

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Artur Amaro - 11811BSI216, Daniel Gonçalves - 12011BCC011 , João Victor
Fernandes de Souza Silva - 11911BSI205, Luiz André de Silva Carvalho -
11911BSI225

Combinando algoritmos bioinspirados

Uberlândia, Brasil

Lista de ilustrações

Figura 1 – Base Iris10 - relação entre acurácia e conjunto de dados para um total da população inicial de 20	5
Figura 2 – Base Iris10 - relação entre acurácia e conjunto de dados para um total da população inicial de 30	6
Figura 3 – Base Iris10 - relação entre acurácia e conjunto de dados para um total da população inicial de 50	7
Figura 4 – Base Iris30 - relação entre acurácia e conjunto de dados para um total da população inicial de 20	8
Figura 5 – Base Iris30 - relação entre acurácia e conjunto de dados para um total da população inicial de 30	9
Figura 6 – Base Iris30 - relação entre acurácia e conjunto de dados para um total da população inicial de 50	10
Figura 7 – Base Iris50 - relação entre acurácia e conjunto de dados para um total da população inicial de 20	11
Figura 8 – Base Iris50 - relação entre acurácia e conjunto de dados para um total da população inicial de 30	12
Figura 9 – Base Iris50 - relação entre acurácia e conjunto de dados para um total da população inicial de 50	13
Figura 10 – Iris: características das flores e suas espécies	14

1 Introdução

O presente projeto compreende um perceptron de uma única camada e de um único neurônio, cujo treinamento de seus pesos é realizado por meio de um algoritmo genético. Este perceptron é capaz de classificar as espécies da base de dados Iris duas a duas.

Em relação ao AG usado para o treinamento dos pesos da rede, ele utiliza como cromossomo um vetor com os 5 pesos em ponto flutuante. A medida de aptidão adotada é a precisão, a seleção é feita por meio de torneio, o crossover é de apenas 1 ponto e a mutação usa a distribuição normal com média 0 e desvio padrão 1 para adicionar valores nos alelos.

2 Implementação

O Perceptron foi implementado na linguagem de programação Python. Instruções para instalação das bibliotecas necessárias e como executar o programa estão disponíveis no README do projeto.

2.1 Dados de entrada

O programa obrigatoriamente deve receber quais as duas espécies que serão usadas para o treinamento do perceptron e, opcionalmente, pode receber o número de gerações, a taxa de crossover, a taxa de mutação, o tamanho da população e a proporção da base que será utilizada para o treinamento.

Por padrão, o número de gerações está definido como 10, a população com 50, a taxa de crossover em 70%, a taxa de mutação em 1% e a proporção para treinamento em 20%.

3 Metodologia dos Testes e Resultados

Para a execução dos testes, foram escolhidas as classes Setosa e Virginica, sendo as mesmas escolhidas para a realização dos testes no trabalho anterior. Os testes são realizados sobre três conjuntos de treino definidos como Iris10, Iris30 e Iris50, sendo estes gerados a partir da base de dados Iris com proporção, respectivamente, de 10%, 30% e 50%.

Para cada conjunto de treino, são realizados 3 testes variando os valores da população inicial em relação a um valor constante do total de gerações. Além disso, para cada conjunto de treino, ao final do treinamento é realizada a inserção da terceira classe. No final do trabalho, na seção 3.4 (Resultados e Discussões), análises são discutidas referentes a inserção da terceira classe e os resultados obtidos em comparação aos resultados obtidos no trabalho anterior.

Abaixo são demonstrados os testes realizados em relação ao conjunto de treino, com gráficos para ilustração das variações da acurácia em relação ao conjunto de dados e os seus respectivos resultados. Por fim, as análises referentes a inserção da terceira classe e discussões acerca da comparação dos resultados obtidos no trabalho anterior em relação a este são apresentadas.

3.1 Testes e resultados para o conjunto de dados Iris10

Para todos os testes, foi considerado um número constante de gerações sendo 20 e variações no tamanho da população inicial.

Para o primeiro teste, com o tamanho da população inicial sendo 20, a acurácia obtida foi de 98.88% sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura [1](#).

