ECON 425 – Homework 2 Write-Up

By: Pujan Thakrar

Background

For this project we look at the math and verbal SAT scores and GPA scores of 105 students. We will fit a supervised learning model (in the form of a linear regression) in order to predict student GPA scores given the SAT scores.

Algorithm

For this project, I loaded the data that contains GPA and SAT scores. Then, I only keep the data that is of interest (i.e. the math SAT, verbal SAT, and GPA scores).

Then, I normalize the data using definition 'rescaleMatrix1()'. This definition computes the mean, minimum, and maximum of each type of score. Then for each observation, I subtract the mean and divide by the range in order to normalize the data. I store the new values into a new 'sat' matrix. Normalization of the data is necessary as we want our data to satisfy standard linear regression assumptions.

Following this, I split the data into training (60 observations) and testing (45 observations) data and define parameters. I call the gradient function which will update estimated parameters in order to find the optimal values, and then compute the cost function.

The cost function is then plotted to determine where it converges (I.e. the minimum value). You can change 'ALPHA' and 'MAX_ITER' to get different curves.

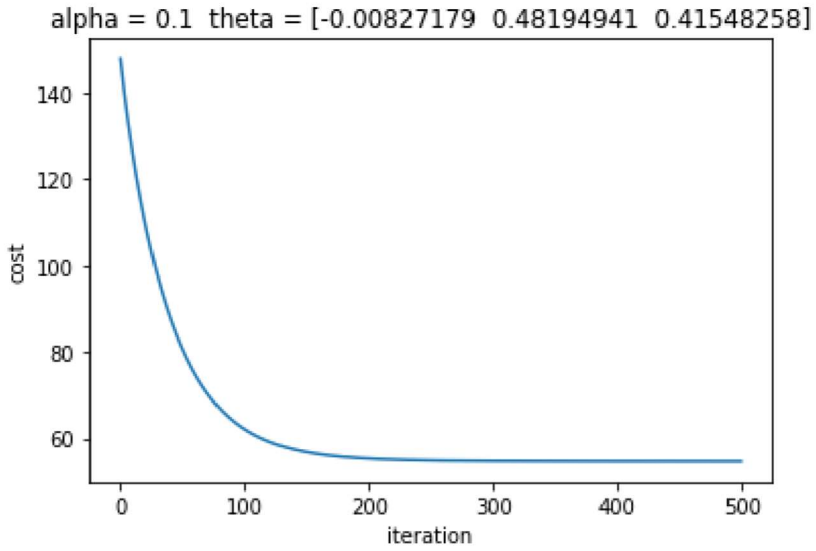
Finally, I test the fitted regression, given by the optimal parameters found from the gradient descent definition, on the testing data and compute the average error and standard deviations.

The algorithm produces the following results for different iterations and alphas.

