James Hahn
CS1675 Machine Learning
Dr. Adriana Kovashka

Results

Below are the L2 errors for the closed-form linear regression solution, as well as the gradient descent solutions with learning rates 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}. Error rates are the average of the error rates from five different simulations. Simulations were ran with 50 iterations across the Red Wine dataset (http://archive.ics.uci.edu/ml/datasets/Wine+Quality), which contains 1599 samples. The training set consisted of 50% of the data, or 800 samples, while the testing set consisted of the remaining samples. No validation set was used in the training process.

Solution	**L2 Error**
closed-form	18.3486
gradient descent, $\eta = 10^{-1}$	18.5002
gradient descent, $\eta = 10^{-2}$	96.9866
gradient descent, $\eta = 10^{-3}$	148.2568
gradient descent, $\eta = 10^{-4}$	158.3877
gradient descent, $\eta = 10^{-5}$	163.2096
gradient descent, $\eta = 10^{-6}$	153.6968

Observations

Clearly, the closed-form solution is the optimal case for our least squares regression algorithm. Additionally, as our learning rate decreases, the error rate increases; they are inversely proportional. This can be explained since the number of iterations for gradient descent is static at 50. Therefore, as the learning rate decreases, the algorithm never has the chance to converge to a globally or locally optimal set of weights. If the number of iterations was increased to 100k, error rates for gradient descent with the lower learning rates, such as 10^{-6}, would actually perform better than learning rate 10^{-1}, but it would take much longer to compute. Finally, we used regular gradient descent instead of stochastic or batch gradient descent, which clearly impacted our error rate; it's higher than the other two algorithms, but much faster.