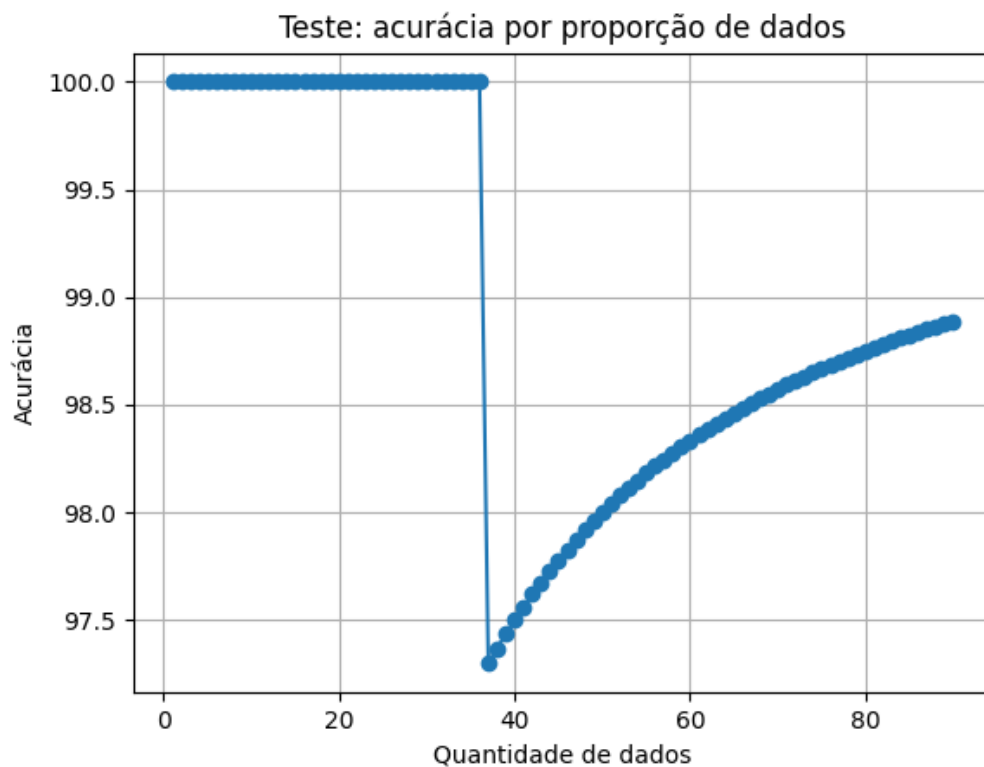


Figura 1 – Base Iris10 - relação entre acurácia e conjunto de dados para um total da população inicial de 20

Já para o segundo teste, com o tamanho da população inicial sendo 30, a acurácia obtida foi de 98.88% sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 2.

