

*Sparse Support Matrix Machines
for the Classification of Corrupted Data*

Muhammad Imran Razzak

School of Computer Science
Faculty of Engg. & IT
University of Technology Sydney
NSW - 2019, Australia

Sparse Support Matrix Machines for the Classification of Corrupted Data

*A thesis submitted in partial fulfilment of the requirements
for the degree of*

Doctor of Philosophy

in
Analytics

by

Muhammad Imran Razzak

to

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

November 2019

AUTHOR'S DECLARATION

I, *Muhammad Imran Razzak* declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science, Faculty of Engineering and Information Technology* at the University of Technology Sydney, Australia, is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Production Note:

SIGNATURE: Signature removed prior to publication.

[Muhammad Imran Razzak]

DATE: 22nd November, 2019

PLACE: Sydney, Australia

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Guandong Xu, for providing continuous support and motivation through out PhD.

DEDICATION

To my family ...

LIST OF PUBLICATIONS

PUBLISHED/ACCEPTED

1. Imran Razzak, Raghil Abu Saris, Michael Blumenstein, and Guandong Xu, "Integrating Joint Feature Selection into Subspace Learning: A Formulation of 2DPCA for Outliers Robust Feature Selection." *Neural Networks (Elsevier)*, doi.org/10.1016/j.neunet.2019.08.030, [**Core rank: A/JCR IF. 7.197/Q1**] [80] .
2. Imran Razzak, Michael Blumenstein, Guandong Xu, Multi-Class Support Matrix Machines by Maximizing the Inter-class Margin for Single Trial EEG Classification, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, doi: 10.1109/TNSRE.2019.2913142 [**Core rank: A*/JCR IF. 3.478/Q1**][73].
3. Imran Razzak, Raghil Abu Saris, Michael Blumenstein, and Guandong Xu. "Robust 2D Joint Sparse Principal Component Analysis With F-Norm Minimization For Sparse Modelling: 2D-RJSPCA." *International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1-7 [**Conference core rank: A**] [85].
4. Imran Razzak, Ibrahim A Hameed, Michael Blumenstein and Guandong Xu, "Sparse Representation and Support Matrix Machines for Epileptic EEG Signal Classification" *Journal of Translational Engineering in Health and Medicine*, doi:10.1109/JTEHM.2019.2942017 [**JCR IF. 2.05/Q2**] [74].
5. Imran Razzak, Muhammad Imran, Guandong Xu, Efficient Brain Tumor Segmentation with Multiscale Two-Pathway-Group Conventional Neural Networks, *IEEE Journal of Biomedical and Health Informatics*, 2018 Oct 4. doi: 10.1109/JBHI.2018.287403 [**Core rank: A*/JCR IF. 4.217/Q1**] [76].
6. Imran Razzak, Muhammad Imran, Guandong Xu, "Big Data Analytics for Preventive Medicine", *Neural Computing and Application*, DOI <https://doi.org/10.1007/s00521-019-04095-y> [**JCR IF. 4.664/Q1**] [82].

-
7. Zafar Saeed, Rabeeh Abbasi, Imran Razzak, Guandong Xu, "Event Detection in Twitter Stream using Weighted Dynamic Heartbeat Graph Approach", IEEE Computational Intelligence Magazine, 14(3): 29-38 (2019) **JCR IF. 5.857/Q1**[92]
 8. Zafar Saeed, Rabeeh Ayaz Abbasi, Imran Razzak, Onaiza Maqbool, Abida Sadaf, Guandong Xu, "Enhanced Heartbeat Graph based Temporal Networks for Emerging Event Detection on Twitter" Vol, 136, pp 115, 132, Expert System with Applications, [**JCR IF. 4.292/Q1**][91]
 9. Zafar Saeed, Rabeeh Abbasi, Imran Razzak, Guandong Xu, Text Stream to Temporal Network - A Dynamic Heartbeat Graph to Detect Emerging Events on Twitter, 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2018. (**Conference core rank: A**)[93]
 10. Amina Naseer, Monahil Rani, Saeeda Naz, Muhammad Imran Razzak, Muhammad Imran, Guandong Xu, "Refining Parkinson's Neurological Disorder Identification Through Deep Transfer Learning", Neural Computing and Applications", doi.org/10.1007/s00521-019-04069-0 [**JCR IF. 4.664/Q1**][53]
 11. Zafar Saeed, Rabeeh Ayaz Abbasi, Maqbool Onaiza, Sadaf Abida, Muhammad Imran Razzak, Daud Ali, Naif Radi Aljohani, Guandong Xu. "Whats Happening around the World? A Survey and Framework on Event Detection Techniques on Twitter", Journal of Grid Computing, Springer, doi.org/10.1007/s10723-019-09482-2 [**JCR IF. 3.288/Q1**] [90]

