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

THE UNIVERSITY
OF ADELAIDE
AUSTRALIA

Contents

List of Figures

List of Tables

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 multipler for the i$^{\text{th}}$ 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*

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* $i, j \in [1 \ldots l]$ 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 forumlation 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*

$sgn(f(\mathbf{x}))$ . . . . . *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.

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