

Modelos de Classificação e Regressão

Logística

Desafio Final

1. Utilizando o data set iris (iris.csv em anexo ou usando sklearn):

Exemplo de como fazer a importação do dataset usando o sklearn

```
from sklearn import datasets  
iris = datasets.load_iris()
```

Responda:

- Faça uma análise inicial sobre esse dataset
- Use o boxplot e o histograma para caracterizar as propriedades de cada uma das espécies existentes.
- Somente olhando esses gráficos, é possível afirmar que uma ou mais propriedades (Sepal_Length, Sepal_Width, Petal_Length, Petal_Width) são suficientes para distinguir as espécies?
- Aplique a regressão logística para avaliar o modelo de classificação.
- Calcule a acurácia, precisão e recall.
- Plote a matriz de confusão com matplotlib ou seaborn.

2. Utilizando o dataset `load_digits`. Exemplo de como fazer a importação do dataset usando o `sklearn`:

```
from sklearn.datasets import  
load_digits digits = load_digits()
```

Responda:

a. Faça uma análise inicial sobre esse dataset

- i.** Quantos dados possui?
- ii.** Existem dados nulos? Se sim quantos?
- iii.** Todos são dados numéricos ou existem colunas com dados categóricos?

b. Aplique a regressão logística para construir e avaliar o modelo de classificação.

c. Calcule a acurácia, precisão e recall com base no desempenho do modelo.

d. Plote a matriz de confusão dos resultados do modelo utilizando `matplotlib` ou `seaborn`.

e. Aplique a regressão logística para avaliar o modelo de classificação dos dígitos de 0 a 9 utilizando o conjunto de dados específico para esse problema (por exemplo, MNIST).

f. Calcule a acurácia, precisão e recall com base no desempenho do modelo para a classificação dos dígitos de 0 a 9.

g. Plote a matriz de confusão dos resultados da classificação dos dígitos de 0 a 9 utilizando `matplotlib` ou `seaborn`.