

1: 29.65625 and 14.875

2: theta = [0.3411937275648189 , 0.7606995613793791] and cost: 22.46325487790739

3: theta = [4.000000000000003 , 2.000000000000018]

The thetas are very different, the thing that needs added is a check for convergence of the cost at the minimum

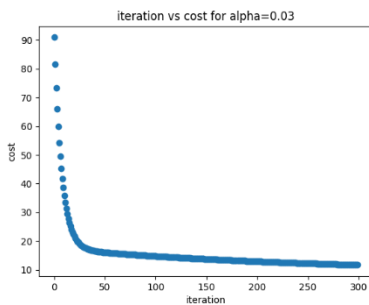
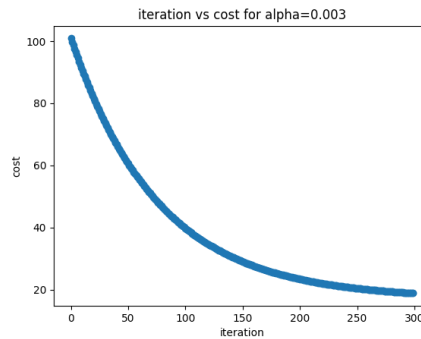
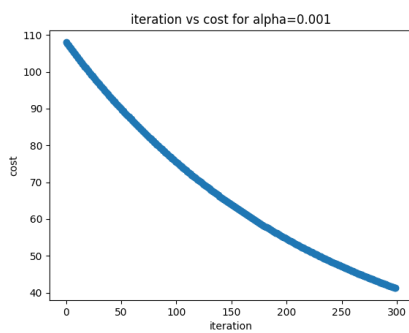
4c: X[178,2] and y[178]

4e: theta computed to = [-5.198563738234863 , 17.086610331376885]

4f: cost: 8.58

4g: when I reran the prediction with the normal equation both gradient descent and normalEqn returned cost = 6.086 and only began deviating at the fourth decimal point

4f:



based on the figures I would say the smaller learning rate is better, however you need more iterations to reach the same level of accuracy that the larger rates reach at a quicker interval. I think for this example the ideal learning rate and iteration combo would be alpha = .003 and iters = 500