Willliam Liu

Write up

1. Normal equation parameters (): [﻿-0.31358288, 0.87211124, 0.47195201]

Gradient descent equation parameters (): ﻿ [ 0.20423419, 0.36261085, 0.37633092]

1. ﻿Residual sum of squares (RSS) of method 1(normal): 0.374727942757

﻿ Residual sum of squares (RSS) of method 2(GD): 0.565909341794

RSS of two models are different. Normal equation worked better fit training data because the dataset is relatively small, so not as computationally expensive. Plus, Gradient descent would work better when there are more features.

1. ﻿MSE of theta\_method1 on testing data: 0.00963677250183

MSE of theta\_method2 on testing data: 0.00723742937981

Method 2 seems to have less error but the numbers seems to be very small. Also, the test dataset is smaller than the training data set. So, that could be a reason for the differences.



﻿ 



Out of the 4 plots, learning rate is the fastest when alpha is 0.1 This means that the higher alpha is, the faster convergence and faster cost function lowers. Vice versa, when alpha is small: (0.001), cost function decreases the slowest, meaning slowest learning rate.