****

**DATA MINING**

**AND**

**DATA WAREHOUSING(Assignment-1)**

**by**

**-Bheemashankar 1PE16CS039**

Program:

|  |
| --- |
| #include <stdio.h> |
|  |

|  |
| --- |
| #include <stdlib.h> |
|  |

|  |
| --- |
| float x[]={1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23}; |
|  |

|  |
| --- |
| float y[]={13953.20,13842.35,13687.35, 13604.60, 13685.15,13647.45,13671.10,13649.05,13651.15,13622.45,13600.05,13751.60,13676.65,13694.25,13617.50,13537.25,13499.15,13542.90, 13509.90,13280.25,13107.75,13208.55,13272.25}; |
|  |

|  |
| --- |
| float out[2]; |
|  |

|  |
| --- |
| int i,j; |
|  |

|  |
| --- |
| float xnew,ynew; |
|  |

|  |
| --- |
| void linear\_reg(); |
|  |

|  |
| --- |
| void gradientDescent(); |
|  |

|  |
| --- |
| int sizex,sizey; |
|  |

|  |
| --- |
| float p; |
|  |

|  |
| --- |
| int n=sizeof(x)/sizeof(float); |
|  |

|  |
| --- |
| FILE \*ftr1,\*ftr2,\*ftr3; |
|  |

|  |
| --- |
| int main() |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| sizex=sizeof(x)/sizeof(float); |
|  |

|  |
| --- |
| sizey=sizeof(y)/sizeof(float); |
|  |

|  |
| --- |
| ftr3=fopen("original.xg","w"); |
|  |

|  |
| --- |
| for(i=0;i<n;i++) |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| fprintf(ftr3,"%.1f\t\t %.4f\n",x[i],y[i]); |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
| fclose(ftr3); |
|  |

|  |
| --- |
| printf("size of x=%d and y=%d\n",sizex,sizey); |
|  |

|  |
| --- |
| linear\_reg(); |
|  |

|  |
| --- |
| gradientDescent(); |
|  |

|  |
| --- |
| return 0; |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| void linear\_reg() |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| float m=0,b=0; |
|  |

|  |
| --- |
| float xsum = 0; |
|  |

|  |
| --- |
| float ysum = 0; |
|  |

|  |
| --- |
| float xmean,ymean,xval,yval; |
|  |

|  |
| --- |
| float num = 0,den = 0; |
|  |

|  |
| --- |
| for(i = 0; i < n; i++) { |
|  |

|  |
| --- |
| xsum += x[i]; |
|  |

|  |
| --- |
| ysum += y[i]; |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
| xmean = xsum /n; |
|  |

|  |
| --- |
| ymean = ysum /n; |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| for(i = 0; i < n; i++) { |
|  |

|  |
| --- |
| num += (x[i] - xmean) \* (y[i] - ymean); |
|  |

|  |
| --- |
| den += (x[i] - xmean) \* (x[i] - xmean); |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| m = num / den; |
|  |

|  |
| --- |
| b = ymean - (m \* xmean); |
|  |

|  |
| --- |
| printf("\n---------------------------------------------------------------------\n"); |
|  |

|  |
| --- |
| printf("The Line Equation after the linear regression is y=%.3fx+%.3f \n",m,b); |
|  |

|  |
| --- |
| printf("\nLinear Regression slope and intercept are:\n"); |
|  |

|  |
| --- |
| printf("\nSlope(m):%f",m); |
|  |

|  |
| --- |
| printf("\nIntercept(b):%f \n",b); |
|  |

|  |
| --- |
| printf("\n\nThe Predicted Opening Values from Nifty MNC Index Data\n\n"); |
|  |

|  |
| --- |
| p = m\*24+b; |
|  |

|  |
| --- |
| printf("1/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m\*25+b; |
|  |

|  |
| --- |
| printf("2/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m\*26+b; |
|  |

|  |
| --- |
| printf("3/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m\*27+b; |
|  |

|  |
| --- |
| printf("4/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m\*28+b; |
|  |

