

Relatório: Projeto de Classificação com Perceptron do Dataset da planta Iris

Matheus Costa Monteiro - 12111BSI281
Thales Atheniel Farias de Godoi - 11511BSI232
Paulo Victor da Silva - 11521BSI219

24 de outubro de 2023

1 Introdução

Este relatório detalha o processo de classificação das espécies de planta Iris usando o algoritmo Perceptron. O conjunto de dados da Iris é uma das bases de dado no estudo de classificação de plantas. O objetivo central é utilizar o modelo do Perceptron para classificar diferentes espécies de Íris com base em suas características morfológicas, incluindo comprimento e largura da sépala e pétala.

O conjunto de dados contém informações sobre três espécies de plantas Iris: Setosa, Versicolor e Virginica. Cada uma dessas espécies é caracterizada por suas dimensões de pétalas e sépalas. O projeto visa empregar o algoritmo do Perceptron, no modelo de uma única camada(Perceptron), para classificar de forma binária ou multiclasse as espécies de Iris a partir dessas características.

O Perceptron é um modelo linear capaz de aprender a separar classes com base em um conjunto de entradas. O treinamento do Perceptron se baseia na atualização de pesos a cada iteração, ajustando-os para minimizar os erros de classificação.

As etapas do projeto incluem a carga e pré-processamento do conjunto de dados, treinamento do modelo Perceptron para diferentes classificações binárias e avaliação da precisão do modelo com base nos dados de teste. Além disso, a visualização dos resultados, como gráficos de iterações do Perceptron e limites de decisão, ajudará a compreender a eficácia do modelo na classificação das plantas Iris.

O trabalho detalhado neste relatório busca explorar o potencial do Perceptron como uma abordagem clássica para a classificação de plantas Iris, discutindo tanto o processo de implementação do algoritmo quanto a análise dos resultados obtidos.

2 Objetivo

O objetivo deste projeto é realizar a classificação binária das espécies de Íris ('Setosa' e 'Versicolor') utilizando o Perceptron como um classificador linear. O Perceptron será treinado para separar as duas classes com base nas características das pétalas e sépalas.

3 Métodos

O projeto foi desenvolvido em Python e fez uso das seguintes bibliotecas:

- **Pandas**: Para carregar e manipular os dados.
- **NumPy**: Para operações numéricas e computações.
- **scikit-learn**: Para dividir os dados em conjuntos de treino e teste, bem como para avaliar a precisão do modelo.
- **Matplotlib**: Para visualização de dados.

4 Implementação do Perceptron

Foi implementada uma classe Perceptron com os seguintes métodos:

- **fit(X, y)**: Realiza o treinamento do Perceptron com o conjunto de características (X) e rótulos (y).

- `net_input(X)`: Calcula a entrada líquida do Perceptron para um conjunto de características X .
- `predict(X)`: Realiza previsões com base nas entradas.

5 Resultados

5.1 Treinamento do Perceptron

Foram treinados dois Perceptrons:

- Um para classificar 'Setosa' e 'Versicolor'.
- Outro para classificar 'Setosa' e 'Virginica'.

Os hiperparâmetros utilizados foram uma taxa de aprendizado de 0.1 e 100 iterações.

5.2 Avaliação dos Modelos

Os modelos foram avaliados com base nos dados de teste, calculando a acurácia. Para a classificação 'Setosa' e 'Virginica', foi feita uma visualização em um gráfico 3D com o limite de decisão.

