Ann Oper Res (2007) 151:325–336
DOI 10.1007/s10479-006-0131-7

Adapting support vector machine methods for horserace odds prediction

David Edelman

Published online: 10 November 2006

Abstract The methodology of Support Vector Machine Methods is adapted in a straightforward manner to enable the analysis of stratified outcomes such as horseracing results. As the strength of the Support Vector Machine approach lies in its apparent ability to produce generalisable models when the dimensionality of the inputs is large relative to the the number of observations, such a methodology would appear to be particularly appropriate in the horseracing context, where often the number of input variables deemed as being potentially relevant can be difficult to reconcile with the scarcity of relevant race results. The methods are applied to a relatively small (200 races in-sample) sample of Australian racing data and tested on 100 races out-of-sample with promising results, especially considering the relatively large number (12) of input variables used.

Keywords Computational Finance · Wagering Markets

1 Background

1.1 Horserace odds prediction

As in Financial markets, considerable attention has been given to the problem of forecasting results for horseracing markets, and other betting markets. In the academic literature, however, the overwhelming emphasis has been on issues analogous to what in the Financial markets is generally referred to as 'Technical Analysis'. In the horseracing context, this amounts to seeing whether financial advantage may be gained by using the Markets Odds alone to forecast racing outcomes, with much of the literature falling into the category of either Odds calibration studies such as 'Favourite-Longshot bias' studies [Hausch et al. (1994)], or 'Insider Trading Adjustment' studies [Shin (1991, 1993)]. Secondary to this in terms of volume of research articles (but by no means less significant in importance) has been the development in the literature of betting strategies which would be applied *in theory* if quality

D. Edelman (✉)
Banking & Finance Unit, University College, Dublin Blackrock, Co. Dublin, Ireland
e-mail: david.edelman@ucd.ie

odds forecasts were assumed to exist [see Breiman (1961), Kelly (1956), and Ziemba et al. (1994)].

In contrast to the case in Financial Markets, however (and due no doubt to the much lower total potential financial gain relative to that achievable in Financial Markets), there has been relatively little published academic work [see Benter (1994), Edelman (2001)] which attempts to combine both Fundamental (i.e., horse, jockey, or trainer history) and Technical Analysis (Market Odds) to produce forecasts leading to profitable betting systems.

In analogy to the various 'Fundamental' variables and indices commonly used for trading in the Financial Markets context, such as Price/Earnings ratio, Market/Book Ratio, Debt/Equity Ratio, Turnover, Momentum, etc., there are a wealth of variables available for consideration in the horseracing context. For instance, variables relating to an individual horse's background, including breeding, age, training and racing history, as well as jockeys' and trainers' histories may all be considered as potential candidates as variables which should be included in modeling, not to mention a wide variety of auxilliary variables such as weather, track, or barrier draw effects. The initial challenge then becomes the problem of 'data overload', where it becomes crucial to limit focus to certain useful summary variables to consider for analysis.

While any serious attempt at designing a prediction model to be used as the basis of a trading system should include examination of as many of these types of variables as possible, the emphaisis here will be on considering only a handful of variables which have been found to be particularly significant, and to argue heuristically, why ignoring some of the others may not be too detrimental to modeling. Also, by limiting focus to certain controlled circumstances, such as the range of race distances considered, we are able to reduce the number of variables which need to be considered.

As an example of applying heuristics, consider horses' breeding information. While clearly important, understanding of breeding generally becomes marginally less important as a horse's race record is observed, as similarly may be argued for the effect a trainer has on a horse. Therefore, to 'first order' we might consider neglecting variables such as these, where proven performance would tend to be much more relevant in a relative sense. An exception to this might be when horses, which begin their careers racing at Sprint (below 1400 m) distances and may be trying longer distances for the first time. To limit the potential effect of this, we restrict our attention to Sprint races. This has the additional desirable effect of minimising the effects of varying intervals between successive runs, as many issues related to conditioning which are so crucial at longer distances becomes much less relevant for Sprint racing.

The next heuristic principle which may be applied to simplfy matters is that, by and large, Betting Markets are near-efficient. This means that the bulk of the information about a horse's racing history prior to its previous run is most likely contained in its Market odds for that event. Therefore, by combining the odds from a horse's last start with the outcome and circumstances of that race and the race today, one may gain a reasonable forecast of its chances today. This suggests a sort of *incremental* approach to forecasting, an approach which will be adopted here.

