**Abdalla Ibrahim Ahmed**

**20201841**

**The code**

@author: babda

"""

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# load the data

df = pd.read\_csv('Regression task data.csv')

# cleaning

from sklearn.impute import SimpleImputer

imputer = SimpleImputer(missing\_values = np.NaN, strategy = "mean")

imputer = imputer.fit(df.iloc[:,:])

df.iloc[:,:] = imputer.transform(df.iloc[:,:])

#handeling outliers for X

Q1 = df.X.quantile(0.25)

Q3 = df.X.quantile(0.75)

Q2 = df.X.quantile(0.50)

IQR = Q3 - Q1

lower\_limit = Q1 - 1.5\*IQR

upper\_limit = Q3 + 1.5\*IQR

df\_no\_outlier\_X = df[(df.X>lower\_limit)&(df.X<upper\_limit)]

#handeling outliers for Y

Q1 = df.Y.quantile(0.25)

Q3 = df.Y.quantile(0.75)

Q2 = df.Y.quantile(0.50)

IQR = Q3 - Q1

lower\_limit = Q1 - 1.5\*IQR

upper\_limit = Q3 + 1.5\*IQR

df\_no\_outlier\_Y = df[(df.Y>lower\_limit)&(df.Y<upper\_limit)]

# split columns

y = df\_no\_outlier\_X.iloc[:,-1].values

X = df\_no\_outlier\_Y.iloc[:,0:1].values

# split to train and test

from sklearn.model\_selection import train\_test\_split

X\_train,X\_test , y\_train ,y\_test = train\_test\_split(X,y,test\_size=0.2,random\_state=0)

# reshape

X\_train= X\_train.reshape(-1, 1)

y\_train= y\_train.reshape(-1, 1)

X\_test= X\_test.reshape(-1,1)

from sklearn.linear\_model import LinearRegression

regressor = LinearRegression()

regressor.fit(X\_train,y\_train)

y\_pred = regressor.predict(X\_train)

y\_pred\_test = regressor.predict(X\_test)

# Visualization

plt.scatter(X\_train, y\_train, color='red')

plt.plot(X\_test, y\_pred\_test , color='blue')

plt.title(' pH and bicarbonate Graph (training set)')

plt.xlabel('pH ')

plt.ylabel(' Bicarbonate ')

plt.show()

plt.scatter(X\_test, y\_test, color='red')

plt.plot(X\_test, y\_pred\_test , color='blue')

plt.title(' pH and bicarbonate Graph(testing set)')

plt.xlabel('pH ')

plt.ylabel(' Bicarbonate ')

plt.show()

**Visualizations**

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated