, obteve-se uma taxa moderada de sucesso na prevenção do *overfitting*. Este resultado motiva a investigação do próximo problema: a aplicação deste método em um problema clássico, a fim de avaliar se as limitações identificadas neste estudo também serão encontradas em outras aplicações.

3.2 Conjunto de Dados Íris

De maneira similar ao problema anterior, uma rede neural artificial (RNA) foi treinada para a aprendizagem do clássico problema de classificação das flores do gênero Íris, cujos dados foram obtidos do *UCI Machine Learning Repository*. Novamente, os modelos resultantes dos treinamentos pelos métodos convencional e condicionado foram submetidos ao conjunto de validação, gerando métricas de erro quadrático médio (EQM) para cada método, o valor-p do teste t entre os erros quadráticos médios e a época de parada para o método condicionado, conforme ilustrado na Tabela 4.

Tabela 4 - Dados comparativos para as Íris

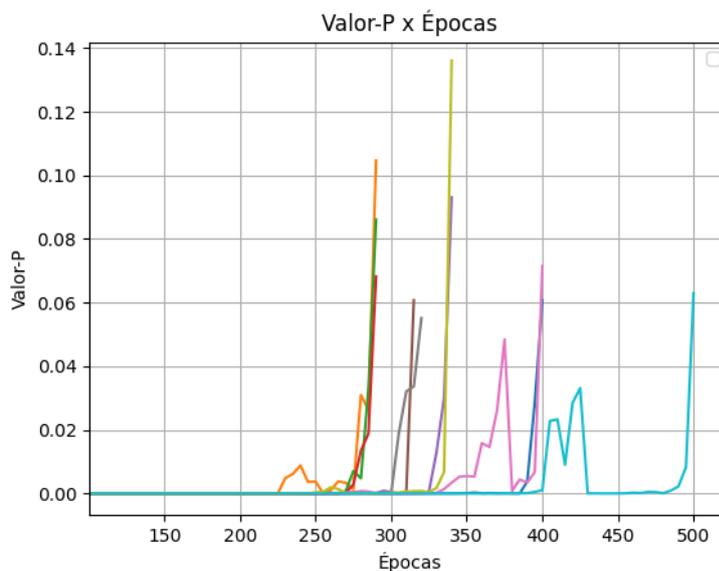
Execução	EQM no Método Condicionado	EQM no Método Convencional	Valor-P do Teste T de Student	Época de Parada
1	0,02309039	0,01843400	0,90083670	400
2	0,00263353	0,00183900	0,94834774	290
3	0,01127246	0,01103709	0,99314294	290
4	0,01944305	0,04525779	0,58077687	290
5	0,01349209	0,00657887	0,79031837	340
6	0,02953869	0,02423343	0,90071506	315
7	0,01452781	0,02071149	0,85746804	400
8	0,02254648	0,01471181	0,82488366	320
9	0,01402950	0,01801891	0,90328164	340
10	0,03668756	0,05915645	0,69252275	500
11	0,02938605	0,02196113	0,85820660	420
12	0,01666545	0,01178627	0,87467984	365
13	0,02042676	0,01270720	0,81718048	375
14	0,00194022	0,00124737	0,94665222	285
15	0,01145520	0,00427808	0,75533312	380
16	0,02504375	0,00856372	0,62475453	500
17	0,03178637	0,02768002	0,92683330	450
18	0,01185348	0,01013902	0,94972963	345
19	0,00324700	0,02301825	0,50815315	315
20	0,01702898	0,02537887	0,82505841	300
21	0,03927302	0,03469046	0,92678600	500
22	0,01631906	0,02658591	0,78703794	370
23	0,00508585	0,00378412	0,93992053	275
24	0,06142106	0,06128899	0,99835950	320
25	0,00328403	0,00145159	0,88462230	305
26	0,00797270	0,00866331	0,97670455	305
27	0,01260261	0,01513507	0,93391404	340
28	0,03065675	0,03235294	0,97060348	315
29	0,01627119	0,00822410	0,77934789	310
30	0,01727804	0,02130360	0,91101499	305

