Project 2

Håkon Olav Torvik, Vetle Vikenes & Sigurd Sørlie Rustad

FYS-STK4155 – Applied Data Analysis and Machine Learning
Autumn 2021
Department of Physics
University of Oslo

October 22, 2021

ABSTRACT: Abstract.

Contents

1 Introduction

2 Theory

Stochastic Gradient Decent

Stochastic gradient decent (SGD) is like the name might indicate, a stochastic approximation to the ordinary gradient decent (GD) method. Both methods are often used to minimize the so-called cost/loss-function. Thus, lets say we have a cost function $C(\boldsymbol{\beta})$ which could be expressed as

$$C(\boldsymbol{\beta}) = \sum_{i=1}^{n} c_i(\mathbf{x}_i, \boldsymbol{\beta}).$$

Where n is the number of datapoints and $\mathbf{x}$ are the datapoints. The gradient with respect to the parameters $\boldsymbol{\beta}$ is then defined as

$$\nabla_\beta C(\boldsymbol{\beta}) = \sum_{i=1}^{n} \nabla_\beta c(\mathbf{x}_i, \boldsymbol{\beta}).$$

The algorithm for GD is then:

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k - \eta \nabla_\beta C(\boldsymbol{\beta}_k), \tag{2.1}$$

where η is what we call the learning rate. This algorithms finds (ideally), with each iteration, new β_{k+1} values which decreases the cost function. This is of course not always the case, and depends on the value of η. If η is to big, our answer can diverge, and if η is very small we need a lot of iterations to reach the minima. Another challenge is that we can get trapped in local minima, meaning that we never reach the global minima.

Another challenge, which is reduced when using SGD, is the large number of computations. Instead of calculating every single derivative in (2.1) we only calculate a substrata.

3 Methods

4 Results

5 Discussion

6 Conclusion

A Appendix

References

[1] Morten Hjorth-Jensen, Computational Physics, Lecture Notes Fall 2015, August 2015, https://github.com/CompPhysics/ComputationalPhysics/blob/master/doc/Lectures/lectures2015.pdf.