SUBMITTED/REVISION

12. Imran Razzak, Muhammad Imran, Michael Blumenstein, Guandong Xu, Robust Sparse Support Matrix Machines for Single Trial EEG Classification, Artificial Intelligence in Medicine *Minor Revision* [**Core rank: A/JCR IF. 3.574/Q1**][75]
13. Imran Razzak, Michael Blumenstein, Guandong Xu, Robust Support Matrix Machine for Classification of Corrupted Data, IEEE Transactions on Pattern Analysis and Machine Intelligence, *In Revision*, [**Core rank: A*/JCR IF. 17.730/Q1**] [78]
14. Imran Razzak, Muhammad Khurram, Muhammad Imran, Michael Blumenstein, and Guandong Xu, Randomized Nonlinear One-Class Support Vector Machines with Bounded Loss Function for Outliers Detection, Future Generation Computer Systems, *Minor Revision* [**Core rank: A/JCR IF. 5.768/Q1**] [33].

-
15. Imran Razzak, Raghil Abu Saris, Michael Blumenstein, Guandong Xu, Robust Two Dimensional Joint Sparse PCA with F-norm Minimization, IEEE Transactions on Image Processing [**Core rank: A*/JCR IF. 6.79/Q1**] [81]
 16. Imran Razzak, Michael Blumenstein, Guandong Xu, Support Matrix Machine via Joint L2 and Nuclear Norm Minimization Under Matrix completion Framework for Classification of Corrupted Data, IEEE Transactions on Pattern Analysis and Machine Intelligence, [**Core rank: A*/JCR IF. 17.730/Q1**][79]
 17. Imran Razzak, Khurram Zafar, Michael Blumenstein, Guandong Xu, One-Class Support Tensor Machines with Bounded Hinge Loss Function for Classification of High-dimensional Data, Future Generation Computer Systems, *Minor Revision* [**Core rank: A/JCR IF. 5.768/Q1**] [77].

TABLE OF CONTENTS

List of Publications	vii
List of Figures	xv
List of Tables	xix
1 Introduction	1
1.1 Background	3
1.2 Motivation	5
1.3 Aims	6
1.4 Objectives	6
1.5 Research Question	6
1.6 Organization and Contributions	7
1.7 Thesis Organization	9
2 Background Knowledge	13
2.1 Notations	13
2.2 Proximal Algorithm	17
2.2.1 Proximal Operator Nuclear Norm	17
2.2.2 Proximal Operator $\ell_{p,q}$ norm	17
2.3 PCA	18
2.4 Support Vector Machines	22
2.4.1 Support Matrix Machine	22
2.4.2 One-Class Support Vector Machines	23
2.4.3 Multiclass Support Vector Machine	25
2.4.4 One-Class Support Tensor Machines	26
3 Related Work	29
3.1 Dimensionality Reduction	29

TABLE OF CONTENTS

3.2	Support Matrix Machines	33
3.2.1	Support Tensor Machines	35
3.3	Summary	36
I	Dimensionality Reduction and Feature Selection	39
4	Joint Feature Selection	41
4.1	Motivation	42
4.2	Outliers Robust 2DPCA	42
4.2.1	Objective Function	43
4.2.2	Optimization	44
4.2.3	Numerical Algorithm	46
4.2.4	Convergence Analysis	46
4.2.5	Connections to Other PCA algorithm	48
4.3	Experimental Results	49
4.3.1	Datasets	49
4.3.2	Parameter Selection	51
4.3.3	Evaluation on Original Dataset	53
4.3.4	Evaluation on Corrupted Dataset	53
4.3.5	Computational Complexity	53
4.3.6	Convergence Verification	54
4.4	Discussion	54
4.4.1	Reconstruction Error	56
4.4.2	Observations	56
4.5	Summary	57
5	Joint Dimensionality Reduction and Sparse Feature Selection	59
5.0.1	Motivation	60
5.1	2D Robust Joint Sparse PCA	60
5.1.1	Objective Function	61
5.1.2	Convergence Analysis	68
5.1.3	Numerical Algorithm	69
5.2	Results and Analysis	70
5.2.1	Datasets	71
5.2.2	Parameter Selection	73

