MATLAB CODE FOR LINEAR REGRESSION

%PROBLEM STATEMENT

%The data consists of 2 columns where the 1st column shows the

%profits(dependent variable)of a restaurant from different cities of

%different population(independent variable).So we are trying to find a

%relation between population of a city and the profit from that city

%data is loaded with comma separeted values

data=load('train.txt');

%size of data is found

d=size(data);

%no of (featues+1) of the data is taken

d=d(2);

%no of training samples

%x and y data is separated

x=data(:,1);

y=data(:,d);

q=ceil(rand(7,1)\*50);

testx=zeros(7,1);

testy=zeros(7,1);

for v=1:length(q);

testx(v,1)=x(q(v));

testy(v,1)=y(q(v));

x(q(v))=[];

y(q(v))=[];

end

nx=x;

%ones column(of m\*1) size is added to make matrix multiplication easier

m=length(x);

x1=ones(m,1);

x=[x1,x];

%initial weights are assinged

theta=zeros(d,1);

%h=a+bx is found

h=x\*theta;

%cost function-mean square error is found

c=sum((h-y).^2);

cost=c/(2\*m);

%learning rate is taken as 0.001

alpha=.001;

%iteration is done to get the correct weights

%theta1=theta1-((alpha/m)\*(x\*theta-y))

%theta2=theta2-((alpha/m)\*(x\*theta-y).\*x(:,2))

%we get the above correction formula by reducing cost function(by

%diffrentiation

for i=1:20000

z=sum(x\*theta-y);

z1=sum((x\*theta-y).\*x(:,2));

theta(1)=theta(1)-((alpha/m)\*z);

theta(2)=theta(2)-((alpha/m)\*z1);

z=0;

z1=0;

end;

%p is the predicted value from weights

p=x\*theta;

plot(nx,y,'rx')

hold on;

%prediction line is then plotted

plot(nx,p)

hold on;

m=length(testx);

x1=ones(m,1);

testnx=[x1,testx(:,1)];

prediction=testnx\*theta;

MSE=(sum(testnx\*theta-testy).^2)/(2\*m)

plot(testx,prediction,'bo')

plot(testx,testy,'gx')

output:

MSE = 0.7078

