

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS - GRADUAÇÃO EM ENGENHARIA ELÉTRICA



EXERCÍCIO VI - BASES REAIS COM ELMs REDES
NEURAIS ARTIFICIAIS

Matheus Barros Oliveira

19 de novembro de 2024

Relatório - Bases reais com ELMs Redes Neurais Artificiais

Relatório sobre ELMs Redes
Neurais Artificiais com
Bases Reais

19 de novembro de 2024

1 Introdução

Extreme Learning Machines (ELMs) são redes neurais de camada única projetadas para lidar com problemas de classificação e regressão de forma eficiente. A estrutura das redes ELM são, na verdade, a mesma estrutura das Redes Perceptron de Múltiplas Camadas [MLP], sendo que a principal diferença entre elas é a forma em que elas são treinadas [de Pádua Braga, 2024]. Diferentemente das redes tradicionais, as ELMs se destacam por seu treinamento extremamente rápido, pois os pesos da camada oculta são definidos aleatoriamente e permanecem fixos, enquanto os pesos da camada de saída são ajustados de forma analítica. Essa abordagem permite que as ELMs capturem relações não lineares complexas, tornando-as ideais para tarefas que exigem alta capacidade de generalização em tempo reduzido.

O Perceptron foi desenvolvido para representar algumas das propriedades fundamentais dos sistemas inteligentes de forma geral, ou seja, o Perceptron representa um sistema nervoso hipotético [Rosenblatt, 1958]. Desde a sua primeira citação no artigo "*The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain*" desenvolvido por F. Rosenblatt em 1958 o Perceptron é uma base fundamental para o estudo e desenvolvimento de redes neurais e algoritmos de inteligência artificial. Sua estrutura e funcionamento são relativamente simples, sendo capaz de realizar classificações lineares ao ajustar pesos com base em exemplos rotulados. Desde forma o Perceptron é capaz de "aprender" a partir de dados de treino.

2 Objetivo

O objetivo do exercício desta semana é aplicar as redes tipo ELM para resolver problemas multidimensionais não-lineares, a partir de conjuntos de dados reais, disponíveis no repositório público *UCI Machine Learning Repository*. Além de se resolver os problemas com dados reais multidimensionais não-lineares deve-se verificar o que ocorre com a acurácia do modelo ao se alterar o número de neurônios. Por fim, espera-se comparar o resultado da ELM com o do modelo Perceptron e desta forma analisar a acurácia dos dois métodos.

3 Materiais e Métodos

Para a execução deste exercício foram utilizados as seguintes ferramentas e base de dados descritas abaixo.

3.1 Materiais

Utilizou-se o RStudio (versão 4.4.1 (2024-06-14 ucrt) instalada em um Windows 11 (versão 23H2) com um processador 12th Gen Intel(R) Core(TM) i5-12500H 2.50 GHz e 32GB de memória RAM.

3.2 Descrição da Base de Dados

1. Breast Cancer (diagnostic)

- **Descrição:** A base de dados *Breast Cancer Wisconsin (Diagnostic)* é utilizada para a classificação de tumores como malignos ou benignos, com base em características calculadas a partir de uma imagem digitalizada de uma amostra de tecido mamário aspirado por agulha fina.

2. Statlog (Heart)

- **Descrição:** A base de dados *Statlog (Heart)* é utilizada para prever a presença ou ausência de uma doença cardíaca com base em atributos médicos.

3.3 Metodologia

Segue abaixo a metodologia realizada durante o experimento:

3.3.1 Passo 1: Carregamento da Base de Dados

Iniciamente realizou-se o carregamento do conjunto de dados para o ambiente R.

3.3.2 Passo 2: Separação do Conjunto de Testes e Treino

O conjunto de dados foi separado em 70% (conjunto de treino) e 30% (conjunto de validação) de forma aleatória.

3.3.3 Passo 3: Conversão das Classes

Foi necessário realizar a conversão das classes de predição das bases de dados para as classes -1 e 1 para o treinamento do modelo ELM e para as classes 0 e 1 para o modelo Perceptron.

