Ullrich, Christian; Seese, Detlef; Chalup, Stephan ‘Foreign exchange trading with support vector machines ‘,30th Annual Conference of the Gesellschaft für Klassifikation e.V.. Advances in Data Analysis: Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation e.V., Freie Universität Berlin, March 8–10, 2006, Part VII (Berlin, Germany 8–10 March, 2006) p. 539-546 (2007)

Available from: http://dx.doi.org/10.1007/978-3-540-70981-7_62

Foreign Exchange Trading with Support Vector Machines

Christian Ullrich[1], Detlef Seese[2], and Stephan Chalup[3]

[1] AIFB, University of Karlsruhe, D-76128 Karlsruhe, Germany
christian.ullrich@bmw.de

[2] AIFB, University of Karlsruhe, D-76128 Karlsruhe, Germany
seese@aifb.uni-karlsruhe.de

[3] School of Electrical Engineering & Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia chalup@cs.newcastle.edu.au

Summary. This paper analyzes and examines the general ability of Support Vector Machine (SVM) models to correctly predict and trade daily EUR exchange rate directions. Seven models with varying kernel functions are considered. Each SVM model is benchmarked against traditional forecasting techniques in order to ascertain its potential value as out-of-sample forecasting and quantitative trading tool. It is found that hyperbolic SVMs perform well in terms of forecasting accuracy and trading results via a simulated strategy. This supports the idea that SVMs are promising learning systems for coping with nonlinear classification tasks in the field of financial time series applications.

1 Introduction

Support Vector Machines (SVMs) have proven to be a principled and very powerful supervised learning system that since its introduction (Cortes and Vapnik (1995)) has outperformed many systems in a variety of applications, such as text categorization (Joachims (1998)), image processing (Quinlan et al. (2004)), and bioinformatic problems (Brown et al. (1999)). Subsequent applications in time series prediction (Müller et al. (1999)) indicate the potential that SVMs have with respect to economics and finance. In predicting Australian foreign exchange rates, Kamruzzaman and Sarker (2003b) showed that a moving average-trained SVM has advantages over an Artificial Neural Network (ANN) based model, which was shown to have advantages over ARIMA models (2003a). Furthermore, Kamruzzaman et al. (2003) had a closer look at SVM regression and investigated how it performs with different standard kernel functions. It was found that Gaussian Radial Basis Function (RBF) and polynomial kernels appear to be a better choice in forecasting the Australian foreign exchange market than linear or spline kernels. Although Gaussian kernels are adequate measures of similarity when the representation dimension of the space remains small, they fail to reach their goal in high dimensional spaces (Francois et al. (2005)). We will examine the general ability of SVMs to correctly classify daily EUR/GBP, EUR/JPY and EUR/USD exchange rate directions. It is more useful for traders and risk managers to predict exchange rate fluctuations than their levels. To predict that the level of the EUR/USD, for instance, is close to the level today is trivial. On the contrary, to determine if the market will rise or fall is much more complex and interesting. Since SVM performance mostly depends on choosing the right kernel, we empirically verify the use of customized p-Gaussians by comparing them with a range of standard kernels. The remainder is organized as follows: Section 2 outlines the procedure for obtaining an explanatory input dataset. Section 3 formulates the SVM as applied to exchange rate forecasting and presents the kernels used. Section 4 describes the benchmarks and trading metrics used for model evaluation. Section 5 gives the empirical results. The conclusion, as well as brief directions for future research, are given in Section 6.

2 Data Selection

The obvious place to start selecting data, along with EUR/GBP, EUR/JPY and EUR/USD is with other leading traded exchange rates. Also selected were related financial market data, such as stock market price indices, 3-month interest rates, 10-year government bond yields and spreads, the prices of Brent Crude oil, silver, gold and platinum, several assorted metals being traded on the London Metal Exchange, and agricultural commodities. Macroeconomic variables play a minor role and were disregarded. All data is obtained from Bloomberg and spans a time period from 1 January 1997 to 31 December 2004, totaling 2349 trading days. The data is divided into two periods. The first period (1738 observations) is used for model estimation and is classified in-sample. The second period (350 observations) is reserved for out-of-sample forecasting and evaluation. Missing observations on bank holidays were filled by linear interpolation. The explanatory viability of each variable has been evaluated by removing input variables that do not contribute significantly to model performance. For this purpose, Granger Causality tests (Granger (1969)) with lagged values up to k=20 were performed on stationary I(1) candidate variables. We find that EUR/GBP is Granger-caused by 11 variables:

