

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

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

**Implementação de um perceptron para
classificar dados da base Iris**

Uberlândia, Brasil

Lista de ilustrações

Figura 1 – Base Iris10 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 10%	5
Figura 2 – Base Iris10 - relação entre erro e epoch para taxa de aprendizado de 10%	6
Figura 3 – Base Iris10 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 20%	7
Figura 4 – Base Iris10 - relação entre erro e epoch para taxa de aprendizado de 20%	8
Figura 5 – Base Iris10 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 30%	9
Figura 6 – Base Iris10 - relação entre erro e epoch para taxa de aprendizado de 30%	10
Figura 7 – Base Iris30 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 10%	11
Figura 8 – Base Iris30 - relação entre erro e epoch para taxa de aprendizado de 10%	12
Figura 9 – Base Iris30 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 20%	13
Figura 10 – Base Iris30 - relação entre erro e epoch para taxa de aprendizado de 20%	14
Figura 11 – Base Iris30 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 30%	15
Figura 12 – Base Iris30 - relação entre erro e epoch para taxa de aprendizado de 30%	16
Figura 13 – Base Iris50 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 10%	17
Figura 14 – Base Iris50 - relação entre erro e epoch para taxa de aprendizado de 10%	18
Figura 15 – Base Iris50 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 20%	19
Figura 16 – Base Iris50 - relação entre erro e epoch para taxa de aprendizado de 20%	20
Figura 17 – Base Iris50 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 30%	21
Figura 18 – Base Iris50 - relação entre erro e epoch para taxa de aprendizado de 30%	22
Figura 19 – Iris: características das flores e suas espécies	23

1 Introdução

O presente relatório se refere à implementação de um Perceptron para a classificação de duas espécies da base de flores Iris. Diferentes valores para os parâmetros de taxa de aprendizado, épocas e proporção da base usada para o treinamento foram considerados. Além disso, foram realizados experimentos considerando as diversas combinações das 3 espécies disponíveis, com a terceira sendo aplicada sobre o modelo já treinado e o resultado analisado.

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

Dadas as duas espécies que serão usadas para o treinamento do neurônio a primeira etapa realizada é a normalização dos dados por meio da técnica min-max. A normalização é aplicada nos dados de todas as três espécies, mas o MIN e o MAX são calculados apenas das duas espécies escolhidas.

2.2 Classe Perceptron

Uma classe foi criada para conter os métodos e as variáveis associadas ao Perceptron. Os pesos (incluindo o bias) são armazenados em um vetor de Floats, o qual é inicializado aleatoriamente com números entre -1 e 1 assim que o objeto da classe é criado.

2.3 Função de Ativação

Para este trabalho simples escolhemos a função de ativação sigmóide. Caso seu retorno seja menor ou igual que 0.5 o neurônio classifica como a primeira espécie passado, e se for maior que 0.5 classifica como a segunda espécie.

3 Metodologia dos Testes e Resultados

Para a execução dos testes foram escolhidas as classes Setosa e Virginica. 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 de treino, são realizados 3 testes variando os valores da taxa de aprendizado em relação a um valor constante da quantidade de iterações. Além disso, para cada conjunto de treino, ao final do treinamento é realizada a inserção da terceira e as análises são discutidas na seção 3.4(Análises sobre a inserção da terceira classe nos testes realizados).

Abaixo são demonstrados os testes realizados em relação ao conjunto de treino, com gráficos para ilustração das variações dos erros em relação ao epoch, variações da acurácia em relação ao conjunto normalizado(proporção da base original) e os seus respectivos resultados. Por fim, as análises referentes a inserção da terceira classe são apresentadas.

3.1 Testes e resultados para o conjunto de dados Iris10

Para todos os testes, foi considerado um valor constante do total de iterações sendo 20 e variações da taxa de aprendizado.

Para o primeiro teste, com uma taxa de aprendizado de 10%. A acurácia obtida foi de 97.77% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 1 e a variação de erro por epoch pela figura 2.

