1. Different learning rates

After playing with the learning rate parameter it was included that when it goes beyond a certain value, value of SSE and also value of the gradient increase and that makes the program to produce incorrect outputs. After observing the behavior of the optimization process of Gradient Descent method, best value for the learning rate parameter was obtained.

The learning rate parameter should be less than 0.00064; otherwise each subsequent iterate increasingly overshoots the minimum, resulting in divergence of the algorithm. On the other hand, choosing smaller values for the learning rate parameter will result in the same SSE optimum value with a higher computation time. Choosing the optimum value for learning rate, results in these minimum values for SEE of training data and test data.

One important point is that the calculated value for SSE does not consider the regularization component since this component depends on the value lambda. Lambda is one the variables that can be tuned thus having a comparison between multiple SSE values which take into account different lambda values is useless.

**E = 0.0001 (convergence criteria)**

**Lambda = 0.1**

**SSE for training data: 10446.326278974173**

**SSE for test data:382445.297371**

**E = 0.0001 (convergence criteria)**

**Lambda = 0.01**

**SSE for training: 9542.138873946133**

**SSE for test data: 1371368.94053**

**E = 0.0001 (convergence criteria)**

**Lambda = 0.001**

**SSE for training: 8024.971961745075**

**SSE for test data: 9577200.5101**

By assessing the results, it can be realized that changing lambda does not affect the value of learning rate, but it heavily changes the optimum value for SSE of both the training and test sets.

1. Learning from the previous part of the optimization we can set the learning rate at 0.00064 and tweak the lambda value.