3.3.4 Passo 4: Aplicação da t-SNE e Normalização dos Dados

A base de dados de treino e testes foram passadas pela técnica de redução t-SNE e depois da redução as bases de dados foram normalizadas para um fator entre 0 e 1.

3.3.5 Passo 5: Visualização dos Dados

Com os dados carregados e tratados, foi necessário realizar a visualização dos conjuntos de treino e validação para se observar a distribuição dos mesmos.

3.3.6 Passo 6: Treinamento e Validação da ELM e Perceptron

Após a avaliação dos conjuntos de treino e teste, foi realizado o treinamento e teste da ELM e do Perceptron. Cada base de dados foi treinada e validada em cinquenta execuções diferentes, e foi realizado o cálculo da acurácia dos modelos tanto para a base de testes quanto para a validação, conforme a média \pm desvio padrão.

Para se avaliar os efeitos da variação de p no *overfitting* e *underfitting*, os modelos da ELM foram treinados com o parâmetro p variando entre (5,10,30,100,500 e 1500).

3.3.7 Passo 7: Visualização de Superfície

Após a realização do treinamento e validação dos modelos foi escolhida um modelo e foi analisado a margem de separação tanto do modelo ELM, quanto para o modelo Perceptron para se observar o impacto da variação do fator p .

4 Resultados

Pode-se observar abaixo os resultados obtidos para os experimentos:

4.1 Base de Dados: Breast Cancer (diagnostic)

Pode-se observar na Figura 1 o conjunto de treinamento e o conjunto de validação da base de dados Breast Cancer (diagnostic) depois de passado pela redução t-SNE e normalização.

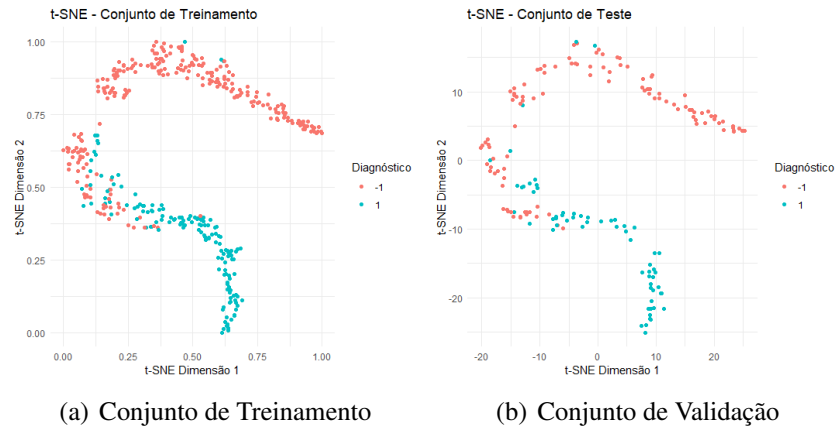


Figura 1: Conjunto de Treinamento e Validação do Conjunto de Dados: Breast Cancer (diagnostic)

Pode-se observar que a base de dados em análise não é linearmente separável. Obversa-se na Figura 2 a superfície de separação calculada para a base de dados de treinamento e na Figura 3 a superfície de separação para a base de validação.