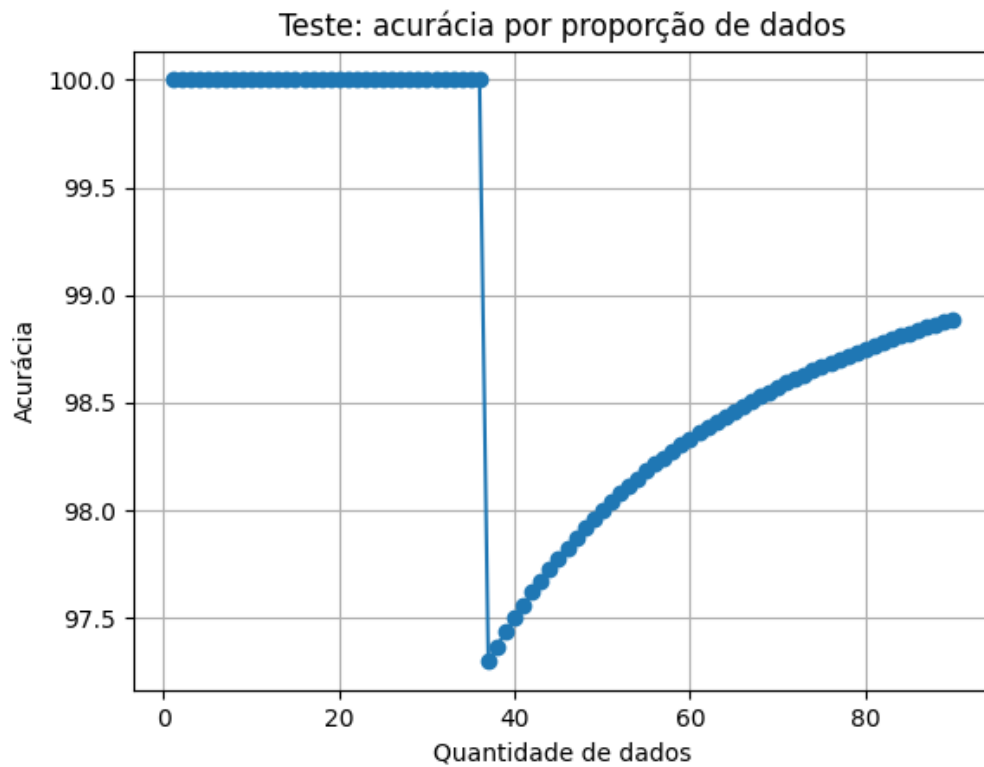


Figura 2 – Base Iris10 - relação entre acurácia e conjunto de dados para um total da população inicial de 30

Por último, foi considerada o tamanho da população inicial sendo 50 e a acurácia obtida foi de 100%, sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 3.

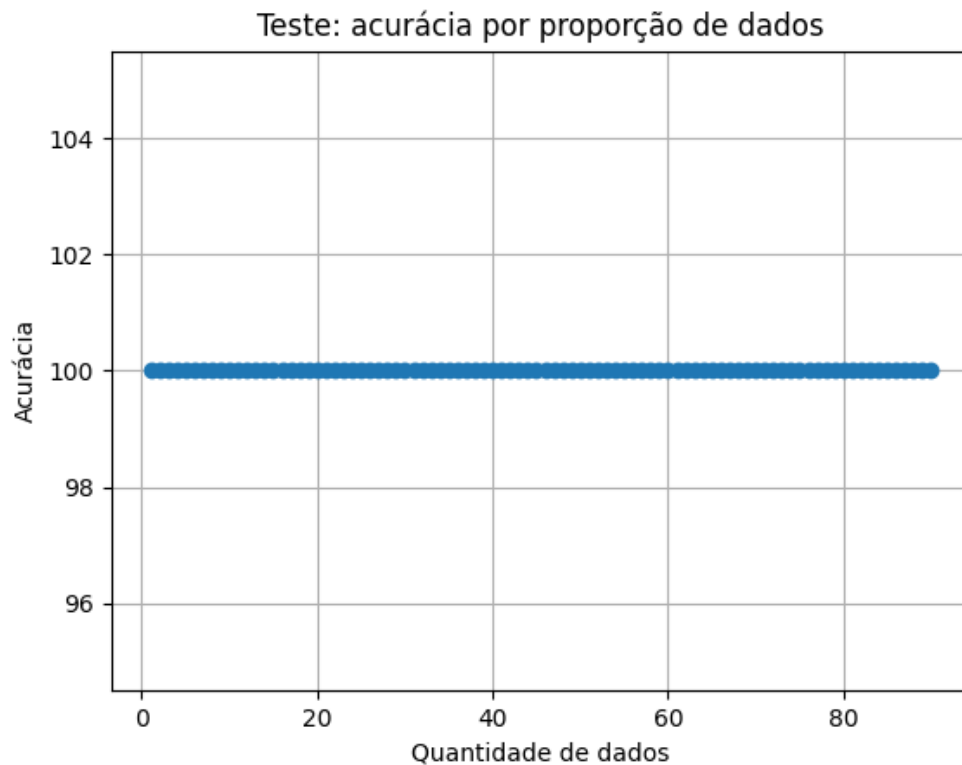


Figura 3 – Base Iris10 - relação entre acurácia e conjunto de dados para um total da população inicial de 50

3.2 Testes e resultados para o conjunto de dados Iris30

Para todos os testes, foi considerado um número constante de gerações sendo 15 e variações no tamanho da população inicial.

Para o primeiro teste, com o tamanho da população inicial sendo 20, a acurácia obtida foi de 94.28% sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 4.

