

**ON DEVELOPING NEW ALGORITHMS FOR MACHINE
LEARNING AND DEEP LEARNING**

RAJESH KUMAR SHARMA



**DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2016**

©Indian Institute of Technology Delhi (IITD), New Delhi, 2016

**ON DEVELOPING NEW ALGORITHMS FOR MACHINE
LEARNING AND DEEP LEARNING**

by

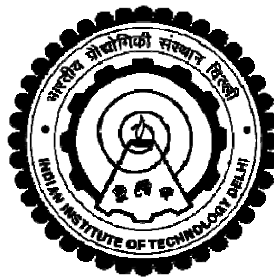
RAJESH KUMAR SHARMA

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

Certificate	i
Acknowledgements	ii
Abstract	iii
List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Introduction to Machine Learning.....	1
1.1.1 Artificial Neural Network.....	1
1.2 Deep Learning	3
1.3 Speeding up of Deep MLP	5
1.4 Methods for varying the Learning Rate in Deep MLP.....	6
1.5 Extreme Learning Machine	8
1.6 Pre-Processing using Feature Selection	9
1.7 Chapters in the Thesis	10
1.7.1 Summary of Chapters	11
2 Improving the Speed of MLP using Parameterized Weights	15
2.1 Parameterized Multilayer Perceptron.....	16
2.2 Forward pass and Backpropagation	17
2.3 Weight update for proposed parameterization	19
2.4 Speed Gain	22
2.5 Results	24
2.5.1 Description of image datasets.....	24
2.5.2 Implementation for image datasets.....	25
2.5.3 Description of Gene Expression datasets	28
2.5.4 Implementation for Gene expression datasets	29
2.6 Conclusions	30
3 Parameterized Weight based Deep Neural Network	31
3.1 Introduction to Deep Learning	31
3.1.1 Denoising Autoencoder	33
3.1.2 Stacked Denoising Autoencoder	35
3.2 Overview of existing work on speeding up Deep Neural Network	37

3.3 Parameterized Weight Deep Neural Network (PWDNN).....	38
3.4 Properties of PWDNN.....	40
3.5 Advantages of PWDNN	44
3.6 Results	47
3.6.1 Datasets.....	47
3.6.2 Implementation Details.....	48
3.6.3 Discussion of the PWDNN results	51
3.6.4 Analysis of Weight Factorization and Periodic-bounded Weights	52
3.7 Conclusions	53
4 Deep Learning with Adaptive Learning Rate using Laplacian Score.....	54
4.1 Deep Learning techniques for supervised training without pre-training.....	54
4.2 Existing methods for computing the Learning rate.....	56
4.3 Proposed Method.....	57
4.4 Illustration of Laplacian Score for Computing the Significance of Neurons.....	59
4.5 Update rule for learning parameter	60
4.6 Algorithm for the proposed method	61
4.7 Results	63
4.7.1 Implementation Details.....	63
4.8. Conclusions	67
5 Adaptive Noise schedule in Denoising Autoencoder for Deep Learning.....	68
5.1 Adaptive Stacked Denoising Autoencoder (ASDA).....	68
5.1.1 Motivation for using noise annealing schedule	69
5.1.2 Motivation for using different noise for different input neurons.....	70
5.1.3 Description of ASDA	72
5.2 Results	75
5.3 Conclusions	77
6 Feature Selection using Denoising Autoencoder	78
6.1 Feature Selection.....	78
6.1.1 Overview of previous work on unsupervised feature selection.....	79
6.2 Multiplicative Autoencoder based Feature Selection (MAFS).....	80
6.2.1 Multiplicative Aggregation Function based Autoencoder.....	81
6.3 Feature Selection using Multiplicative aggregation function Autoencoder.....	86

