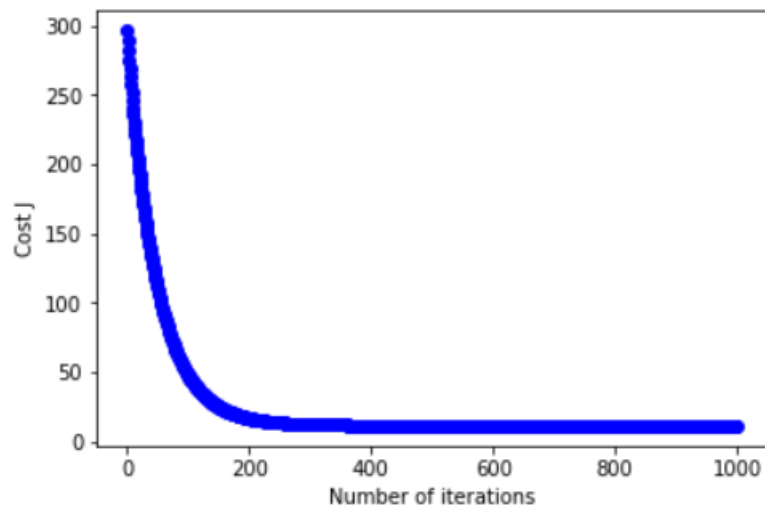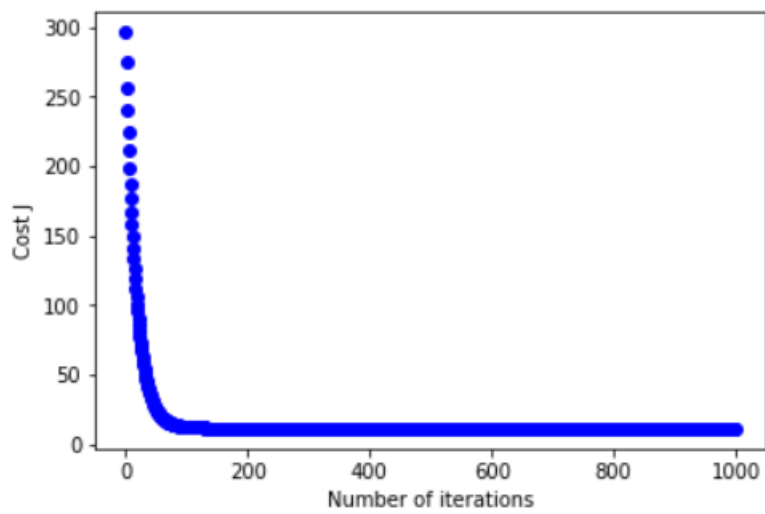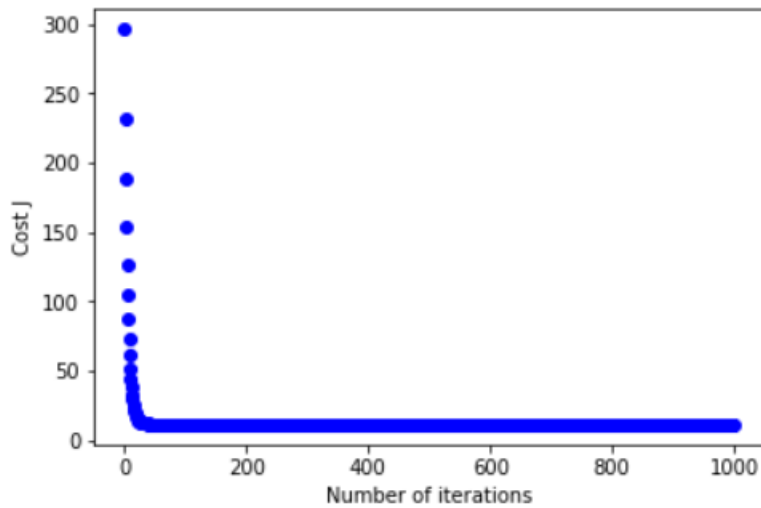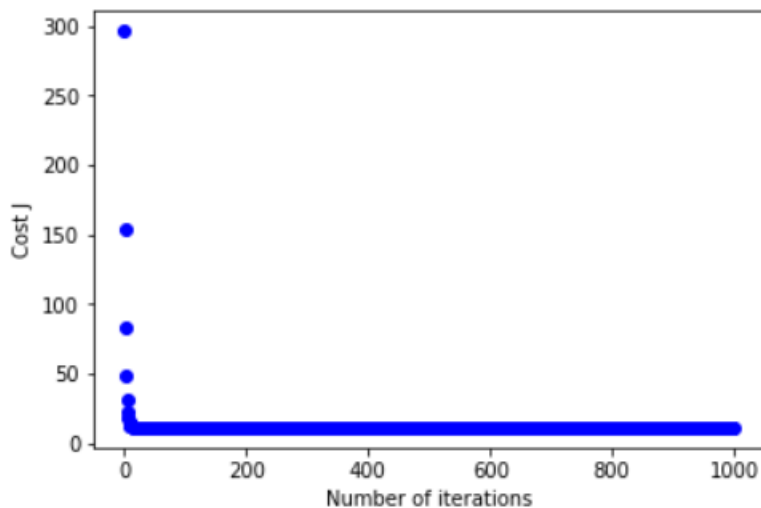learning rate = 0.01



learning rate = 0.03

learning rate = 0.1



learning rate = 0.3



Learning rate = 0.3 with 1000 iterations is good for this problem.

As learning rate becomes larger, J converges faster. With a smaller learning rate, the model requires more iterations for J to converge. However, when the learning rate is 0.3, it only need 100 iterates for J to converge.