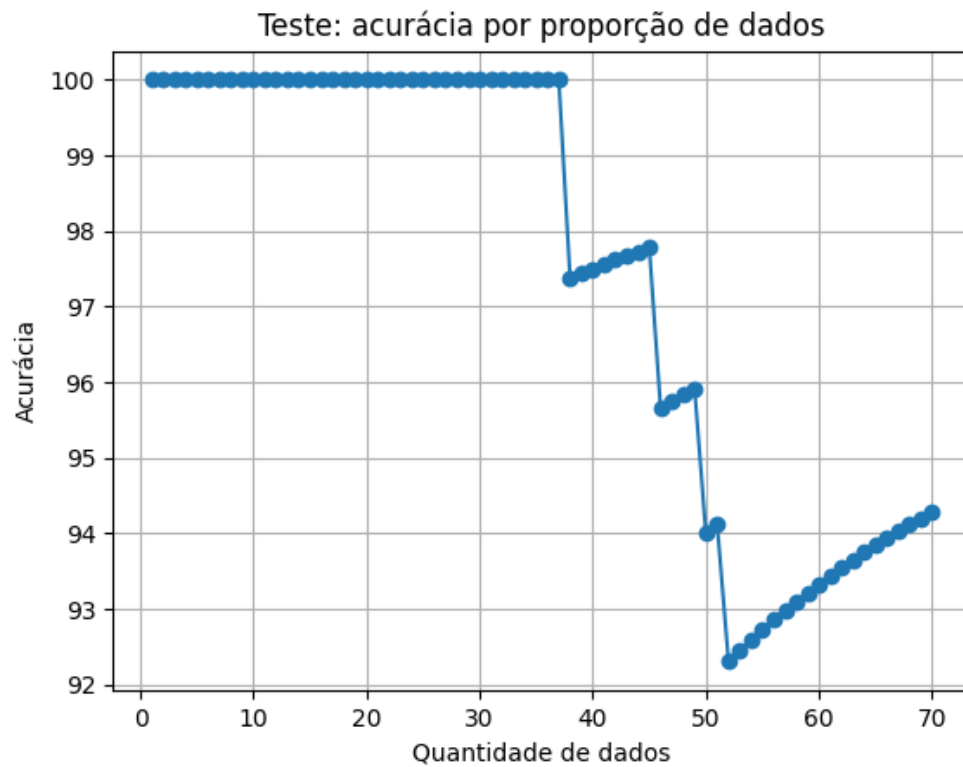


Figura 4 – Base Iris30 - relação entre acurácia e conjunto de dados para um total da população inicial de 20

Já para o segundo teste, com o tamanho da população inicial sendo 30, a acurácia obtida foi de 98.57% sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 5.