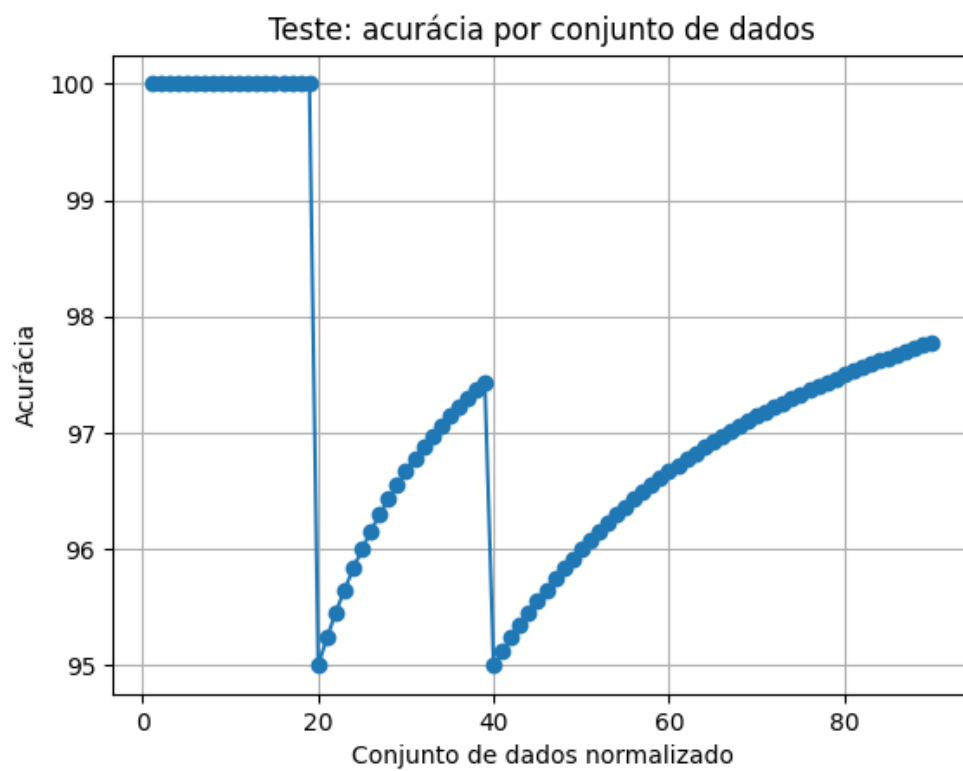


Figura 1 – Base Iris10 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 10%

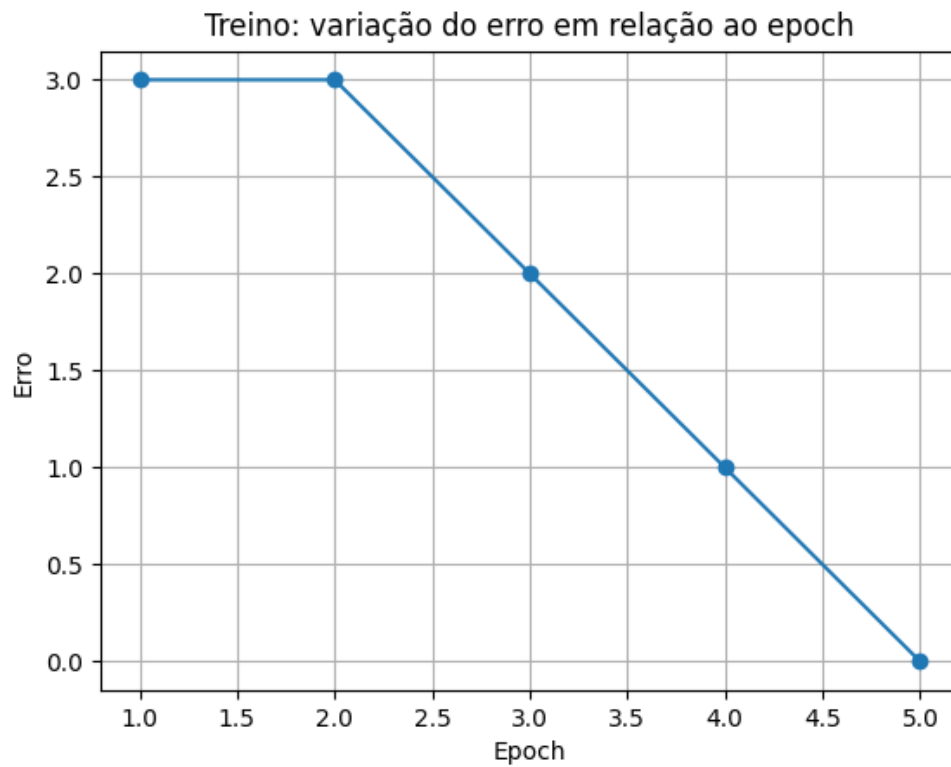


Figura 2 – Base Iris10 - relação entre erro e epoch para taxa de aprendizado de 10%

Já para o segundo teste, com uma taxa de aprendizado de 20%. A acurácia obtida foi de 84.44% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 3 e a variação de erro por epoch pela figura 4.



Figura 3 – Base Iris10 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 20%

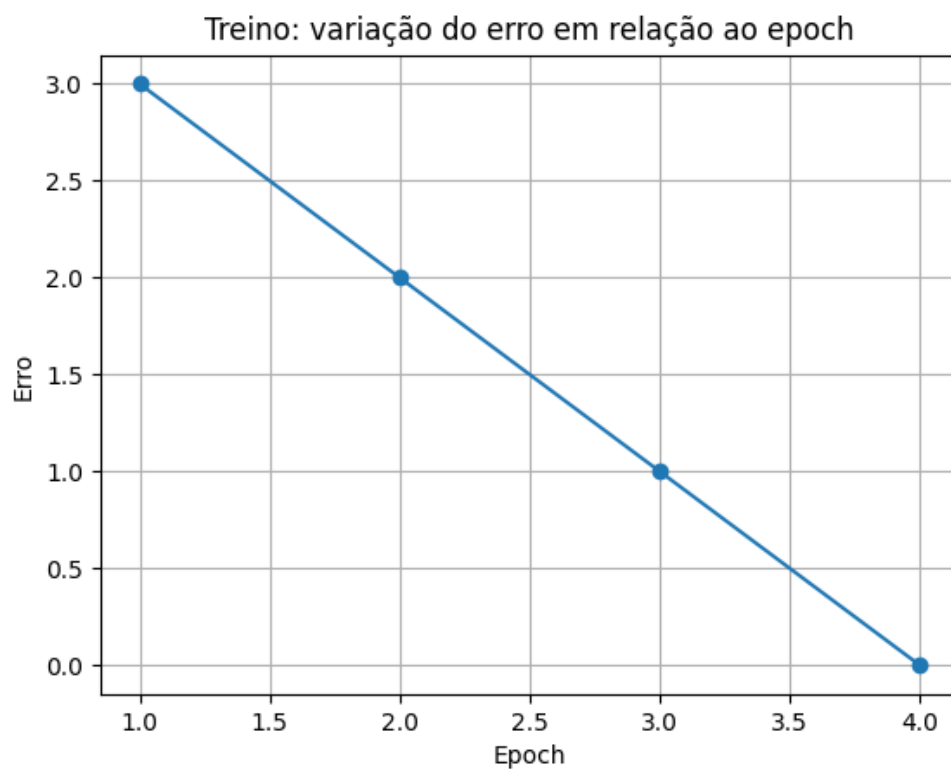


Figura 4 – Base Iris10 - relação entre erro e epoch para taxa de aprendizado de 20%

Por último, foi considerado uma taxa de aprendizado de 30%. A acurácia obtida foi de 86.66% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 5 e a variação de erro por epoch pela figura 6.