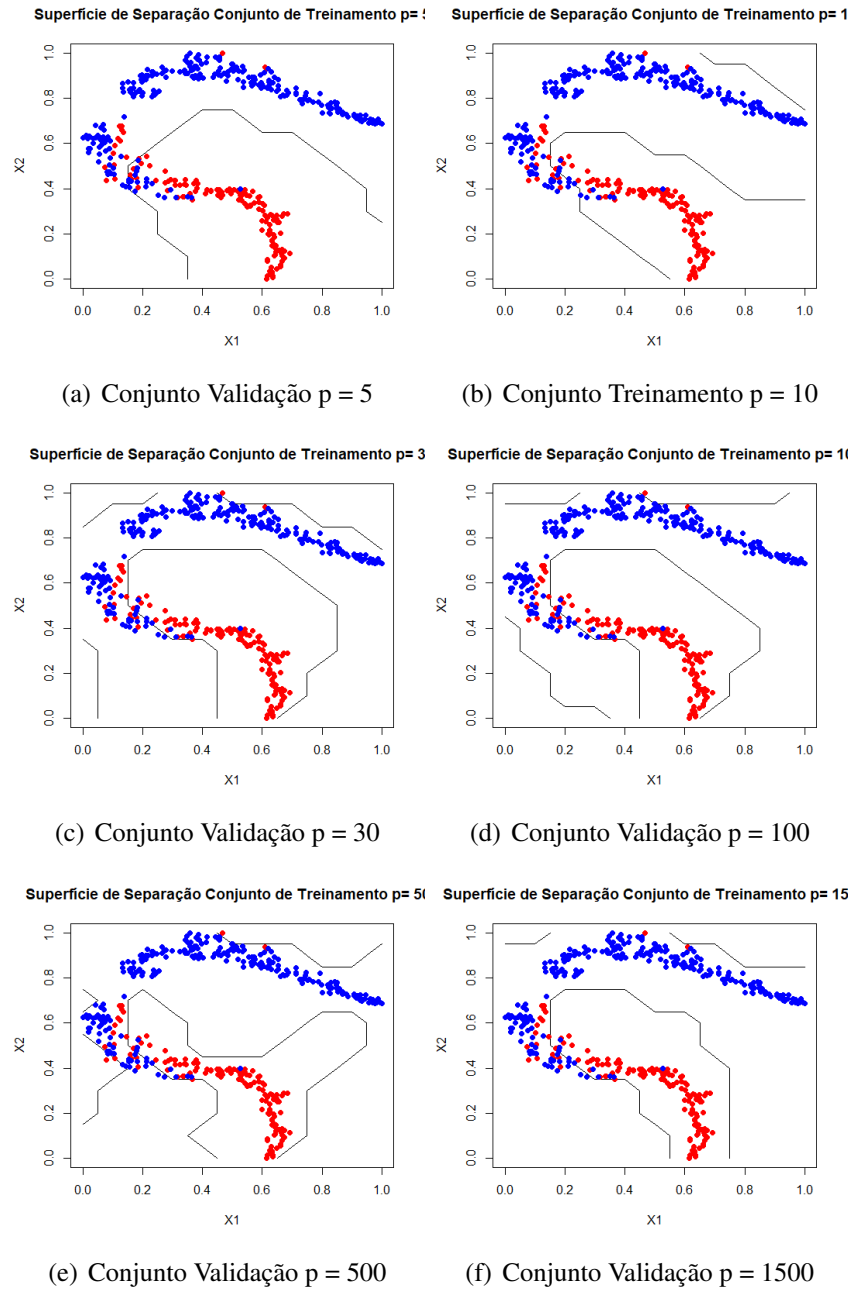


Figura 2: Gráficos das Superfícies de Separação (ELM) do Conjunto de Dados de Treinamento: Breast Cancer (diagnostic)

