INDIAN INSTITUTE OF TECHNOLOGY ROPAR

PROJECT REPORT

WEATHER PREDICTION SYSTEM

GE103 COURSE

INTRODUCTION TO PROGRAMMING

2019(SEMESTER-1)

UNDER THE GUIDANCE OF-

DR. SUDARSHAN IYENGAR

(PROFESSOR)

SUBMITTED BY-

RHYTHM JAIN

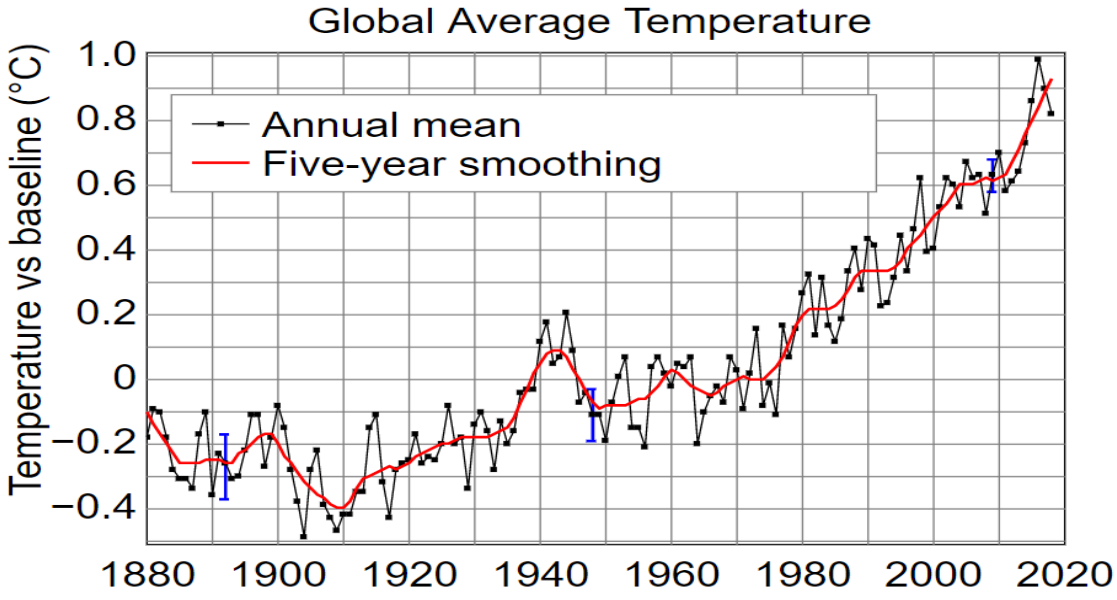
(2019CSB1111)

INSPIRATION:

As we all know that the Earth's climate is changing. Some of this change is due to natural variations that have been taking place for millions of years, but increasingly, human activities that release heat-trapping gases into the atmosphere are warming the planet by contributing to the "greenhouse effect."

As as result, the average temperature of the earth is increasing on an alarming rate year by year. Summers have record breaking temperature,every year.

Through my program,I want to spread awareness about climate change and that there is an urgent need for climate control.



INTRODUCTION:

I have developed a python program called Weather Prediction Program.

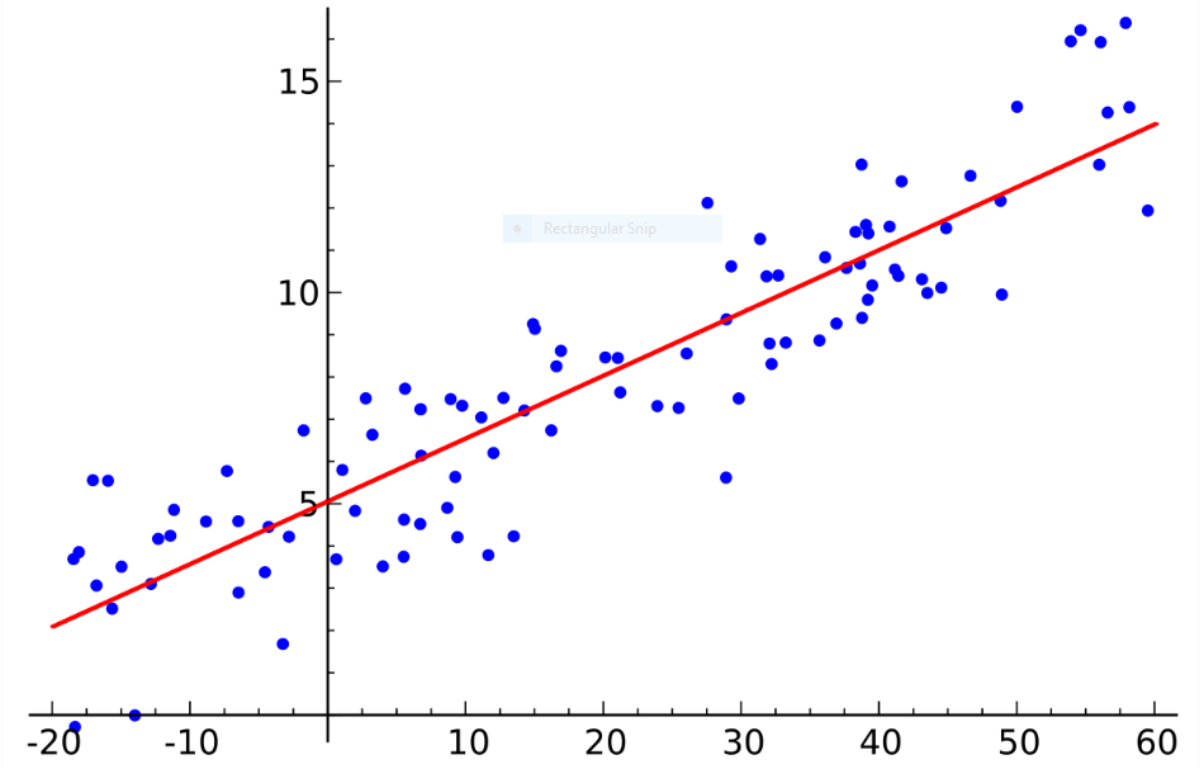
As the name suggests , the program can predict the average temperature of desired months and year .

This program uses Simple Linear Regression Machine Learning by analysing previous years data(1901 to 2017).

WORKING :

SIMPLE LINEAR REGRESSION:

Regression is a method of modelling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables.

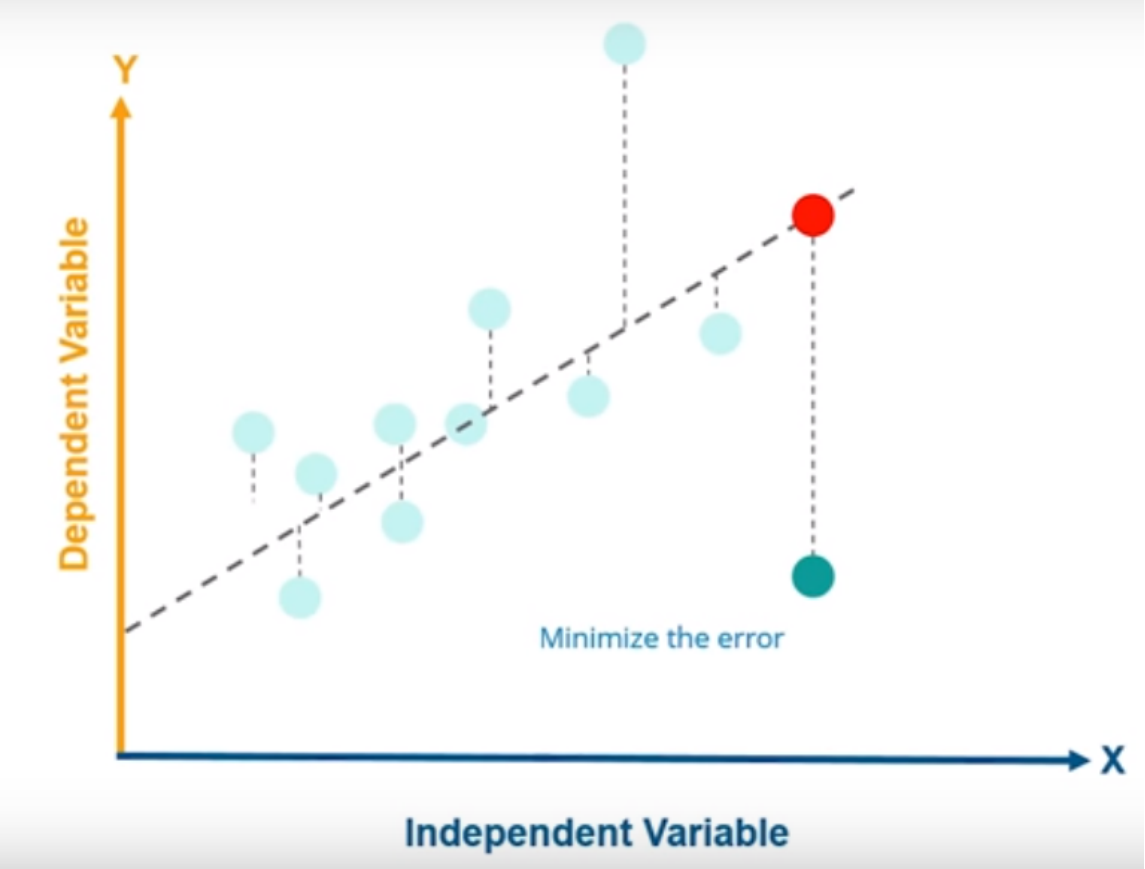


Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable. The red line in the above graph is referred to as the best fit straight line.

The line can be modelled based on the linear equation, y= mx + c

The core idea is to obtain a line that best fits the data. The best fit line is the one for which total prediction error (all data points) are as small as possible. The error is the distance between the point to the regression line.

Main aim is to minimize this error to obtain the most accurate value of m and c.



Caculations used:

Suppose we have array(X) and array(Y)

mean_x=mean of all the values of array(X)

mean_y=mean of all the values of array(Y)

num=$\sum$(X[i]-mean_x)(Y[i]-mean_y)

den=$\sum$(X[i]-mean_x)^2

m(slope)=num/den

c=mean_y – m(mean_x)

And,we get a best fit straight line.

MATPLOTLIB:

To plot graph,I have used ‘matplotlib library’ by importing ‘matplotlib.pyplot as plt’.

Matplotlib is a plotting library for
the Python programming language and its numerical mathematics extension NumPy.

matplotlib.pyplot is a plotting library used for 2D graphics in python programming language. It can be used

in python scripts, shell, web application servers and other graphical user interface toolkits.

NUMPY:

I have used numpy to make arrays for easy manipulation of data.

NumPy is a package in Python used for Scientific Computing. NumPy package is used to perform different operations. The ndarray (NumPy Array) is a multidimensional array used to store values of same datatype. These arrays are indexed just like Sequences, starts with zero

DICTIONARY:

I Have used dictionary to make my data well organised for easier retrieval of data.

Dictionary in Python is an unordered collection of data values, used to store data values like a map, which unlike other Data Types that hold only single value as an element, Dictionary holds key:value pair. Key value is provided in the dictionary to make it more optimized.

The Code

Firstly, I have opened my database and loaded the data in a dictionary for easier manipulation.

Secondly ,I have defined some functions which I have used later in my program.

I am attaching my screenshots of code from numpy.

```
import numpy as np

import matplotlib.pyplot as plt

import string as str

print("-------------------------------------------------------------------------------
print("Welcome to Future Weather Prediction.")
print("This program can predict future weather on the basis of previous years data.")
n=input("Press Enter to continue.")
print("\n")

dic={}

l1=[]



#Reading Dataset.
with open("Mean_temp.csv", "r") as file:

    data = file.readlines()
    for line in data:
        #print(line)
        word = line.split(',')
        #print(word)
        l1.append(word)


l2=l1[1:len(l1)]
```

```
#Creating Dictionary
for i in range(len(l1[0])):
    #print(l1[0][i])
    lnew=[]
    for j in range (len(l2)):
        lnew.append(l2[j][i])
    #print(lnew)
    dic[l1[0][i]]=lnew

#Function to calculate Mean.
def mean_list(X):

    j = 0
    for i in X:
        j=j+i
        ans=j/len(X)
    return(ans)


def bestfitline(x):

    #Making Array.
    X=np.array(dic['YEAR']).astype(np.float)
    Y=np.array(dic[x]).astype(np.float)

    # Mean
    mean_x=mean_list(X)
    mean_y=mean_list(Y)

    n=len(X)


    #Using the formula to calculate m and c.
    numerator=0
    denominator=0
    for i in range(n):
        numerator=numerator+(X[i]-mean_x)*(Y[i]-mean_y)
        denominator=denominator+(X[i]-mean_x)**(2)
    m=numerator/denominator
    c=mean_y-(m*mean_x)
```

```
    #Plotting values and regression line
    max_x=np.max(X) + 100
    min_x=np.min(X) - 100

    #Calculating line values x and y
    x=np.linspace(min_x,max_x,1000)
    y=c + m*x

    #Plotting line and graph
    fig = plt.figure()
    plt.plot(x,y,c='#0000A0',label="Regression Line")
    plt.scatter(X,Y,c='#ffa500',label="Scatter Plot")
    plt.xlabel("YEARS")
    plt.ylabel("Average Temperature")
    plt.legend()
    plt.draw()
    plt.pause(5)  # <-------
    raw_input=("<Hit Enter To Close>")
    plt.close(fig)

#Funtion that can find previous years data.
def prvs_temp(month,year):
    for i in range(len(dic['YEAR'])):
        if year==dic['YEAR'][i]:
            return(dic[month][i])
#Funtion that prints previous years data.
def if_1():

    print("Note that the the dataset has data from year 1901 to 2017.")
    print("\n")
    year = input("Enter the Year in 4 number format:")
    print("Year is", year, ".")
    print("\n")

    for yr in dic['YEAR']:

        if yr == year:
            month = input("Enter the first three letters of the month in capitals:")
            print("Month is", month, ".")
            print("\n")
            print("The average temperature for the month of", month, "in the year", year, "is",
                  prvs_temp(month, year),"Celcius.")
            print("----------------------------------------------------------------------------------------
```

```
#Function to predict temperature for a provided month and year
def pred_temp(month,year):
    #Making Array
    X=np.array(dic['YEAR']).astype(np.float)
    Y=np.array(dic[month]).astype(np.float)

    # Mean
    mean_x=mean_list(X)
    mean_y=mean_list(Y)
    n=len(X)
    #Using the formula to calculate m and c
    numerator=0
    denominator=0
    for i in range(n):
        numerator=numerator+(X[i]-mean_x)*(Y[i]-mean_y)
        denominator=denominator+(X[i]-mean_x)**(2)
    m=numerator/denominator
    c=mean_y-(m*mean_x)
    x=int(year)
    y=c+m*x
    return y
#Function to print predicted temperature.
def if_2():

    print("Note that the program can predict data upto year 2100,that is for 21 century only!")
    print("\n")

    year = input("Enter the Year in 4 number format:")
    if year>'2017':
        print("Year is", year, ".")
        print("\n")

        month = input("Enter the first three letters of the month in capitals:")
        print("Month is", month, ".")
        print("\n")
        print("The predicted average temperature for the month of", month, "in the year", year,
              "is", pred_temp(month, year),"Celcius.")
    else:
        print("Data already exists for year",year,".")
    print("-------------------------------------------------------------------------------------------")

#To take input from user about what he wants the program to do.
main_input=int(input("If you want to know previous year Average Temperature ,enter 1. "
                "\nIf you want to predict the Average Temperature ,enter 2. "
                "\nIf you want to plot the Graph of Average Temperature vs Year along "
                    "with the 'best fit line',enter 3:"))

print("Input is",main_input,".")
print("\n")

if main_input==1:
    if_1()

if main_input==2:
    if_2()

if main_input==3:
    month=input("Enter the first three letters of the month in capitals for which you "
                "want to get graph:")
    bestfitline(month)
```

RESULT:

We can use machine learning to predict weather conditions by analysing past year data.

There is a regular trend of increasing temperature yearly.

Linear regression is a simple and fun algorithm .