A template for the arxiv style

A Preprint

Ekaterina Statkevich
MIPT
statkevichk@bk.ru

Aleksandr Beznosikov
MIPT

Abstract

This work considers a regularization for such algorithms as Adam, OASIS for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The main difference from gradient descent is that Adam's and OASIS algorithm use information about previous gradients to update parameters of the model.

Keywords First keyword · Second keyword · More

1 Introduction

At first we should consider unified description of the methods $w_{k+1} = w_k - \eta_k D_k^{-1} \nabla f(w_k)$ These methods are examples of the last stage of development of gradient descent. They are faster due to keeping information about previous iterations, because of which they choose a more efficient direction of descent and will work quicker in the case of a ill-conditioned quadratic problem. Also, simple gradient descent will "wobble"in case of, for example, a saddle point, and these methods, thanks to the hessian in the case of OASIS and the gradient square in the case of Adam, will be able to hit it.

Список литературы

Stochastic Gradient Methods with Preconditioned Updates (Abdurakhmon Sadiev, Aleksandr Beznosikov, Abdulla Jasem Almansoori, Dmitry Kamzolov, Rachael Tappenden, Martin Takáč)

Doubly Adaptive Scaled Algorithm for Machine Learning Using Second-Order Information (Majid Jahani, Sergey Rusakov, Zheng Shi, Peter Richtárik, Michael W. Mahoney, Martin Takáč)

Adam: A Method for Stochastic Optimization (Diederik P. Kingma, Jimmy Ba)