1a.

Given that y = +

This equation is of the form

y = +

y follows Normal Distribution of the form y

The likelihood probability will follow Gaussian distribution and hence

P(y / ) =

P(y / ) =

1b.

The Likelihood equation is given by

L(Y / X,W) = )

Using the equation of ) from 1a we have the equation for likelihood as

L(Y / X,W) =

Log Likelihood is given by

Log(L(Y / X,W)) = Log( )

We know the log(ab) = log(a) + log(b) and log = x using these two we have

Log(L(Y / X,W)) = ) +

Ignoring the term and since they are independent of we will have

Log(L(Y / X,W)) =

Conditional Log Likelihood is given by the equation

1c.

The MLE is given by the equation

= ( )

Since we have a - for the equation The max of will be minimum of

Multiplying it by 1/2 we have minimum of is equal to minimum of This is implying that maximizing the likelihood is equal to minimizing the Least Square error.

2a.

The Least Square Error is given by the equation.

LSE = +

Partial Derivative with respect to is given by

= ( + )

Since = 2(a-bx) \* (a-bx) = 2(a-bx) \* -b , since (a-bx) = b

() = \*-

Since = 2x

() = () = = λ

Since is not present here () = 0

Using all the above equations obtained

() = (-1)

() = \*

() = \*(-)

() = λ

() = λ

= + 0 = (-1)

= \*) + λ

= \* + λ

Gradient Descent update rules is given by the equation

= - Ƞ( )

I am considering the cost function = LSE = +

Using the equations obtained above we get

= - ( )

= - ( \*(- + λ )

= - ( \*(- + λ )

2b.

The MAP estimate is given by the equation

= ( log()P() ))

Taking ) from question 1 and we are given that w1,w2 ~ hence equation for

P() = , using these equations we have

log()P() ) =

log( \*)

We know the log(ab) = log(a) + log(b) and log = x using these two we have

log()P() ) =

) + + log() +

Neglecting all the terms independent of w, we get

= ( ) neglecting the constants

= ( )

Since we have a minus at the start of each terms this is same as

= ( )

Since are constants, If we add the for the second term we get,

= ( )

This is equivalent to minimizing the +

Where λ =

3a.

The code for the same is submitted.

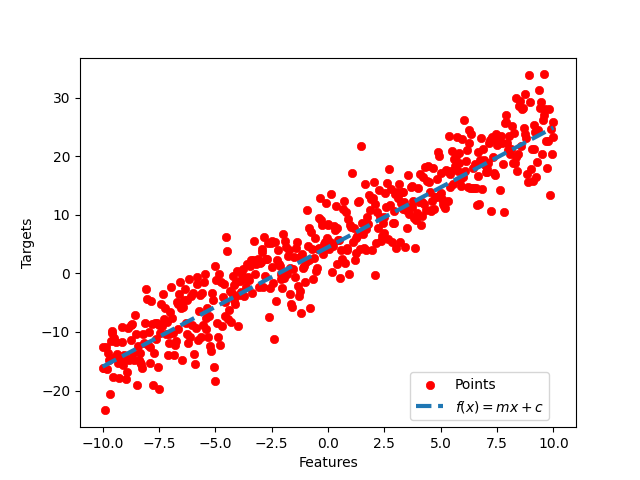
3b.

The code for the same is submitted

3c.

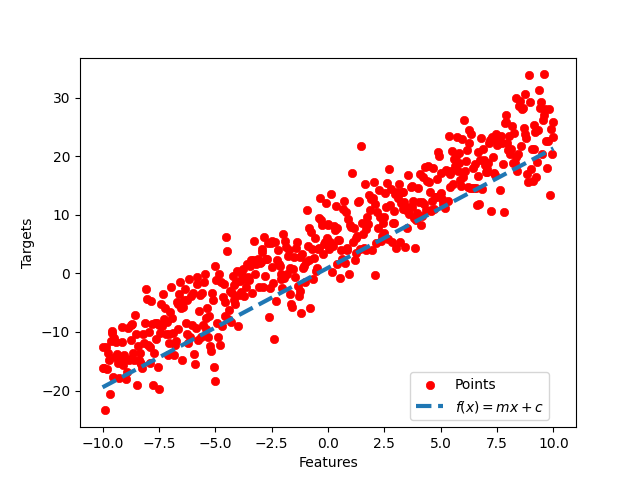
The Plots and data obtained are as follows

For Learning Rate 0.01 :



The value of Loss obtained at the end of 100th epoch is = 17.469

For Learning rate 0.001:



The value of Loss obtained at the end of 100th epoch = 34.815

From the data above we can say that the Learning rate 0.01 is working better compared to 0.001. The Learning rate 0.01 is neither too low nor to high hence within 100 epochs it almost got the optimal values and if we observe the curve obtained for 0.01 it fitted well with the data points and hence MSE is less for 0.01. For the case of 0.001 it is very slow and it actually need more epochs to converge and give the optimal values.