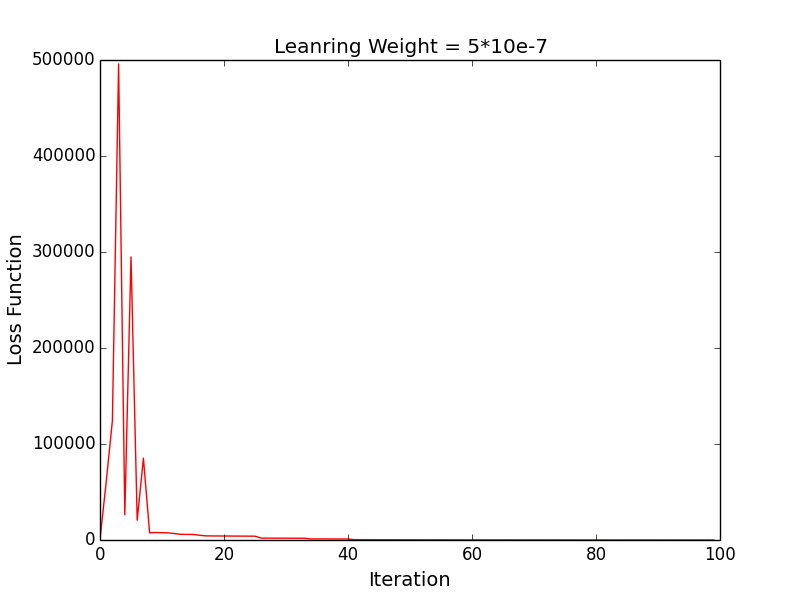
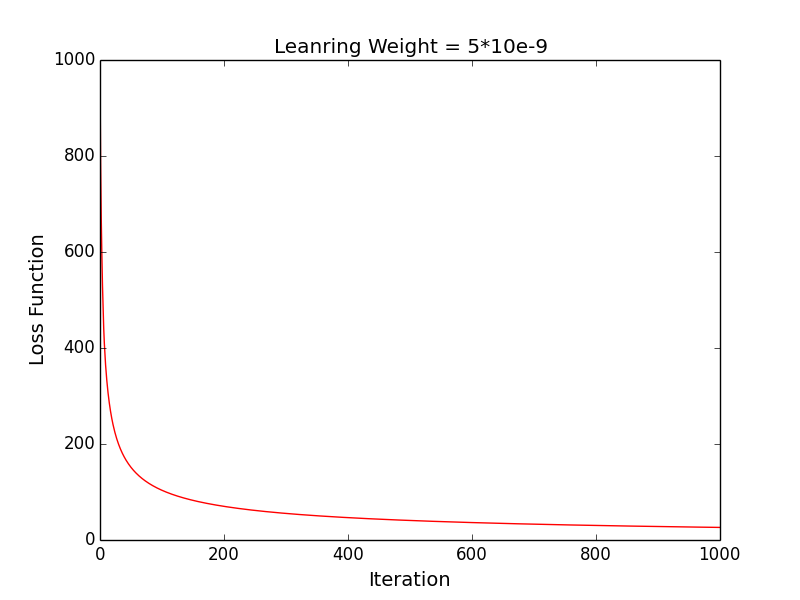
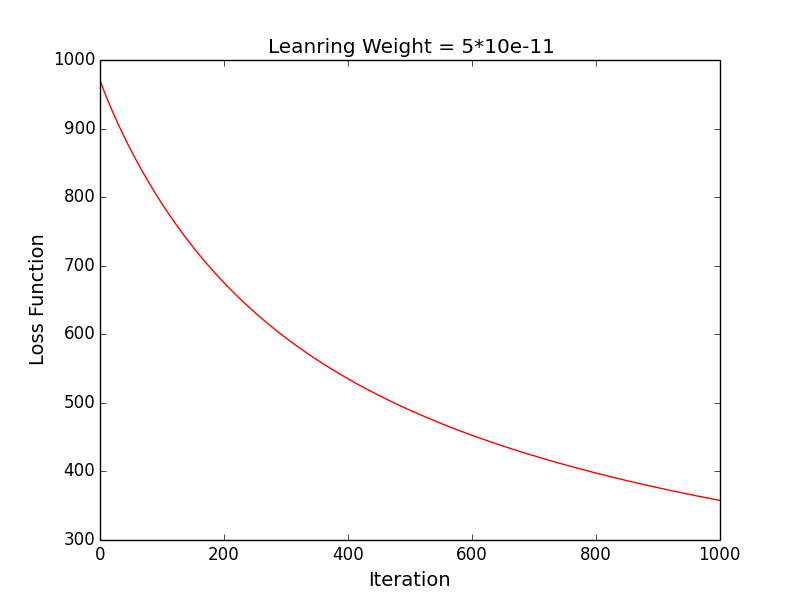
1. Implement with different learning rates of and plot objective function (Loss function) as a function of iteration. The plots are show below.





While we set learning rate as , we observed that the objective function fluctuated at small iteration; if we set learning rate as , we observed the objective function decreases slowly, which means slow convergence during gradient descent. When we set learning rate as , the objective function converged fast and no fluctuation so we choose as our final learning rate.

1. For each gradient iteration, plot the training accuracy and the test accuracy as a function of the number of gradient descent iterations by using the learning rate as . The plots are shown below.

3.

**Gradient of new objective function:**

d =

**Pseudo Code for modified algorithm:**

Given training example (xi, yi), I = 1, …, N

Let w 🡨 (0,0,0….,0)

Repeat until convergence encountered:

4.