Considerando a hipótese nula H_0 : “*Não há diferença significativa entre os EQMs*” e um nível de significância $\alpha = 0,05$, a hipótese nula foi aceita para todas as execuções. Isso revela a consistência do método condicionado em produzir resultados estatisticamente similares àqueles gerados pelo método convencional, ao mesmo tempo em que reduz o número de épocas e, portanto, o tempo de treinamento. A média de épocas executadas no método condicionado foi de 352 épocas, com desvio padrão de 66, o que representa uma redução de 16% a 43% no tempo de treinamento em relação às 500 épocas executadas no método convencional.

Em 37% das trinta execuções, os erros quadráticos médios obtidos foram menores do que os do método convencional, conforme indicado pelos valores destacados em verde na Tabela 4. Este valor é similar ao obtido no problema anterior, onde se encontrou um percentual de 33% para todas as execuções, o que pode indicar uma tendência nas capacidades do método condicionado. A média de todos os EQMs no método convencional foi de 0,0193, com desvio padrão de 0,0154, enquanto o método condicionado obteve um EQM de 0,0189, com desvio padrão de 0,0129. Portanto, além de reduzir o tempo de treinamento, o método condicionado também apresentou uma redução média dos erros de validação, sendo um resultado positivo para sua aplicação.

Assim como no problema anterior, o comportamento do valor-p do ANOVA com medidas repetidas ao longo das épocas de treinamento para todas as execuções do método condicionado pode ser observado na Figura 3. Entretanto, diferentemente do problema anterior, os erros apresentaram distribuição normal, não havendo necessidade de aplicar métodos secundários, como normalização ou pontuação padrão, para lidar com picos de valores iniciais. De modo geral, os valores são estáveis e crescem gradualmente até atingirem um valor superior ao nível de significância $\alpha = 0,05$, resultando na parada do treinamento. Observa-se que o intervalo entre o ponto em que os valores-p começam a crescer significativamente e o ponto de parada ocorre, em média, em uma janela de 100 épocas. Este crescimento normalmente ocorre entre 250 e 350 épocas executadas, sugerindo uma tendência para este problema.

Figura 3 - Valor-p ao longo das épocas de treinamento para as Íris



Por fim, os erros quadráticos médios da primeira execução deste problema estão apresentados na Tabela 5. Os dados foram espaçados em janelas de 10 épocas para simplificar a visualização. Conforme mencionado anteriormente, a escala dos valores iniciais de erro não é tão discrepante em relação aos valores próximos à época de parada: enquanto o problema anterior apresenta uma razão entre o erro inicial e o erro final da ordem de $7,6 \times 10^3$, neste problema, a razão é de aproximadamente $5,5 \times 10^1$. Considerando apenas o primeiro e o último erro da primeira janela de treinamento, com 25 valores agrupados de 5 em 5, o problema anterior atinge uma razão de $4,3 \times 10^3$, enquanto este atinge apenas $1,7 \times 10^0$.

Pelo fato de as discrepâncias entre os erros quadráticos médios não serem significativas para este problema, não foi necessário aplicar métodos de normalização ou controle de *outliers*, evidenciando o impacto exclusivo do ANOVA com medidas repetidas no treinamento condicional. Também para efeito de comparação com o problema anterior, a progressão do erro médio quadrático para este problema está ilustrada em forma de gráfico na [x].

Em resumo, os dados apresentados indicam que a utilização do ANOVA com medidas repetidas como critério de parada no treinamento de uma rede neural artificial mostrou-se eficaz na redução do tempo de treinamento, sem comprometer a qualidade da rede treinada. Apesar das limitações, os resultados permitem concluir que a abordagem adotada foi capaz, inclusive, de aprimorar o treinamento, possivelmente por evitar o *overfitting* em um número moderado de casos.