Results

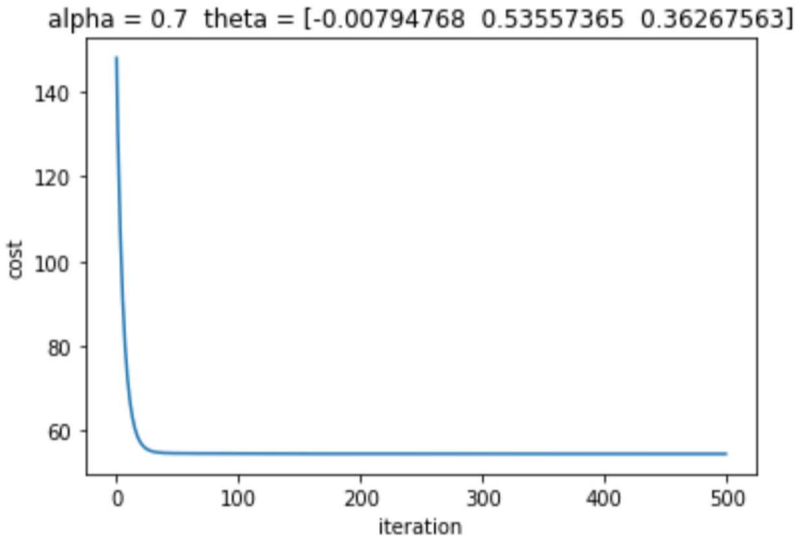
Changing alpha and fixing iterations

For alpha = 0.1 and max_iter = 500

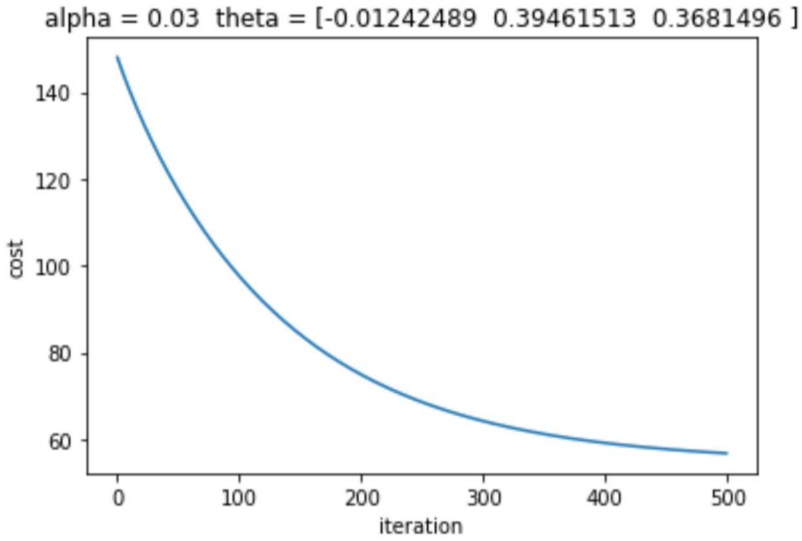



```
results: 0.17326009831040687 (0.12729258175503566)
```

For alpha = 0.7 and max_iter = 500



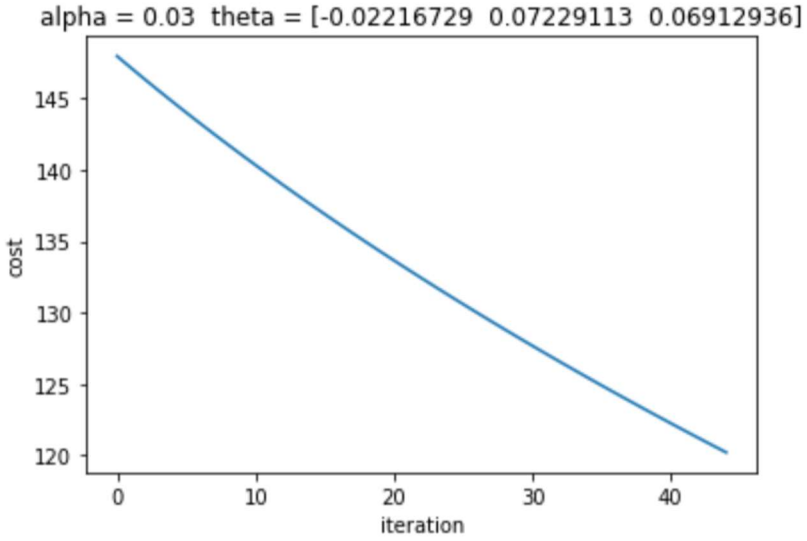

```
results: 0.1746711527870884 (0.1272265387827901)
```




```
results: 0.16007231938551486 (0.12158257572740852)
```

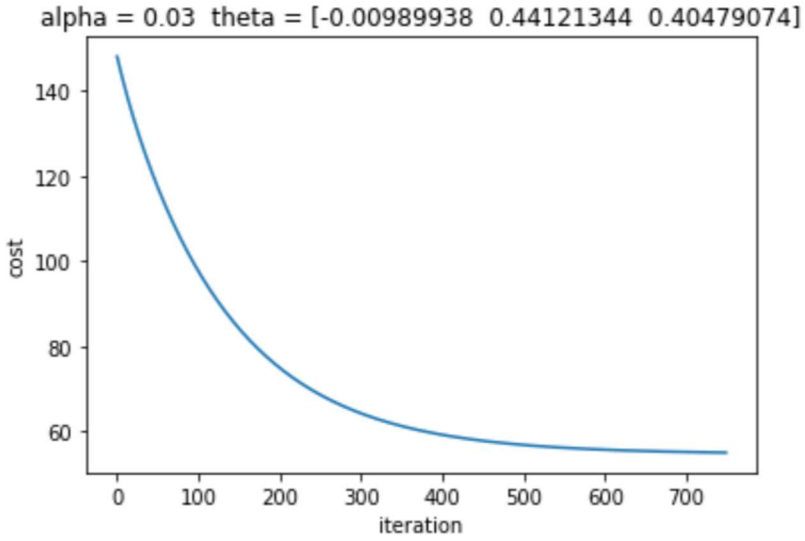
<u>Changing the number of iterations and fixing alphs</u>

For alpha = 0.03 and max_iter = 45




```
results: 0.1576453020632505 (0.1273488229842768)
```

For alpha = 0.03 and max_iter = 750




```
results: 0.16795316958181583 (0.12444542689420444)
```

Further discussion

Given the results, there are a few things to note. First, the change in alpha will result in a change in the curvature for the convergence graph. Specifically, if the alpha is higher there will be a steeper curve while a smaller alpha will result in a slower or more gradual convergence. Changing the number of iterations will simply extend the curve out given an increase in iterations, or shorten a curve with a decrease in iterations. Changing either the number of iterations or the alpha doesn't significantly impact the average error or standard deviation. Notice that the curves tend to asymptote to zero.