The primary challenge to applying such logic (even at a fixed distance range) is that horses' previous starts may have been in races of different levels of difficulty, or *class*. As a proxy for *class*, we use total race prizemoney on offer, since raising the stakes is generally the most effective way race organisers have of ensuring that as many racehorse owners as possible are interested in entering their best horses in an event.

Hence, by including only a few such variables as are discussed above, relating to the current and previous race, and including, most significantly, Odds Market information, one may hope to build quite a reliable forecasting model for a horse's chances of winning a race.

The specific variables to be used used will be be discussed in more detail in a later section, but methodolgically, the approach follows that of Benter (1994, 2003)), whereby a regression of Horses' Finishing Position (i.e., first = 1, second = 2, etc.) scaled to the interval [−.5,5] (referred to as *normalised finishing position*) as a linear function of the variables mentioned above (and more specifically later) to produce a 'Winningness' Index. This Index is then used as an input to a Multinomial Logit regression, stratified by Race, with outcome variable being 1 for the winning horse and 0 otherwise. The resulting probability forecasts are then compared to the payoff Odds of the respective betting markets, and optimal bets are determined. Before explaining this final step, it is worth clarifying here that the primary distinction between the approach taken here and that of Benter, is in the method by which the Winningness Index is determined; whereas Benter used linear regression on the fundametal variables to approximate Normalised Finishing Position, the approach taken here is to apply Support Vector Machine methods (using a slightly different version of Normalised Finishing Position) to produce the Winningness Index.

As will be explained in more detail in the next subsection, in all other respects we follow the approach of Benter, where we assume that Indices have indeed been computed, and will be used in betting.

Thus, it may be determined whether a set of variables and model type appear to give rise to a profitable betting system, where the truest indication may be gotten by applying the methods to unseen data.

1.2 From forecasts to betting

Before proceeding further, a brief word on the methodology which would be applied to translate the aforementioned 'Winningness' Indices resulting from regression (or, in the sequel, Support Vector Machine) methods into actual bets to apply in practice. Here, this methodology will be applied to the various methods for producing the Winningness Index and the Results compared on the basis of financial performance on unseen data.

If a Winningness Index forecasts are available, then, logically, the remaining steps may be divided into two parts: probability regularisation, and staking.

First, let u_{ij} denote the binary (0-1) variable indicating whether horse j in race i has won the race (1) or not (0), where i ranges over the number of races analysed in the sample, and j ranges over the number of horses in race i. $\hat{y}_{ij}$ denotes the Winningness Index forecast produced by the model being employed. Then, as a function of the single 'regularisation' parameter α, the following likelihood is maximised:

$$L(\alpha) = \prod_{i=1}^{N_{Races}} \frac{\prod_{j=1}^{n_i} \exp(u_{ij}\hat{y}_{ij}\alpha)}{\sum_{j'=1}^{n_i} \exp(\hat{y}_{ij'}\alpha)}$$

Thus probability forecasts $\hat{p}_{ij}$ are then expressed as

$$\hat{p}_{ij} = \frac{\exp(\hat{y}_{ij}\hat{\alpha})}{\sum_{j'=1}^{n_i} \exp(\hat{y}_{ij'}\hat{\alpha})},$$

where $\hat{\alpha}$ is the result of the previous Maximum Likelihood procedure.

Next, assuming this step has been performed, we follow Kelly (1956) and apply the so-called 'Kelly criterion' of maximising the expected logarithmic return for each race, given the Market Odds payoffs and corresponding forecasts. Breiman (1961) showed that for repeated independent betting opportunities of this type, the strategy of myopically (here, race by race) optimising the expected logarithm of wealth maximises one's long-run growth in wealth.

Here, this amounts to solving for each race i the set of bets $b_{i1}, \ldots, b_{in_i}$; $b_{ij} \geq 0$ all j, such that quantity

$$\sum_{j=1}^{n_i} \hat{p}_{ij} \log \left\{ 1 + b_{ij} \times t_{ij} - \left(\sum_{j=1}^{n_i} b_{ij} \right) \right\}$$

is a maximum, where t_{ij} denotes the gross return if horse j were to win. This optimisation is well-posed, and may be carried out by any of a number of routines for constrained nonlinear optimisation.

