

**Support Vector Machines with
Dual Error Extensions for
Target Detection and
Object Recognition**

by

Hong Gunn Chew

B.E. (FIRST CLASS HONOURS), B.Sc,
The University of Tasmania

Thesis submitted for the degree of

Doctor of Philosophy

in

School of Electrical and Electronic Engineering,
Faculty of Engineering,
Computer and Mathematical Sciences
The University of Adelaide, Australia

2013

© 2013
Hong Gunn Chew
All Rights Reserved



Contents

Contents	iii
List of Figures	ix
List of Tables	xi
List of Acronyms	xiii
Glossary	xv
Abstract	xix
Statement of Originality	xxi
Acknowledgments	xxiii
Chapter 1. Introduction	1
1.1 The Target Detection Problem	1
1.2 Support Vector Machines	2
1.3 Aims of this Thesis	3
1.4 Outline of Thesis	4
1.5 Contributions of Thesis	5
1.5.1 Publication	8
Chapter 2. Dual-C Support Vector Machines	9
2.1 Introduction	9
2.2 Support Vector Machines	10
2.2.1 Formulation of C -parameter support vector machine (C -SVM)	10
2.2.2 Wolfe Dual	14

Contents

2.2.3	Decision function	15
2.2.4	Karush-Kuhn-Tucker conditions	16
2.2.5	Kernel discussions	17
2.3	Training SVMs with Uneven Class Sizes	18
2.3.1	Understanding the 2-D SVM training classification plots	21
2.4	Dual Error Weightings	21
2.4.1	Revisiting uneven class training set sizes	23
2.5	Experiments on dual- C -parameter support vector machine ($2C$ -SVM)	24
2.5.1	Experiment datasets	25
2.5.2	Converting images into data vectors	25
2.5.3	Training biasing	26
2.5.4	Uneven class size correction	28
2.5.5	Selection of error parameter C	31
2.6	Chapter Summary	32
Chapter 3. Dual-ν Support Vector Machines		33
3.1	Introduction	33
3.2	ν Support Vector Machines	34
3.2.1	Formulation of ν -parameter support vector machine (ν -SVM)	34
3.2.2	Wolfe Dual	35
3.2.3	Decision function	38
3.3	Significance of ν on the SVM	39
3.3.1	Training SVMs with Uneven Class Sizes	41
3.4	Dual- ν Support Vector Machines	43
3.4.1	Deriving the dual- ν -parameter support vector machine (2ν -SVM) problem	45
3.4.2	Wolfe Dual	46
3.4.3	Significance of ν_+ and ν_-	48
3.5	Experiments on 2ν -SVM	49
3.5.1	Ease of choosing ν	50

3.5.2	Training biasing	51
3.5.3	Uneven class size correction	52
3.5.4	Working range of ν	53
3.6	Chapter Summary	55
Chapter 4. The Relationship between 2ν-SVM and $2C$-SVM		57
4.1	Introduction	57
4.2	The 2ν -SVM and $2C$ -SVM Formulations	58
4.2.1	Restating 2ν -SVM and $2C$ -SVM problems	58
4.2.2	Moving solutions between the Dual and the Primal	61
4.3	Relationship between 2ν -SVM and $2C$ -SVM	63
4.3.1	Relating 2ν -SVM solutions to $2C$ -SVM	63
4.3.2	Relating $2C$ -SVM solutions to 2ν -SVM	66
4.3.3	Discussion on relationship	68
4.4	Experiments	69
4.4.1	Results discussion	71
4.5	Chapter Summary	71
Chapter 5. SVM Training Implementation		75
5.1	Introduction	75
5.2	2ν -SVM Initialisation Problem	76
5.2.1	Iterative decomposition optimisation method	77
5.2.2	Iterative decision variable initialisation	78
5.2.3	Choosing the update variable	80
5.2.4	Constraints of initialisation	80
5.2.5	Algorithm	83
5.3	2ν -SVM	84
5.3.1	Iterative decomposition training	85
5.3.2	Choosing optimising points	88
5.3.3	Constraints of 2ν -SVM	89

