

L^AT_EX Guidelines for Simple, Two-Column Papers

Edward A. Lee

6.337 Project Proposal
April, 19th 2017

University of California at Berkeley
Berkeley, CA, 94720, USA

eal@eecs.berkeley.edu

Abstract

Abstract for algorithms used and performance on the bench mark problem.

loss between two steps is called loss decrement. The training algorithm stops when the specified criteria is met. There are multiple training algorithms that are available to train a neural net. We consider three different algorithms to train a neural net; Gradient descent, Quasi Newton (BFGS), Adam-Optimizer.

1 Introduction

Machine learning and optimization has found home in some of the most important machine learning (ML) applications. The fields of machine learning and mathematical programming are becoming increasingly intertwined. Optimization problems lie at the heart of most machine learning approaches. Many optimization algorithms that are applied in machine learning problems converge at different rates. Many of the learning algorithms also have different memory footprint and different computational complexity. Studying how different optimization algorithms work when applied in ML applications is therefore an interesting field of research.

The work here explores different optimization schemes that are used in training weight matrices in a Neural net that uses logistic regression. The learning problem in neural nets is formulated in terms of the minimization of a loss function. We can combine the weights and the biases in a neural net in the loss function weight vector w . The loss function is then $f(w)$ with w^* as its minima. The learning problem in neural nets is basically searching for a weight vector w^* such that the value of the loss function $f(w)$ is minimum at w^* . This means that if we evaluate the value of the gradient of $f(w)$ at w^* , the gradient is zero.

The loss function in general is a multi-dimensional non-linear function. In order to minimize the loss function, typically iterative optimization methods are used. The iterative algorithm calculates the loss at each iteration, the change of

The algorithms mentioned above are applied on a simple problem first to make sure that they are functioning properly. They are then used on a benchmark problem which is the classification of handwritten digits in a MNIST data set using a neural net that uses logistic regression. The next section describes each of the algorithms in detail along with the pseudocode and their performance characteristics.

2 Algorithms

2.1 Gradient Descent

2.1.1 Pseudocode

2.1.2 Performance characteristics

2.2 Adam-Optimizer

2.2.1 Pseudocode

2.2.2 Performance characteristics

2.3 BFGS

2.3.1 Pseudocode

2.3.2 Performance characteristics

2.4 Benchmark Problem

2.5 Numerical results and discussion