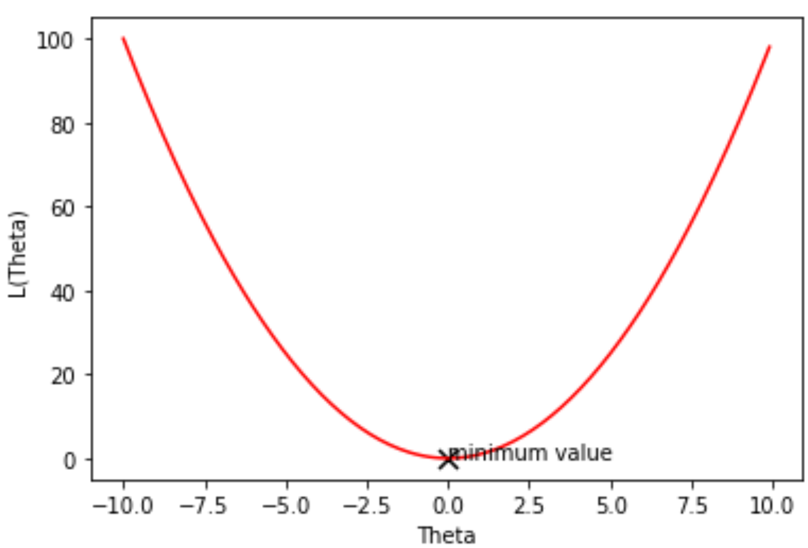
Question 1

Plot θ vs. L(θ), where L(θ) =θ 2 . θ varies from -10 to +10 with step size of 0.1. Now locate the minimum value of L(θ) with corresponding θ value from plot

In [ ]:
```
import numpy as np
import matplotlib.pyplot as plt
theta=np.arange(-10,10,0.1)

L=theta**2
L=np.array(L)
index=np.argwhere(L == np.min(L))
minimumvalue=np.min(L)
anstheta=theta[index[0][0]]
print("minimum value of L(theta) = "+str(minimumvalue)+" and the c
orresponding theta value is "+str(anstheta))
plt.plot(theta,L,color='red')
plt.xlabel("Theta")
plt.ylabel("L(Theta)")
plt.annotate("minimum value",(minimumvalue,anstheta))
plt.scatter(minimumvalue,anstheta, marker="x",c="black",s=80)
plt.show()
```

```
minimum value of L(theta) = 1.2621774483536189e-27 and the corre
sponding theta value is -3.552713678800501e-14
```



Observation For Question 1:

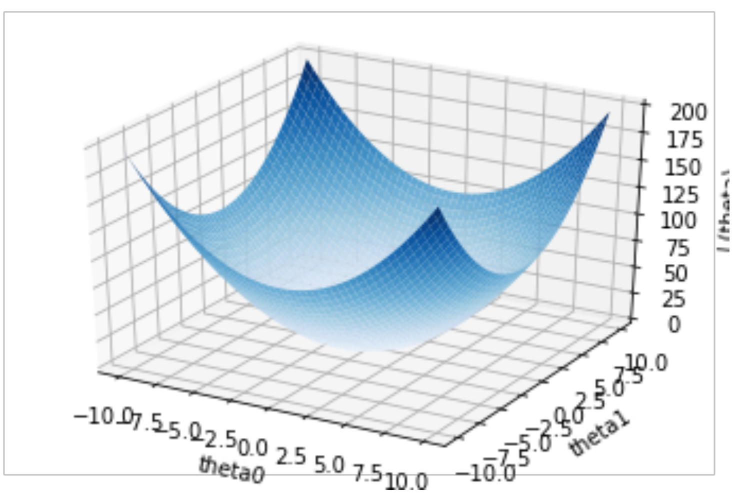
As shown in figure, L will approach to its minimum value 0 , when theta is 0.

Question 2

Plot θ vs. L(θ), where L(θ) =θ1^2 + θ2^2. θ1and θ2 vary from -10 to +10 with step size of 0.1 . Locate the minimum value of L(θ1, θ2) with corresponding θ1 , θ2 values from the plot.

