Problem 1 (50 points)

Develop a code that run linear regression with gradient decent algorithm for each of the explanatory variables in isolation. In this case, you assume that in each iteration, only one explanatory variable (either X1, or X2, or X3) is explaining the output. Basically, you need to do three different training, one per each explanatory variable. For the learning rate, explore different values between 0.1 and 0.01 (your choice). Initialize your parameters to zero (theta to zero).

1. Report the linear model you found for each explanatory variable.

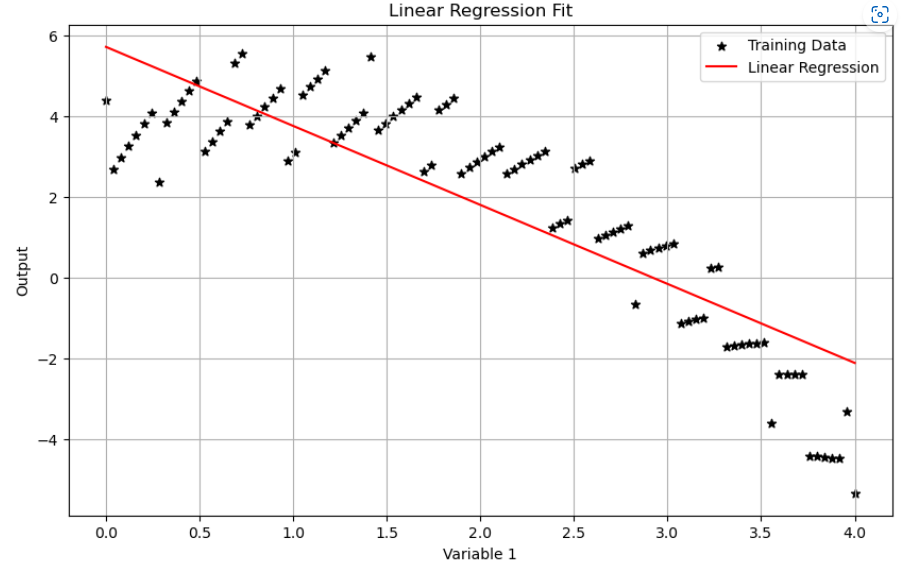
We can suppose that y=θ+θ1\*x, so we can substitute the value into the function. So wo got differentθandθ1:

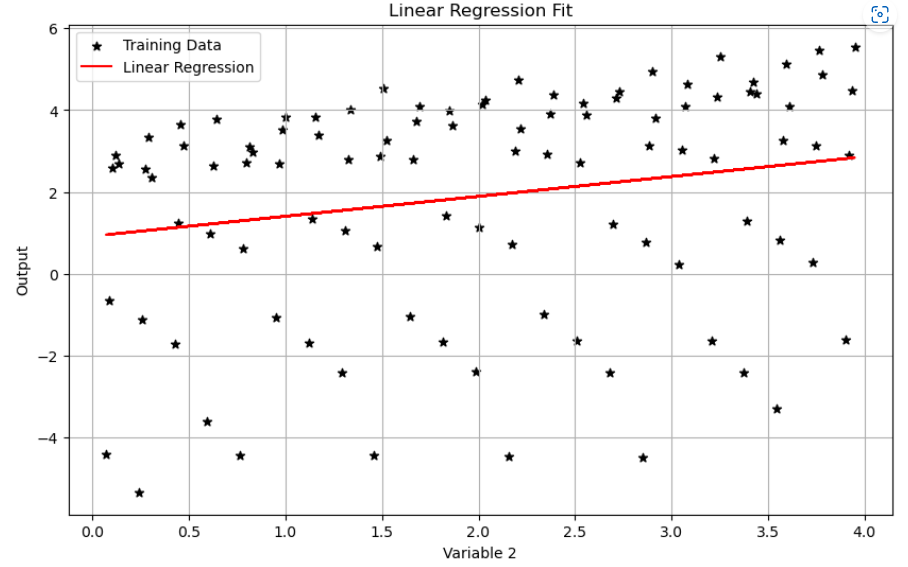
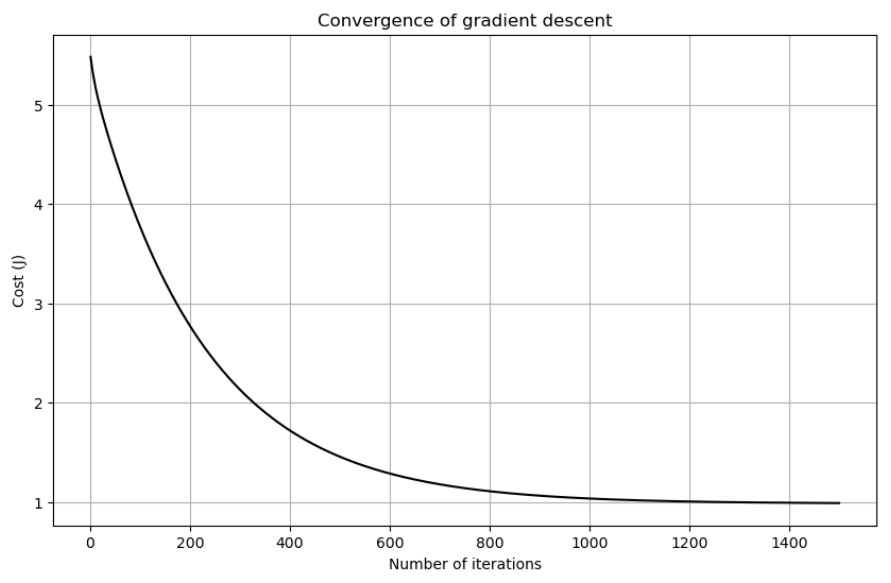
Y=-1.9568206x1+5.71850653