It may be shown that if the optimal solution results in a single positive bet, on horse j_0, say, then

$$b_{ij_0} = \frac{t_{ij_0} \hat{p}_{ij_0} - 1}{t_{ij_0} - 1},$$

whereas in the general case a characterisation of the relative bet size may be gotten from

$$b_{ij} = \left(\hat{p}_{ij} - \frac{c_i}{t_{ij}} \right)^+$$

for all j, where $(\cdot)^+$ denotes the positive part of its argument, and c_i is a constant (positive, typically slightly less than unity) which must be determined iteratively.

It should be noted that other betting and staking strategies such as Fractional Kelly [Ziemba et al. (1994)] may be used to optimise with respect to utility functions other than the logarithm, but as is shown by Edelman (2000), optimisation according to the (pure) Kelly criterion is very close to optimisation of the Sharpe ('mean-to-standard deviation') Ratio [Sharpe (1994)], which is one of the most widely accepted risk-adjusted measures of Investment performance used in Finance.

1.3 Support Vector Machines

The recent Support Vector Machine approach from Vapnik's Statistical Learning Theory [Vapnik (2001)] has received increasing attention in pattern-matching and recognition applications.

In its simplest form, a Support Vector Machine is a supervised learning method for discriminating between two separable groups $\{(x; y)\}$, where the (scalar) target variable y is equal to either $+1$ or -1, the (vector) input variable x is arbitrary, and a so-called 'separating hyperplane', or plane in x-space which separates positive and negative cases, is of the form

$$(w \cdot x) + b = 0,$$

where b is an adapted bias ('intercept') constant and w is the vector with the minimum L_2 norm such that

$$((w \cdot x_i) + b)y_i \geq 1, \quad \text{for all } i$$

where x_i and y_i denote the observed input and output variables for the i^{th} training pattern, and i ranges from 1 to the number of observations n.

Vapnik has shown that for any set of separable patterns, or training patterns which may be separated by such a hyperplane, this problem is well-defined, and may be solved in a straightforward manner using standard Quadratic Programming Methods. As Vapnik argued, it may be shown that for any separable SVM problem the solution depends only on a set $\{x_i; y_i\}$ of 'support vectors' for which

$$((w \cdot x_i) + b)y_i = 1,$$

often a set with far fewer elements than the entire training set. In fact, it may be said that it is such a reduction in complexity which may explain the superior generalisation properties over neural networks and other methods which SVM methods have achieved in many applications (see Vapnik (2001) for a discussion).

The first generalisation is for the case in which misclassification errors are allowed, in which case the objective becomes the minimisation of

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$$

subject to

$$((w \cdot x_i) + b)y_i \geq 1 - \xi_i, \quad \xi_i \geq 0,$$

where ξ_i denotes the distance of prediction i to the classification boundary, commonly referred to as a "slack variable", with C being a "slack" constant, suitably chosen (virtually always by cross-validation in practice) to adequately penalise misclassification error while avoiding overfitting.

An inspired variation on these methods was the application to the general regression case, where the formulation is to optimise the balance between error with respect to the ε-insensitive linear loss function (see Vapnik (1995), p. 183),

$$\sum_{i=1}^{n} |y_i - (w \cdot x_i) - b|_\varepsilon = \sum_{i=1}^{n} (|y_i - (w \cdot x_i) - b| - \varepsilon)^+$$

($(\cdot)^+$ denoting the 'positive part' of its argument), and the norm of w. The quadratic programming solution to this problem via solution of a $2n$-dimensional dual problem (where n denotes the number of observations) is given in Vapnik (2001) and elsewhere. In the dual formulation, the required $2n$ variables are Lagrange parameters relating to the conditions

$$y_i - (w \cdot x_i) - b \geq \varepsilon,$$

$$y_i - (w \cdot x_i) - b \leq -\varepsilon,$$

or

$$-\varepsilon < y_i - (w \cdot x) - b < \varepsilon.$$

Since Support Vector Machines are known to perform well with high dimensionality in the input variable, use of *feature maps*

$$x \longrightarrow h(x)$$

is possible. A Feature map of an input vector is a vector-valued transformation of the input onto another (in practice, often much larger) 'feature' space, typically preceeding a drastic reduction in dimensionality effected by the Support Vector Machine methodology.

In this case, we have chosen a Radial Basis Function feature map with Gaussian Kernel, with centers at the various datapoints, and bandwidth chosen according to a 'leave-one-out' cross-validation criterion, where in this instance for bandwidth γ

$$K(x; x_j) = \exp\left(-\frac{1}{2\gamma^2}\|x - x_j\|^2\right)$$

If the centers of the Kernels are in fact taken to be the set of input variables themselves, then this feature map amounts to, for each input observation, a sort of 'feature closeness' index to each and every other variable in the sample.