|  |
| --- |
| printf("5/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ftr1=fopen("linear\_reg.xg","w"); |
|  |

|  |
| --- |
| fprintf(ftr1,"TitleText: NSE\_predictions\_Linear Regression(Day vs NSE\_Opening)\nXUnitText: Days\nYUnitText: Opening\_Value\n\n\n"); |
|  |

|  |
| --- |
| fprintf(ftr1,"\"Original Data\"\n"); |
|  |

|  |
| --- |
| for(i=0;i<n;i++) |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| fprintf(ftr1,"%d\t\t %.4f\n",(i+1),y[i]); |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
| fprintf(ftr1,"\n\"Fit Line\"\n"); |
|  |

|  |
| --- |
| for(i=1;i<32;i++) |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| fprintf(ftr1,"%d\t\t %f\n",i,(m\*i+b)); |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
| fclose(ftr1); |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| void gradientDescent() { |
|  |

|  |
| --- |
| float xin,yin,guess; |
|  |

|  |
| --- |
| float m1=0.0; |
|  |

|  |
| --- |
| float b1=0.0; |
|  |

|  |
| --- |
| float error; |
|  |

|  |
| --- |
| float learning\_rate=0.002; |
|  |

|  |
| --- |
| for(j=0;j<1000;j++) |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| for (i = 0; i < n; i++) |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| xin = x[i]; |
|  |

|  |
| --- |
| yin = y[i]; |
|  |

|  |
| --- |
| guess = (m1 \* xin )+ b1; |
|  |

|  |
| --- |
| error = yin-guess; |
|  |

|  |
| --- |
| m1 += learning\_rate\*(error \* xin ); |
|  |

|  |
| --- |
| b1 += (learning\_rate\*error); |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
| printf("\n-------------------------------------------------------------------------------------\n"); |
|  |

|  |
| --- |
| printf("\nThe Line Equation after Gradient decent is y=%.3fx+%.3f \n",m1,b1); |
|  |

|  |
| --- |
| printf("\nGradient decent slope and intercept are:\n"); |
|  |

|  |
| --- |
| printf("\nSlope(m):%f",m1); |
|  |

|  |
| --- |
| printf("\nIntercept(b):%f \n",b1); |
|  |

|  |
| --- |
| printf("\n\nThe Predicted Opening Values from Nifty MNC Index Data\n\n"); |
|  |

|  |
| --- |
| p = m1\*24+b1; |
|  |

|  |
| --- |
| printf("1/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m1\*25+b1; |
|  |

|  |
| --- |
| printf("2/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m1\*26+b1; |
|  |

|  |
| --- |
| printf("3/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m1\*27+b1; |
|  |

|  |
| --- |
| printf("4/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
| p = m1\*28+b1; |
|  |

|  |
| --- |
| printf("5/Feb/2019:%f\n",p); |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ftr2=fopen("gradient\_decent.xg","w"); |
|  |

|  |
| --- |
| fprintf(ftr2,"TitleText: NSE\_predictions\_Gradient\_Decent(Day vs NSE\_Opening)\nXUnitText: Days\nYUnitText: Opening\_Value\n\n\n"); |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| fprintf(ftr2,"\"Original Data\"\n"); |
|  |

|  |
| --- |
| for(i=0;i<n;i++) |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| fprintf(ftr2,"%d\t\t %.4f\n",(i+1),y[i]); |
|  |