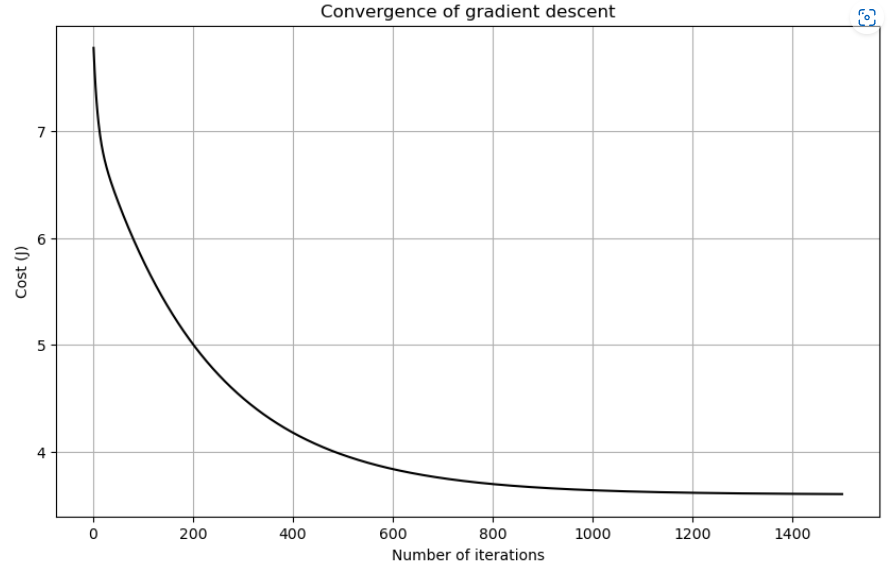
Y=0.48530143x+0.92183754

Y=-0.49304729x+2.80205172

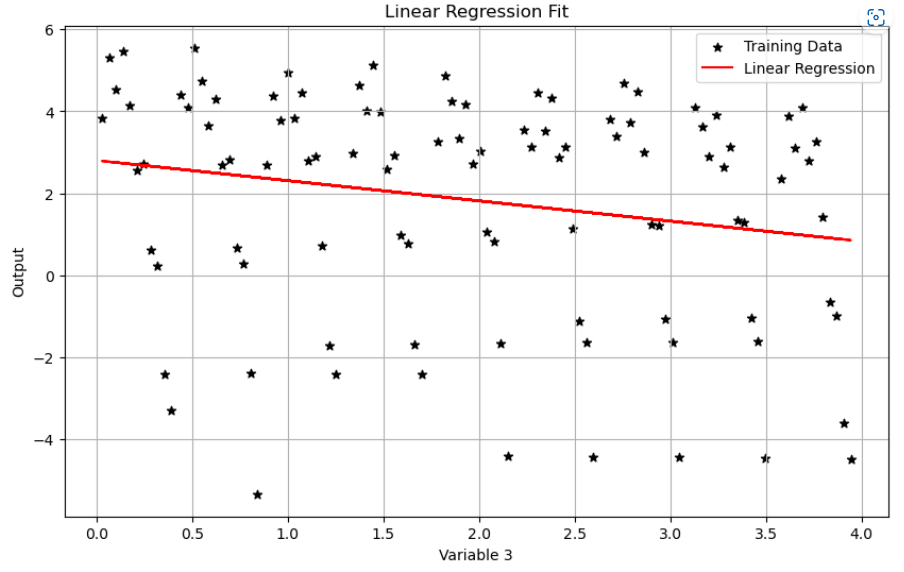
1. Plot the final regression model and loss over the iteration per each explanatory variable.

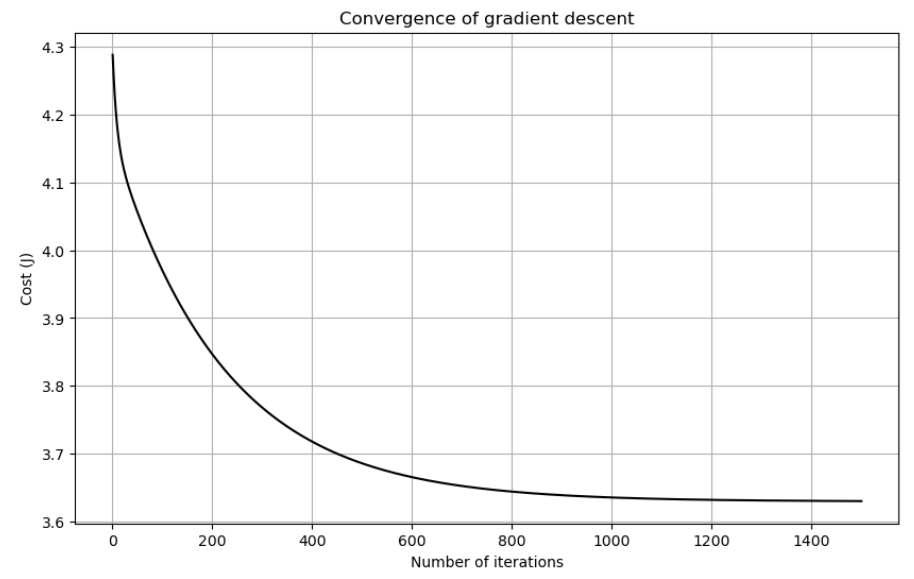
Plot of X1:

Plot of X2:



Plot of X3:





1. Which explanatory variable has the lower loss (cost) for explaining the output (Y)?

At the 0.01 learning rate:

cost\_history (Variable 1) = [5.48226715 5.44290965 5.40604087 ... 0.99063932 0.99061433 0.99058944]

cost\_history (Variable 2) = [7.78019751 7.67266456 7.57620204 ... 3.60380823 3.60378856 3.60376899]

cost\_history (Variable 3) = [4.28817888 4.27128079 4.25606956 ... 3.63008237 3.63007954 3.63007671]

the lowest cost of X1,X2 and X3 is 0.99058944, 3.60376899 and 3.63007671.