In [ ]:
```
from mpl_toolkits import mplot3d
import numpy as np
import matplotlib.pyplot as plt
fig = plt.figure()
ax = plt.gca(projection ='3d')
x = np.arange(-10,10,0.1)
y = np.arange(-10,10,0.1)
x,y= np.meshgrid(x,y)
z=x**2+y**2
coor=np.argwhere(z == np.min(z))
print("The minimum value for L(theta)="+str(np.min(z)))
print("theta0="+str(x[coor[0][0]][coor[0][1]])+"\n theta1="+str(y
[coor[0][0]][coor[0][1]]))
ax.plot_surface(x,y,z,cmap='Blues')
ax.set_xlabel('theta0')
ax.set_ylabel('theta1')
ax.set_zlabel('L(theta)')
plt.show()
```

```
The minimum value for L(theta)=2.5243548967072378e-27
theta0=-3.552713678800501e-14
 theta1=-3.552713678800501e-14
```



Observation For Question 2:

As shown in figure, the graph for L vs theta is a 3D curve having a global minimum.Now , minimum value of L will occur at theta0=theta1=0.

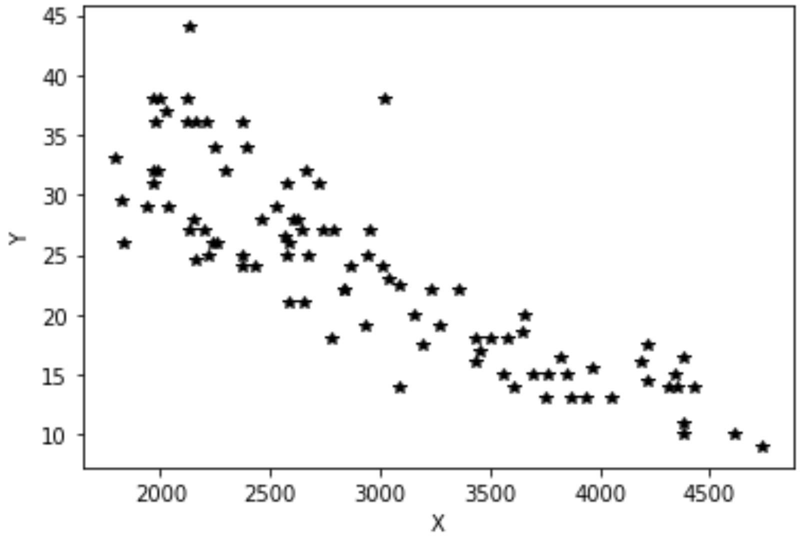
Question 3

Plot for L(θ) = sigma[{y i – (θ0 + θ1 . x (i)) }^2], where m is the number of input examples and x(i), y(i) are the values taken from given data file. Obtain the minimum value of L(θ) with corresponding θ0, θ1 values from the plot.

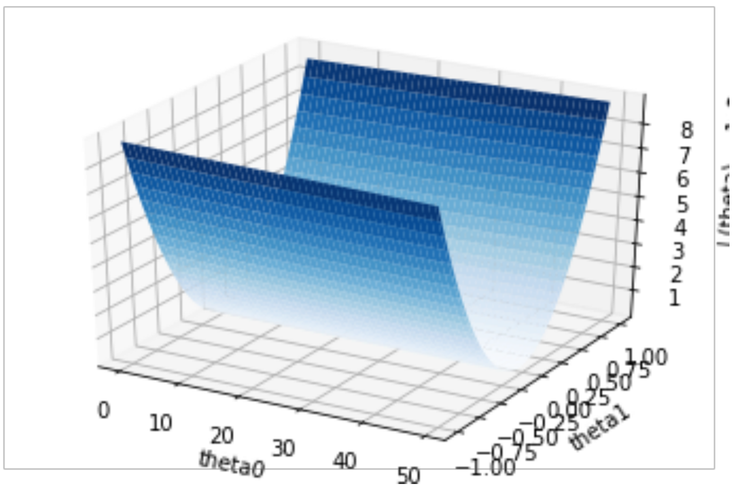
In [ ]:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv('Assign1.csv')
X =df.iloc[:,0:1]
Y=df.iloc[:,1:2]
X=np.array(X)
Y=np.array(Y)
plt.plot(X,Y,'*',color='black')
plt.xlabel("X")
plt.ylabel("Y")
plt.show()
theta0=np.arange(0,50,0.5)
theta1=np.arange(-1,1,0.001)
z=np.zeros((len(theta0),len(theta1)))
fig = plt.figure()
ax = plt.gca(projection ='3d')
for i in range(len(theta0)):
    for j in range(len(theta1)):
        for k in range(len(X)):
            z[i][j]=z[i][j]+((Y[k]-(X[k]*theta1[j])-theta0[i])**2)
theta1,theta0= np.meshgrid(theta1,theta0)
coor=np.argwhere(z == np.min(z))
print("The minimum value for L(theta)="+str(np.min(z)))
print("theta0="+str(theta0[coor[0][0]][coor[0][1]])+"\n theta1="+s
tr(theta1[coor[0][0]][coor[0][1]]))
ax.plot_surface(theta0,theta1,z,cmap='Blues')
ax.set_xlabel('theta0')
ax.set_ylabel('theta1')
ax.set_zlabel('L(theta)')