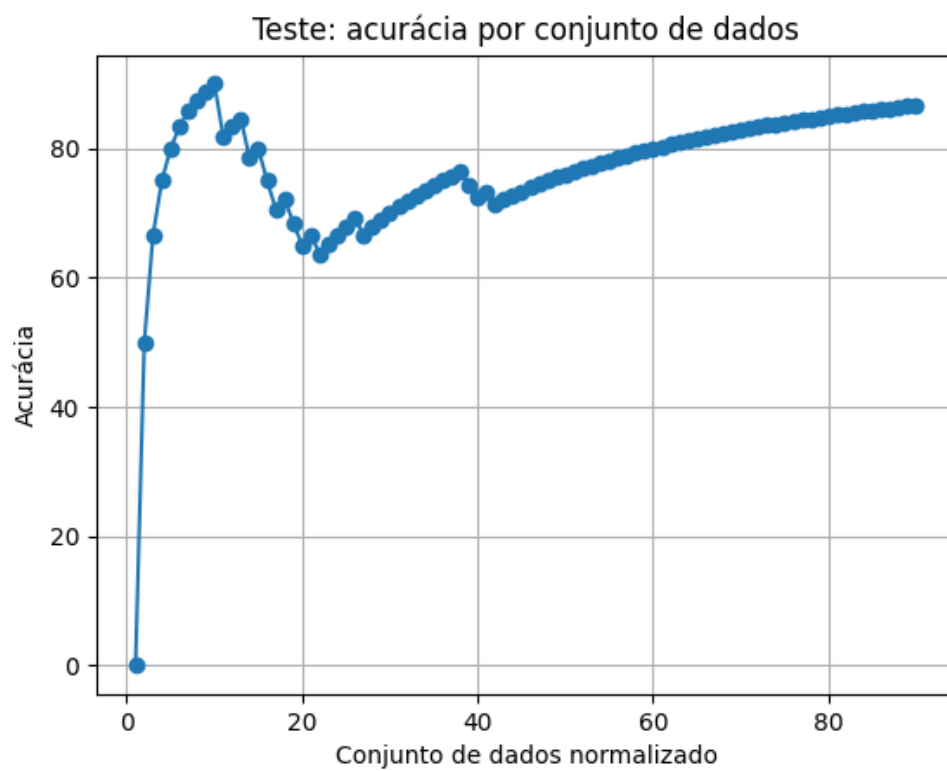


Figura 5 – Base Iris10 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 30%

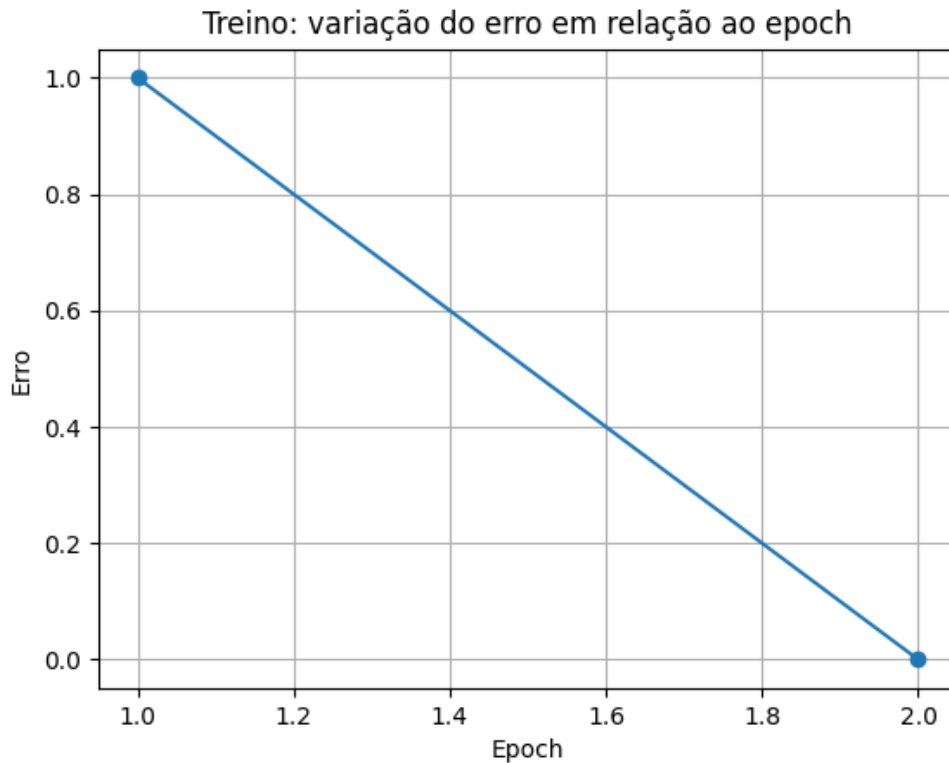


Figura 6 – Base Iris10 - relação entre erro e epoch para taxa de aprendizado de 30%

Portanto, tais oscilações no valor da acurácia deve-se ao tamanho da base e a forma como a taxa de aprendizagem afeta no treinamento, visto que quanto menor a taxa de aprendizagem, menos ajustes nos pesos serão realizados pelo modelo a cada iteração, explicando assim a maior acurácia para quando a taxa de aprendizagem é de 10%.

3.2 Testes e resultados para o conjunto de dados Iris30

Para todos os testes, foi considerado um valor constante do total de iterações sendo 50 e variações da taxa de aprendizado.

Para o primeiro teste, com uma taxa de aprendizado de 10%. A acurácia obtida foi de 94.28% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 7 e a variação de erro por epoch pela figura 8.