5.2.3	Evaluation on Original Datasets	77
5.2.4	Robustness against Outliers	78
5.2.5	Reconstruction Error	78
5.2.6	Computational Complexity	79
5.2.7	Observations	79
5.3	Summary	80
 II Regualizer Optimization		 83
6	Support Matrix Machine	85
6.1	Motivation	86
6.2	The proposed RSSM	86
6.2.1	Objective Function	86
6.2.2	Theoretical Justification	90
6.2.3	Empirical Risk Minimization	90
6.2.4	Numerical Algorithm	92
6.2.5	Convergence Analysis	95
6.2.6	Computational Complexity	95
6.3	Experimental Evaluation	96
6.3.1	Image Classification	97
6.3.2	EEG Classification	101
6.3.3	Parameter Selection	103
6.4	Discussion	103
6.5	Summary	106
7	Support Matrix Machine with Matrix Recovery Framework	109
7.1	Motivation	110
7.2	Problem Formulation	110
7.3	SMM with Matrix Recovery Framework	111
7.4	Dataset	117
7.4.1	Caltech Face Dataset	117
7.4.2	INRIA person dataset	118
7.4.3	BCI Competition	118
7.5	Result and Discussion	121
7.6	Conclusion	123

TABLE OF CONTENTS

8	MultiClass Support Matrix Machines	125
8.1	Motivation	126
8.2	Maximizing Inter-Class Margins for SMM	126
8.2.1	Objective Function	127
8.2.2	Learning Algorithm	128
8.2.3	Theoretical Justification	133
8.3	Experimental Evaluation	134
8.3.1	Dataset	134
8.3.2	Evaluation Metrics	135
8.3.3	EEG Preprocessing and Feature Extraction	135
8.3.4	Results	136
8.3.5	Parameter Setting	136
8.3.6	Computational Complexity	137
8.3.7	Discussion	138
8.4	Summary	139
	III Hinge Loss Optimization	145
9	One Class Support Tensor Machines	147
9.1	Motivation	148
9.2	Randomized Kernel Bounded One-Class STM	148
9.2.1	Bounding Loss Function	149
9.2.2	Optimization	150
9.2.3	Randomized Feature Embedding	154
9.2.4	Convergence	157
9.3	Experiments	158
9.3.1	Dataset	159
9.3.2	Results and Discussion	160
9.3.3	Parameter Setting	164
9.3.4	Computational Complexity	168
9.4	Summary	168
10	Conclusions and Future Direction	173
	Bibliography	179

LIST OF FIGURES

FIGURE	Page
1.1 Real-world data in the form of matrix	3
1.2 Organisation of thesis, key contributions (publications) are marked with ★	7
2.1 (a) Original dataset (b)PC1 vs PC2 (c) Original data along a pair of lines (d) PC2 plotting shows small loss since it it contributes the least to the variation in the data set. We can notice more variation in PC1 as compared to PC2	19
2.2 3D Visualization of PCA. We can notice more variatin in PC1 as compared to PC2 and PC3	20
2.3 (top) Linearly separable data and (bottom) non linearly separable data	23
2.4 Simulation of one class support vector machine on linearly separable data point	25
2.5 Simulation of multiclass support vector machine	27
4.1 Sample images of CMU PIE, ORL, Yale and AR First two rows real dataset, Row 3 contaminated with block and Row 4 is contaminated with salt and chapter noise 15%	50
4.2 Classification performance at different value of λ for real (left) and contaminated (right) datasets	51
4.3 Comparative evaluation on real datasete (AR, Yale, ORL, and CMUIPIE)	54
4.4 Comparative evaluation on corrupted datasete (AR, Yale, ORL, and CMUIPIE)	55
4.5 Convergence curve of ORPCA on four datasets	55
5.1 Illustration of SPCA (left) and 2D-JSPCA (right) using 10 x 11 matrix: white block represents zero loadings and color block represents different features	61
5.2 Illustration of 2D-JSPCA, JSPCA, PCA on 350 data points including 70 outliers: Results shows robustness of 2D-JSPCA against outliers	63

