CS168 Project 3

Henry Lin, Kaylee Bement

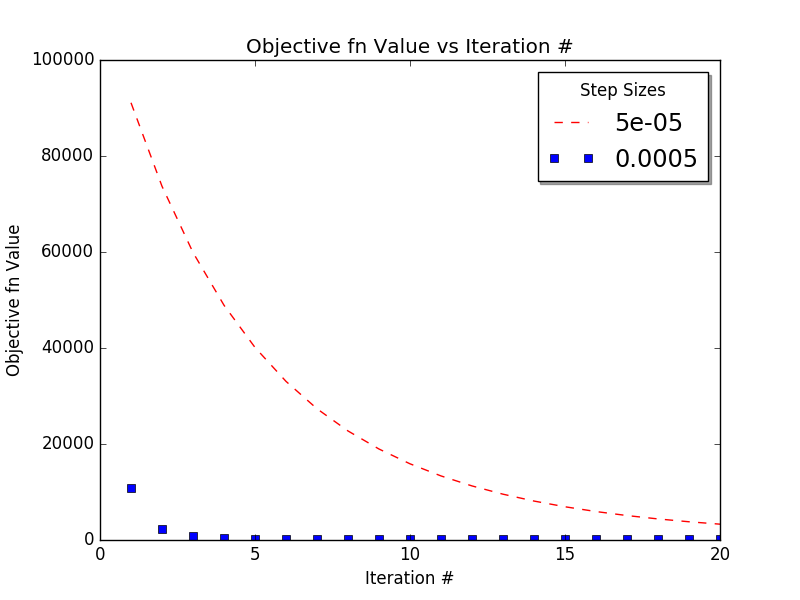
1.

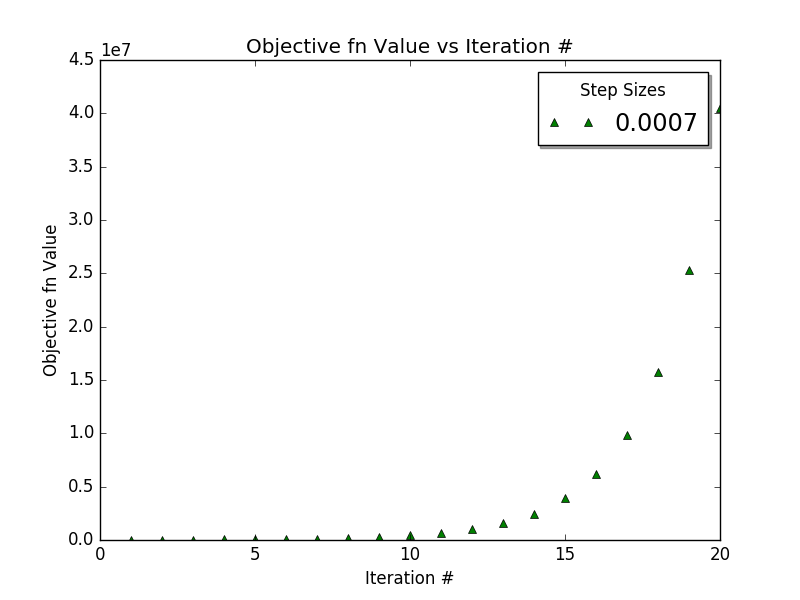
A. Objective function value: 220.34264027

Objective function value with a consisting of all 0s: 91827.65959497

Since there is randomness in some of the function values, these numbers fluctuate, but are close to the numbers given.

B. See code.py





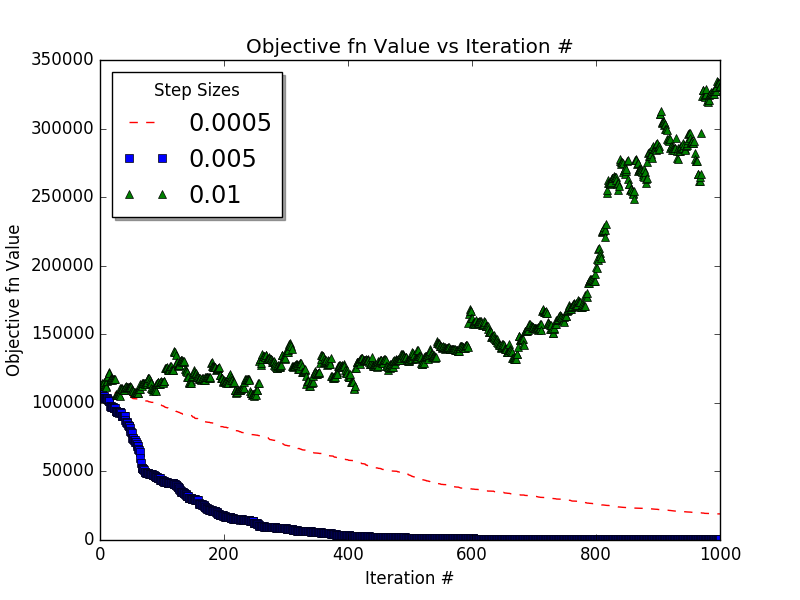
Gradient of f at at:

Optimal step size: 0.0005 with final objective function value of 233.09586745

The step size has a great impact on the convergence of gradient descent. If the step size is too small, gradient descent might be heading in the right direction, but never hit convergence, such as with step size = 0.00005, or it might find a local minimum and converge at the wrong value. If the step size is too large, gradient descent might overstep the minimum and step towards an incorrect local minimum, such as with step size 0.0007. However, step size 0.0005 seemed to work well, and gave us an objective function value close to the one in part A.

C. See code.py

Optimal step size: 0.005 with final objective function value of 499.53798953



Step size influences the convergence of stochastic gradient descent in the same way it affects regular gradient descent, but it takes more iterations for SGD to converge. Regular gradient descent has a better final value than SGD, but in SGD, each data point is only used once on average since there are 1000 iterations to choose a random point and there are 1000 data points, while in regular gradient descent, each data point is used 20 times since every point is used in each iteration.