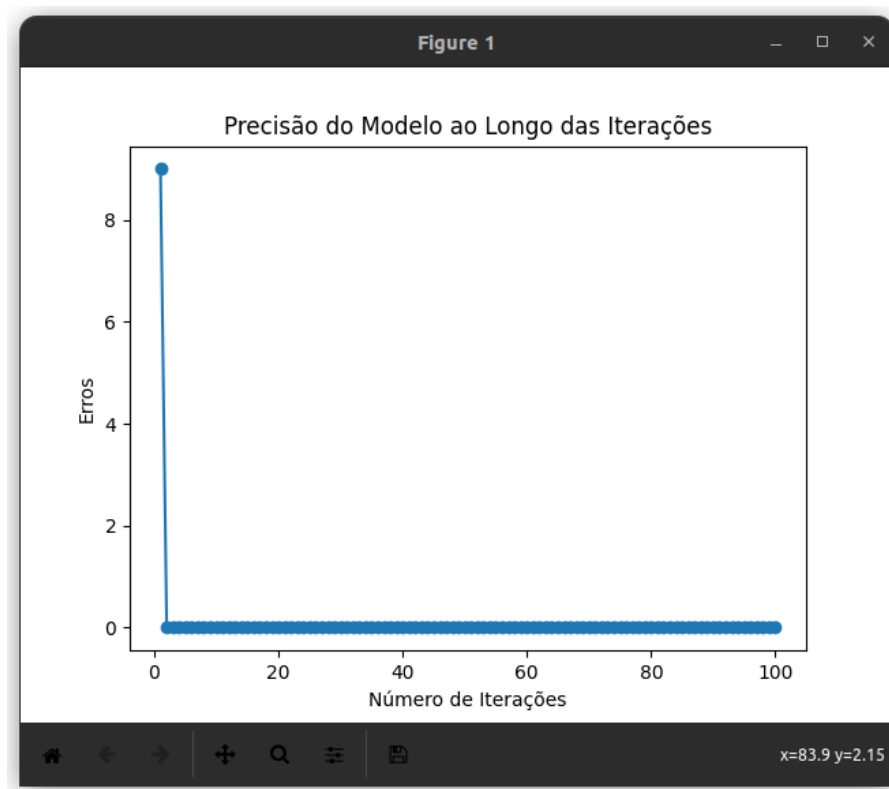


Figura 1: Gráfico da Precisão do Modelo ao Longo das Iterações

6 Implementação do Perceptron para Classificação dos 3 Tipos de Planta Iris

Nesta seção, discutiremos a tentativa de utilizar o algoritmo Perceptron para classificar as três espécies de plantas Iris presentes no conjunto de dados. A abordagem inicial consistia em realizar uma classificação multiclasse, visando identificar as espécies 'Setosa', 'Versicolor' e 'Virginica' com base nas características morfológicas de pétalas e sépalas.

Apesar da natureza robusta do Perceptron em problemas de classificação, logo ficou evidente que a classificação multiclasse para esse conjunto de dados apresentava desafios significativos. A principal questão foi a complexidade do problema e a não linearidade dos limites de decisão entre as três classes.

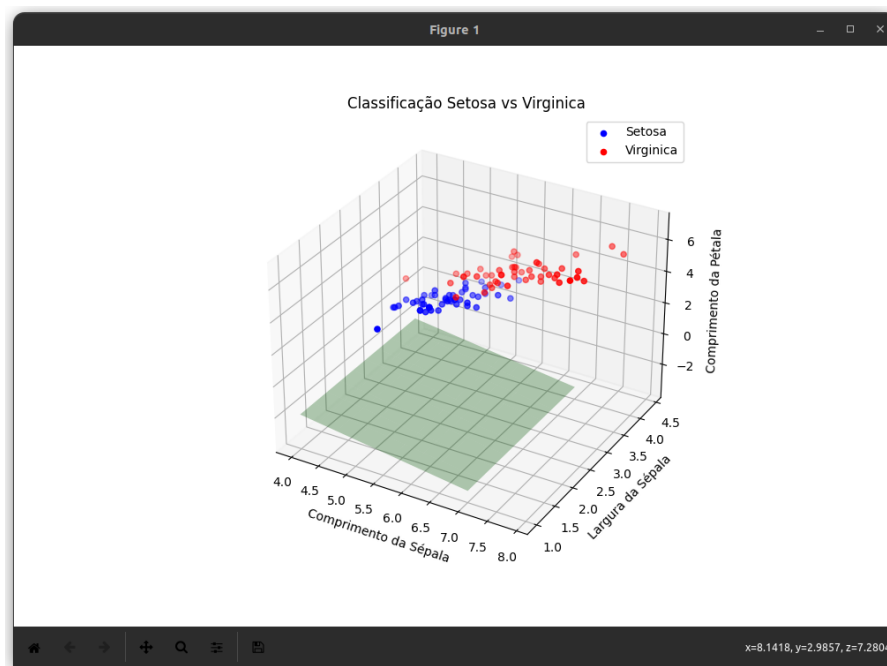


Figura 2: Classificação Setosa vs Virginica (Gráfico 3D)

