Dustin Oakes

Economics 425

Professor Liu

29 January 2021

Gradient Descent Assignment

Part 1.

Placeholder 1: Adjusting ALPHA and MAX_ITER

Holding the number of iterations constant, increasing alpha results in a steeper curve, as the cost decreases quickly (Figure 1). Decreasing alpha to smaller amounts like 0.0001 resulted in a less steep curve, as the cost decreased less quickly (Figure 2). When alpha was made small enough, the number of iterations had to be increased to see where the asymptote of the curve actually is (Figure 3).

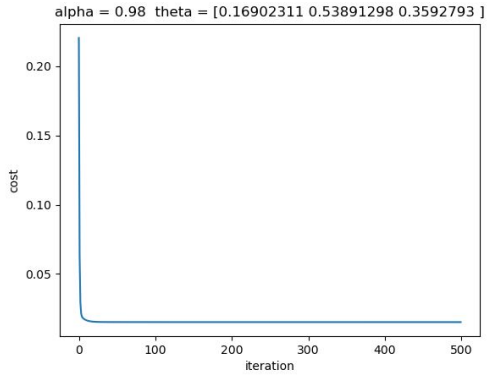


Figure 1

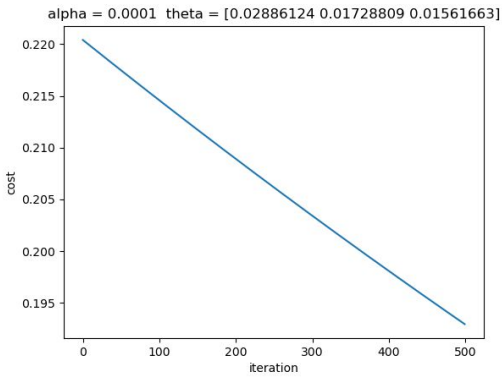


Figure 2

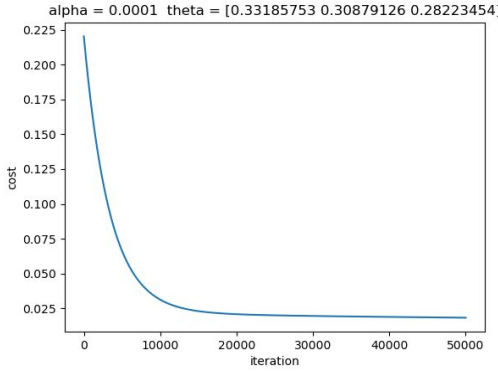


Figure 3

In terms of the results of the regression, as alpha was decreased the standard deviation and error also decreased. However, it should be noted that as alpha was made sufficiently small, the results seemed to increase, but this only happened because more iterations should have been run (alpha was made too small for the limited number of iterations).

Placeholder 2: Changing Normalization Function

Using the RescaleNormalization function resulted in each data point having the minimum subtracted and then being divided into the range of all data points as a method of normalization. This affected our regression by making both the standard dev and the error values extremely small, but they have to be evaluated differently as the results from the default RescaleMatrix as the units are not comparable anymore. Using the MeanNormalization resulted in similar, but not as drastically small, regression results to the RescaleNormalization function. In this function the mean, not the minimum, is subtracted from each data point before being divided into the range, which resulted in a slightly different scale. The convergence curves looked very similar to each other no matter what method was used.

Placeholder 3: Calculating R-Squared

Calculating the r-squared value returned a rather low result of 0.23.