- EUR/USD, JPY/USD and EUR/CHF exchange rates
- IBEX, MIB30, CAC and DJST stock market indices
- the prices of platinum and nickel
- 10-year Australian and Japanese government bond yields

We identify 10 variables that significantly Granger-cause EUR/JPY:

- EUR/CHF exchange rate
- IBEX stock market index

- the price of silver
- Australian 3-month interest rate
- Australian, German, Japanese, Swiss and US government bond yields
- UK bond spreads

For EUR/USD, the tests yield 7 significant explanatory variables:

- AUD/USD exchange rate
- SPX stock market index and
- the prices of copper, tin, zinc, coffee and cocoa

3 SVM Classification Model and Kernels

3.1 SVM Classification Model

We will focus on the task of predicting the rise ("+1") or fall ("-1") of daily EUR/GBP, EUR/JPY and EUR/USD exchange rates. We apply the C-Support Vector Classification (C-SVC) algorithm as described in Boser et al. (1992) and Vapnik (1998), and implemented in R packages "e1071" (Chang and Lin (2001)) and "kernlab" (Karatzoglou et al. (2004)): Given training vectors $x_i \in R^n (i = 1, 2, \ldots, l)$, in two classes, and a vector $y \in R^l$ such that $y_i \in \{+1, -1\}$, C-SVC solves the following problem:

$$min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \tag{1}$$

$$\begin{aligned} y_i \left(w^T \phi(x_i) + b \right) &\geq 1 - \xi_i \\ \xi_i &\geq 0, i = 1, 2, \ldots, l \end{aligned}$$

The dual representation is given by

$$min_\alpha \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \tag{2}$$

$$\begin{aligned} 0 \leq \alpha_i \leq C, i = 1, 2, \ldots, l \\ y^T \alpha = 0 \end{aligned}$$

where e is the vector of all ones, C is the upper bound, Q is a lxl positive semidefinite matrix and $Q_{ij} \equiv y_i y_j K(x_i, x_j)$. $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel, which maps training vectors x_i into a higher dimensional, inner product, feature space by the function ϕ. The decision function is

$$f(x) = sign \left(\sum_{i=1}^{l} y_i y_j K(x_i, x) + b \right) \tag{3}$$

Training a SVM requires the solution of a very large quadratic programming optimization problem (QP) which is solved by using the Sequential Minimization Optimization (SMO) algorithm (Platt (1998)). SMO decomposes a large QP into a series of smaller QP problems which can be solved analytically. Time consuming numerical QP optimization as an inner loop can be avoided.

3.2 Kernel Functions

How to find out which kernel is optimal for a given learning task is a rather unexplored problem. Under this circumstance, we compare a range of kernels with regards to their effects on SVM performance. Standard kernels chosen include the following:

- Linear: $k(x, x') = \langle x, x' \rangle$
- Polynomial: $k(x, x') = (scale \cdot \langle x, x' \rangle + offset)^{degree}$
- Laplace: $k(x, x') = exp(-\sigma \, \|x - x'\|)$
- Gaussian radial basis: $k(x, x') = exp(-\sigma \, \|x - x'\|^2)$
- Hyperbolic: $k(x, x') = tanh(scale \cdot \langle x, x' \rangle + offset)$
- Bessel: $k(x, x') = \frac{Bessel^n_{v+1}(\sigma \|x - x'\|)}{(\|x - x'\|^{-n(v+1)})}$

Also, the use of customized p-Gaussian kernels $K(x_i, x_j) = exp\,(-d(x_i, x)^p / \sigma^p)$ with parameters p and σ is verified. The Euclidean distance between data points is defined by $d(x_i, x) = (\sum_{i=1}^{n} |x_i - x|^2)^{1/2}$. Compared to RBF-kernels, p-Gaussians include a supplementary degree of freedom in order to better adapt to the distribution of data in high-dimensional spaces. p and σ depend on the specific input set for each exchange rate return time series and are calculated as proposed in (Francois et al. (2005)):

$$p = \frac{ln\left(\frac{ln(0.05)}{ln(0.95)}\right)}{ln\left(\frac{dF}{dN}\right)} ; \sigma = \frac{d_F}{(-ln(0.05))^{1/p}} = \frac{d_N}{(-ln(0.95))^{1/p}} \tag{4}$$