Devido à natureza intrinsecamente linear do Perceptron, a capacidade de traçar fronteiras de decisão não lineares entre as espécies tornou-se limitada. Isso resultou em um desempenho insatisfatório do modelo, com dificuldades na correta classificação das três espécies.

Além disso, o tempo de treinamento necessário para alcançar um desempenho aceitável na classificação das três espécies também se mostrou proibitivamente longo, dada a complexidade do problema e a quantidade de iterações necessárias para ajustar os pesos do modelo.

Diante desses desafios, a abordagem foi redirecionada para uma classificação binária, onde o Perceptron foi mais eficaz na separação das espécies de plantas Iris.

Apesar do revés na tentativa inicial de classificação multiclasse, os desafios encontrados contribuíram para uma compreensão mais aprofundada das limitações do Perceptron em problemas mais complexos e multidimensionais de classificação.

7 Avaliação dos resultados na Classificação Apenas pelo Comprimento da Pétala

Durante a primeira tentativa de classificação das espécies de plantas Iris, observamos uma limitação no método de classificação, que baseava-se exclusivamente no comprimento da pétala para diferenciar entre as espécies. Essa abordagem, embora inicialmente simples, provou-se inadequada para realizar distinções precisas entre as diferentes espécies.

O uso único do comprimento da pétala para a classificação resultou em fronteiras de decisão lineares, o que levou a uma separação insuficiente entre as espécies Iris. Por conta disso, os limites de classificação acabaram sendo imprecisos, o que dificultava a correta distinção entre as diferentes classes de plantas.

Essa limitação se tornou evidente ao explorar o conjunto de dados e perceber que outras características, como a largura da sépala e o comprimento da pétala, desempenham papéis cruciais na diferenciação das espécies Iris. A dependência apenas do comprimento da pétala tornou o modelo de classificação menos preciso e, consequentemente, menos eficaz na separação das espécies.

Essa restrição inicial nos levou a reavaliar e reformular nossa abordagem para incluir outras características e, assim, criar um modelo de classificação mais abrangente e preciso para lidar com as distintas espécies de plantas Iris.

8 Conclusão

Os modelos de Perceptron utilizados para a classificação das espécies de plantas Iris demonstraram uma acurácia razoável na diferenciação entre as classes 'Setosa' e 'Versicolor' e 'Setosa' e 'Virginica'. No entanto, uma

observação atenta dos resultados revela a limitação de um modelo de classificação linear para distinguir com precisão as espécies quando as características estão mais misturadas.

Um destaque importante é o comportamento do gráfico da Figura 1, que mostra a evolução da taxa de erro do modelo ao longo das iterações. Conforme o algoritmo de treinamento do Perceptron é executado, é perceptível a queda gradual do erro, culminando em zero após um número específico de iterações. Esse comportamento ressalta a capacidade do modelo em aprender a separar as classes, mesmo com um limite de decisão linear.

Entretanto, como evidenciado no gráfico 3D, a limitação de um modelo linear é claramente visível. A tentativa de separar 'Setosa' de 'Virginica' mostra que uma abordagem linear pode não ser suficiente para a correta separação dessas espécies.

Portanto, concluímos que a abordagem do Perceptron, embora valiosa, pode ser limitada para lidar com classificações mais complexas. A utilização de modelos mais avançados, como SVMs ou redes neurais, pode ser mais adequada para classificações mais intrincadas que possam não ser facilmente separadas por um limite de decisão linear.