It is this (regression/RBF) version of the Support Vector Machine Methodology which we shall employ here in the context of horseracing, with only the minor modification that multiple biases are allowed, resulting in an effective stratified analysis by race.

Before presenting the details of the stratified SVM analysis, however, a brief summary of the analogous Ordinary Least Squares approach (see Benter (1994, 2003)) which is often adopted in practice is helpful.

In general terms, the modeling objective in horserace forecasting is to model Win probabilities for horseraces as a function of input variables. The first attempts to fit such a model were via direct application of the Multinomial Logit [Bolton and Chapman (1986)] to the input and output data. Specifically, if x_{ij} is used to denote the input vector for the j^{th} runner in the i^{th} race, and y_{ij} the corresponding binary win-loss outcome variable, then the corresponding probabilistic prediction p_{ij}, as a function of coefficient vector β, is given by

$$p_{ij} = \frac{\exp((x_{ij} \cdot \beta))}{\exp((x_{i1} \cdot \beta)) + \exp((x_{i2} \cdot \beta)) + \cdots + \exp((x_{in_i} \cdot \beta))}.$$

Fitting is carried out via Maximum Likelihood, with large-sample (though, notably, not small-sample) unbiasedness following, along with the usual consistency and asymptotic efficiency for Maximum Likelihood Estimators under standard regularity assumptions.

It turns out, however, that applying this method directly to the win-loss outcomes y_{ij} results in drastic information loss, as only the Win outcomes are used, with the relative performance of the remaining runners in each race being ignored. Following several attempts to address this shortcoming, a two-step method [Benter (1994, 2003)] was employed. First, a stratified linear regression using all of the finishing-order information ('placings') summarised in a *normalised finishing position* variable as outcome was fitted, followed by a 'regularisation' step, in which the predicted values from the first regression were used as inputs to the Multinomial Logit described previously, resulting in forecasts p_{ij} of the desired probabilistic nature. [It is perhaps worth noting that, were it not for the information loss associated with fitting Win/Loss probabilities and outcomes in this particular context, recent work of Platt (1999) on Support Vector Machines with probabilistic outcomes might be relevant.]

In analogy to standard regression with Ordinary Least Squares, Stratified Regression in this context merely entails allowing a separate intercept term for each race, or equivalently, centering all variables (input and output) by race and enforcing zero intercept. In the SVM context, this entails fitting of a seperate bias constant for each stratum (race), which in fact characterises the difference between the method presented here and the usual SVM regression method.

In this context, the Normalised Finishing Position outcome variable discussed previously for the Linear Regression may be defined formally as

$$y_{ij} = \frac{fp_{ij}}{n_i + 1} - .5, \quad j = 1, 2, \ldots, n_i$$

where fp_{ij} denotes the *finishing position* or rank-finishing order (winner = 1, second = 2, etc.) for the i^{th} race, where the i^{th} race has n_i runners. An alternative which may be considered preferable is the *Normal Scores finishing position*, will be employed here:

$$y_{ij} = \Phi^{-1}\left(\frac{fp_{ij}}{n_i + 1}\right),$$

$\Phi^{-1}(\cdot)$ denoting the inverse function of the cumulative Standard Normal (Gaussian) distribution: for $0 < \alpha < 1$,

$$\int_{-\infty}^{\Phi^{-1}(\alpha)} \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} dz \equiv \alpha$$

It should be noted that, provided all runners are included,

$$\sum_{j=1}^{n_i} y_{ij} = 0, \quad i = 1, 2, \ldots$$

automatically follows.

In analogy, then, to the first regression in the linear case, where the coefficient vector w is chosen to minimise

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} (y_{ij} - (w \cdot x_{ij}) - b_i)^2,$$

the Support Vector approach will begin by minimising the ε-insensitive total absolute error

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} |y_{ij} - (w \cdot h(x_{ij})) - b_i|_{\varepsilon},$$

to produce predictions $\hat{y}_{ij}$. Then the final 'regularisation' step of producing probabilistic predictions from the y's is effected, through estimation of the 'strength coefficient' β via Maximum Likelihood where

$$p_{ij} = \frac{\exp((\hat{y}_{ij} \cdot \beta))}{\exp((\hat{y}_{i1} \cdot \beta)) + \exp((\hat{y}_{i2} \cdot \beta)) + \cdots + \exp((\hat{y}_{in_i} \cdot \beta))}.$$