In the case of EUR/USD, for example, we are considering 1737 8-dimensional objects. We calculate 1737x1737 distances and compute the 5% (d_N) and 95% (d_F) percentiles in that distribution. In order to avoid the known problem of overfitting, we determine robust estimates for C and *scale* (σ) for each kernel through 20-fold cross validation.

4 Benchmarks and Evaluation Method

Letting y_t represent the exchange rate at time t, we forecast the variable

$$sign(\Delta y_{t+h}) = sign(y_{t+1} - y_t) \tag{5}$$

where $h = 1$ for a one-period forecast with daily data. The naïve model ($sign(\hat{y}_{t+1}) = sign(y_t)$) and univariate ARMA($p$, q) models are used as benchmarks. ARMA(p, q) models with p autoregressive (AR) terms and q moving averages (MA) are given by

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} \quad (6)$$

where $\epsilon_t \sim$ i.i.d. (0, σ^2). Simple models, that were estimated according to Box and Jenkins (1976), provide the best testing results while preserving generalization ability for forecasting (s-step-ahead predictions for $s \leq q$ are given in parentheses):

- $c = -3.58E-05, \beta_1 = -0.0535$, and $\beta_3 = -0.0559$ ($\hat{y}_{t+s} = \hat{c} + \hat{\beta}_s \epsilon_t + \hat{\beta}_{s+2}\epsilon_{t-2}$) for the EUR/GBP series
- $c = -7.84E-05$ and $\beta_1 = 0.0288$ ($\hat{y}_{t+s} = \hat{c} + \hat{\beta}_s \epsilon_t$) for the EUR/JPY series
- $c = -8.32E-05, \alpha_1 = -0.5840$ and $\beta_1 = 0.5192$ ($\hat{y}_{t+s} - \hat{c} = \hat{\alpha}_1(\hat{y}_{T+s-1} - \hat{c}) + \hat{\beta}_s \epsilon_t$) for the EUR/USD series

Out-of-sample forecasts are evaluated statistically via confusion matrices and practically via trading simulations. The reason for this twofold evaluation procedure is that trading decisions driven by a model with a small statistical error may not be as profitable as those driven by a model that is selected using financial criteria. In case of the latter, return predictions $\hat{y}_{t+1}$ are first translated into positions. Next, a decision framework is established that tells when the underlying asset is bought or sold depending on the level of the price forecast. We define a single threshold τ, which is set to $\tau = 0$ and use the following mechanism:

$$I_t = \begin{cases} 1 & if \quad \hat{y}_t < y_{t-1} - \tau \\ -1 & if \quad \hat{y}_t > y_{t-1} + \tau \\ 0 & if \quad \text{otherwise} \end{cases}, \text{ with } I_t = \begin{cases} 1 & if \quad \text{the position is long} \\ -1 & if \quad \text{the position is short} \\ 0 & if \quad \text{the position is neutral} \end{cases} \quad (7)$$

The gain or loss π_t on the position at time t is $\pi_t = I_{t-1}(y_t - y_{t-1})$. Since financial goals are user-specific, we examine the models' performances across nine Profit and Loss (P&L) related measures:

- Cumulated P&L: $PL_T^C = \sum_{t=1}^{T} \pi_t$
- Sharpe ratio: $SR = \frac{PL_T^A}{\sigma_T^A}$, with annualized P&L $PL_T^A = 252 \frac{1}{T} \sum_{t=1}^{T} \pi_t$, and annualized volatility $\sigma_T^A = \sqrt{252}\sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(\pi_t - \bar{\pi})^2}$
- Maximum daily profit: $Max(\pi_1, \pi_2, \ldots, \pi_T)$
- Maximum daily loss: $Min(\pi_1, \pi_2, \ldots, \pi_T)$
- Maximum drawdown: $MD = Min(PL_t^C - Max_{i=1,2,\ldots,t}(PL_i^C))$
- Value-at-Risk with 95% confidence: $VaR = \mu - Q(\pi, 0.05)$ with $\mu = 0$
- Net PL_T^C: $NPL_T^C = \sum_{t=1}^{T}(\pi_t - I_t \cdot TC)$, where $I_t = \begin{cases} 1 & \text{if} \quad \pi_{t-1} \cdot \pi_t < 0 \\ 0 & \text{else} \end{cases}$

