**2) Multiple Linear Regression –**

**Description:** In reality, there are multiple variables that predict the Co2emission. When more than one independent variable is present, the process is called multiple linear regression. For example, predicting co2emission using FUELCONSUMPTION\_COMB, EngineSize and Cylinders of cars. The good thing here is that Multiple Linear Regression is the extension of simple linear regression model.

**Code:**

import pandas as pd

import numpy as np

**# data import**

df = pd.read\_csv("data-genrator/FuelConsumptionCo2.csv")

cdf = df[['ENGINESIZE','CYLINDERS','FUELCONSUMPTION\_CITY','FUELCONSUMPTION\_HWY','FUELCONSUMPTION\_COMB','CO2EMISSIONS']]

**# split data**

msk = np.random.rand(len(df)) < 0.8

train = cdf[msk]

test = cdf[~msk]

from sklearn import linear\_model

regr = linear\_model.LinearRegression()

x = np.asanyarray(train[['ENGINESIZE','CYLINDERS','FUELCONSUMPTION\_COMB']])

y = np.asanyarray(train[['CO2EMISSIONS']])

regr.fit (x, y)

**# The coefficients**

print ('Coefficients: ', regr.coef\_)

**# testing**

y\_hat= regr.predict(test[['ENGINESIZE','CYLINDERS','FUELCONSUMPTION\_COMB']])

x = np.asanyarray(test[['ENGINESIZE','CYLINDERS','FUELCONSUMPTION\_COMB']])

y = np.asanyarray(test[['CO2EMISSIONS']])

print("Residual sum of squares: %.2f"

% np.mean((y\_hat - y) \*\* 2))

**# Explained variance score: 1 is perfect prediction**

print('Variance score: %.2f' % regr.score(x, y))

**Output:**

Coefficients: [[10.98416244 7.17093657 9.84521564]]

Residual sum of squares: 481.46

Variance score: 0.88