An Object-Oriented Support Vector Machine Toolbox for Matlab

Gavin C. Cawley *

September 15, 2000

Abstract

1 Introduction

1.1 Support Vector Classification

The support vector machine (Boser et al., 1992; Cortes and Vapnik, 1995), given labelled training data

$$\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{\ell}, \quad \boldsymbol{x}_i \in \boldsymbol{X} \subset \mathbb{R}^d, \quad y_i \in \boldsymbol{Y} = \{-1, +1\},$$

constructs a maximal margin linear classifier in a high dimensional feature space, $\Phi(\boldsymbol{x})$, defined by a positive definite kernel function, $k(\boldsymbol{x}, \boldsymbol{x}')$, specifying an inner product in the feature space,

$$\Phi(\boldsymbol{x}).\Phi(\boldsymbol{x}') = k(\boldsymbol{x}, \boldsymbol{x}').$$

A common kernel is the Gaussian radial basis function (RBF),

$$k(\boldsymbol{x}, \boldsymbol{x}') = e^{-||\boldsymbol{x}-\boldsymbol{x}'||^2/2\sigma^2}.$$

The function implemented by a support vector machine is given by

$$f(\boldsymbol{x}) = \left\{ \sum_{i=1}^{\ell} \alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) \right\} - b. \tag{1}$$

To find the optimal coefficients, $\boldsymbol{\alpha}$, of this expansion it is sufficient to maximise the functional,

$$W(\boldsymbol{\alpha}) = \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} y_i y_j \alpha_i \alpha_j k(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{2}$$

*G. C. Cawley is with the School of Information Systems, University of East Anglia, Norwich, Norfolk, U.K. NR4 7TJ. E-mail: `gcc@sys.uea.ac.uk` .

in the non-negative quadrant,

$$0 \leq \alpha_i \leq C, \qquad i = 1, \ldots, \ell, \tag{3}$$

subject to the constraint,

$$\sum_{i=1}^{\ell} \alpha_i y_i = 0. \tag{4}$$

C is a regularisation parameter, controlling a compromise between maximising the margin and minimising the number of training set errors. The Karush-Kuhn-Tucker (KKT) conditions can be stated as follows:

$$\alpha_i = 0 \quad \Longrightarrow \quad y_i f(\boldsymbol{x}_i) \geq 1, \tag{5}$$
$$0 < \alpha_i < C \quad \Longrightarrow \quad y_i f(\boldsymbol{x}_i) = 1, \tag{6}$$
$$\alpha_i = C \quad \Longrightarrow \quad y_i f(\boldsymbol{x}_i) \leq 1. \tag{7}$$

These conditions are satisfied for the set of feasible Lagrange multipliers, $\boldsymbol{\alpha}^0 = \{\alpha_1^0, \alpha_2^0, \ldots, \alpha_\ell^0\}$, maximising the objective function given by equation 2. The bias parameter, b, is selected to ensure that the second KKT condition is satisfied for all input patterns corresponding to non-bound Lagrange multipliers. Note that in general only a limited number of Lagrange multipliers, $\boldsymbol{\alpha}$, will have non-zero values; the corresponding input patterns are known as support vectors. Let $\mathcal{I}$ be the set of indices of patterns corresponding to non-bound Lagrange multipliers,

$$\mathcal{I} = \{i \;:\; 0 < \alpha_i^0 < C\},$$

and similarly, let $\mathcal{J}$ be the set of indices of patterns with Lagrange multipliers at the upper bound C,

$$\mathcal{J} = \{i \;:\; \alpha_i^0 = C\}.$$

Equation 1 can then be written as an expansion over support vectors,

$$f(\boldsymbol{x}) = \left\{ \sum_{i \in \{\mathcal{I}, \mathcal{J}\}} \alpha_i^0 y_i k(\boldsymbol{x}_i, \boldsymbol{x}) \right\} - b. \tag{8}$$

For a full exposition of the support vector method, see the any of the excellent books (Vapnik, 1995; Vapnik, 1998; Cristianini and Shawe-Taylor, 2000) or tutorial articles () available.

2 Training Algorithms

3 Model Selection

4 Summary

5 Acknowledgements

References

Boser, B., Guyon, I., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory*, pages 144–152, Pittsburgh. ACM.

Cortes, C. and Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20:1–25.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines (and other kernel-based learning methods).* Cambridge University Press, Cambridge, U.K.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory.* Springer-Verlag, New York.

Vapnik, V. N. (1998). *Statistical Learning Theory.* Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications and Control. Wiley, New York.