So variable 1 has the lower loss (cost) for explaining the output (Y).

1. Based on your training observations, describe the impact of the different learning rates on the final loss and number of training iteration.

At the 0.01 learning rate the costs are 0.99,3.60 and 3.63;

At the 0.02 learning rate the costs are 0.98500018, 3.59937205 and 3.62945198;

At the 0.03 learning rate the costs are 0.98499309, 3.59936603 and 3.62945113;

At the 0.05 learning rate the costs are 0.98499308, 3.59936602 and 3.62945112;

At the 0.1 learning rate the costs are 0.98499308, 3.59936602 and 3.62945112.;

 As we can see, the learning rate should not as small as possible. When we choose the learning rate at 0.01, it has the highest cost and when we choose the learning rate at 0.1 or 0.05, it has the lowest cost. That means the step of 0.01 is too smell to get the point.

And going to the epoch, I lock learning rate to 0.01. For 500 iterations, the cost is 1.45786855 for X1, 1000 is 1.03643608, 1500 is 0.99058944, 2000 is 0.9856019 and 2500 is 0.98505931. We all add 500 epoch each but the increase become less and less.

Problem 2 (50 points)

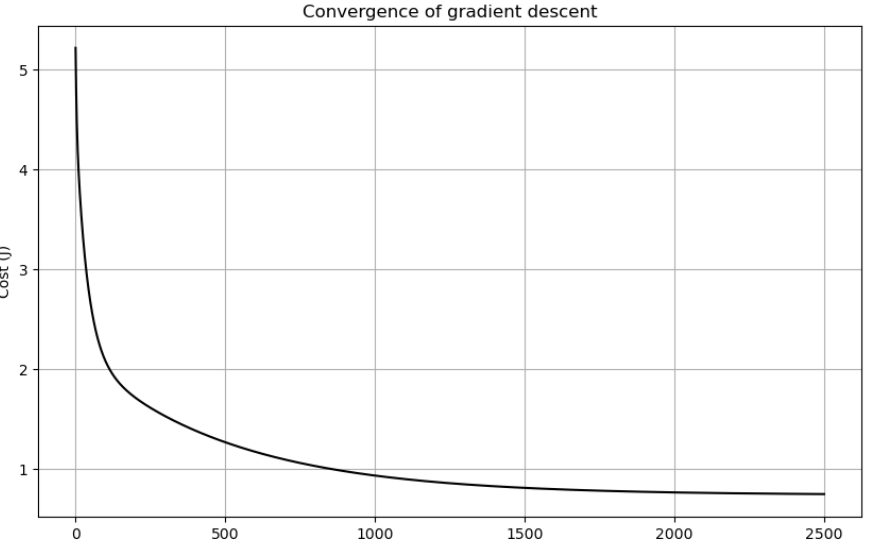
This time, run linear regression with gradient descent algorithm using all three explanatory variables. For the learning rate, explore different values between 0.1 and 0.01 (your choice). Initialize your parameters (theta to zero).

1. Report the final linear model you found the best.

Similar to problem 1, we suppose the function as y=θ1\*x1+θ2\*x2+θ3\*x3+θ0 and we can substitute the value into the function to getθ. The linear model that was found:

Y=-1.94311861X1+0.60344978X2-0.20272198X3+4.88518623

1. Plot loss over the iteration.



1. Based on your training observations, describe the impact of the different learning rates on the final loss and number of training iteration.

I tried 0.01, 0.02, 0.03, 0.05 and 0.1. I found that the smaller the learning rate, the slower the convergence, which means more iterations are needed

1. Predict the value of y for new (X1, X2, X3) values (1, 1, 1), for (2, 0, 4), and for (3, 2, 1)

Y for (1, 1, 1) is 3.34279542

Y for (2, 0, 4) is 0.1880611

Y for (3, 2, 1) is 0.06000798