- Average gain/loss ratio: $\frac{AG}{AL} = \frac{(\text{ Sum of all } \pi_t > 0)/\#\text{up}}{(\text{ Sum of all } \pi_t < 0)/\#\text{down}}$
- Trader's advantage: $TA = 0.5\left(1 + \left(\frac{(WT \cdot AG) + (LT \cdot AL)}{\sqrt{(WT \cdot AG^2) + (LT \cdot AL^2)}}\right)\right)$ with WT:= number of winning trades, LT:= number of losing trades, AG:= average gain in up periods, and AL:= average loss in down periods

Accounting for transaction costs (TC) is important for assessing trading performance in realistic ways. An average cost of 3 pips (0.0003) per trade, for a tradable amount of typically 10 to 20 million EUR is considered a reasonable guess and incorporated in NPL_T^C. A model is operationally superior compared to another if it exhibits a larger number of superior performance measures.

5 Empirical Results

Accuracy rates for the out-of-sample period are depicted in bar charts as shown in Figure 1. Figures 2 through 4 give results of the trading simulation as described in Section 4. Dominant strategies are represented by the maximum value(s) in each row and are written in bold. We observe the following:

- Statistically, both the naïve and the linear model are beaten by SVM with a suitable kernel choice. The SVM approach is statistically justified.
- Hyperbolic SVMs deliver superior performance for out-of-sample prediction across all currency pairs. In the case of EUR/GBP, the Laplace and hyperbolic SVM perform equally well. In the cases of EUR/JPY and EUR/USD, hyperbolic kernels outperform the other models more clearly. This makes hyperbolic kernels promising candidates to map all sorts of financial market return data into high dimensional feature spaces.
- Operational evaluation results confirm statistical ones in the case of EUR/GBP. The hyperbolic and Laplace SVM give the best results along with the RBF-SVM. For EUR/JPY and EUR/USD, statistical superiority of hyperbolic SVMs cannot be confirmed. Operational evaluation techniques not only measure the number of correctly predicted exchange rate ups and downs but also include the magnitude of returns. Consequently, if local extremes can be exploited, forecasting methods with less statistical performance may yield higher profits than methods with greater statistical performance. In the case of EUR/USD, the trader would have been better off applying a p-Gaussian SVM to maximize profit. In regards to EUR/JPY, no single model is able to outperform the naïve strategy. The hyperbolic SVM, however, dominates two performance measures.
- p-Gaussian SVMs perform reasonably well in predicting EUR/GBP and EUR/USD return directions. For these two currency pairs, p-Gaussian data representations lead to better generalization than Gaussians due to an additional degree of freedom p.

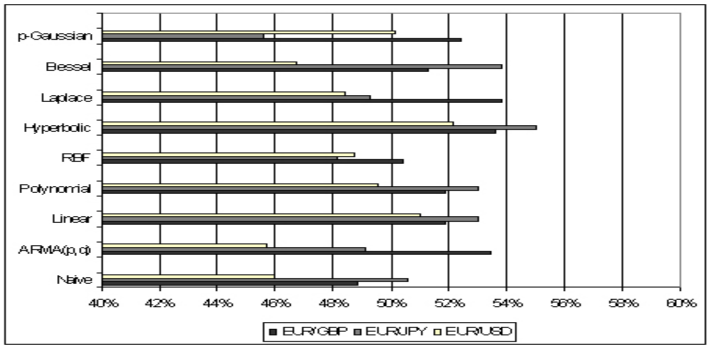

Fig. 1. Classification performance EUR/GBP, EUR/JPY, EUR/USD.

Table 1. Operational performance for EUR/GBP, EUR/JPY, and EURUSD.

