Stochastic Gradient Descent with Momentum and Line Searches

David Nardi
MSc student in Artificial Intelligence, Univeristy of Florence

26th January 2024

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Contents

1 Introduction **2**

2 Mini-batch gradient descent variants **2**

2.1 Fixed step-size 2

1 Introduction

2 Mini-batch gradient descent variants

2.1 Fixed step-size

Mini-batch Gradient Descent with fixed step-size

```
dati $w^0 \in \mathbb{R}^n$, $f(w) = \sum_{i=1}^{N} \log\big(1 + \exp(-y^{(i)} w^T x^{(i)})\big) + \lambda\|w\|^2$, $k = 0$ e $\{\alpha_k\} \mid \alpha_k = \alpha$
while ($\|\nabla f(w^k)\| > \varepsilon$)
 shuffle $\{1, \ldots, N\}$ in $N/M$ blocchi $B_1, \ldots, B_{N/M}$ di dimensione $1 < |B_t| = M \ll N$
 $y_0 = w^k$
 for $t = 1, \ldots, N/M$
  $y_t = y_{t-1} - \alpha_k \frac{1}{M} \sum_{j \in B_t} \nabla f_j(y_{t-1})$
 end for
 $w^{k+1} = y_{N/M}$
 $k = k + 1$ fine epoca
end while
```