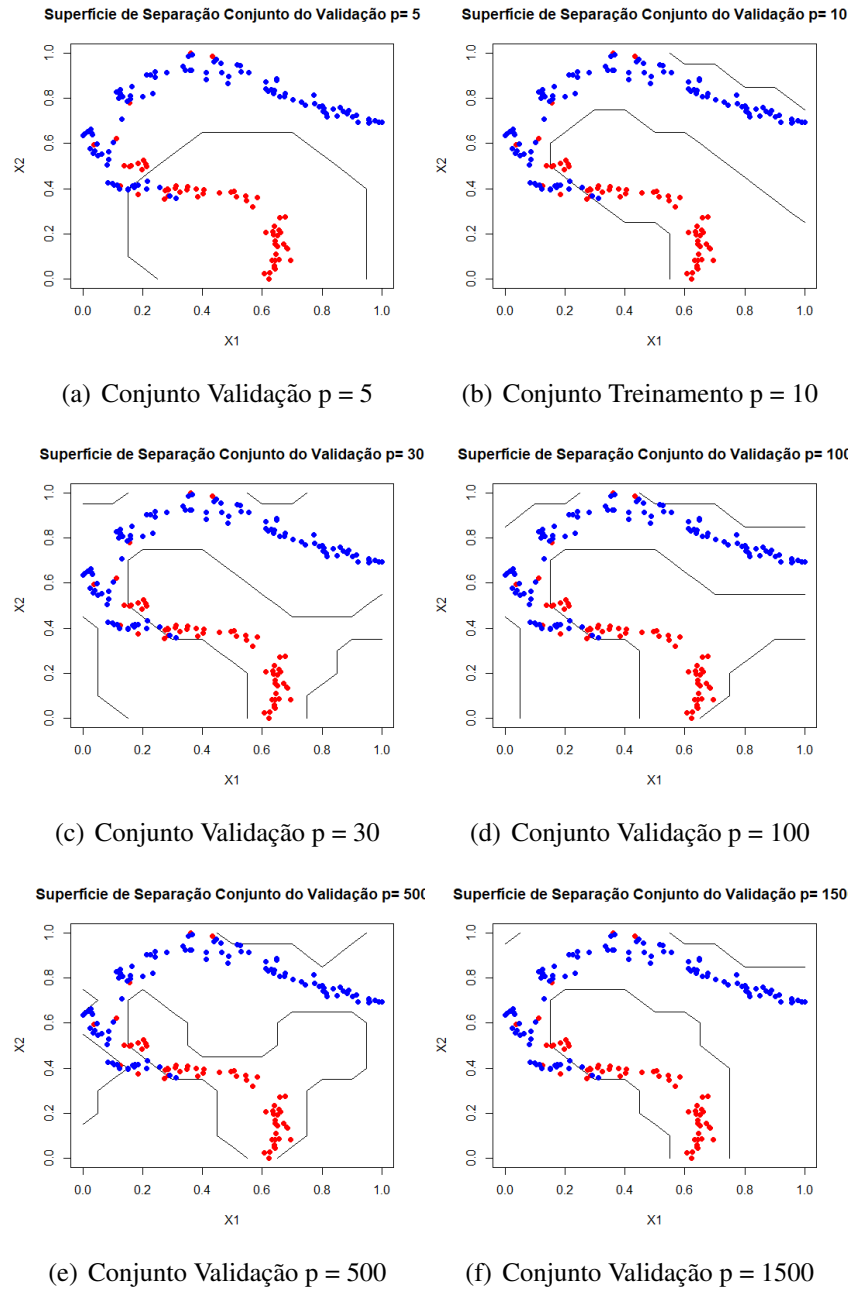
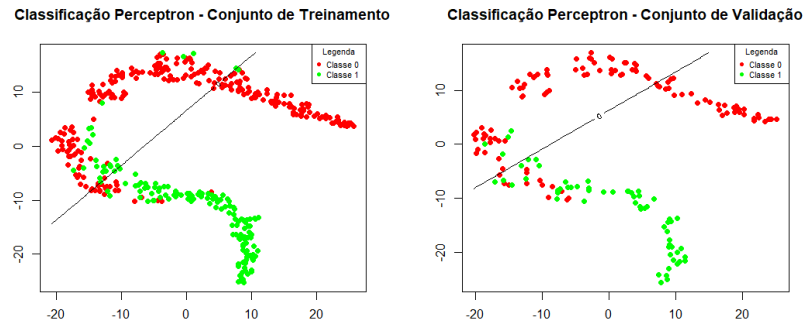


Figura 3: Gráficos das Superfícies de Separação (ELM) do Conjunto de Dados de Validação: Breast Cancer (diagnostic)

Pode-se verificar na Figura 4 a superfície de separação calculada para o Perceptron.



(a) Superfície de Separação do Perceptron: Conjunto de Treinamento (b) Conjunto Validação $p = 500$

Figura 4: Superfície de Separação do Perceptron: Conjunto de Validação

Tabela 1: Resultados de Acurácia para Diferentes Valores de p

p	Acurácia Conjunto de Treinamento	Acurácia Conjunto de Teste
5	90.88% \pm 0.0075	89.12% \pm 0.0169
10	91.72% \pm 0.0033	92.19% \pm 0.0043
30	92.76% \pm 0.0055	93.06% \pm 0.0049
50	93.01% \pm 0.0038	92.72% \pm 0.0044
100	93.02% \pm 0.0021	92.53% \pm 0.0032
500	95.23% \pm 0.0004	90.33% \pm 0.0029
1500	92.43% \pm 0.0016	92.95% \pm 0.0008

