ECGR 4105: Intro to Machine Learning HW 1

Axel Leon Vasquez

Student ID: 801182414

September 16, 2024

Link to GitHub

Problem 1

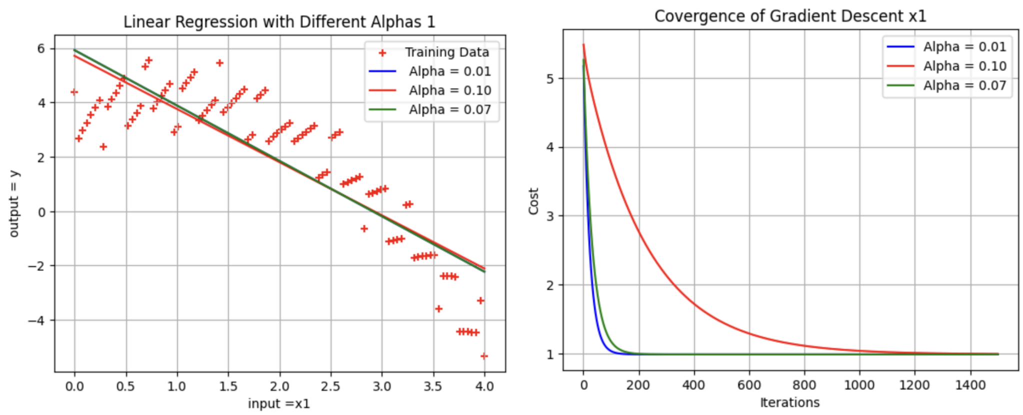
1. Model Results:

X1 Model:

Alpha and Theta used for x1 mode:

- $\alpha = 0.01, 0.10, 0.07$
- $\Theta_0 = 5.92794892$
- $\Theta_1 = -2.03833663$

Figure 1: X1 Models

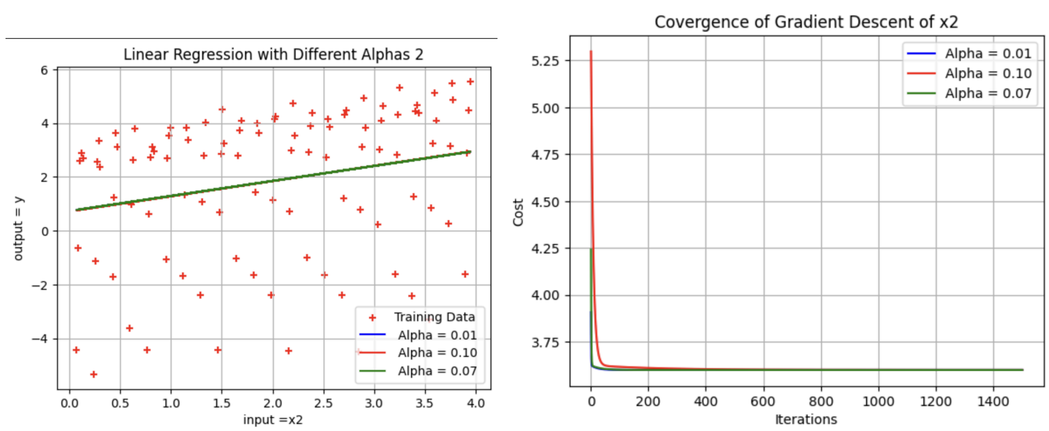


X2 Model:

Alpha and Theta used for x1 mode:

- $\alpha = 0.01, 0.10, 0.07$
- $\Theta_0 = 0.73606043$
- $\Theta_1 = 0.55760761$

Figure 2: X2 Models

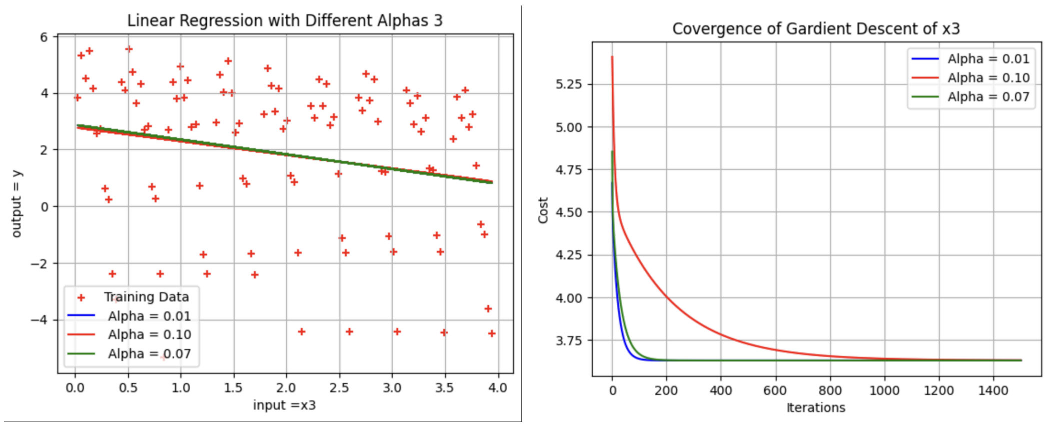


X3 Model:

Alpha and Theta used for x1 mode:

- $\alpha = 0.01, 0.10, 0.07$
- $\Theta_0 = 2.8714221$
- $\Theta_1 = -0.52048288$

Figure 3: X3 Models



2. **Cost History**

   The cost in linear regression aims to find the best-fit line that predicts the target variable Y based on the explanatory variables X1, X2, and X3. To achieve this, we minimize the cost function to measure how well the model fits. The lowest cost output was model X1, which was 0.98499308. This is lower compared to the other variables X2 and X3. Based on X2 and X3, both have a nonlinear relationship. To compute the cost function J represented the formula of

   $J(\Theta) = \frac{1}{2m}\sum_{i=1}^{m}(\gamma^{(i)} - y^{(i)})^2$, this function measures how far off the predicted values γ are from the actual values y, as it can be seen in the table below :

Variable	Cost
X1	0.98499308
X2	3.59936602
X3	3.62945112

3. **Learning Rates on the number of training iterations**

   The learning rate in the gradient descent algorithm controls the size of the steps taken towards minimizing the cost function. The choice of learning rate directly impacts the number of training iterations needed for the model to converge and reach a point where the cost function no longer decreases significantly. The model rate appears to report 0.10 as the lowest loss of the explanatory values, it can be seen in the table below:

Variable	Learning Rate (Alpha)	Loss
X1	0.01	0.98499308
X1	0.07	0.848484845
X1	0.10	0.848484830
X2	0.01	3.59936602
X2	0.07	3.59936602
X2	0.10	3.599660218
X3	0.01	3.62945112
X3	0.07	3.62945112
X3	0.10	3.62945112

Problem 2

1. **Best Linear Model:**

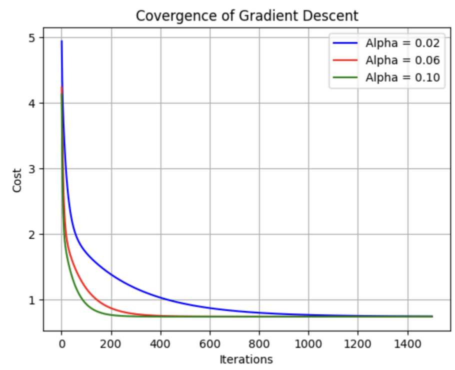
   The theta values found for the lowest cost is:
   - $\alpha = 0.10$
   - $\Theta_0 = 5.31416563$
   - $\Theta_1 = -2.00371905$
   - $\Theta_2 = 0.53256359$
   - $\Theta_3 = -0.26560164$

   Cost calculated after 1,500 iterations for $\alpha = 0.10$

$$J = 0.73846424$$

2. Plot Model of Cost Loss Using All Explanatory Variables

Figure 4 : Plot Loss Over Iteration



3. Learning Rate

The learning rate α is a crucial parameter in the gradient descent that controls the size of the steps taken toward minimizing the cost function. In this case, in the previous problem, while calculating the learning rate, the gradient descent was at the lowest at 0.10 as shown in the table below :

Learning Rate (Alpha)	Loss Function
0.01	0.83846302
0.05	0.73846424
0.10	0.73842424

4. Predictions

In this analysis, we utilized gradient descent to perform linear regression with three explanatory variables (X1, X2, and X3). We experimented with different learning rates 0.02, 0.06, and 0.10 to observe their effects on convergence speed and stability. The results revealed that the learning rate of 0.06 provided a balanced approach, achieving effective convergence without the instability associated with higher rates. We then manually computed predicted values for new inputs using the final learned coefficients theta from our model. For the input values (1, 1, 1), (2, 0, 4), and (3, 2, 1), the predictions demonstrated the model's ability to generalize and provide estimates based on the learned relationships from the training data. The final predictions for these inputs were calculated as y1, y2, and y3, as can be shown below :

$$Y1([1,\ 1,\ 1]) = 3.57740937429$$
$$Y2([2,\ 0,\ 4]) = 0.24432117283$$
$$Y3([3,\ 2,\ 1]) = 0.102534217289$$