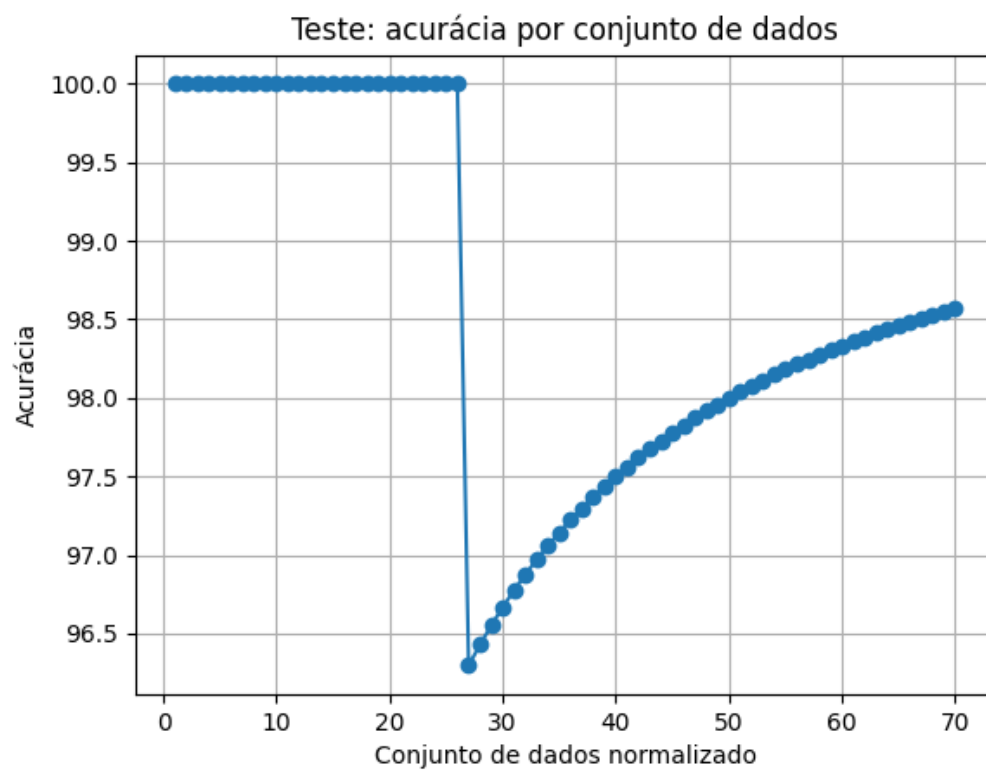


Figura 7 – Base Iris30 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 10%

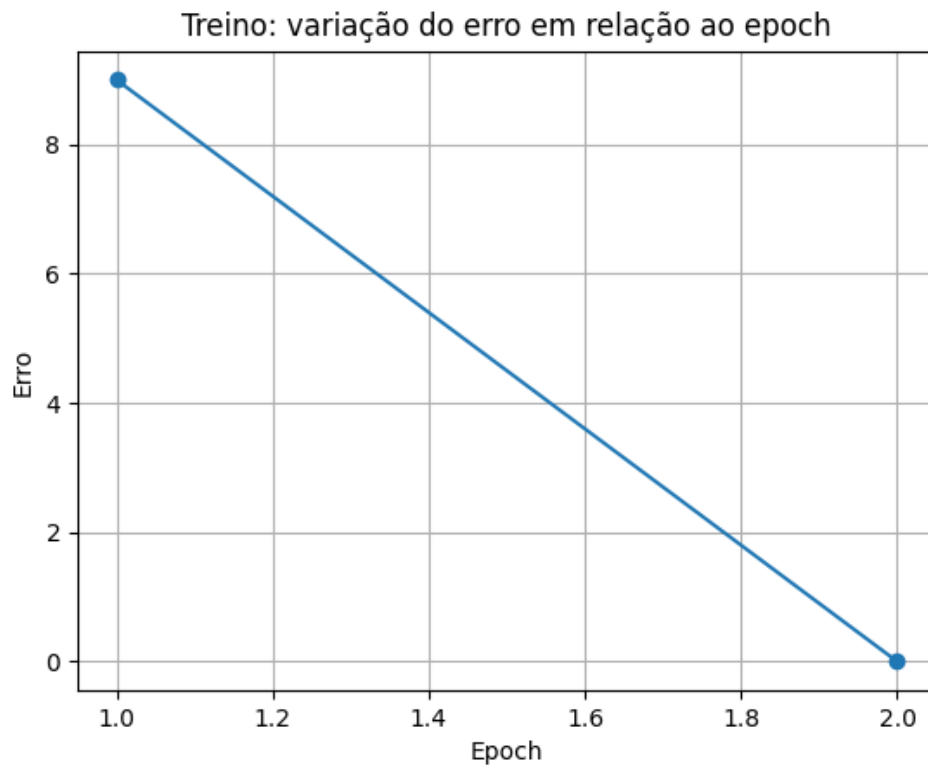


Figura 8 – Base Iris30 - relação entre erro e epoch para taxa de aprendizado de 10%

Já para o segundo teste, com uma taxa de aprendizado de 20%. A acurácia obtida foi de 97.14% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 9 e a variação de erro por epoch pela figura 10.