Tabela 2: Resultados de Acurácia para o Perceptron (sem regularização)

Conjunto	Acurácia
Treinamento	71.94% \pm 0.0896
Teste	72.95% \pm 0.1110

4.1.1 Análise dos Resultados

Ao analisar os resultados obtidos a partir das 50 execuções para cada valor de p , é possível observar um comportamento característico das redes neurais em relação ao balanceamento entre capacidade de modelagem e generalização. Quando p é pequeno, como nos casos de $p = 5$ e $p = 10$, a rede neural apresenta uma acurácia crescente no conjunto de teste, atingindo um valor máximo de 92.19% com $p = 10$. Isso indica que, nesse intervalo, a rede consegue capturar a estrutura subjacente dos dados sem superajustar.

Conforme p aumenta para 30 e 50, a acurácia do conjunto de teste começa a estabilizar, e o desempenho do conjunto de treinamento continua a melhorar. No entanto, o ganho na acurácia de teste torna-se menos significativo, o que sugere que a rede neural pode estar começando a memorizar os dados de treinamento em vez de generalizar bem, isso é um indicativo de overfitting.

Para valores maiores de p , como 100, 500 e 1500, o fenômeno de overfitting se torna mais evidente. A acurácia do conjunto de treinamento permanece alta, enquanto a acurácia do conjunto de teste diminui. Por exemplo, com $p = 500$, a acurácia de teste cai para 90.33%, e, embora o custo computacional aumente consideravelmente, os resultados no conjunto de teste não melhoram. O comportamento inconsistente para $p = 500$ e $p = 1500$ reforça que aumentar a complexidade da rede além de certo ponto não traz benefícios.

Perceptron teve um desempenho ruim, com acurácias médias de 71.94% no treinamento e 72.95% no teste. O desvio padrão elevado revela uma alta inconsistência nos resultados. O modelo não conseguiu capturar os padrões necessários, destacando sua inadequação para este conjunto de dados.

4.2 Base de Dados: Statlog (Heart)

Pode-se observar na Figura 5 o conjunto de treinamento e o conjunto de validação da base de dados Statlog (Heart) depois de passado pela redução t-SNE e normalização.

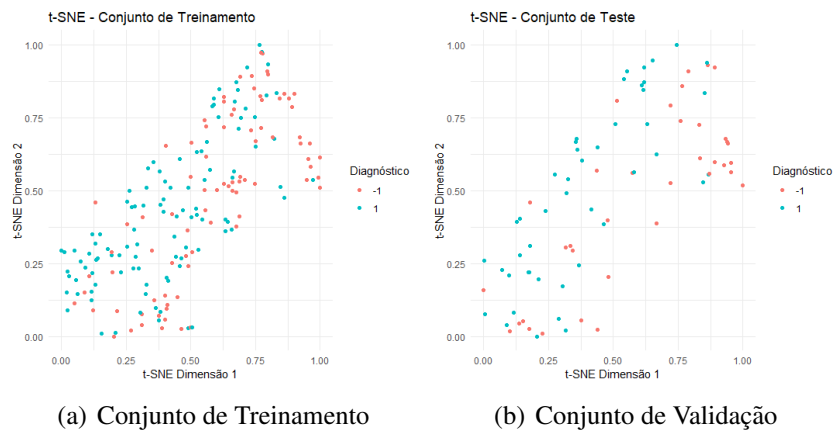


Figura 5: Conjunto de Treinamento e Validação do Conjunto de Dados: Statlog (Heart)

Pode-se observar que a base de dados em análise não é linearmente separável. Obversa-se na Figura 6 a superfície de separação calculada para a base de dados de treinamento e na Figura 7 a superfície de separação para a base de validação. Já na Figura 8 observa-se a superfície de seperação calculada para o Pecptron.

