Regressions And Stuff

1 Introduction

In this paper, we discuss various methods of calculating regression. First, we discuss the benfits of using Stochastic Gradient Descent versus Batch Gradient Descent. Next, we analyze the performance of these methods in performing a Least Squares Linear regression. Finally, we discuss blah blah

2 Overview

3 Gradient Descent

Gradient descent is an iterative optimization algorithm; in other words, it calculates the parameters $\mathbf{x}$ that minimize a given objective function $f(\mathbf{x})$. The algorithm repeatedly translates an initial *guess* $\mathbf{x}_0$ in a direction proportional to the negative *gradient* $\nabla f(x)$.

In every step of the itreration, we update our guess as following:

$$\mathbf{x}_{n+1} = \mathbf{x} - \lambda \nabla f(\mathbf{x}_n)$$

where λ is the *step size* of the iteration. The algorith terminates upon the following convergence condition:

$$|f(\mathbf{x}_{n+1}) - f(\mathbf{x}_n)| < \delta$$

where δ is the convergence *threshold*. Upon convergence, the algorithm returns a final guess of $\mathbf{x}_{\text{opt}} = \mathbf{x}_{n+1}$.

In this section, we compare two methods of gradient descent: *batch gradient descent*, which computes a gradient over all samples for each iteration, and *stochastic gradient descent*, which uses pointwise gradients instead.

To analyze the perforamnce of the different algorithms, we use the *Gaussian function*:

$$f(x) = -\frac{1}{\sqrt{(2\pi)^n|\sigma|}} \exp\left[-\frac{1}{2}(x-u)^T \sigma^{-1}(x-u)\right]$$

In addition, we use the *quadratic bowl function*:

$$f(x) = \frac{1}{2}x^T A x - x^T b$$

Finally, we use the *least squares error* function:

$$J(\theta) = |X\theta - y|^2$$

We analyze the affects of step size (λ), threshold (δ), and initial guess $\mathbf{x}_0$ on the performance and accuracy of the gradient descent algorithm.

3.1 Batch Gradient Descent

3.2 Stochastic Gradient Descent

3.3 Effect of Finite Gradient