Contents

5.3.4	Algorithm	91
5.4	2C-SVM	93
5.4.1	Iterative decomposition training	93
5.4.2	Choosing optimising points	95
5.4.3	Constraints of 2C-SVM	96
5.4.4	Algorithm	98
5.5	Implementation Considerations	100
5.5.1	Kernel calculations	101
5.5.2	Initialisation calculations	101
5.5.3	Optimisation process	103
5.6	Performance Experiments	103
5.6.1	Results	104
5.7	Chapter Summary	105
Chapter 6. Multi-Category Classification		109
6.1	Introduction	109
6.2	Multi-category Classification Applications	110
6.2.1	Handwritten digit recognition	110
6.2.2	The handwritten digit dataset	111
6.3	Using 2ν -SVMs for Multi-Class Problems	111
6.3.1	Adapting one-against-rest strategy for SVM	113
6.4	Experiments and Results	115
6.4.1	Iterative selection of error parameters	116
6.4.2	Comparison of 2ν -SVM and C -SVM for one-against-rest	119
6.4.3	Misclassification by 2C-SVM and 2ν -SVM	121
6.5	The Reliability Metric	122
6.5.1	Reliability Metric performance	128
6.6	Chapter Summary	130
Chapter 7. Conclusion		133

7.1	Thesis Conclusions	133
7.2	Future Work	135
7.2.1	Optimising SVM training and testing using GPUs and FPGAs	135
7.2.2	Kernels from pre-processing	136
7.2.3	Relaxed constraints to reduce SV count	137
7.3	Chapter Summary	137
Appendix A. Karush-Kuhn-Tucker Conditions		139
A.1	C Support Vector Machine	139
A.2	Dual- C Support Vector Machine	140
A.3	ν Support Vector Machine	141
A.4	Dual- ν Support Vector Machine	142
Appendix B. Datasets		143
B.1	D_1 - Synthetic Dataset 1	143
B.2	D_2 - Synthetic Dataset 2	143
B.3	D_3 - Handwritten Digit Dataset	145
B.4	D_4 - Target Detection Dataset	147
Appendix C. Proofs		149
C.1	Proposition C.1	149
C.2	Proposition C.2	150
Bibliography		153

List of Figures

2.1	SVM solution of a linearly separable example.	11
2.2	SVM solution of a linearly in-separable example.	13
2.3	Geometric positions of support vectors.	18
2.4	Examples of Kernel mappings from 2-D data space	19
2.5	Bias effect with uneven training class size using C -SVM	21
2.6	Process of constructing a data vector from an image	26
2.7	Using C s to bias the SVM for a 2-D problem	26
2.8	Using C s to correct biasing due to uneven class sizes	29
2.9	Histogram of classification values ($f(\mathbf{x})$) of training points for D_1	30

3.1	ν -SVM training	36
3.2	Using ν s to bias the SVM for a 2D problem	51
3.3	Distribution of classification values ($f(\mathbf{x})$) of training points for D_1	54

5.1	Comparison of initialisation and optimisation times for the ad hoc and the iterative methods of initialising the decision variables	107
5.2	Comparison of training times for $2C$ -SVM using SMO, and for $2C$ -SVM and 2ν -SVM using the proposed implementation algorithm.	108

6.1	Process of constructing a data vector from an image	115
6.2	Comparison of test errors for C -SVM ($C = 100$) and 2ν -SVM ($\nu = 0.01$).	119

List of Figures

6.3	A digit classification with (a) 2ν -SVM and (b) C -SVM.	120
6.4	Images of digits misclassified by $2C$ -SVMs.	121
6.5	Images of digits misclassified by 2ν -SVMs.	122
6.6	Decision flow with reliability metric	123
6.7	Decision results for test sample 1.	124
6.8	Decision results for an image with multiple positive values.	124
6.9	Decision results for an image with all negative values.	125
6.10	Classification results for an image with the best two values at ± 1 . . .	126
6.11	Classification results for an image with two large positive values. . . .	126
6.12	Classification results for an image with the two highest values between ± 1	127
6.13	Classification results for an image with two positive values.	127
6.14	Characteristic of Reliability Metric	128
6.15	Reliability plots for Reliability Metric ($I_r = f_1 \cdot (-f_2)$). Histograms of (a) classification performance and (b) classification accuracy, and plots of (c) classification performance for data above and below the confidence index and (d) the ratio of data samples above the confi- dence index.	129

B.1	2-D visualisation of Dataset 1	144
B.2	2-D visualisation of Dataset 2	144
B.3	Sample of handwritten digits	146
B.4	Sample of Target Detection Images	148

List of Tables

2.1	Classification performance effects of biasing with $2C$ -SVM	27
2.2	Classification performance effects of biasing correction with $2C$ -SVM	31
3.1	Classification performance effects of biasing with 2ν -SVM	52
3.2	Classification performance effects of inherent biasing correction with 2ν -SVM	54
4.1	Classification performance comparison	70
4.2	Parameter transformation starting from $\nu_+ = \nu_- = 0.01$	73
5.1	Training times for the ad hoc and the iterative methods of initialising the decision variables	106
6.1	Classification performance comparison between 2ν -SVM and $2C$ -SVM	117
6.2	Training time comparison between 2ν -SVM and $2C$ -SVM	118
6.3	Classification values comparison between 2ν -SVM and C -SVM	120
6.4	Characteristic of Reliability Metric	127

List of Acronyms

BSV	bounded support vector
C -SVM	C -parameter support vector machine SVM using a single C for error weighting.
$2C$ -SVM	dual- C -parameter support vector machine SVM using a C for error weighting for each class.
KKT	Karush-Kuhn-Tucker KKT conditions are utilised in constrained optimisation.
ν -SVM	ν -parameter support vector machine SVM using a single ν for error weighting.
2ν -SVM	dual- ν -parameter support vector machine SVM using a ν for error weighting for each class.
QP	quadratic programming
RBF	radial basis function kernel function based on the radial basis function, $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(\frac{\ \mathbf{x}_i - \mathbf{x}_j\ }{\sigma^2})$.
SAR	synthetic aperture radar
SMO	sequential minimal optimization
SVM	support vector machine
UBSV	unbounded support vector

Glossary

- α_i Lagrange multiplier for the i^{th} data point The weight of the training data vector on the SVM solution. page 14
- \mathcal{B} Number of BSVs in the SVM page 39
- b Bias of the hyperplane The distance of the hyperplane from the origin. page 13
- \mathcal{B}_+ Number of BSVs in the positive class page 19
- \mathcal{B}_- Number of BSVs in the negative class page 19
- positive bias ... bias where percentage of positive errors to positive vectors is lower than the percentage of negative errors to negative vectors page 26
- negative bias ... bias where percentage of negative errors to negative vectors is lower than the percentage of positive errors to positive vectors page 26
- C Error parameter/weighting for C-SVM page 12
- C_+ Error parameter for the positive class in 2C-SVM page 21
- C_- Error parameter for the negative class in 2C-SVM page 21
- C_i Upper bound of α_i in 2C-SVM and 2ν -SVM See Equation(2.34) for 2C-SVM, and Equations(3.58) and(3.59) for 2ν -SVM. page 43
- D_1 Dataset 1 Synthetic dataset of two Gaussian distribution with data points in each class. See Appendix B.1. page 25
- D_2 Dataset 2 Synthetic dataset of two Gaussian distribution with data points in the positive class and data points in the negative class. See Appendix B.2. page 25
- D_3 Dataset 3 Benchmark dataset of handwritten digits from MNIST[51]. See Appendix B.3. page 25

Glossary

- D_4 *Dataset 4* Real world dataset of synthetic aperture radar images for detection of vehicles. See Appendix B.4. *page 25*
- $f(\mathbf{x})$ *Classification function* Returns the position of the test vector \mathbf{x} away from the hyperplane. *page 21*
- hyperplane $(\mathbf{w} \cdot \mathbf{x} + b = 0)$ Linear hyperplane that separates the positive and negative space in some high dimensional feature space. *page 12*
- i, j *index of the data point in the training set* $i, j \in [1 \dots l]$ unless otherwise stated. *page 11*
- $K(\cdot, \cdot)$ *kernel function* The dot product of two vectors in the feature space. $K(\cdot, \cdot) = \Phi(\cdot) \cdot \Phi(\cdot)$. *page 14*
- l *Number of data vectors in the training set* *page 12*
- l_+ *Number of positively labelled data vectors in the training set* *page 20*
- l_- *Number of negatively labelled data vectors in the training set* *page 20*
- L_d *Wolfe Dual formulation Lagrangian objective function* *page 61*
- margin errors .. *Data points that are on the wrong side of the margin* $(\mathbf{w} \cdot \mathbf{x}_i + b < y_i)$. See also BSV. *page 17*
- margin hyperplanes $(\mathbf{w} \cdot \mathbf{x}_i + b = \pm\rho)$ where $\rho = 1$ for 2C-SVM *page 11*
- ν *Error parameter for ν -SVM* *page 33*
- ν_+ *Error parameter for the positive class in 2ν -SVM* *page 34*
- ν_- *Error parameter for the negative class in 2ν -SVM* *page 34*
- $\Phi(\cdot)$ *Mapping function* Mapping function from the data space to the feature space. See also $K(\cdot, \cdot)$. *page 12*
- ρ *Position of the margin hyperplanes in ν -SVM* $\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b = \pm\rho$. See also margin hyperplanes *page 34*
- \mathcal{S} *Number of SVs in the SVM* *page 39*