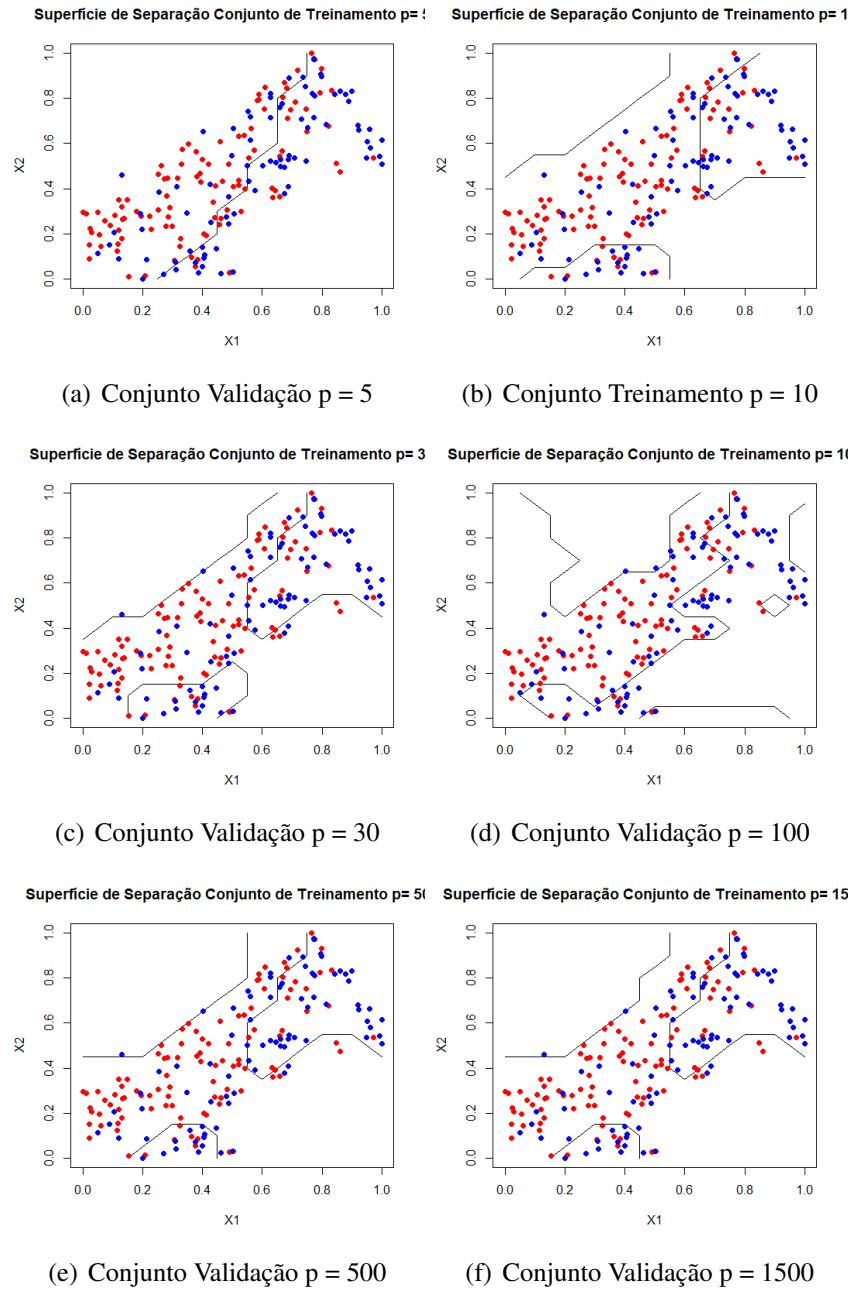


Figura 6: Gráficos das Superfícies de Separação (ELM) do Conjunto de Dados de Treinamento: Statlog (Heart)

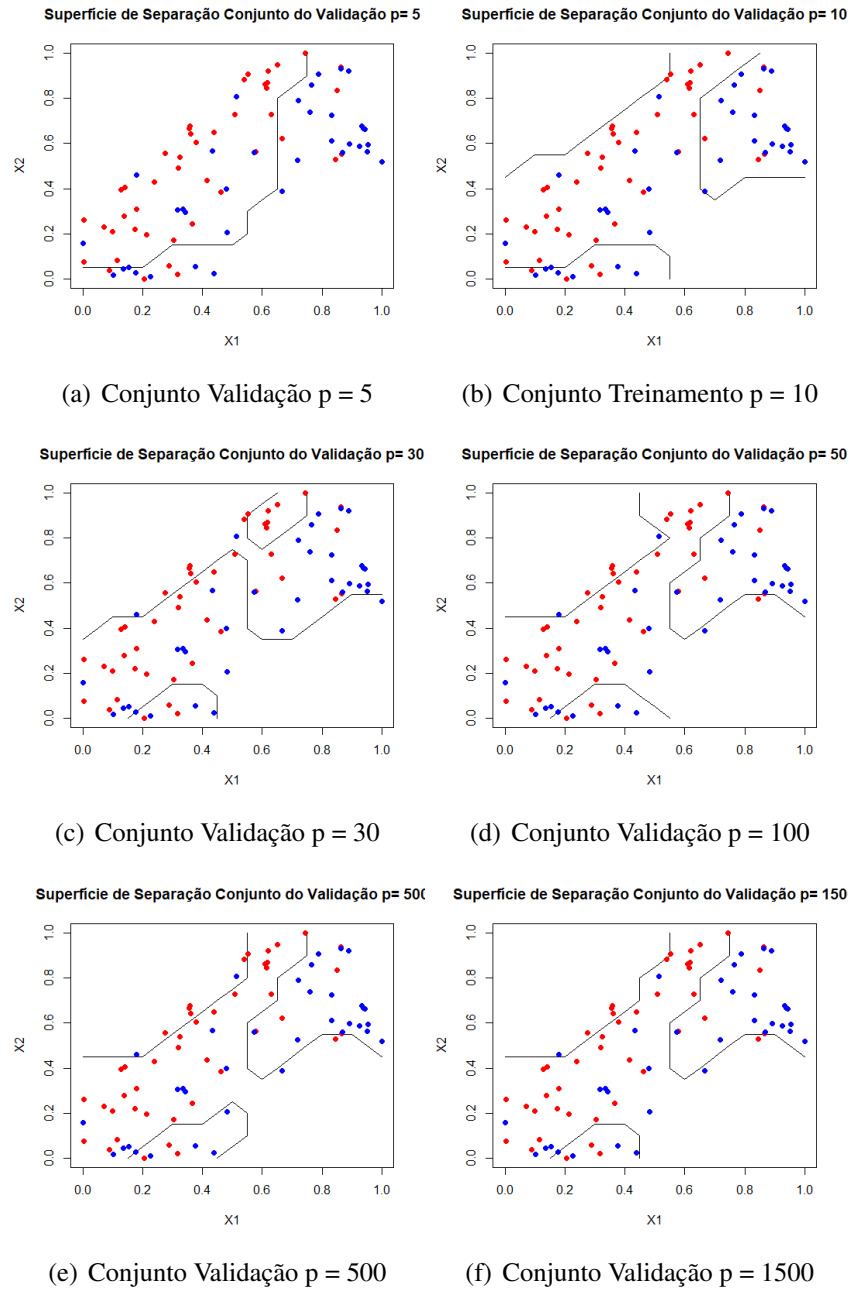
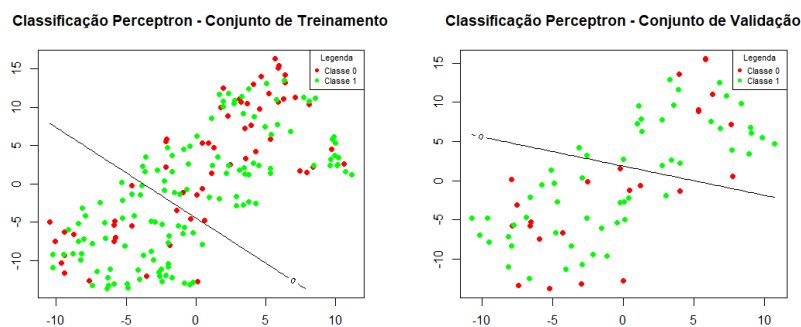


Figura 7: Gráficos das Superfícies de Separação (ELM) do Conjunto de Dados de Validação: Statlog (Heart)

Pode-se verificar na Figura 8 a superfície de separação calculada para o Perceptron.