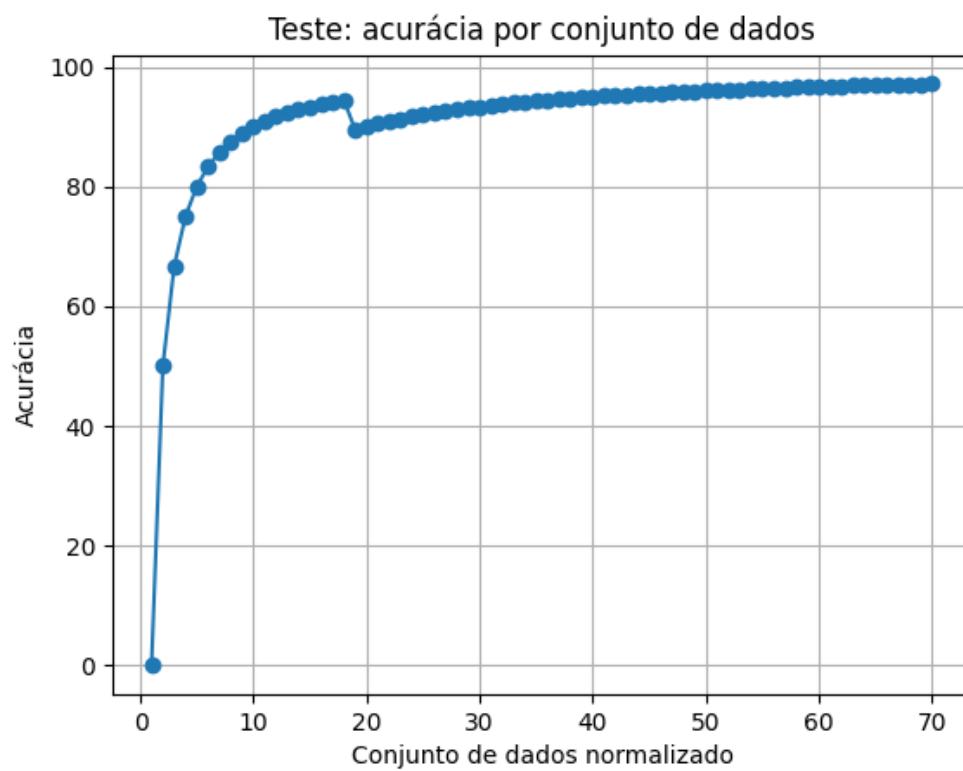


Figura 9 – Base Iris30 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 20%

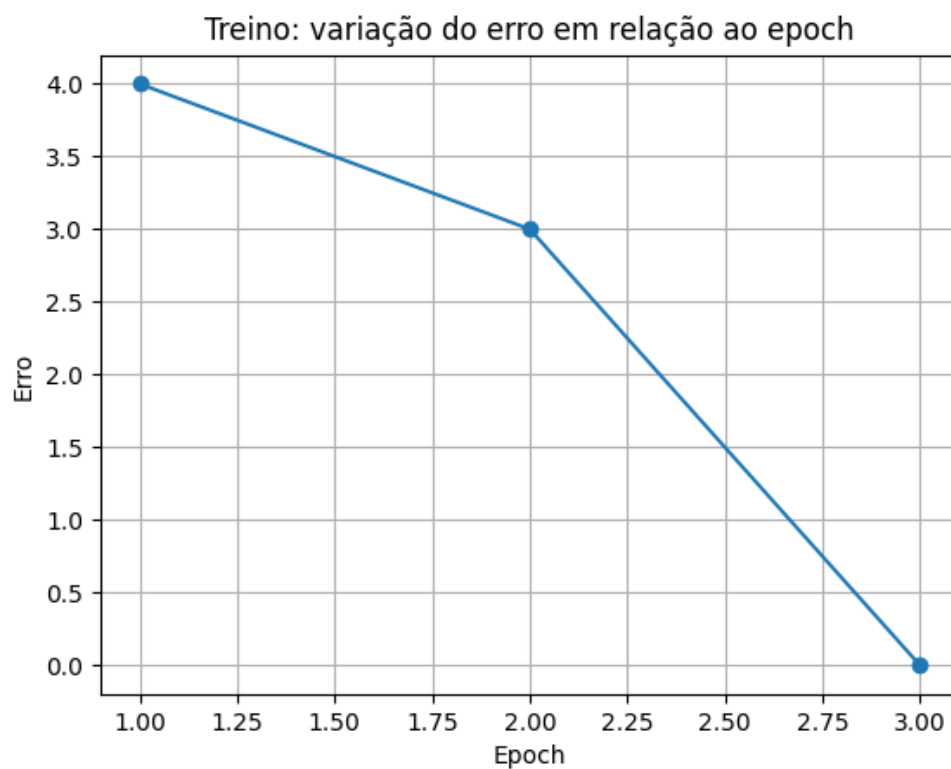


Figura 10 – Base Iris30 - relação entre erro e epoch para taxa de aprendizado de 20%

Por último, foi considerado uma taxa de aprendizado de 30%. A acurácia obtida foi de 98.57% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 11 e a variação de erro por epoch pela figura 12.

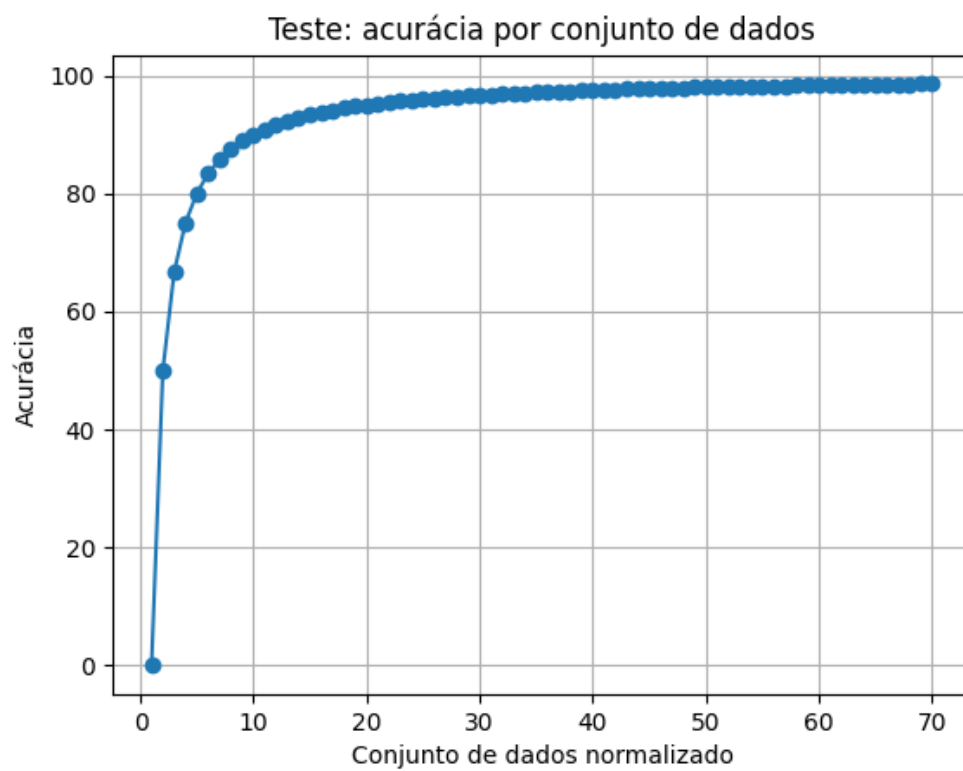


Figura 11 – Base Iris30 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 30%

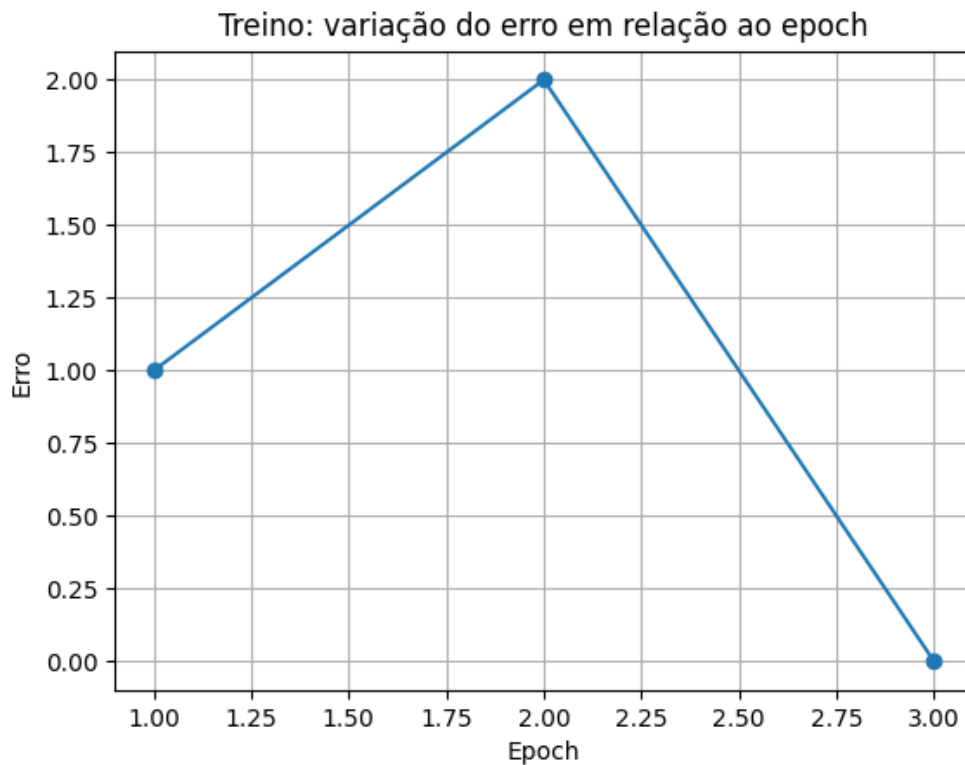


Figura 12 – Base Iris30 - relação entre erro e epoch para taxa de aprendizado de 30%

Assim, é possível concluir que a alteração do tamanho do conjunto de dados afetou a acurácia consideravelmente quando comparado aos resultados obtidos nos testes realizados com a base Iris10, além também das variações da taxa de aprendizagem e épocas. Logo, uma maior quantidade de dados ajuda o modelo a aprender padrões mais robustos e a generalizar melhor.

3.3 Testes e resultados para o conjunto de dados Iris50

Para todos os testes, foi considerado um valor constante do total de iterações sendo 10 e variações da taxa de aprendizado.

Para o primeiro teste, com uma taxa de aprendizado de 10%. A acurácia obtida foi de 98% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 13 e a variação de erro por epoch pela figura 14.

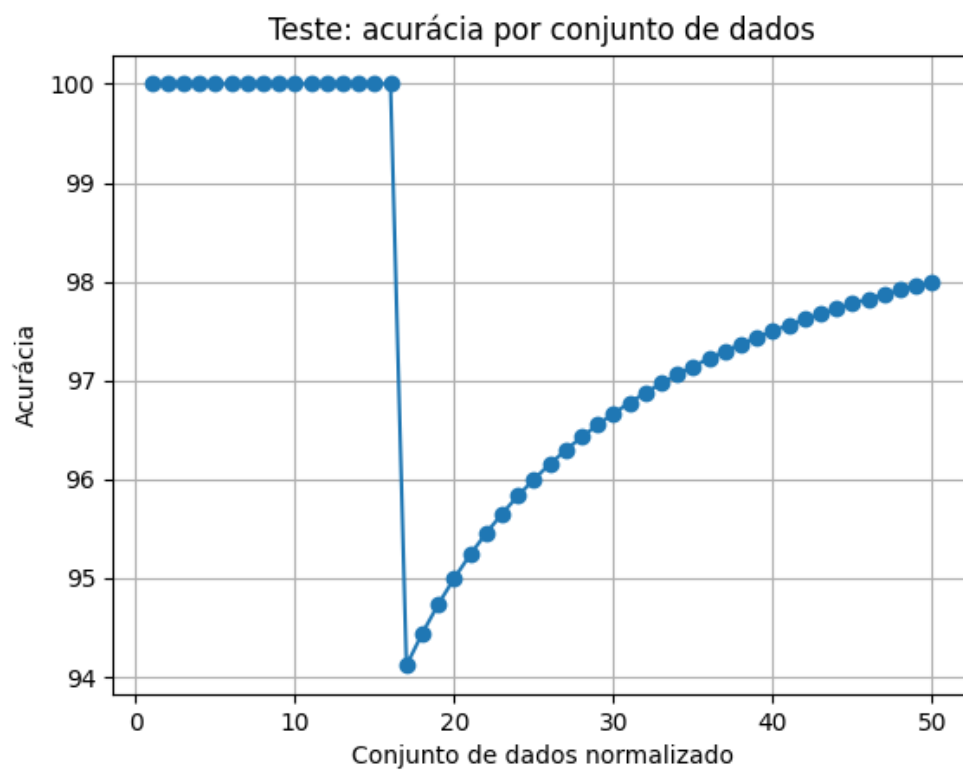


Figura 13 – Base Iris50 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 10%

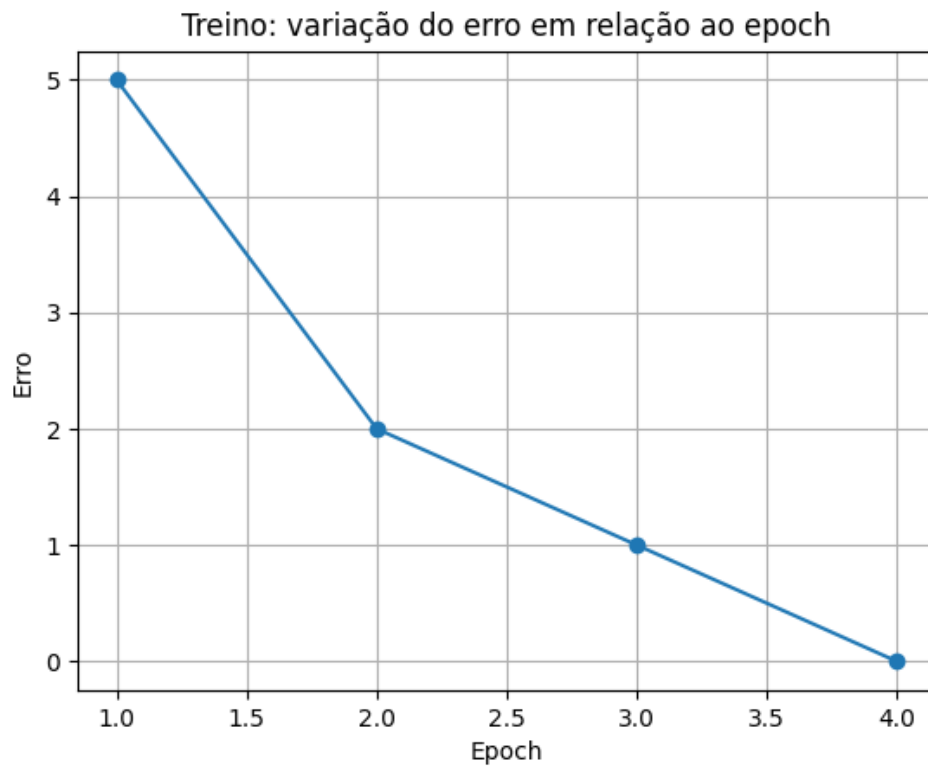


Figura 14 – Base Iris50 - relação entre erro e epoch para taxa de aprendizado de 10%

Já para o segundo teste, com uma taxa de aprendizado de 20%. A acurácia obtida foi de 98% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 15 e a variação de erro por epoch pela figura 16.