5.3	Sample images of CMU PIE, COIL20, Yale [98] and AR [50] First two rows real dataset, row 3 contaminated with block and Row 4 is contaminated with salt and pepper noise 15%	71
5.4	Comparative evaluation at different value of λ_a (left column) and λ_b (right column) for real (top row) and contaminated (bottom row) datasete	72
5.5	Comparative evaluation on real datasete (AR, Yale, ORL, FERET, COIL20 and CMUIPIE	74
5.6	Comparative evaluation on corrupted datasete (AR, Yale, ORL, FERET, COIL20 and CMUIPIE	75
5.7	Comparison:Reconstruction Error versus features numbers (a) AR (b) ORL (c) Yale (d) COIL20	77
6.1	Four matrices with special structures: (a) sparse; (b) low-rank; (c) sparse and low-rank using ℓ_1 . (d) sparse and low-rank using $\ell_{2,1}$ (proposed). Various colors denote different numerical values and white color represents zero.	87
6.2	Convergence curve of RSSM	95
6.3	Sample images from Caltech Face dataset. Face images shows that the dataset is challenging due to different face appearance, expressions and lighting conditions etc	98
6.4	Comparative evaluation (accuracy) based on average classification accuracy on Caltech Face dataset	98
6.5	Comparative evaluation (accuracy) based on average classification accuracy on contaminated Caltech Face dataset	99
6.6	Sample images from INRIA person dataset. The human detection is challenging due to similar appearance of persons and human statues	99
6.7	Comparative evaluation (accuracy) based on average classification accuracy on INRIA person dataset	100
6.8	Comparative evaluation (accuracy) based on average classification accuracy on contaminated INRIA person dataset	100
6.9	Comparative evaluation (accuracy) based on average classification accuracy on BCI dataset	106
7.1	Motivation for joint low rank plus matrix recovery based classification for missing plus corrupted data	110
7.2	Effect of different parameters (τ , α_1 and α_2) values	119

7.3	Comparative evaluation of SVM, SMM, MSMM and SMMRe on IVa:top left to bottom right (left-hand vs right hand, left-hand vs feet, left-hand vs tongue, right-hand vs feet, right-hand vs tongue, feet vs tongue)	120
7.4	Comparative evaluation (accuracy) based on average classification accuracy on real (top) contaminated (bottom) Caltech Face dataset	122
7.5	Comparative evaluation (accuracy) based on average classification accuracy on real (top) Corrupted (bottom) INRIA person dataset	123
7.6	Convergence curve of SSMRe (objective function value (y-axis) vs iteration (x-axis)	124
8.1	Illustration of multiclass support matrix machine: For four classes, we need three parameters W_1, W_2, W_3 , and W_4 to maximize the inter-class margins . .	126
8.2	Illustration of proposed framework equipped with M-SSM for EEG signal classification	127
8.3	Convergence process of M-SMM on subject k3b and l1b of IIIa dataset	137
8.4	Behaviour of τ on on the classification performance for IIa and IIIa datasets	138
9.1	Randomized projection of matrix data	156
9.2	Performance comparison of proposed R1STM-BH with state of the art methods on Iris dataset	161
9.3	Performance comparison of proposed R1STM-BH with state of the art methods on Lungs dataset	161
9.4	Performance comparison of proposed R1STM-BH with state of the art methods on the task of face recognition (ORL dataset)	162
9.5	Performance comparison of proposed R1STM-BH with state of the art methods on contaminated ORL dataset)	162
9.6	Performance comparison of proposed R1STM-BH with state of the art methods with different level of corruption on ORL dataset	163