(a) Superfície de Separação do Perceptron: Conjunto de Treinamento (b) Conjunto Validação $p = 500$

Figura 8: Superfície de Separação do Perceptron: Conjunto de Validação

Tabela 3: Resultados de Acurácia para Diferentes Valores de p

p	Acurácia Conjunto de Treinamento	Acurácia Conjunto de Teste
5	67.26% \pm 0.0100	74.27% \pm 0.0152
10	69.02% \pm 0.0067	76.47% \pm 0.0181
30	71.45% \pm 0.0132	71.95% \pm 0.0346
50	71.93% \pm 0.0102	70.49% \pm 0.0309
100	75.26% \pm 0.0052	63.14% \pm 0.0224
500	71.30% \pm 0.0074	70.27% \pm 0.0145
1500	70.93% \pm 0.0060	70.12% \pm 0.0114

Tabela 4: Resultados de Acurácia para o Perceptron (sem regularização)

Conjunto	Acurácia
Treinamento	47.04% \pm 0.0867
Teste	44.05% \pm 0.0516

4.2.1 Análise dos Resultados

Os resultados da segunda base de dados mostram um comportamento diferente e mais desafiador. Com valores pequenos de p , como 5 e 10, o modelo atinge uma acurácia de teste relativamente alta, chegando a 76.47% com $p = 10$. Isso sugere que, inicialmente, o modelo consegue aproveitar a complexidade adicional para melhorar a generalização. No entanto, essa tendência não se mantém à medida que p continua a crescer.

Para $p = 30$ e $p = 50$, a acurácia de teste começa a cair, apesar de a acurácia de treinamento continuar a aumentar. Esse comportamento é um indicativo claro de overfitting. Além disso, o desvio padrão mais elevado na acurácia de teste demonstra uma instabilidade crescente no desempenho do modelo, o que indica que o modelo está se tornando mais sensível às variações nos dados de treino.

Quando p é aumentado ainda mais para 100, 500 e 1500, a queda na acurácia de teste se torna evidente. O caso mais crítico ocorre com $p = 100$, onde a acurácia de teste despenca para 63.14%, evidenciando que o modelo perdeu a capacidade de generalizar. Mesmo que a acurácia do conjunto de treinamento continue alta, o desempenho no conjunto de teste não justifica a complexidade adicional. Com $p = 500$ e $p = 1500$, a acurácia de teste permanece baixa, e o modelo não se recupera, o que demonstra que o aumento de p prejudica a generalização.

O desvio padrão maior para valores intermediários de p reforça essa instabilidade no desempenho, indicando que o modelo se torna imprevisível e menos confiável em termos de resultados.

Para este conjunto de dados, o Perceptron foi treinado e validado com os dados não normalizados, pois ao se utilizar os dados normalizados o desempenho foi muito baixo. Com acurácias de 47.04% no treinamento e 44.05% no teste, e alta variabilidade nos resultados o Perceptron demonstra não ser adequado para lidar com a complexidade deste conjunto de dados.

5 Conclusão

Este trabalho analisou o impacto da variação do número de neurônios p nas Extreme Learning Machines (ELM) em relação à acurácia nos conjuntos de treinamento e teste. Observou-se que, para valores pequenos e moderados de p , as ELM alcançaram um equilíbrio razoável entre acurácia e generalização, com uma acurácia de teste relativamente alta. No entanto, conforme p aumentou, a acurácia no conjunto de treinamento continuou subindo, enquanto a acurácia no conjunto

de teste diminuiu, indicando o fenômeno de overfitting, onde o modelo se ajusta excessivamente aos dados de treinamento e perde sua capacidade de generalizar.

Além disso, comparativamente, as ELM demonstraram um tempo de treinamento significativamente mais rápido do que o Perceptron, mesmo para valores elevados de p . Isso reforça a vantagem computacional das ELM em problemas que exigem eficiência e desempenho. No entanto, a escolha de p continua sendo crucial para garantir que o modelo mantenha um bom equilíbrio entre precisão e generalização.

Referências

[de Pádua Braga, 2024] de Pádua Braga, A. (2024). *Aprendendo com Exemplos: Princípios de Redes Neurais Artificiais e de Reconhecimento de Padrões*. Escola de Engenharia da UFMG, Departamento de Engenharia Eletrônica.

[Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.