plt.show()
```




```
The minimum value for L(theta)=1595.5695840001015
theta0=47.5
 theta1=-0.007999999999999119
```



Observation For Question 3:

By plotting Y(given) vs X(given) graph , the points for y values were in the range of 100 and for x it was 10000 and value of y mmostly decreases a x increases, so, from that we can predict that the value for theta 1 will be in range of -1 to 0,here I had plotted for theta1 for the range -1 to 1 and theta 0 for the range of 0 to 50. The figure shows the graph of L vs theta,where minimum is achieved at theta 1=-0.008 and theta 0= 47.5 and the minimum value is 1595.

Question 4

Apply Pseudo Inverse (Least Squares (LS)) approach to get θ vector for the cost function (objective function) L(θ) given in example 3.Verify whether θ1 ,θ2 obtained are same as that found in example 3.

In [ ]:
```
X =df.iloc[:,0:1]
Y=df.iloc[:,1:2]
X=np.array(X)
Y=np.array(Y)
X2=X
X=np.c_[ np.ones(len(X)),X ]
XT=X.transpose()

temp=np.dot(XT,X)
temp=np.linalg.pinv(temp)
temp2=np.dot(XT,Y)
theta=np.dot(temp,temp2)
print("theta = ")
print(theta)
print(" here theta 0 = "+str(theta[0]))
print(" here theta 1 = "+str(theta[1]))
```

```
theta =
[[ 4.92376299e+01]
 [-8.61193478e-03]]
 here theta 0 = [49.23762989]
 here theta 1 = [-0.00861193]
```

Observation For Question 4:

Here, pseudo inverse formula is used to compute theta from X and Y.Formula:: Theta=(X^T *X)^-1* X^T *Y .The theta obtained from pseudo inverse is close but not equal to value obtained from example 3.There is difference because in example 3 we have taken step for theta 0 as 0.5 and for theta 0.001,so it will not check for smaller values so its minimum will not be equal to pseudo inverse minimum as pseudo inverse will be having more precision.So,by pseudo inverse will be more accurate than by example 3.

Question 5

Calculate the value of L(θ) using the θ vector obtained by Pseudo Inverse (as done in Example 4). Now Assume any θ vector (other than the one obtained in Example 4) and compute the new L(θ) value. Comment on why the Pseudo Inverse is also called LS method.

In [ ]:

```
X =df.iloc[:,0:1]
Y=df.iloc[:,1:2]
X=np.array(X)
Y=np.array(Y)
theta=np.zeros((2,1))
theta[0]=49.23
theta[1]=-0.00861
X=np.c_[ np.ones(len(X)),X ]
temp=np.matmul(X,theta)
temp=temp-Y
temp=np.matmul(temp.transpose(),temp)
print("for theta 0=49.23 and theta 1= -0.00861 (THETA obtained fro
m pseudo inverse), L(Theta) = "+str(np.sum(temp)))

theta[0]=49
theta[1]=-0.05
temp=np.matmul(X,theta)
temp=temp-Y
temp=np.matmul(temp.transpose(),temp)
print("for theta 0=49 and theta 1= -0.05 (Random theta value close
to theta obtained from pseudo inverse), L(Theta) = "+str(np.sum(te
mp)))
```

```
for theta 0=49.23 and theta 1= -0.00861 (THETA obtained from pse
udo inverse), L(Theta) = 1572.6509287476
for theta 0=49 and theta 1= -0.05 (Random theta value close to t
heta obtained from pseudo inverse), L(Theta) = 1523348.190000000
2
```

Observation For Question 5:

Here the value for L(theta) is obtained lesser than example 3 as stated above.Also ,I had taken theta 1 and theta 0 closer to value obtained by pseudo inverse.But the difference is large as for theta=[49.3 -0.00861]^T ,L = 1572 , but for theta=[49 -0.05]^T ,L = 1523348, which shows that slope near minima is very high.So for small change in theta there is very large change in L.Here pseudo inverse is called Least square because it will return theta for which the error or loss function(which is the measured by taking sum of square of each entries) will be minimum.Thus it is also called LS(least square method).