Figura 15 – Base Iris50 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 20%

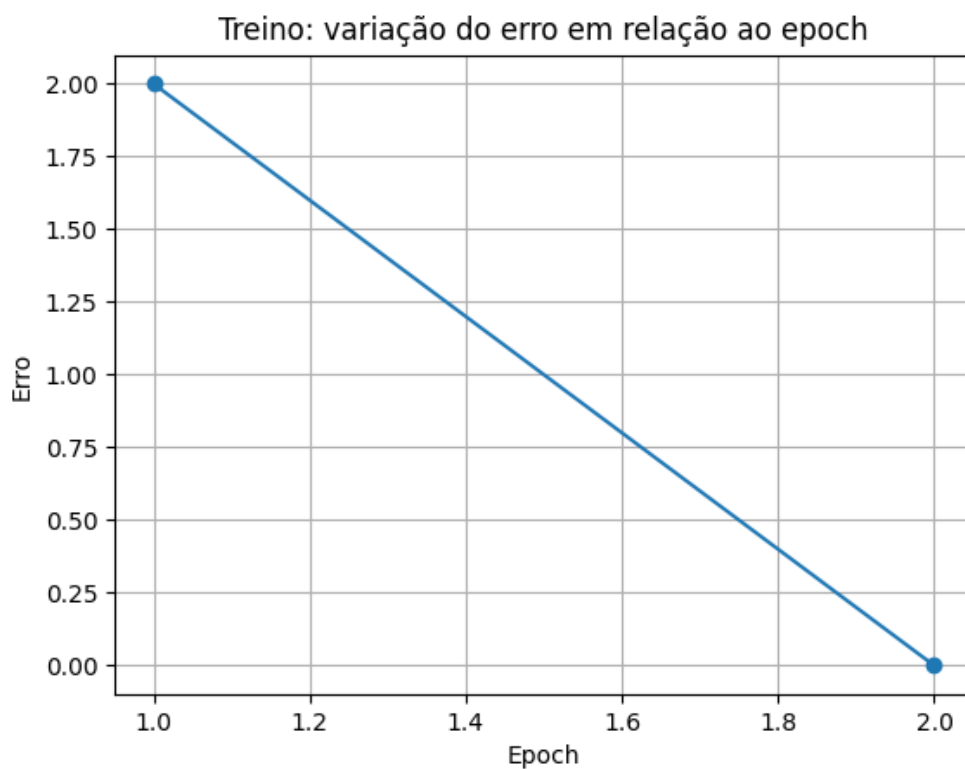


Figura 16 – Base Iris50 - relação entre erro e epoch para taxa de aprendizado de 20%

Por último, foi considerado uma taxa de aprendizado de 30%. A acurácia obtida foi de 98% sendo possível observar a variação da acurácia em relação ao conjunto de dados normalizado pela figura 17 e a variação de erro por epoch pela figura 18.

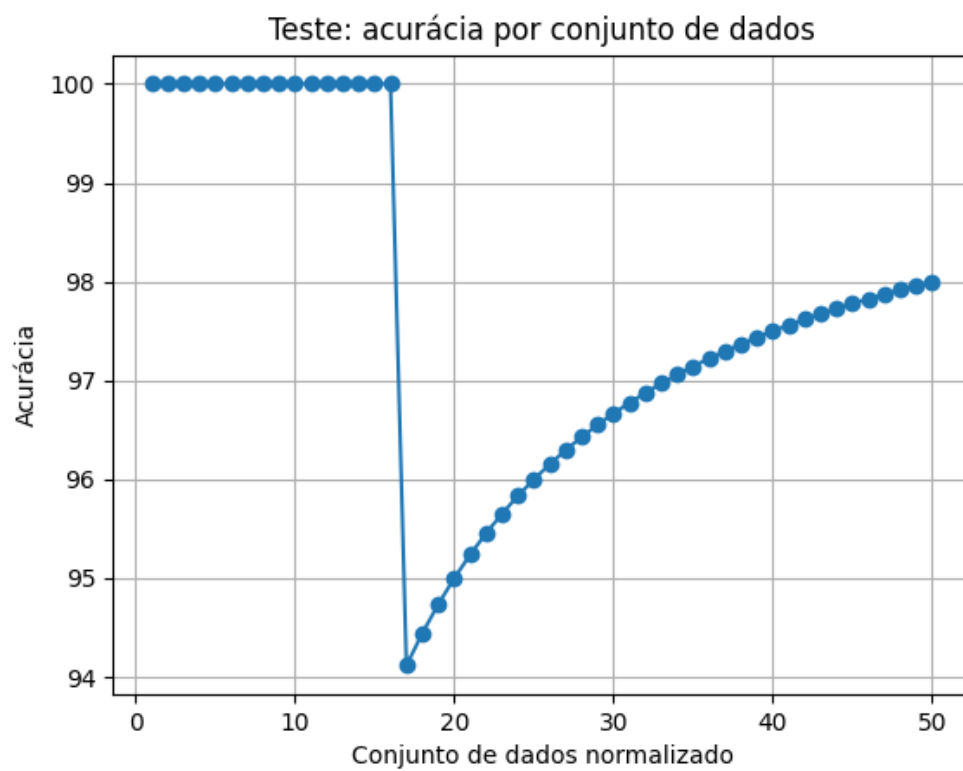


Figura 17 – Base Iris50 - relação entre acurácia e conjunto de dados normalizado para taxa de aprendizado de 30%