EUR/GBP	Naive	MA(1,3)	Linear	Polynomial	RBF	Hyperbolic	Laplace	Bessel	p-Gaussian
Cumulative P&L	-0,00750	-0,00953	-0,09360	-0,09360	-0,03896	**0,10360**	0,01546	-0,04114	0,05958
Sharpe ratio	-0,07966	-0,10112	-0,99367	-0,99367	-0,41354	**1,09938**	0,16407	-0,43671	0,63235
Maximum daily profit	0,01492	0,01492	**0,01684**	**0,01684**	**0,01684**	0,01492	**0,01684**	0,01385	0,01232
Maximum daily loss	-0,01684	-0,01684	-0,01492	-0,01492	**-0,01385**	-0,01684	**-0,01385**	-0,01684	-0,01684
Maximum drawdown	-0,03811	-0,03811	-0,03619	-0,03619	**-0,03496**	-0,03811	-0,03512	-0,03564	-0,03811
VaR (alpha = 0.05)	-0,00695	-0,00734	-0,00752	-0,00752	-0,00728	-0,00698	**-0,00691**	-0,00744	-0,00694
Net Cumulative P&L	-0,06120	-0,01013	-0,12750	-0,12750	-0,09026	**0,05590**	-0,01964	-0,09214	0,01428
Avg gain/loss ratio	**1,05178**	0,85038	0,80370	0,80370	0,91714	1,03981	0,89932	0,88235	1,01891
Trader's Advantage	0,00000	**1,00000**	0,53003	0,53003	0,48716	0,48144	0,58986	0,39350	0,43507

EUR/JPY	Naive	MA(1)	Linear	Polynomial	RBF	Hyperbolic	Laplace	Bessel	p-Gaussian
Cumulative P&L	**0,05441**	-0,11333	-0,09477	-0,09477	-0,21907	-0,13867	-0,28671	-0,31145	-0,24980
Sharpe ratio	**0,38680**	-0,80435	-0,67432	-0,67432	-1,55679	-0,98622	-2,03603	-2,21115	-1,77460
Maximum daily profit	**0,02187**	0,02187	0,02068	0,02068	0,02068	0,02174	0,02068	0,02068	0,02050
Maximum daily loss	**-0,02050**	-0,02174	-0,02187	-0,02187	-0,02187	-0,02187	-0,02187	-0,02187	-0,02187
Maximum drawdown	-0,08535	-0,08659	-0,06479	-0,06479	-0,08672	**-0,06197**	-0,08672	-0,06479	-0,08672
VaR (alpha = 0.05)	**-0,01003**	-0,01144	-0,01092	-0,01092	-0,01111	-0,01081	-0,01127	-0,01145	-0,01130
Net cumulative P&L	**0,00281**	-0,11363	-0,15267	-0,15267	-0,27607	-0,19837	-0,34461	-0,36185	-0,30260
Avg gain/loss ratio	**1,04111**	0,92829	0,89996	0,89996	0,88278	0,86458	0,83323	0,83752	0,82177
Trader's advantage	0,00000	0,00000	0,43005	0,43005	0,43247	**0,43647**	0,41154	0,40350	0,40139

EUR/USD	Naive	ARMA(1,1)	Linear	Polynomial	RBF	Hyperbolic	Laplace	Bessel	p-Gaussian
Cumulative P&L	-0,18070	-0,22255	-0,13259	-0,13259	-0,00927	0,04797	-0,10055	-0,16166	**0,10182**
Sharpe ratio	-1,23452	-1,52256	-0,90434	-0,90434	-0,06296	0,32520	-0,68505	-1,10372	**0,68905**
Maximum daily profit	0,01962	0,01962	0,01667	0,01667	0,01962	**0,01962**	0,01889	0,01869	0,01889
Maximum daily loss	-0,01889	-0,01889	-0,01962	-0,01962	**-0,01869**	-0,01889	-0,01962	-0,01962	-0,01962
Maximum drawdown	-0,04172	**-0,04112**	-0,04484	-0,04484	-0,04391	-0,04410	-0,04484	-0,04484	-0,04484
VaR (alpha = 0.05)	-0,01247	-0,01179	-0,01260	-0,01260	-0,01176	**-0,01085**	-0,01183	-0,01165	-0,01116
Net cumulative P&L	-0,23680	-0,22345	-0,17429	-0,17429	-0,05967	-0,00003	-0,14525	-0,21056	**0,05112**
Avg gain/loss ratio	0,94708	0,93486	0,88117	0,88117	1,03619	0,96269	0,94573	0,94569	**1,10874**
Trader's advantage	0,00000	0,31863	**0,62531**	**0,62531**	0,56826	0,55311	0,58379	0,42194	0,49915

