Indian Institute of Technology, Indore

Adam Optimiser: Analysis and Implementation

Proposal (CS 357)

Department of Computer Science and Engineering

Submitted by –

Keshav Goyal (150001014) Dhruv Chadha (150001009)

Under the Guidance of **Dr. Kapil Ahuja**

Introduction

Stochastic gradient-based optimization is of core practical importance in many fields of science and engineering. Many problems in these fields can be cast as the optimization of some scalar parameterized objective function requiring maximization or minimization with respect to its parameters. If the function is differentiable w.r.t. its parameters, gradient descent is a relatively efficient optimization method, since the computation of first-order partial derivatives w.r.t. all the parameters is of the same computational complexity as just evaluating the function. Often, objective functions are stochastic.

Algorithm:

Adam is a method for efficient stochastic optimization that only requires first-order gradients with little memory requirement. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients. The name Adam is derived from adaptive moment estimation. The method is straightforward to implement, is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients, and is well suited for problems that are large in terms of data and/or parameters.

The method combines the advantages of two recently popular optimization methods: the ability of AdaGrad to deal with sparse gradients, and the ability of RMSProp to deal with non-stationary objectives

Aim

Implementation:

We will implement our own version of Adam optimiser, and try to perform all the necessary background mathematics involved. We will also compare its performance to the popular library implementations of the algorithm.

Analysis:

We will investigate different popular machine learning models, including logistic regression, multilayer fully connected neural networks and deep convolutional neural networks. Using large models and datasets, we will demonstrate how Adam can efficiently solve practical deep learning problems.