ON DEVELOPING NEW ALGORITHMS FOR MACHINE LEARNING AND DEEP LEARNING

RAJESH KUMAR SHARMA

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2016

ON DEVELOPING NEW ALGORITHMS FOR MACHINE LEARNING AND DEEP LEARNING

by

RAJESH KUMAR SHARMA

Department of Mathematics

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

Indian Institute of Technology Delhi

OCTOBER 2016

Certificate

This is to certify that the thesis entitled **On Developing New Algorithms for Machine Learning and Deep Learning**, being submitted by **Rajesh Kumar Sharma** to the **Department of Mathematics, Indian Institute of Technology Delhi,** for the award of the degree of **Doctor of Philosophy**, is a bonafide research work done under my guidance and supervision.

The thesis has reached the standard fulfilling the requirements of the regulations relating to the degree. The results obtained in the thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Date:

Dr (Mrs.) B. Chandra
Professor
Department of Mathematics
Indian Institute of Technology Delhi
Hauz Khas, New Delhi -110016
India

Acknowledgements

I convey my deepest gratitude to my supervisor Prof. B. Chandra for her extra ordinary guidance, patient discussions, active support, and constant encouragement in completing this thesis. I thank her for introducing me to the exciting topic of Deep Learning.

Her in-depth knowledge, valuable comments and suggestions provided high-quality foundation for the present thesis. I am indebted to her for her untiring efforts, constant support and valuable guidance. No words can ever describe the efforts of Prof. B. Chandra for motivating me to set yardsticks and inspired me to achieve higher goals.

I thank the chairman Student Research Committee and its members for their valuable suggestions. I also take this opportunity to convey my gratitude to the chairman HPC committee and its members without which all this high end computations would not have been possible. I would also like to acknowledge the financial support provided by CSIR during my PhD work.

I thank my parents for their constant support.

Rajesh Kumar Sharma

Abstract

The thesis aims to address some important issues in the area of Machine Learning and Deep Learning. Achieving speed and accuracy at the same time is a difficult task in any classification task. The thesis proposes various new methods to enhance the performance of Machine Learning and Deep Learning techniques where the emphasis is not only on improving the classification accuracy, but also on reducing the training time.

Neural Network is one of the widely used Machine Learning techniques for pattern classification. Multilayer Perceptron (MLP) being one of the widely used Neural Network techniques, an attempt has been made in this thesis to reduce the size of weight matrices which requires high computational cost. In order to achieve this, the weight matrix is factorized into two smaller matrices of lower rank. This reduces the training time of MLP drastically. Trigonometric functions have also been used to restrict the weights in a certain range, which makes the proposed parameterization inherently max-norm regularized and leads to improved classification accuracy in comparison to the standard MLP.

Deep MLP requires higher computational cost due to the use of large number of hidden layers. Hence, the proposed parameterization has been extended to Deep MLP by introducing a speedup parameter which controls not only the speed of training but also the classification accuracy. Therefore, it maintains a balance between speedup and classification accuracy. It has been shown on benchmark datasets that the proposed parameterization provides eight fold reduction in training time along with significant improvement in the classification accuracy of Deep MLP.

Performance of Deep MLP heavily depends on the choice of appropriate learning rate for backpropagation. Optimizing the learning rate as a hyper-parameter requires training of several

Deep MLPs, which is computationally expensive. A method has been proposed in the thesis to compute the learning rate in an adaptive manner during the training itself. The proposed method uses error gradient and Laplacian score to compute the learning rate during every iteration. This removes the necessity of optimizing the learning rate as a hyper-parameter and leads to drastic reduction in the classification error of Deep MLP.

Initializing a Deep MLP by stacking the Denoising Autoencoders has a limitation that the noise level is kept fixed for the training of Denoising Autoencoder. The thesis introduces an adaptive noise schedule for training the Denoising Autoencoder where the noise level of individual input neurons is adapted during every training iteration. The performance of the proposed adaptive noise schedule is superior on several benchmark datasets as compared to the method with fixed noise level.

The concept of selecting relevant features aligns with the aim of the thesis to improve the classification accuracy while reducing the training time. Hence a novel feature selection method has also been proposed which uses Denoising Autoencoder with correlation based multiplicative aggregation function to select relevant features in an unsupervised manner. The features selected by the proposed unsupervised method not only preserve the structure of the dataset, but also outperform other unsupervised feature selection methods with regard to the classification accuracy.

Extreme Learning Machine (ELM) is a method in Machine Learning for training the single hidden layer MLP. ELM is orders of magnitude faster than backpropagation but has a limitation that data dependent training of input to hidden weights is not performed. In order to improve the performance of ELM, a method has been proposed to initialize the weights in ELM by using the weights of trained Denoising Autoencoder. A new method has been proposed for training the Denoising Autoencoder in parallel for each minibatch which reduces the execution time of the proposed method drastically. Another ELM has also been proposed in the complex domain which removes the drawbacks of previously proposed complex-valued ELM in the literature.

Contents

List of Figures

List of Tables