

Linear Regression with Artificial Generated Dataset

Import Library

In [1]:

```
import pandas as pd
import numpy as np
```

Generate Dataset

In [2]:

```
from sklearn.datasets import make_regression
```

In [3]:

```
#without coeficient of underline model
x,y=make_regression(n_samples=5000,n_features=5,coef=False,bias=12,noise=10,random_state=2529)
```

In [4]:

```
#with coeficient of underline model
x,y,w=make_regression(n_samples=5000,n_features=5,coef=True,bias=12,noise=10,random_state=2529)
```

In [5]:

```
x.shape,y.shape
```

Out[5]:

```
((5000, 5), (5000,))
```

In [6]:

```
w #coefficients of x
```

Out[6]:

```
array([88.62484371, 84.86464695, 33.71253849, 78.0071541 , 30.9016715 ])
```

In [8]:

```
x.shape,y.shape
```

Out[8]:

```
((5000, 5), (5000,))
```

Get the First Five Rows of Targer Variable(y) and Features(x)

In [9]:

```
x[0:5]
```

Out[9]:

```
array([[ -1.10189822,  -0.15778588,  -1.39532267,  -1.28964247,  -0.09851181],
       [ -1.51051825,  -0.36667239,  -0.52617689,   2.05966724,   0.31605398],
       [  0.09163246,  -0.84170091,   1.27986583,  -0.92279759,  -1.60830728],
       [  0.82068382,  -0.38509437,   1.87463472,  -1.30607754,  -0.82246444],
       [  1.01174832,   0.46058375,   0.15592574,   0.06295633,   0.26270053]])
```

In [10]:

```
y[0:5]
```

Out[10]:

```
array([-254.36307479,  -6.95976655, -143.91305472,  -11.68742279,
        160.3922063  ])
```

Get Shape of DataFrame

In [11]:

```
x.shape,y.shape
```

Out[11]:

```
((5000, 5), (5000,))
```

Get Train Test Split

In [12]:

```
from sklearn.model_selection import train_test_split
```

In [13]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=2529)
```

In [14]:

```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

Out[14]:

```
((3500, 5), (1500, 5), (3500,), (1500,))
```

Get Linear Regression Model Train

In [16]:

```
from sklearn.linear_model import LinearRegression
```

In [17]:

```
model=LinearRegression()
```

In [18]:

```
model.fit(x_train,y_train)
```

Out[18]:

```
LinearRegression()
```

Get Intercept and Coefficents

In [19]:

```
model.intercept_
```

Out[19]:

```
12.002962698571839
```

In [20]:

```
model.coef_
```

Out[20]:

```
array([88.52759919, 84.88346683, 33.54442312, 78.20238836, 31.12119717])
```

Get Model Prediction

In [22]:

```
y_pred=model.predict(x_test)
```

In [23]:

```
y_pred
```

Out[23]:

```
array([ 56.14844057,  -1.07452182, -216.38586159, ..., 128.12933609,
       -10.0993807 ,   14.75377678])
```

Get Model Evaluation

In [24]:

```
from sklearn.metrics import mean_squared_error,mean_absolute_error,mean_absolute_percentage_error,r2_score
```

In [26]:

```
mean_squared_error(y_test,y_pred)
```

Out[26]:

```
94.56240560250635
```

In [27]:

```
mean_absolute_error(y_test,y_pred)
```

Out[27]:

```
7.723506813122925
```

In [28]:

```
mean_absolute_percentage_error(y_test,y_pred)
```

Out[28]:

```
0.4218034871083074
```

In [29]:

```
r2_score(y_test,y_pred)
```

Out[29]:

```
0.9958664337826227
```