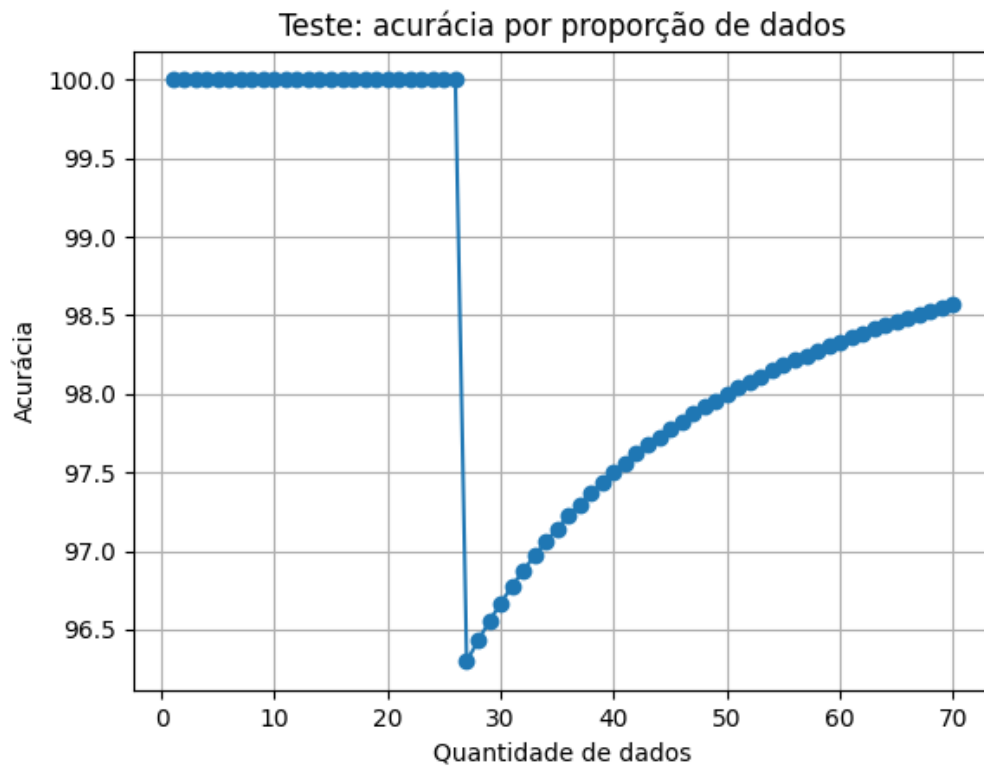


Figura 5 – Base Iris30 - relação entre acurácia e conjunto de dados para um total da população inicial de 30

Por fim, no terceiro teste, foi considerado o tamanho da população inicial sendo 50 e a acurácia obtida foi de 100%, sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 6.

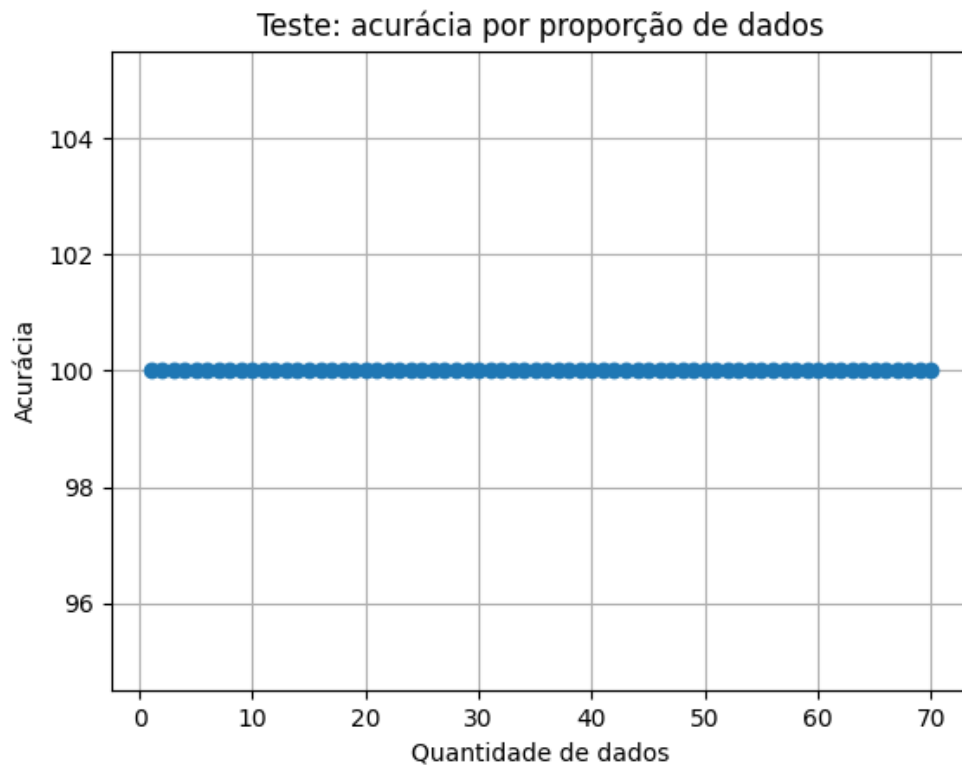


Figura 6 – Base Iris30 - relação entre acurácia e conjunto de dados para um total da população inicial de 50

3.3 Testes e resultados para o conjunto de dados Iris50

Para todos os testes, foi considerado um número constante de gerações sendo 10 e variações no tamanho da população inicial.

Para o primeiro teste, com o tamanho da população inicial sendo 20, a acurácia obtida foi de 94% sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 7.