Figura 4 - Evolução do EQM no problema das Íris

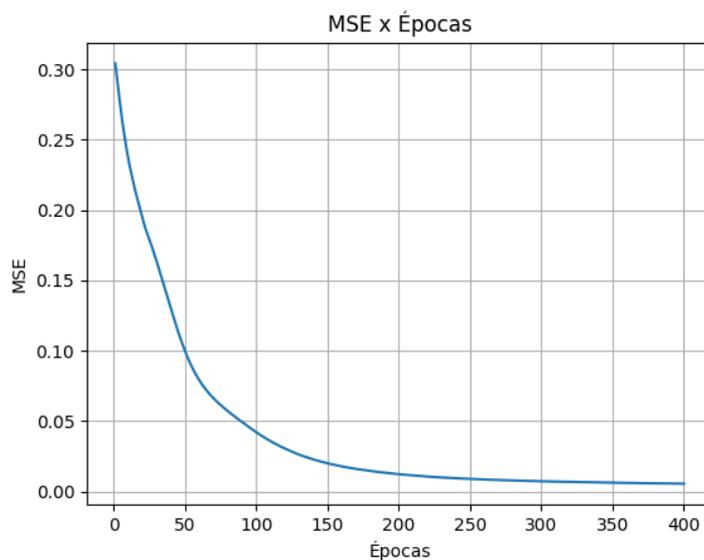


Tabela 5 – Progressão do EQM para as Íris

Época	EQM	Época	EQM
1	0,30422890	210	0,01145280
10	0,23716141	220	0,01066887
20	0,19452959	230	0,01000993
30	0,16362824	240	0,00945538
40	0,13018017	250	0,00898968
50	0,09976073	260	0,00855552
60	0,07903890	270	0,00818254
70	0,06625599	280	0,00786271
80	0,05711884	290	0,00756309
90	0,04936500	300	0,00730520
100	0,04211197	310	0,00706243
110	0,03576883	320	0,00685826
120	0,03059380	330	0,00664419
130	0,02624364	340	0,00646792
140	0,02278659	350	0,00628680
150	0,02007748	360	0,00613883
160	0,01787249	370	0,00595517
170	0,01608249	380	0,00583397
180	0,01463360	390	0,00573618
190	0,01337927	400	0,00557236
200	0,01234519		

CONCLUSÃO

Este trabalho investiga a aplicação da análise de variância (ANOVA) com medidas repetidas como critério de parada no treinamento de redes neurais artificiais (RNA), visando otimizar o processo de treinamento e atenuar a ocorrência de *overfitting*. Experimentos realizados em dois problemas diferentes – a modelagem polinomial quadrática usando ruído gaussiano e a classificação das flores do gênero Íris – fornecem informações valiosas sobre a eficácia e as limitações do método.

Os resultados obtidos mostram que o método proposto (denominado “método condicional”) é capaz de reduzir moderadamente o tempo de treinamento para ambos os problemas. Para problemas polinomiais, o número de épocas de treinamento foi reduzido em 47% a 75%, e para problemas de classificação das flores do gênero Íris, foi reduzido em 16% a 43%. A redução do tempo de treinamento é alcançada sem afetar significativamente a qualidade do modelo final, conforme evidenciado pela análise estatística dos erros quadráticos médios (EQM) coletados.

Um aspecto a ser notado nos resultados é a consistência da abordagem condicional na produção de modelos com desempenho estatisticamente semelhante às abordagens tradicionais. No problema Íris, todas as trinta execuções aceitam a hipótese nula de que não há diferença significativa entre os erros quadráticos médios dos dois métodos. Além disso, numa proporção significativa de execuções (37% para o Íris e 40% para o polinômio), a abordagem condicional produz erros quadráticos médios menores do que a abordagem tradicional.

No entanto, o estudo também revelou algumas limitações e desafios na aplicação desta abordagem. Em problemas polinomiais, a presença de *outliers* nos erros iniciais do treinamento requer uma etapa adicional de normalização utilizando a pontuação padrão ou Z-Score para avaliá-los. Isto sugere que a eficácia do método pode variar dependendo das características específicas do problema e dos dados relevantes.

