DATA MINING
AND
DATA WAREHOUSING(Assignment-1)

by
-Bheemashankar   1PE16CS039

Program:

```c
#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()
{
```
-1PE16CS039

```c
        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);
```

```c
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");
```

```c
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:

```
bsjakkani@bsjakkani-Lenovo-ideapad-FLEX-4-1470: ~/6th_sem/DMassign
bsjakkani@bsjakkani-Lenovo-ideapad-FLEX-4-1470:~$ cd 6th_sem/DMassign
bsjakkani@bsjakkani-Lenovo-ideapad-FLEX-4-1470:~/6th_sem/DMassign$ gcc Predict.cbs
-Lenovo-ideapad-FLEX-4-1470:~/6th_sem/DMassign$ ./a.out
size of x=23 and  y=23

--------------------------------------------------------------------
The Line Equation after the linear regression is y=-24.389x+13871.445

Linear Regression slope and intercept are:

Slope(m):-24.388721
Intercept(b):13871.445312


The Predicted Opening Values from Nifty MNC Index Data

1/Feb/2019:13286.116211
2/Feb/2019:13261.727539
3/Feb/2019:13237.338867
4/Feb/2019:13212.950195
5/Feb/2019:13188.561523

--------------------------------------------------------------------
```

```
--------------------------------------------------------------------

The Line Equation after Gradient decent is y=-25.541x+13865.518

Gradient decent slope and intercept are:

Slope(m):-25.540836
Intercept(b):13865.517578


The Predicted Opening Values from Nifty MNC Index Data

1/Feb/2019:13252.537109
2/Feb/2019:13226.997070
3/Feb/2019:13201.456055
4/Feb/2019:13175.915039
5/Feb/2019:13150.374023
bsjakkani@bsjakkani-Lenovo-ideapad-FLEX-4-1470:~/6th_sem/DMassign$
```
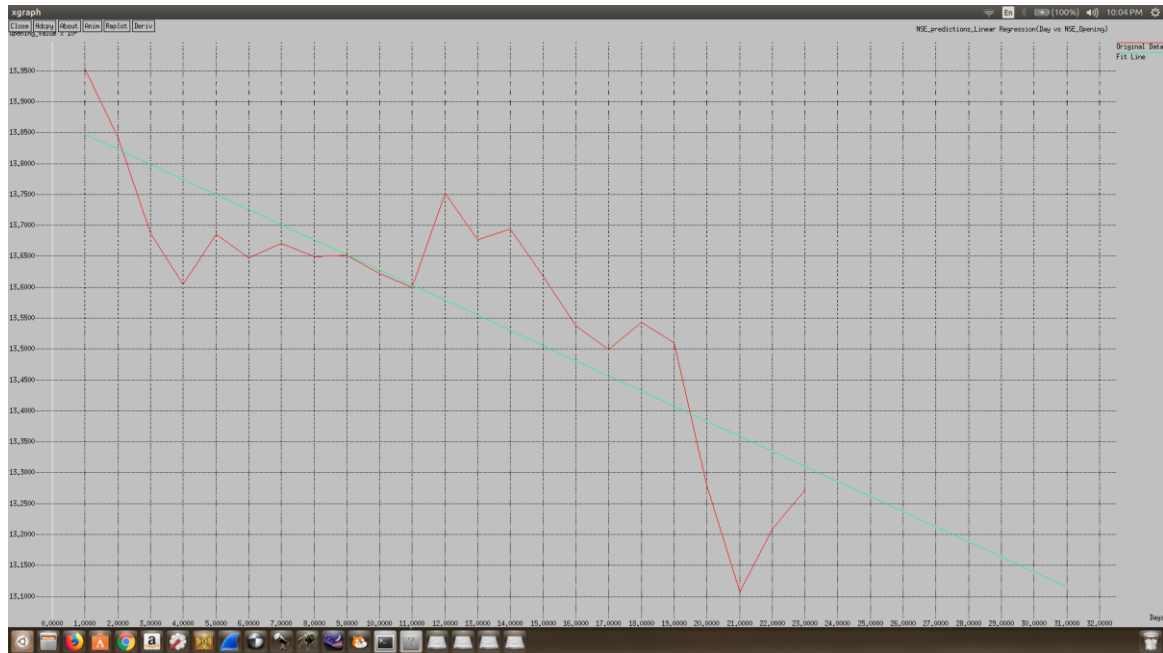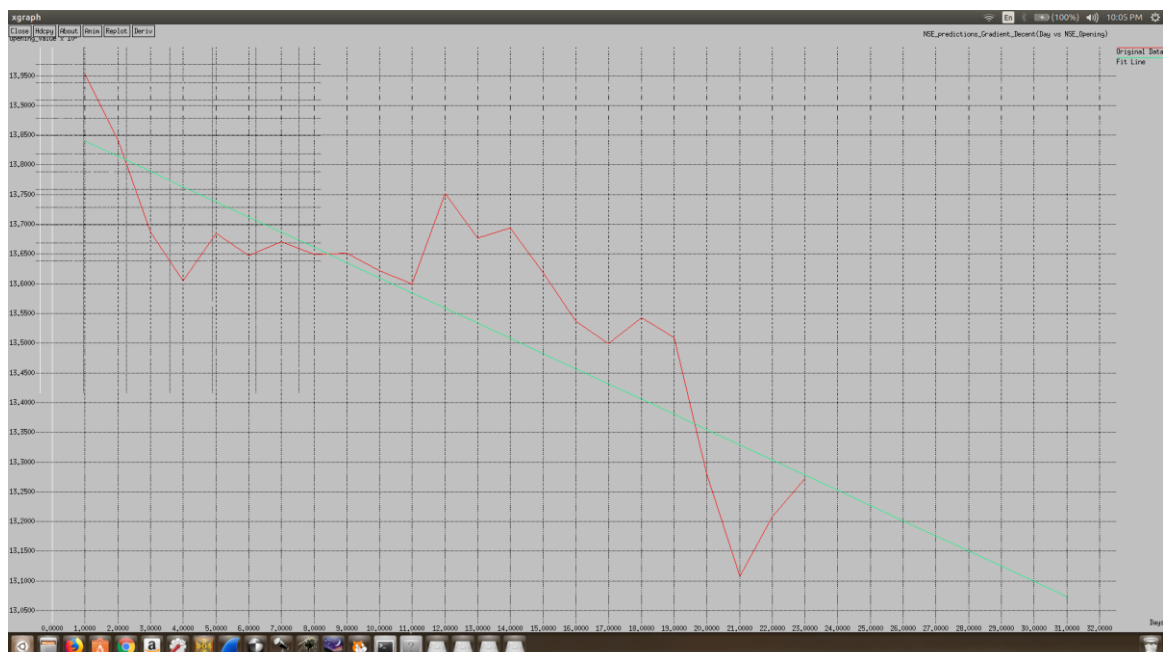
Graph:

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:

Historical Data for NIFTY MNC						
For the period 01-01-2019 to 31-01-2019						
Date	Open	High	Low	Close	Shares Traded	Turnover (₹ Cr)
01-Jan-2019	13953.20	13953.20	13835.00	13881.65	26267392	1344.17
02-Jan-2019	13842.35	13842.35	13669.65	13689.60	70964713	2340.88
03-Jan-2019	13687.35	13780.20	13567.60	13581.85	49021505	1898.49
04-Jan-2019	13604.60	13645.55	13505.80	13584.20	49740144	1957.16
07-Jan-2019	13685.15	13713.75	13606.55	13620.85	55204447	1729.99
08-Jan-2019	13647.45	13647.95	13557.30	13606.80	54337980	1675.58
09-Jan-2019	13671.10	13676.50	13555.55	13622.35	50784402	2038.33
10-Jan-2019	13649.05	13669.35	13603.95	13641.15	52639283	1754.57
11-Jan-2019	13651.15	13659.80	13581.00	13620.30	33545185	1440.69
14-Jan-2019	13622.45	13628.90	13520.55	13576.70	32286434	1318.42
15-Jan-2019	13600.05	13743.60	13599.35	13730.90	40029859	1739.09
16-Jan-2019	13751.60	13785.90	13638.75	13649.50	33263755	1554.56
17-Jan-2019	13676.65	13697.20	13601.55	13659.95	26815746	1514.92
18-Jan-2019	13694.25	13718.05	13569.85	13599.30	29184192	1475.38
21-Jan-2019	13617.50	13667.40	13530.95	13554.15	27668108	1438.14
22-Jan-2019	13537.25	13540.35	13427.55	13491.05	49622234	1927.71
23-Jan-2019	13499.15	13585.60	13479.95	13503.50	33647191	1820.45
24-Jan-2019	13542.90	13547.90	13442.20	13473.55	54213260	1900.99
25-Jan-2019	13509.90	13557.45	13222.95	13267.00	47744896	3358.94
28-Jan-2019	13280.25	13280.30	13066.35	13128.35	42108791	2703.56
29-Jan-2019	13107.75	13188.75	13053.65	13159.75	42351941	1866.93
30-Jan-2019	13208.55	13248.60	13158.30	13228.10	38651582	1554.91
31-Jan-2019	13272.25	13331.10	13217.05	13320.65	79836599	2648.05

-1PE16CS039