|  |
| --- |
| } |
|  |

|  |
| --- |
| fprintf(ftr2,"\n\"Fit Line\"\n"); |
|  |

|  |
| --- |
| for(i=1;i<32;i++) |
|  |

|  |
| --- |
| { |
|  |

|  |
| --- |
| fprintf(ftr2,"%d\t\t %f\n",i,(m1\*i+b1)); |
|  |

|  |
| --- |
| } |
|  |

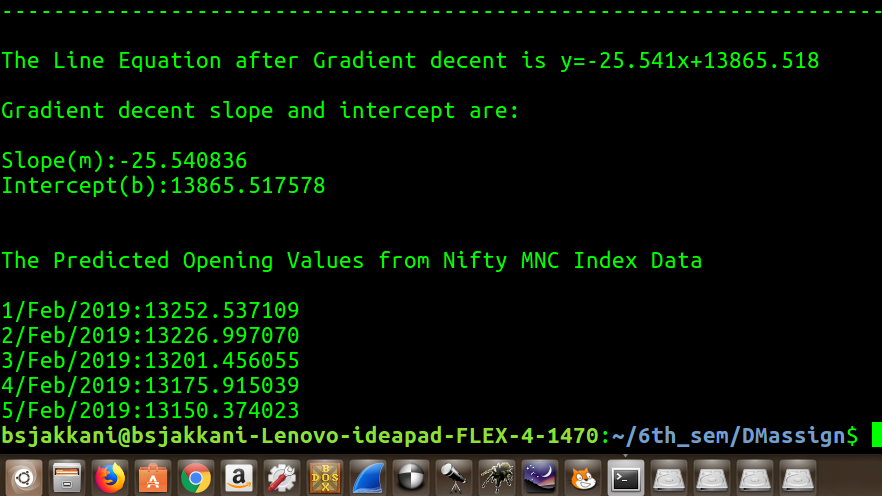
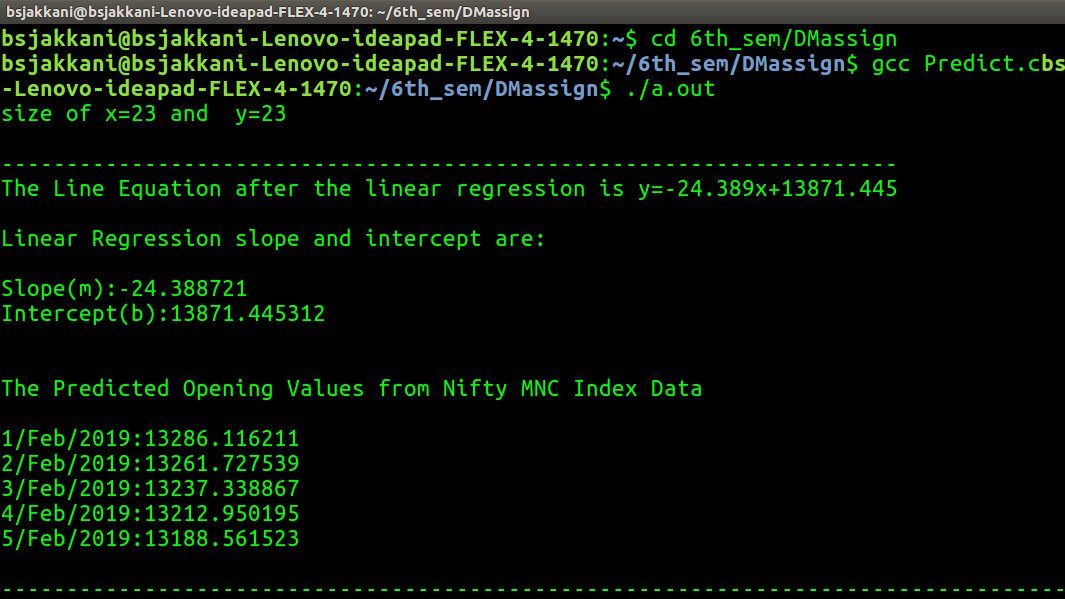
|  |
| --- |
| fclose(ftr2); |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