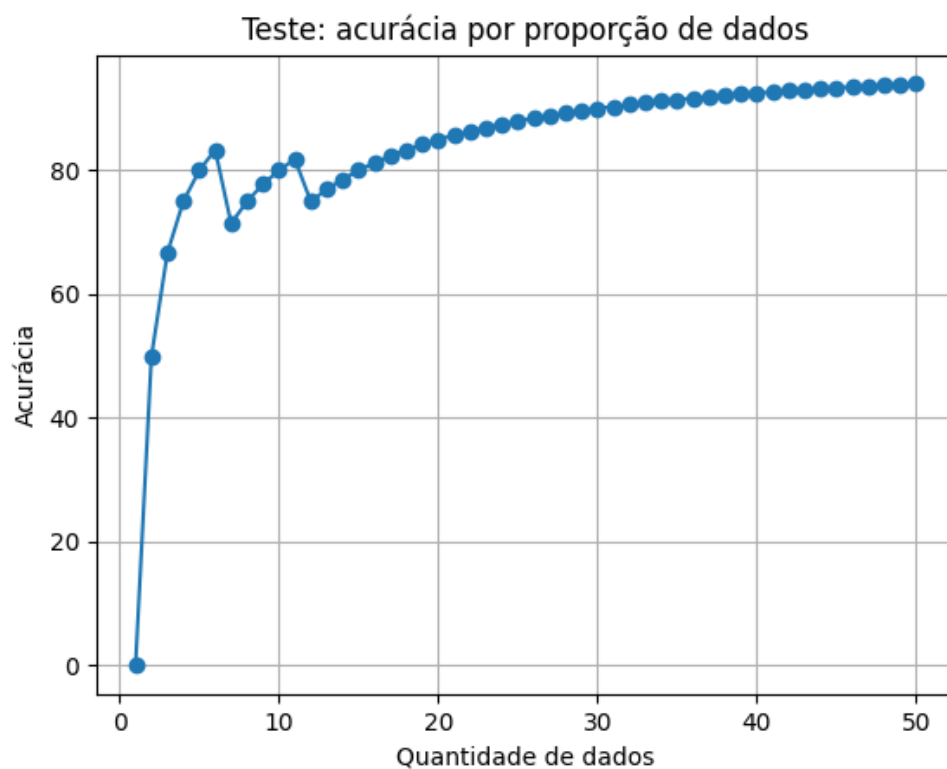


Figura 7 – Base Iris50 - relação entre acurácia e conjunto de dados para um total da população inicial de 20

Já para o segundo teste, com o tamanho da população inicial sendo 30, a acurácia obtida foi de 96% sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 8.