6 Conclusion

The results support the general idea that SVMs are promising learning systems for coping with nonlinear classification and regression tasks in financial time series applications. Future research will likely focus on improvements of SVM models, such as examination of other kernels, adjustment of kernel parameters and development of data mining and optimization techniques for selecting the appropriate kernel. In light of this research, it would also be interesting to see if the dominance of hyperbolic SVMs can be confirmed in further empirical investigations on financial market return prediction.

References

BOSER, B.E., GUYON, I.M. and VAPNIK, V.N. (1992): A training algorithm for optimal margin classifiers. In: D. Haussler (Ed.): *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory.* ACM Press, 144–152.

BOX, G.E.P. and JENKINS, G.M. (1976): *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

BROWN, M., GRUNDY, W., LIN, D., CRISTIANINI, N., SUGNET, C., FUREY, T., ARES, M. and HAUSSLER, D. (1999): Knowledge-based analysis of microarray gene expression data using support vector machines. *Technical Report.* University of California, Santa Cruz.

CHANG, C.C. and LIN, C.J. (2001): LIBSVM: a library for support vector machines (version 2.31). *Technical Report.* Department of Computer Sciences and Information Engineering, National Taiwan University, Taipei, Taiwan.

CORTES, C. and VAPNIK, V. (1995): Support vector network. *Machine Learning, 20, 273–297.*

FRANCOIS, D., WERTZ, V. and VERLEYSEN, M. (2005): About the locality of kernels in high-dimensional spaces. *ASMDA 2005 - International Symposium on Applied Stochastic Models and Data Analysis.* Brest, France, 238–245.

GRANGER, C.W.J. (1969): Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37, 424–438.*

JOACHIMS, T. (1998): Text categorization with support vector machines. *Proceedings of European Conference on Machine Learning (ECML).*

KAMRUZZAMAN, J. and SARKER, R.A. (2003a): Forecasting of currency exchange rate: a case study. *Proceedings of the IEEE International Conference on Neural Networks & Signal Processing (ICNNSP).* Nanjing.

KAMRUZZAMAN, J. and SARKER, R.A. (2003b): Application of support vector machine to Forex monitoring. *Third International Conference On Hybrid Intelligent Systems (HIS).* Melbourne.

KAMRUZZAMAN, J., SARKER, R.A. and AHMAD, I. (2003): SVM Based Models for Predicting Foreign Currency Exchange Rates. *Proceedings of the Third IEEE International Conference on Data Mining (ICDM).*

KARATZOGLOU, A., HORNIK, K., SMOLA, A. and ZEILEIS, A. (2004): kernlab - An S4 package for kernel methods in R. *Journal of Statistical Software, 11, 9.*

KUAN, C.M. and LIU, T. (1995): Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics, 10, 347–364.*

MÜLLER, K.R., SMOLA, A., RÄTSCH, G., SCHÖLKOPF, B., KOHLMORGEN, J. and VAPNIK, V. (1999): Using Support Vector Machines for Time Series Prediction. In: B. Schölkopf , C.J.C. Burges and A.J. Smola (Eds.): *Advances in Kernel Methods.* MIT Press, 242–253.

PLATT, J.C. (1998): Sequential minimal optimization: A fast algorithm for training support vector machines. *Technical Report.* Microsoft Research.

QUINLAN, M.J., CHALUP, S.K. and MIDDLETON, R.H. (2004): Application of SVMs for colour classification and collision detection with AIBO robots. *Advances in Neural Information Processing Systems, 16, 635–642.*

SCHÖLKOPF, B. (2001): The kernel trick for distances. *Technical Report, MSR 2000-51.* Microsoft Research, Redmond, WA.

VAPNIK, V. (1998): *Statistical Learning Theory.* Wiley.