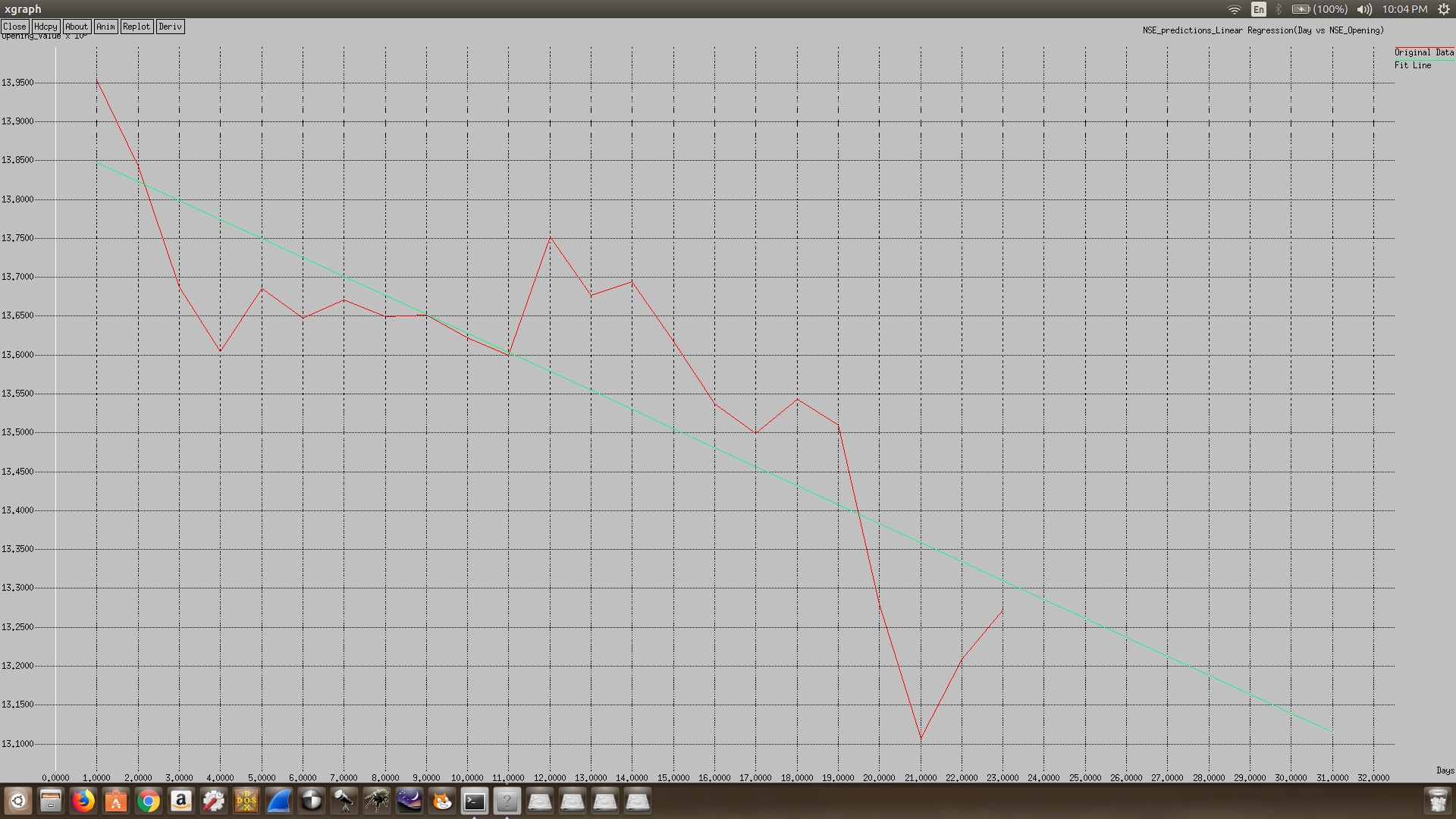
}

Output:

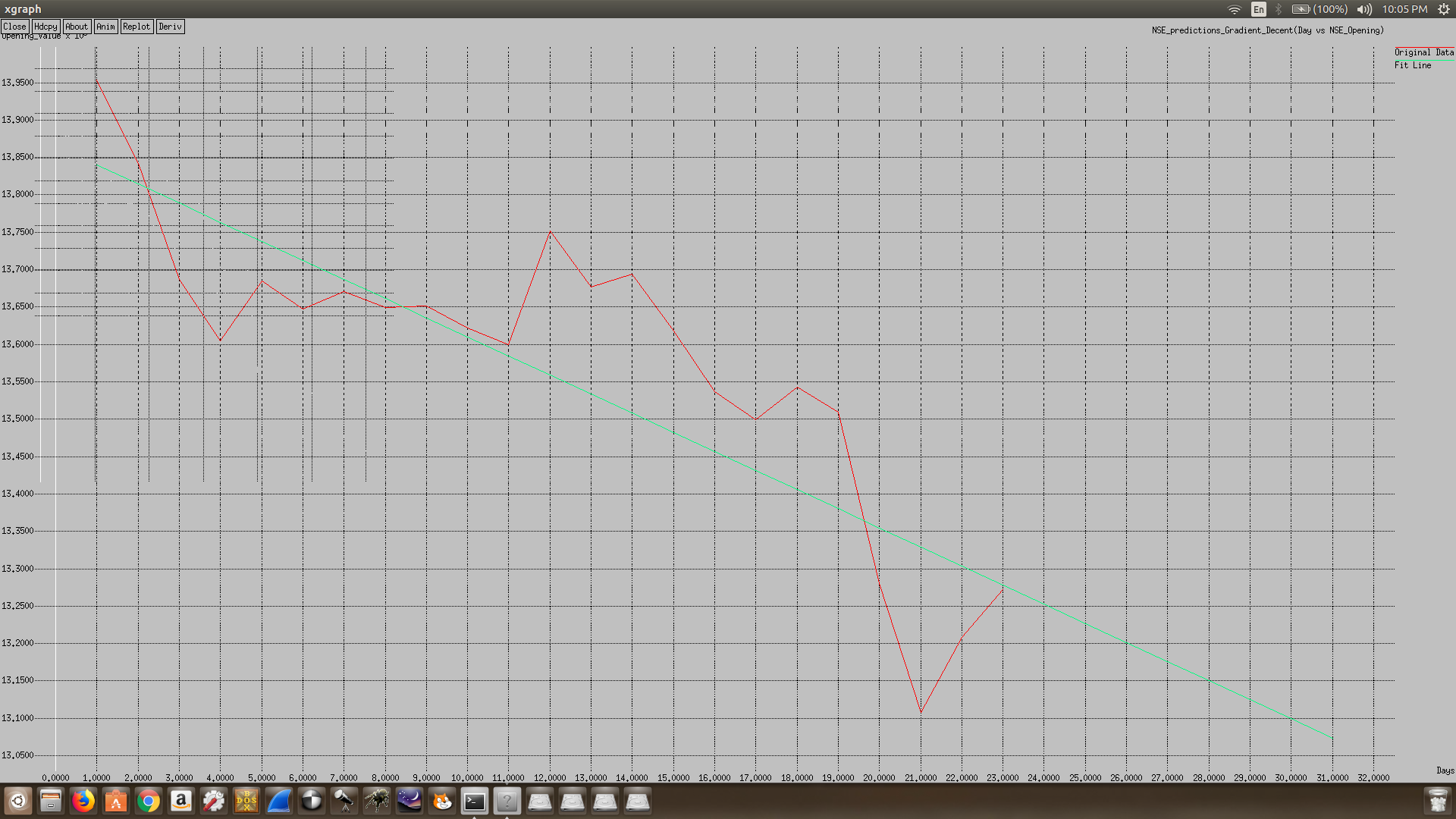


Graph:

1.Linear Regression



2.Gradient Descent



Problem Statement:

Use Nifty MNC NSE data for a month (from Jan 1st-Jan 31st, 2019) for one of the indices.

(i) Model the ‘opening’ value data with linear regression (i.e., estimate parameters) using

a. Gradient descent

b. Least squares approach and

compare the resulting parameters .The estimation procedure is to be coded up using C (no library functions are allowed) (ii) Use the estimated parameters to predict the ‘opening’ value on Feb 1st, 2019 and Feb 5th, 2019

Dataset:

