nhn5dfexg

April 8, 2024

[1]:
```
# Aim: To perform Simple Linear Regression and Find out Coefficient of it.
```

[2]:
```
# Name: Mandar K Satpute
# Sub: Big Data Analytics (ET-Lab 2)
# Section : B
# Roll no: 54
```

[3]:
```
import os
```

[4]:
```
import pandas as pd
```

[5]:
```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import matplotlib.pyplot as plt
```

[6]:
```
from sklearn.linear_model import LogisticRegression  # for Logistic Regression␣
 ↪algorithm
from sklearn.model_selection import train_test_split
from sklearn.datasets import load_iris
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score
from sklearn.metrics import mean_squared_error, r2_score
```

[7]:
```
iris=load_iris()
X = iris.data  # Features
y = iris.target
dir(iris)
```

[7]:
```
['DESCR',
 'data',
 'data_module',
 'feature_names',
 'filename',
 'frame',
 'target',
 'target_names']
```

[8]:
```
os.getcwd()
```

[8]:
```
'C:\\Users\\hp\\Desktop\\BDA practicals(ET-2)'
```

[9]:
```
df=pd.read_csv("C://Users//hp//Desktop//IRIS.csv")
```

[10]:
```
df.head()
```

[10]:
```
   sepal_length  sepal_width  petal_length  petal_width      species
0           5.1          3.5           1.4          0.2  Iris-setosa
1           4.9          3.0           1.4          0.2  Iris-setosa
2           4.7          3.2           1.3          0.2  Iris-setosa
3           4.6          3.1           1.5          0.2  Iris-setosa
4           5.0          3.6           1.4          0.2  Iris-setosa
```

[11]:
```
df.tail()
```

[11]:
```
     sepal_length  sepal_width  petal_length  petal_width         species
145           6.7          3.0           5.2          2.3  Iris-virginica
146           6.3          2.5           5.0          1.9  Iris-virginica
147           6.5          3.0           5.2          2.0  Iris-virginica
148           6.2          3.4           5.4          2.3  Iris-virginica
149           5.9          3.0           5.1          1.8  Iris-virginica
```

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

[13]:
```
df.describe()
```

[13]:
```
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
min        4.300000     2.000000      1.000000     0.100000
25%        5.100000     2.800000      1.600000     0.300000
```

```
50%        5.800000     3.000000      4.350000     1.300000
75%        6.400000     3.300000      5.100000     1.800000
max        7.900000     4.400000      6.900000     2.500000
```

[14]:
```
df.isnull()
```

[14]:
```
     sepal_length  sepal_width  petal_length  petal_width  species
0           False        False         False        False    False
1           False        False         False        False    False
2           False        False         False        False    False
3           False        False         False        False    False
4           False        False         False        False    False
..            …          …           …          …      …
145         False        False         False        False    False
146         False        False         False        False    False
147         False        False         False        False    False
148         False        False         False        False    False
149         False        False         False        False    False

[150 rows x 5 columns]
```

[15]:
```
df.isna().sum()
```

[15]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

[16]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
 ↪random_state=40)
```

[17]:
```
model = LinearRegression()

# Train the model using the training sets
model.fit(X_train, y_train)

# Make predictions using the testing set
y_pred = model.predict(X_test)

# Coefficients
print('Coefficients:', model.coef_)
```

```
Coefficients: [-0.1502982  -0.04339123  0.25345042  0.58205165]
```

[18]:

```
mse = mean_squared_error(y_test, y_pred)
print('Mean squared error: %.2f' % mse)

# Calculate coefficient of determination (R^2 score)
r2 = r2_score(y_test, y_pred)
print('Coefficient of determination (R^2 score): %.2f' % r2)
```

```
Mean squared error: 0.04
Coefficient of determination (R^2 score): 0.94
```

[ ]:

```

```