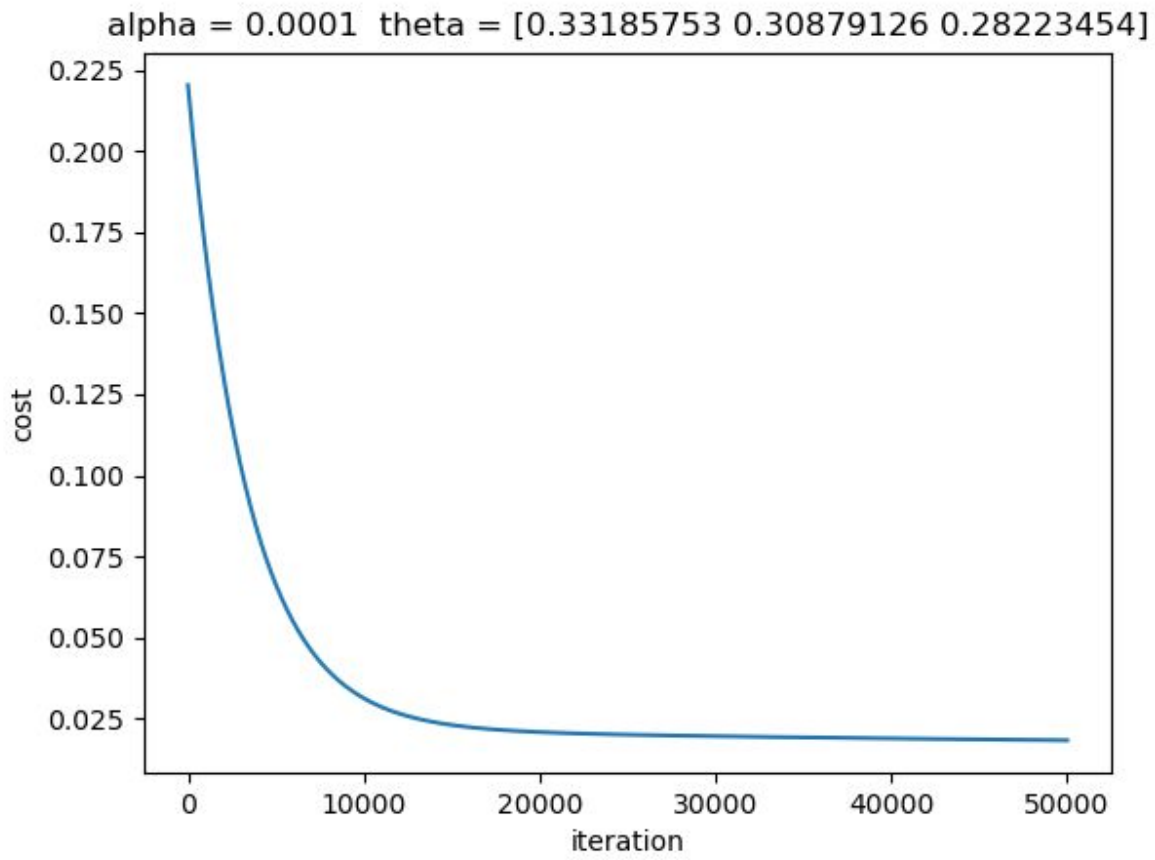
Placeholder 4/5: GD Theta/Cost

See code for implementations

Part 1 Results (alpha = 0.0001, iterations = 50000) - Table 1

Method	rescaleMatrix	rescaleNormalize	meanNormalize
Error	0.147015	0.000364	0.015823
Standard Dev	0.117227	0.000306	0.009357

Part 1 Results(alpha = 0.0001, iterations = 50000) - Figure 4



**Note: All convergence curves looked the same across normalization methods*
See Page 1 (Fig. 1-3) for how convergence can change

Part 2.

Results of Various Regression Types (R-Squared)			
Regression Type	Optimal ALPHA	Optimal MAX_ITER	Best R-Sq.
Linear Regression	0.0001	1000000	0.90
Ridge Regression	0.0001	1000000	0.98
LASSO	0.00001	5000000	-0.10
Elastic Net	0.00001	5000000	-0.10

By my analysis, the ridge regression is the best in terms of r-squared value. However, I think there is more I could do with the LASSO and ElasticNet given the seemingly poor outcomes. Lowering the alpha and increasing the number of iterations seemed to really boost the r-squared scores of the Linear and Ridge regressions, but there was not much improvement to be had on the LASSO or ElasticNet (I tried adding more and more iterations but at that point it was too computationally consuming to yield any reasonably quick results).