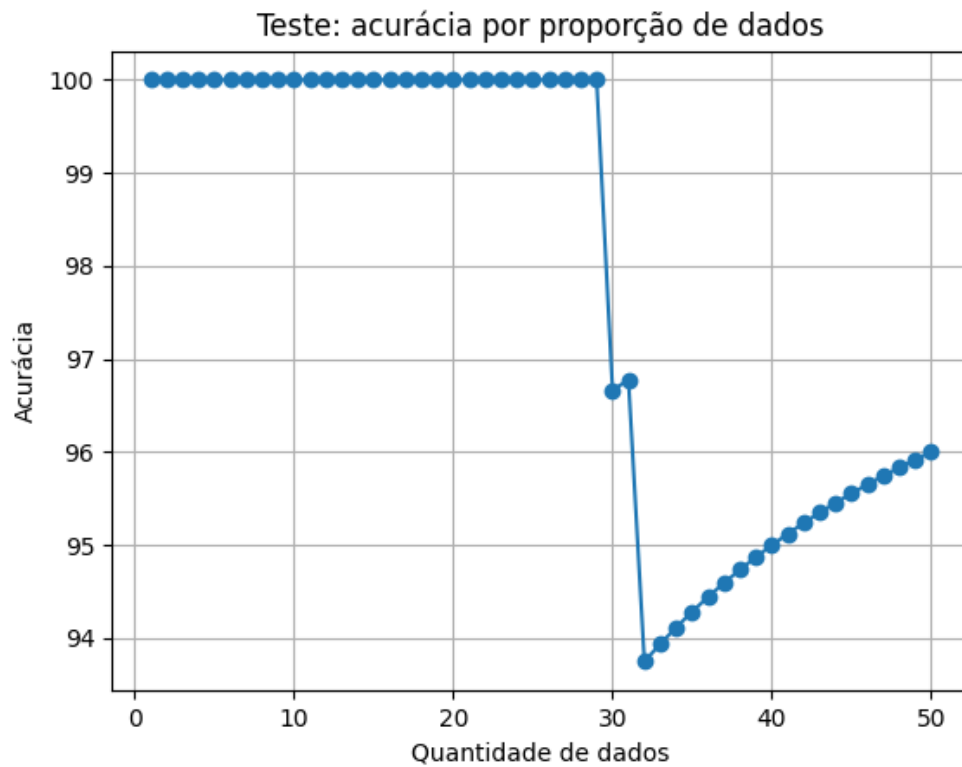


Figura 8 – Base Iris50 - relação entre acurácia e conjunto de dados para um total da população inicial de 30

Por último, foi considerada o tamanho da população inicial sendo 50 e a acurácia obtida foi de 100%, sendo possível observar a variação da acurácia em relação a proporção do conjunto de dados na figura 9.

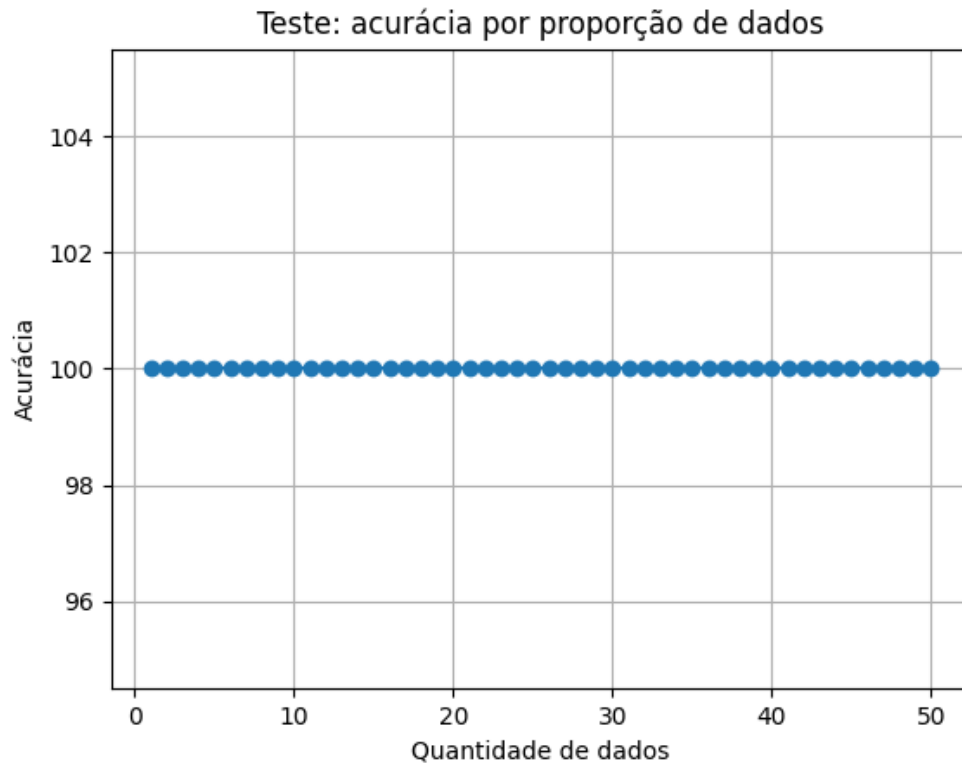


Figura 9 – Base Iris50 - relação entre acurácia e conjunto de dados para um total da população inicial de 50

3.4 Resultados e Discussões

Portanto, conforme pode ser observado nos resultados obtidos para cada teste realizado, a acurácia oscila em relação ao tamanho da população inicial, levando em consideração o contexto deste trabalho, mesmo alterando a quantidade de gerações. Além disso, com base nos testes realizados e nos resultados obtidos, quanto menor a quantidade da população inicial menos ajustes nos pesos são realizados e, desta forma, implica em uma acurácia menor. Em outras palavras, quanto menor a população inicial, mais baixa é a acurácia.

Além disso, ao inserir a terceira classe (Versicolour), toda a base para a classe Versicolour foi classificada como Virginica, o motivo desta classificação pode ser observado na figura 10. Tal classificação se dá devido as características em comum entre as duas classes.

Concluí-se, assim, que os resultados obtidos neste trabalho não divergem dos resultados obtidos no trabalho anterior de maneira significativa, visto que este apresenta resultados superiores para quando o tamanho da população inicial é maior, chegando a uma taxa de acurácia de 100%. Além disso, os resultados obtidos aqui demonstram uma con-

sistência em relação a variação dos parâmetros, o que não é possível observar no trabalho anterior. Outrossim, ao inserir a terceira classe, no trabalho anterior teve uma divergência em relação a quantidade classificada corretamente para a base de dados Iris10, o que não ocorreu neste, onde todos os registros foram classificados corretamente como Virginica.

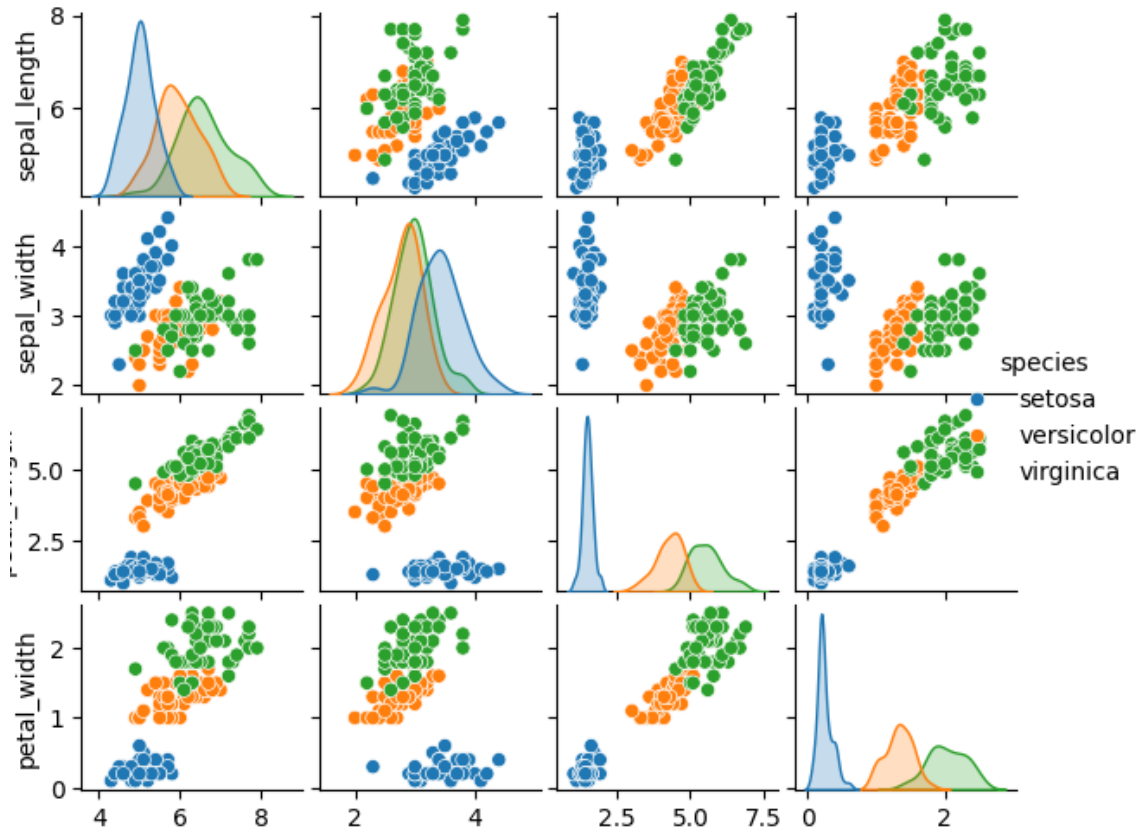


Figura 10 – Iris: características das flores e suas espécies