LIST OF TABLES

TABLE	Page
2.1 Notations and their description	14
4.1 Algorithmic procedure of ORPCA	46
4.2 Average classification accuracy (accuracy \pm corresponding standard deviation) on real dataset at optimal result of ORPCA	52
4.3 Comparative evaluation based on average classification accuracy ((accuracy \pm corresponding standard deviation)) on contaminated datasets at optimal result of ORPCA	52
4.4 Average Reconstruction Error ($\times 10^{-3}$) and corresponding standard deviation of each approach on the Extended Yale B,AR, and CMU PIE databases	57
5.1 Algorithmic procedure of 2D Joint PCA	70
5.2 Comparative evaluation based on average classification accuracy on real dataset at optimal result of 2DJSPCA	73
5.3 Comparative evaluation based on average classification accuracy on contaminated dataset at optimal result of 2DJSPCA	76
6.1 Algorithmic procedure of sparse support matrix machine	89
6.2 Summary of dataset.	97
6.3 Classification performance (accuracy) of different algorithms on dataset BCI 2b.	104
6.4 Comparative evaluation based on average classification accuracy on BCI 2a .	105
6.5 Comparative evaluation based on average classification accuracy on BCI III-IVa	105
7.1 Algorithmic procedure of proposed sparse support matrix machine under matrix recovery framework (SMMRe)	116
7.2 Summary of dataset.	118

7.3	Classification performance (accuracy) of different algorithms on dataset BCI 2b.	119
7.4	Comparative evaluation based on average classification accuracy on BCI 2a .	119
8.1	Algorithmic procedure of sparse support matrix machine	141
8.2	kappa/error rate %: classification performance of different algorithms on data-set IIIa	142
8.3	kappa/error rate%: classification performance of different algorithms on dataset IIa	142
8.4	Comparative evaluation of classification performance of different algorithms on IIIa data-set	142
8.5	Comparative evaluation of classification performance of different algorithms on IIa data-set	142
8.6	Comparison of average training and testing time (in seconds) on IIIa and IIa data-sets	143
9.1	Algorithmic procedure of OCSTM-BH	154
9.2	Average accuracy (%) and ACU (%) on Breast Cancer dataset with differnt training samples	160
9.3	Average accuracy (%) and ACU (%) on corrupted Breast Cancer dataset with different training samples	160
9.4	Average %age of test AUC on different datasets with sample size 2	165
9.5	Performance comparision of proposed R1STM-BH with state of the art methods on the task of handwritten digit recognition (MNIST dataset)	166
9.6	Computational and Space complexity analysis of proposed approach with state of the art methods	167
9.7	Performance evaluation (Accuracy, AUC and number of iteration) of R1STM-BH with different methods on different training sample size	170
9.8	Comparative evaluation of training time (sec), test time (sec) and number of iterations on Breast Cancer dataset	171

ABSTRACT

Data acquisition has improved substantially over recent years, with devices acquiring data at faster rates and increased resolution. The interpretation process, however, has only recently begun to benefit from computer technology and still struggling especially for high dimensional and noisy data. We are still short of tools to convert all such data to useful information. Traditional support vector machines (SVMs) require data to reshape each matrix into a vectors, which ultimately results in losing the important structural information of the originally featured matrix. On the other-hand, the classification of high dimensional domains poses significant challenges. In contrast, modern classification approaches such as support matrix machine assume that all entities within each input matrix can serve as the explanatory features for its label. These methods are able to capture explanatory features by regularizing the regression matrix to be low-rank. However, in real-world, the data is noisy and most of the features may be redundant as well as may be useless, which in turn affect the classification performance. Thus it is important to perform robust feature selection under robust metric learning to filter out redundant features and ignore the noisy data points for more interpretable modelling. To overcome this challenge, in this work, we have adapted two different approaches. The first problem we address is the issue of dimensionality reduction. In our first approach, we introduce two-dimensional outliers-robust principal component analysis (ORPCA) by imposing the joint constraints on the objective function (**chapter 4**). ORPCA relaxes the orthogonal constraints and penalizes the regression coefficient, thus, it selects most important features and in the meantime, it ignores the same features that have already been selected in other principal components. To overcome the data redundancy, we further extend ORPCA and introduced additional sparsity-inducing regularization that relaxes the orthogonal constraints resulting the joint features selection (**chapter 5**). The introduced regularization terms penalizes all regression coefficients corresponding to single feature as a whole to features jointly. Hence, 2D-JSPCA approximates to high-dimensional data in flexible manner as it has more freedom to learn low-dimensional space efficiently.