- \mathcal{S}_+ *Number of SVs in the positive class page 41*
- \mathcal{S}_- *Number of SVs in the negative class page 41*
- $\text{sgn}(f(\mathbf{x}))$ *Decision function Returns the classification of the test vector. page 15*
- $\{\mathbf{x}_i, y_i\}$ *i^{th} data tuple of the training set page 12*
- \mathbf{w} *Normal vector of the hyperplane See also hyperplane page 13*
- w_δ *δ^{th} element of the vector \mathbf{w} See also \mathbf{w} . page 14*
- \mathbf{x}_i *i^{th} data vector of the training set page 12*
- ξ_i *Slack variable for the i^{th} data point page 13*
- y_i *Class label of the i^{th} data vector of the training set page 12*

Abstract

The Support Vector Machine (SVM) is a binary classification paradigm based on statistical learning. It is an important tool in object detection and pattern recognition, as well as signal processing, and the SVM algorithm has been shown to outperform existing classification algorithms. There are two main drawbacks with the original SVM formulation: the classification performance of the tuning regularisation parameter, and the computational and time cost of training phase. This Thesis focuses on the development of a new formulation of 2ν -SVM, with two cost parameters ν_+ and ν_- , to mitigate these drawbacks. In addition, the computation implementation of training process is designed and explained, and the relationship between the new and the original formulations is developed mathematically and discussed.

As with many statistical learning problems, the formulation of the original SVM (C -SVM) uses a regularisation parameter (C) to balance the generalisation performance of the classifier. The regularisation parameter places equal weightings on the number of training errors of both class labels. The equal weightings can be undesirable, and can be detrimental to the overall performance of the resulting classifier. In this Thesis, SVMs are extended to dual-error parameters ($2C$ -SVMs and 2ν -SVMs) that improve the classification performance, in particularly where the training class sizes are greatly different, and allow classification biasing based on *a priori* information. We discuss the creation of the new formulation and show results that provide indications of the improvements, and illustrate the use of 2ν -SVMs for multi-category classification. In addition, we describe a metric that measures the reliability of a multi-category classification for the one-against-rest strategy.

We introduce a novel implementation for training 2ν -SVMs. The weakness of SVMs is the computation cost of training multiple SVMs to produce an optimised classifier. Each training involves the optimisation of a quadratic programming problem that is non-trivial. Optimisation algorithms for C -SVMs have proved to reduce the computational cost involved, but these algorithms cannot be directly applied to 2ν -SVMs. The novel implementation improves on the existing methods while reducing the computational requirements.

Abstract

The relationship between the new 2ν -SVM formulation and the original describes the link between the objective functions of the two formulations. The mathematical link shows that the new formulation provides the same functionality of the original formulation, while improving on the classification and computation performances.

The resulting extension and implementation of 2ν -SVMs provides a strong and robust method for producing binary classifiers as well as multi-category classifiers. The research detailed here has resulted in 2ν -SVMs that have improved the computation and classification performance of SVMs. These improvements contribute to the broad area of multi-dimensional signal processing, such as for signal detection and image classification. The robustness of the 2ν -SVM in forming working classifiers from unprocessed data, and the reduction of number of training cycles allows users to quickly and effectively produce results for their classification problems.

Statement of Originality

This work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of the thesis, when deposited in the University Library, being available for loan, photocopying, and dissemination through the library digital thesis collection, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Thesis Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed

Date

Acknowledgments

I would like to acknowledge and thank my supervisors, Emeritus Professor Robert Bogner and Associate Professor Cheng-Chew Lim in the School of Electrical and Electronic Engineering at the University of Adelaide. They have been a great source of wisdom, guidance and support throughout the candidature.

The University of Adelaide provided financial support in the form of the Australian Postgraduate Awards. The Cooperative Research Centre for Sensor Signal and Information Processing provided additional support for living and travel expenses, as well as research contacts and connections to the Defence Science and Technology Organisation (DSTO). I sincerely thank these organisations, Dr Nick Redding (DSTO) and Dr David Crisp (DSTO) for their support and collaboration.

I would like thank my colleagues in the school for all the friendship and support, many of which have become my closest friends. The many non-research activities together have been good distractions from the daily grind.

On the personal front, I would like to thank the unwavering support of my parents, without which I would not be able to see this adventure to completion. In addition, to all my friends around the world, thanks for all encouragement over the years.

Hong-Gunn