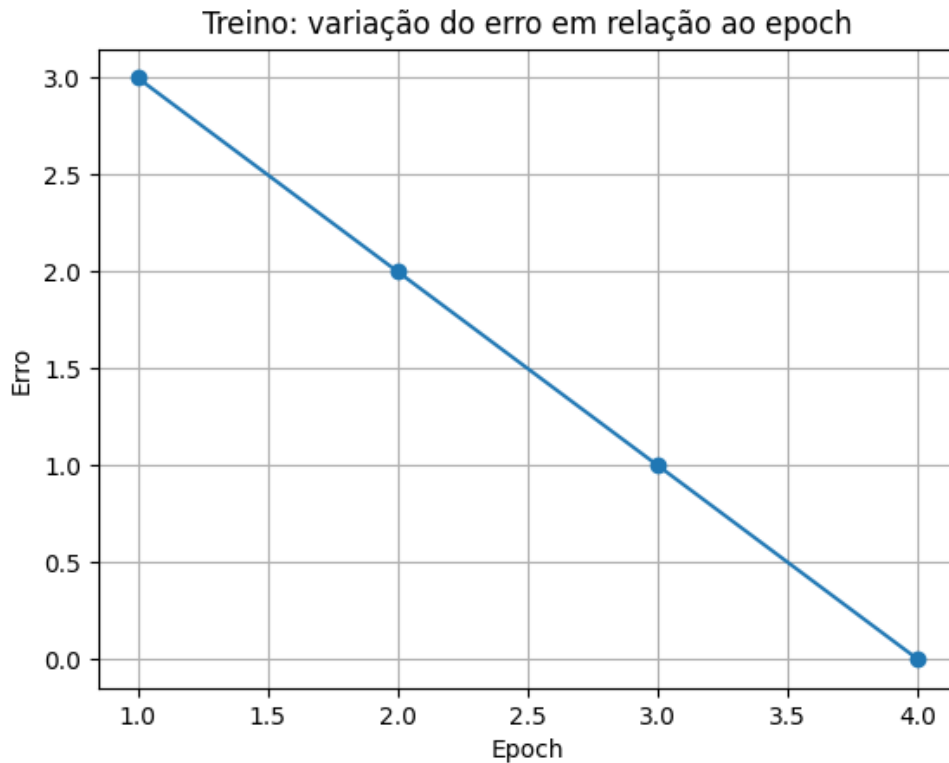


Figura 18 – Base Iris50 - relação entre erro e epoch para taxa de aprendizado de 30%

Sendo assim, é possível concluir que com um base de 50% obtém-se dados altamente informativos e bem distribuídos implicando assim em uma acurácia constante independente da taxa de aprendizagem.

3.4 Análises sobre a inserção da terceira classe nos testes realizados

Ao inserir a terceira classe(Versicolour), com exceção dos testes realizados com a base de treino Iris10 onde foram obtidos 44 registros classificados como Virginica e 6 como Sentosa para a taxa de aprendizagem de 10% e 43 registros classificados como Virginica e 7 como Sentosa para a taxa de aprendizagem de 20%, toda a base para a classe Versicolour foi classificada como Virginica, o motivo desta classificação pode ser observado na figura 19. Tal classificação se dá devido as características em comum entre as duas classes.

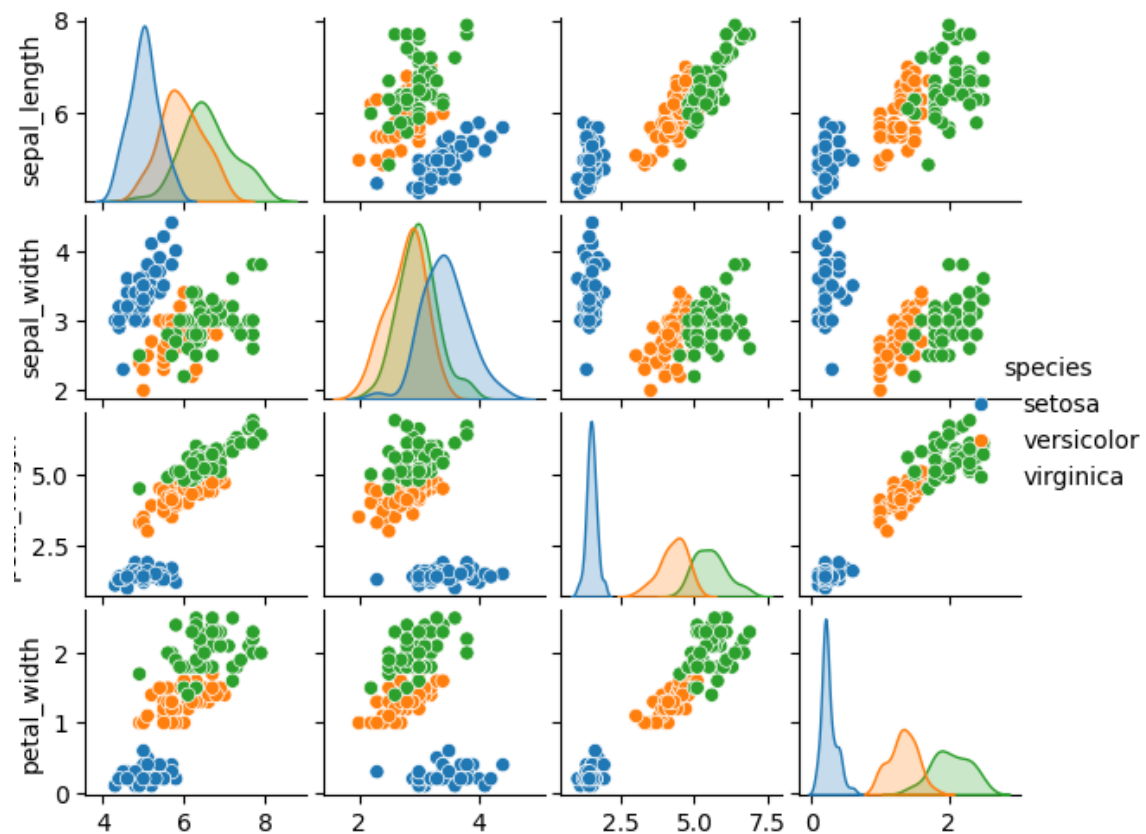


Figura 19 – Iris: características das flores e suas espécies