6.3.1 Masking procedure	87
6.3.2 Detailed description of Multiplicative Autoencoder based Feature Selection (MAFS).....	87
6.4 Results	90
6.4.1 Feature Selection for Image Datasets	92
6.4.2 Feature Selection for Gene Expression Dataset	94
6.5 Conclusions	96
7 Improving the performance of Extreme Learning Machine and its Variations	97
7.1 Overview of Extreme Learning Machine	98
7.2 Parallel-Denoising Autoencoder-Extreme Learning Machine (P-DAE-ELM).....	99
7.2.1 Detailed Description of P-DAE-ELM	100
7.2.2 Parallelization of P-DAE-ELM	105
7.2.3 Results	106
7.3 Multi Projection Complex ELM (MP-CELM).....	107
7.3.1 Detailed Description of MP-CELM	108
7.3.2 Results	112
7.4 Conclusions	114
8 Conclusions	116
References	118
List of Symbols Used.....	125
Publications	129
About the Author	130

List of Figures

2.1 Multilayer Perceptron	16
2.2 Update values of incoming weights to various hidden neurons	21
2.3 Update values of incoming weights to various output neurons	21
2.4 Testing accuracy for various values of K, with 500 hidden neurons.....	27
2.5 Testing accuracy for various values of K, with 1000 hidden neurons.....	27
3.1 Denoising Autoencoder	35
3.2 Greedy Layer-wise pre-training using Denoising Autoencoders.....	36
3.3 Fine-tuning of Deep Neural Network using class labels	36
3.4 Kernel density estimate of weights	41
3.5 Training, validation and testing errors of Rotation dataset for 1000 epochs.....	45
3.6 Plots of misclassification error and SGR for various values of alpha.	51
4.1 Algorithm for proposed adaptive learning rate.....	62
5.1 Effect of noise on search neighborhood and reconstruction accuracy.....	70
5.2 Relation between noise annealing schedule and noise of individual input neurons	71
5.3 Graphical depiction of ASDA.....	73
5.4 Algorithm for training Denoising Autoencoder with Adaptive noise	74
6.1 Multiplicative aggregation function based autoencoder	82
6.2 Multiplicative Autoencoder based Feature Selection (MAFS).....	90
7.1 Plots of sigmoid and modified sigmoid functions and their inverse.....	101
7.2 Graphical depiction of P-DAE-ELM.	104
7.3 The proposed P-DAE-ELM algorithm.....	104
7.4 Direct extension of real-valued input to complex domain.....	109

7.5 Graphical depiction of MP-CELM	111
7.6 Algorithm for MP-CELM.....	111

List of Tables

2.1 Description of Image Dataset.....	24
2.2 Classification accuracy (%) and runtime (sec) of MLP and PMLP for 500 neurons in the hidden layer.....	25
2.3 Classification accuracy (%) and runtime (sec) of MLP and PMLP for 1000 neurons in the hidden layer.....	26
2.4 Description of Gene Expression Datasets.....	29
2.5 Results for 500 hidden neurons.....	29
2.6 Results for 1000 hidden neurons.....	30
3.1 Description of various datasets	48
3.2 Classification error of various datasets	49
3.3 Results for weight factorization and periodic-bounded weights.....	52
4.1 Classification error (in %) and run time (in seconds) for Rectified Linear Network.	65
4.2 Classification error (in %) and run time (in seconds) for Maxout Network.....	65
4.3 Classification error (in percentage) and run time (in seconds) for R-EBP-P.	65
5.1 Classification error (%) for various datasets.....	76
6.1 Details of image datasets	92
6.2 Jaccard Index of image datasets.....	93
6.3 Classification accuracy of Multilayer Perceptron for image datasets.....	93
6.4 Details of gene expression datasets.....	94
6.5 Jaccard Index of gene expression datasets.....	95
6.6 Classification accuracy of Multilayer Perceptron for gene expression datasets	95
7.1 Classification Accuracy (in %) of testing dataset for optimum hyper-parameters.....	107
7.2 Description of datasets.....	112

7.3 Classification accuracy for various datasets	113
7.4 p-value for t-test	114