A análise do comportamento dos valores-p durante o período de treinamento fornece informações valiosas sobre o processo de convergência do modelo. No problema Íris, observa-se um padrão mais estável e progressivo de crescimento do valor-p, enquanto no problema polinomial o comportamento é mais instável e requer ajustes adicionais.

Ressalta-se que, embora a proposta tenha se mostrado moderadamente eficaz para os problemas escolhidos, a utilização de um único tipo de modelo e apenas do

ANOVA com medidas repetidas na condição de parada é um limitante na generalização do método desenvolvido neste trabalho a outros contextos e modelos.

Em resumo, este trabalho demonstra que o ANOVA com medidas repetidas pode ser uma ferramenta eficaz para otimizar o treinamento de redes neurais artificiais, fornecendo critérios de parada estatisticamente fundamentados. Espera-se que o método proposto reduza o tempo de treinamento sem afetar significativamente a qualidade do modelo e, em alguns casos, até melhore o desempenho do modelo.

Pesquisas futuras poderiam explorar a aplicação deste método a uma gama mais ampla de problemas e arquiteturas de redes neurais, e investigar melhorias no método para resolver as limitações identificadas. Além disso, seria valioso comparar este método com outros métodos de parada precoce e explorar sua combinação com técnicas complementares de otimização de treinamento, bem como sua utilização em conjunto com outras técnicas.

Esta pesquisa contribui para o campo do aprendizado de máquina ao propor uma nova abordagem para enfrentar os desafios fundamentais no treinamento de redes neurais artificiais, abrindo caminho para o desenvolvimento de técnicas de treinamento de modelos de Inteligência Artificial mais eficientes e poderosas.

REFERÊNCIAS

AGARWAL, Rajeev; WIRAHADI, Bambang. **Application of Z-Score for Outlier Detection in Research Data**. *Journal of Mathematical and Fundamental Sciences*, v. 48, n. 3, p. 159-173, 2016.

BERGSTRA, James; BENGIO, Yoshua. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, n. 2, p. 281-305, 2012.

CHOLLET, François. **Deep Learning with Python**. 2nd ed. Shelter Island, NY: Manning Publications, 2021.

DUA, Dheeru; GRAFF, Casey. **Iris Dataset**. UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences. Disponível em: <https://archive.ics.uci.edu/ml/datasets/Iris>. Acesso em: agosto, 2024.

FIELD, A. (2013). **Discovering Statistics Using IBM SPSS Statistics**. SAGE Publications.

FISHER, Ronald A. **The Use of Multiple Measurements in Taxonomic Problems**. *Annals of Eugenics*, v. 7, n. 2, p. 179-188, 1936.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge, MA: MIT Press, 2016.

KOHAVI, Ron. A study of cross-validation and bootstrap for accuracy estimation and model selection. **In: Proceedings of the 14th International Joint Conference on Artificial Intelligence**. San Mateo, CA: Morgan Kaufmann, 1995. p. 1137-1143.

MONTGOMERY, Douglas C. **Design and Analysis of Experiments**. 8th ed. Hoboken, NJ: Wiley, 2012.

NG, Andrew Y. Feature selection, L1 vs. L2 regularization, and rotational invariance. **In: Proceedings of the Twenty-First International Conference on Machine Learning**. ACM, 2004. p. 78.

PRECHELT, Martin. Early stopping-but when? **In: Neural Networks: Tricks of the Trade**. Springer, Berlin, Heidelberg, 1998. p. 55-69.

SJÖBERG, Jonas; LJUNG, Lennart. Overtraining, regularization, and validating neural network models. **In: 1992 IJCNN International Joint Conference on Neural Networks**. IEEE, 1992. p. 315-322.

SRIVASTAVA, Nitish; HINTON, Geoffrey; KRIZHEVSKY, Alex; SUTSKEVER, Ilya; SALAKHUTDINOV, Ruslan. Dropout: a simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, v. 15, n. 1, p. 1929-1958, 2014.