Since the nuclear norm is the best convex approximation of the matrix rank over the unit ball of matrices, this makes it more tractable to solve the resulting optimization problem. Inspired by this, in our second approach, we propose a new model to address the classification problem of high dimensionality data by jointly optimizing the both regularizer terms ($\|\cdot\|_{2,1}$ and $\|\cdot\|_*$) and hinge loss. In our first approach (**chapter 6**), we combine the hinge loss and regularization terms as spectral elastic net penalty. The regulariza-

tion term which promotes the structural sparsity and shares similar sparsity patterns across multiple predictors. It is a spectral extension of the conventional elastic net that combines the property of low-rank and joint sparsity together, to deal with complex high dimensional noisy data. Furthermore, it also leverages the structural information as well as the intrinsic structure of data and avoids the inevitable upper bound. The optimization problem for the RSMM is convex, non-smooth and non-differentiable, however, the combination of hinge loss, $\ell_{2,1}$ -norm and nuclear norm makes the problem nontrivial to be solved directly. To tackle this issue, we split the problem into sub-problems with the *Generalized Forward-Backward* (GFB) splitting approach to solve the optimization problem efficiently.

Support matrix machine is fragile to the presence of outliers: even few corrupted data points can arbitrarily alter the quality of the approximation, What if a fraction of columns are corrupted? Combining the recovery along with feature selection and classification could significantly improve the performance. We assume that the data consists of a low rank clean matrix plus a sparse noise matrix. We extended our work and present support matrix machine (**chapter 7**) based on matrix recovery framework under the incoherence and ambiguity conditions and able to recover intrinsic matrix of higher rank and recover data with much denser corruption. We perform matrix recovery, feature selection and classification through joint minimization of $\ell_{2,1}$ and nuclear norm. We assume that the data consists of a low rank clean matrix plus a sparse noise matrix i.e. the data matrix can be decomposed as $X = L + S$. S is the column-sparse matrix that corresponds to corrupted columns, thus at most αn columns are non zeros, L corresponds to non corrupted matrix, thus $rank(L) = r$ and $(1 - \alpha)n$ columns of matrix L are non zeros, corresponding to the outliers. Since the objective function is convex, non-smooth and non-differentiable, however, the combination of hinge loss, $\ell_{2,1}$ -norm and nuclear norm makes the problem nontrivial to be solved directly. To decouple the hinge loss and nuclear norm with respect to W in SMMRe, we have introduced an *auxiliary variable, and applied Lagrange multiplier*.

Although, above both methods takes full advantage of low rank assumption to exploit the strong correlation between columns and rows of each matrix and able to extract useful features, however, are originally built for binary classification problems. To improve the robustness against data that is rich in outliers, we further extend this problem and present a novel multiclass support matrix machine (**chapter 8**) by utilizing the maximization of the inter-class margins (i.e. margins between pairs of classes). The proposed model is a combination of binary hinge loss for models fitting, and elastic net penalty as a regularization on regression matrix. The binary hinge loss uses C matrices to simulate one-vs-one classifier of all classes rather than $\frac{c(c-1)}{2}$ models. The optimization problem is convex but non-smooth and non-differentiable, thus, stochastic gradient descent and the Nesterov methods cannot be applied (i.e. in convex optimization setting, sub gradient of the nuclear norm function cannot be used in standard descent approaches and as a result solving it directly is difficult). Thus, an alternative approach is required to solve it, we devise an alternating direction method (*GFB splitting*) that can handle an arbitrary non-differentiable with a proximal operator.

Several non-convex and bounded loss function has been presented to substitute the

hinge loss function in order to suppress the affect of outliers and improve the robustness of support vector machines. However, there is no work done for the improvement of one-class tensor machines. Furthermore, computational complexity of traditional support tensor machines is high and increases with the increase of training samples. Thus, it limits the applicability of OCSTM for large dataset. We consider one class support tensor machines and introduce a scalable algorithm for large dataset by replacing the traditional hinge loss with bounded loss function resulting in reduction of classification error caused by outliers (**chapter 9**). For larger dataset, we further used randomized features rather than finding the optimized support tensors which results in not only improving the robustness against outliers as well as significantly reduces the training time. To solve the corresponding optimization problem, we have presented *half quadratic optimization* to transform the objective function to same like traditional OCSTM, followed by solving it like a typical OCSTM optimization problem.

We demonstrate the significance and advantage of our methods on different available benchmark datasets such as person identification, face recognition and EEG classification. Results showed that our methods achieved significantly better performance both in terms of time and accuracy for solving the classification problem of highly correlated matrix data as compared to state-of-the-art methods.

