Universität Hamburg
Department Informatik
Knowledge Technology, WTM

Comparison of Gradient Descent Optimization Methods for Neural Networks

Seminar Paper

Neural Networks

Ali Saleh

Matr.Nr. 6517831

3saleh@informatik.uni-hamburg.de

24.04.2017

Abstract

Gradient Descent is the most widely used optimization method for neural networks training. This paper aims to explore different algorithms to for gradient optimization. Using standard datasets and different neural network architectures time and complexity of the different algorithms will be compared and analyzed.

1 Introduction

Introducing the gradient, and the need to optimize it.

2 Gradient Computing And Optimization

Introducing the methods of computing gradient and at least 3 different algorithms for optimizing it. Those 3 algorithms will be

- Stochastic Gradient Descent is the standard gradient optimization algorithm, it will be used as the base line to compare other performances with.
- Adam (Adaptive Moment Estimation) [Diederik P. Kingma, 2015]
- Adagrad (Adaptive Subgradient Methods for Online Learning and Stochastic Optimization) [John Duchi, 2011]

3 Results of Optimization Methods

Build two neural network models (RNN, CNN) and use the 3 chosen optimization algorithms to train them on different datasets. Then show the results and do some analysis of those results.

The chosen datasets are:

- CIFAR-10 [Krizhevsky, 2009]: is a standard dataset of labeled images. The dataset has 10 classes of images. The dataset have been used in different architecture and is large enough to show difference with different algorithms.
- SVHN [Yuval Netzer, 2011]: is a dataset of images. It is similar to MNIST but the images comes from house numbers in Google Street View images. The dataset is much harder to classifym, and is order of magnitudes bigger than MNIST dataset. This will make it interesting to investigate and also big enough to show different algorithms performance.

3.1 Discussion

Discuss the pros and cons of each of the methods used, and how to choose between them.

4 Conclusion

Conclude the paper with final findings.

References

[Diederik P. Kingma, 2015] Diederik P. Kingma, J. L. B. (2015). Adam: A method for stochastic optimization. *ICLR*.

[John Duchi, 2011] John Duchi, Elad Hazan, Y. S. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *JMLR*.

[Krizhevsky, 2009] Krizhevsky, A. (2009). Learning multiple layers of features from tiny images.

[Yuval Netzer, 2011] Yuval Netzer, Tao Wang, A. C. A. B. B. W. A. Y. N. (2011). Reading digits in natural images with unsupervised feature learning.