

Comparison of Gradient Descent, Heavy Ball GD & NAGD algorithms for binary classification.

A . Suresh Varma - 21848
I . B . V . Lakshmana Raju - 21190

Introduction



Problem Statement : The objective of this project is to compare convergence rates of various gradient based optimization methods for a binary classification problem.

Dataset Used : Diabetes.csv 750 Samples with 8 Features

Algorithms Used: Gradient Descent , Heavy Ball Gradient Descent ,Nesterov Accelerated GD

Loss Function is L_2 regularized cross entropy loss

$$f(w) = 1/m \sum_{i=1}^m y_i x_i^T w - \log(1 + \exp(x_i^T w)) + \lambda/2 ||w||^2 .$$

Here $f(w)$ is L - Smooth function with $L = \lambda + \max ||x_i||^2 / 4$, $i = 1, \dots, m$ (#samples) and μ - strongly convex with $\mu = \lambda$

From the given data set we obtained $L = 18.1$ & $\mu = \lambda = 0.1$.

Updated Equations & Convergence Rates

1. Gradient Descent

$$w_{t+1} = w_t - \eta \cdot \nabla w_t$$

2. Heavy Ball (Momentum Based) Gradient Descent

$$Update_t = \beta * Update_{t-1} + \eta * \nabla W_t$$

$$W_{t+1} = W_t - Update_t$$

3. Nesterov Accelerated Gradient Descent

$$y_{t+1} = w_t - \eta_t \nabla f(w_t)$$

$$z_{t+1} = z_t - \eta_t ((t+1)/2) \cdot \Delta f(w_t)$$

$$w_{t+1} = ((t+1)/(t+3))y_{t+1} - (2/(t+3))z_{t+1}$$

Method	Convergence
Gradient Descent	$k.O(\log(1/\epsilon))$
Heavy Ball GD	$\sqrt{k}.O(\log(1/\epsilon))$
Nestro GD	$\sqrt{k}.O(\log(1/\epsilon))$

Intuitive Implementation of NAGD



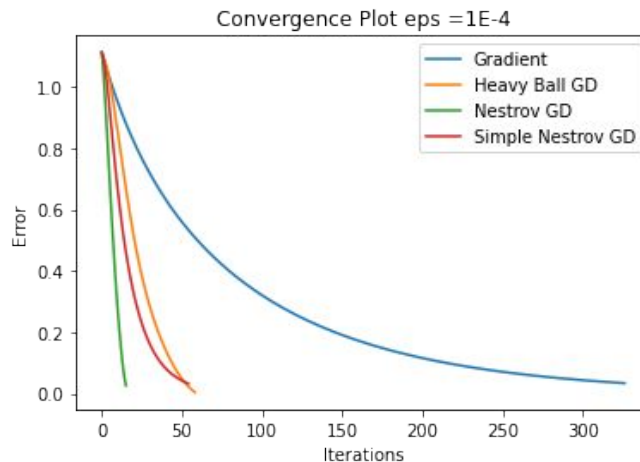
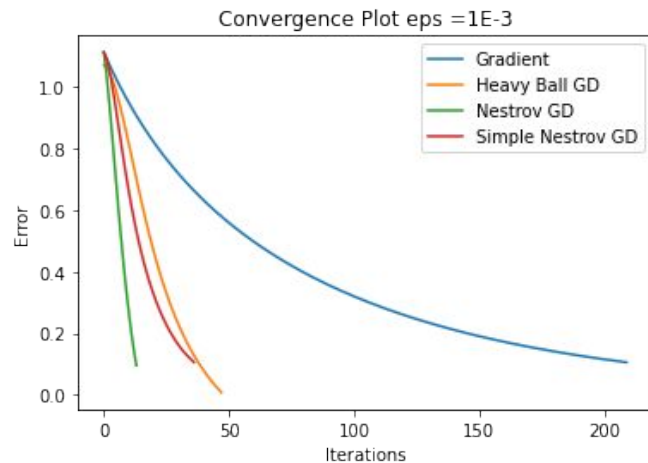
1. Implemented simple NAGD with the following update equations

$$\begin{aligned}w_{lookahead} &= w_t - \beta \cdot update_{t-1} \\update_t &= \beta \cdot update_{t-1} + \eta \nabla w_{lookahead} \\w_{t+1} &= w_t - update_t\end{aligned}$$

Observations

1. As expected Nesterov accelerated Gradient Descent having better convergence rate than other 2 methods .
2. The practical convergence rates of these algorithms are found to differ from theoretical analysis which may be due to imbalance in data set.

Results



E	GD Iterations	Heavy Ball GD Iterations	NAGD Iterations	Simple NAGD Iterations
1E-03	209	47	13	36
1E-4	326	58	15	54

Method	Step Size(α)	Momentum(β)
Gradient Descent	1/L	-
Heavy Ball GD	1/L	0.9
NAGD	0.3	-
Simple NAGD	1/L	0.8



Thank You