

In 1] : !pip install pandas

Requirement already satisfied: pandas in c:\users\dsu-cse-513-013\anaconda3\lib\site-packages (2.2.2)
Requirement already satisfied: numpy>=1.26.0 in c:\users\dsu-cse-513-013\anaconda3\lib\site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\dsu-cse-513-013\anaconda3\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\dsu-cse-513-013\anaconda3\lib\site-packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\dsu-cse-513-013\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\dsu-cse-513-013\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

import pandas

print(pandas.__version__)

2.2.2

| C-1 : import pandas as pd

a=pd.readcsv("C:\\Users\\DSU-CSE-513-013\\Downloads\\Irisdatasample.csv",encodings"utf-8",n rows=20)
df=pd.DataFrame(a) df

Unnamed: 0 SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Species

0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	NaN	1.4	0.2	NaN
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	??	3.1	1.5	0.2	Iris-setosa
4	5	5	3.6	###	0.2	Iris-setosa
5	6	5.4	3.9	NaN	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	4.3	3.0	1.1	0.1	Iris-setosa
14	15	5.8	4.0	1.2	0.2	Iris-setosa
15	16	5.7	4.4	1.5	0.4	Iris-setosa
16	17	5.4	3.9	1.3	0.4	Iris-setosa
17	18	5.1	3.5	1.4	0.3	Iris-setosa
18	19	5.7	3.8	1.7	0.3	Iris-setosa
19	20	5.1	3.8	1.5	0.3	Iris-setosa

!J1 : import pandas as pd import
numpy as np import seaborn as
sns import matplotlib.pyplot as
plt

import pandas as pd
from sklearn.datasets import load_iris from sklearn.preprocessing import
StandardScaler from sklearn.model_selection import train_test_split from
sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

Load the Iris dataset iris = load_iris()

Create a DataFrame with the features and target
df = pd.DataFrame(data=iris.data, columns=iris.feature_names) df['target'] =
iris.target

Display the first 5 rows print("First 5 rows of the dataset:")

```
print(df.head())
```

First 5 rows of the dataset:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)\
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

target 0

```
0
1  0
2  0
3  0
4  0
```

```
In [3]: print("\n dataset information")
print(df.info())
```

dataset information <class 'pandas.core.frame.DataFrame'> RangeIndex: 150 entries, 0 to 149 Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	sepal length (cm)	150 non-null	float64
1	sepal width (cm)	150 non-null	float64
2	petal length (cm)	150 non-null	float64
3	petal width (cm)	150 non-null	float64
4	target	150 non-null	int32

dtypes: float64(4), int32(1)

memory usage: 5.4 KB

None

```
In [4]: print("\n summary statistics")
print(df.describe())
```

summary statistics			
	sepal length (cm)	sepal width (cm)	petal length (cm) \
count	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000
std	0.828066	0.435866	1.765298
min	4.300000	2.000000	1.000000
25%	5.100000	2.800000	1.600000
50%	5.800000	3.000000	4.350000
75%	6.400000	3.300000	5.100000
max	7.900000	4.400000	6.900000
	petal width (cm)	target	
count	150.000000	150.000000	
mean	1.199333	1.000000	
std	0.762238	0.819232	
min	0.100000	0.000000	
25%	0.300000	0.000000	
50%	1.300000	1.000000	
75%	1.800000	2.000000	
max	2.500000	2.000000	

```
In [25]: print("\n class distribution")
print(df['target'].value_counts())
```

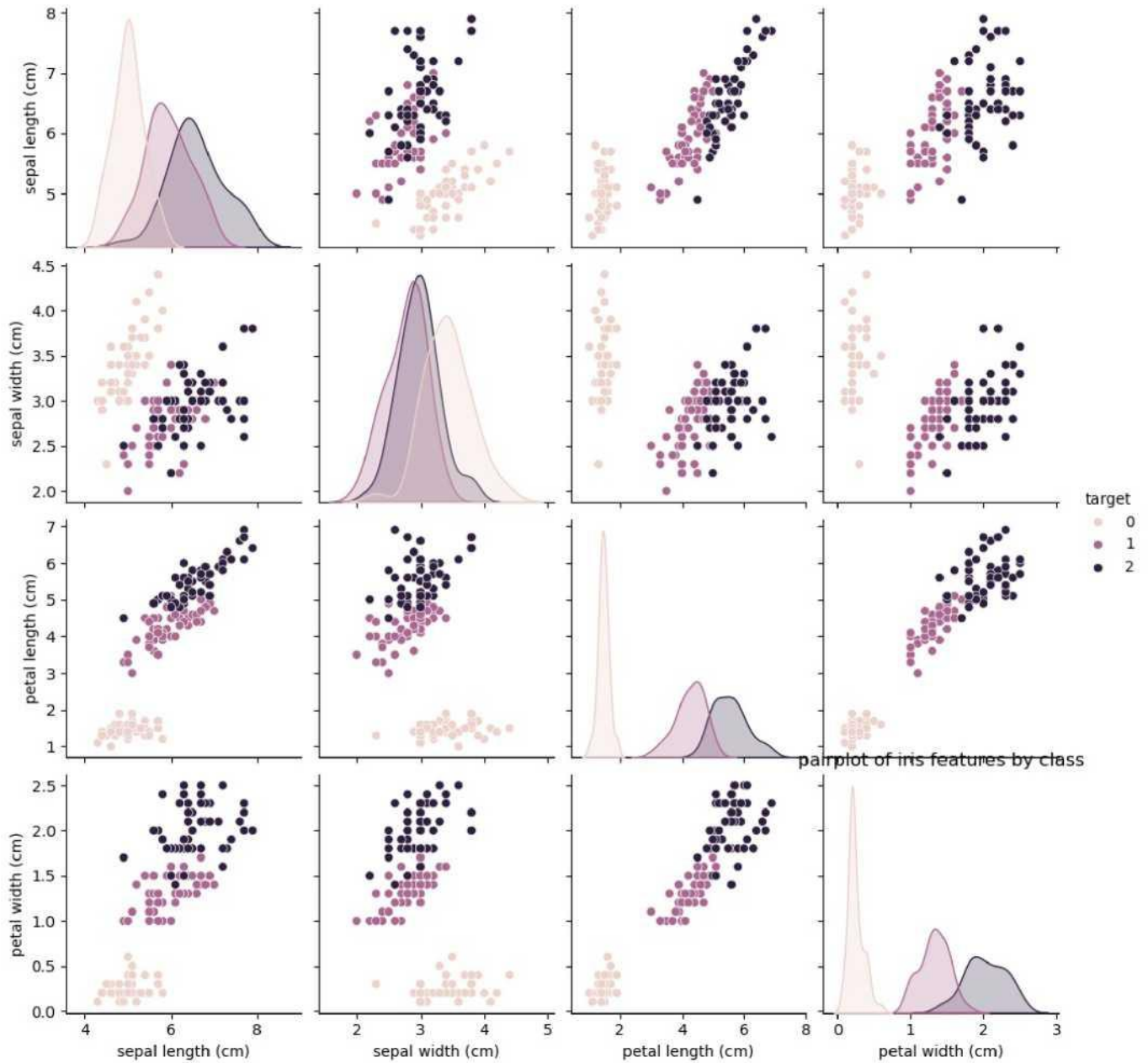
class distribution

target

```
0    50
1    50
2    50
```

Name: count, dtype: int64

```
27]: sns.pairplot(df, hue="target")
plt.title("pairplot of iris features by class")
plt.show()
```



In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js