[Alternatively, in the final step, if Market Odds[1] are to be used, the specification of p_{ij} is slightly different:

$$p_{ij} = \frac{\exp(\hat{y}_{ij}\beta_1 + \log(1 + odds_{ij})\beta_2)}{\exp(\hat{y}_{i1}\beta_1 + \log(1 + odds_{i1})\beta_2) + \cdots + \exp(\hat{y}_{in_i}\beta_1 + \log(1 + odds_{in_i})\beta_2)}.\Bigg]$$

In order to determine the betting efficacy of the model, it is of interest to apply a betting rule and assess profitability, both within the sample used for parameter-fitting, and in a holdout sample. As mentioned earlier, while there are a number of approaches which could be employed, including Markowitz (1952), Kelly (1956), fractional Kelly [Ziemba et al. (1994)], and Fixed-Return [Scott (1985)], a comparative analysis will not be carried out, as the emphasis here is on demonstrating the usefulness of the Support Vector Machine approach rather than to present a comprehensive guide to betting. Hence only the (full) Kelly betting staking approach will be employed, as it is the simplest (though arguably over-sensitive to Misspecification or Model Risk in practice), which entails betting so as to maximise logarithmic growth rate for each race.

2 Results

The data to be analysed are Metropolitan race results from Australia during 1995 (Australian Associated Press (AAP) Racing Services, Data Subscription), with such an early dataset chosen to prove concept in an exploratory sense, where more recent data is kept for 'holdout' purposes in analysing more well-developed models, to be presented in future studies.

One of the oft-mentioned features of Support Vector Machines is their ability to fit generalisable models with a large number of covariates relative to the amount of training patterns.

To see how effective the performance is in the present case, we shall use the results of only 200 races with 12 input variables, and test the results on a holdout sample of 100 races.

While a proper handicapping model consists of a large number of input variables, fit over a large number of races, we shall consider a simple set of input covariates here to demonstrate the method, relating only to runners' single previous start. These are

- finishing position of previous race
- bookmakers' odds of previous race
- total prizemoney in previous race
- final call position of previous race
- indicator, no call position of previous race
- distance of previous race
- odds × prizemoney, previous race
- weight carried in previous race

[1] Odds of x-to-1 denoting a potential gross return of $1 + x$ for a successful wager.

- days since previous race
- total prizemoney in current race
- weight carried in current race
- distance of current race

At first it may seem strange that only information from the recent past is considered. However, while much can undoubtedly be gained from earlier information, the inclusion of *odds* at last start may be considered to be a Market Summary of the horse's status at last start, which may be considered to include *implicit* information about a horse's earlier history. Thus, it is hoped that for the purposes of the present study, such an 'engineering approximation' will not prove to be a serious drawback. Another seemingly important omission might ar first appear to be information about the relative competitive difficulty or 'class' of a horse's previous races. However, the inclusion of previous race prizemoney (the level of prizemoney generally increasing for higher-class racing) is widely accepted as an (albeit imperfect) proxy for this.

On the other hand, a variable which *is* a serious omission may be said to be the most telling piece of information of all, namely the odds offered in the current race. This omission is deliberate, however, and is due to the fact that the manner in which bettors and bookmakers frame odds may not be very stable over time. Hence, we fit an essentially 'Fundamental' Model first, and combine with up-to-date Market information (in the form of bookmakers' odds) during the 'regularisation' step (i.e., to produce probabilities) as described above.

The Support Vector Machine resulting from these inputs with adjusted finishing position (i.e., the inverse Normal score of the ratio of the finishing position over the number of runners plus 1, as explained in the previous section) as an outcome variable yields a 48% correlation between fitted and predicted values in the training sample and 45% correlation in the holdout set, figures which compare favourably with the corresponding figures of approximately 39% and 38% for the model obtained by applying linear methods to the same data. The optimisation routine used to solve the approximately 4000-dimensional Quadratic Programming problem was a modification of the Sequential Minimal Optimisation algorithm by Platt (1998), and was run on a 1.2 Ghz AMD Athlon with 500Mb RAM in about two minutes.

Unfortunately, as in the case of Artificial Neural Networks, the (many) resulting coefficients of an SVM model are virtually impossible to interpret on their own, containing no discernible information about which input variables are the most influential, nor which combinations of variables interact. Hence, the long list of coefficients will not be presented here. One possible approach to address the issue of the lack of intuition provided by the model involves the use of Artificial Intelligence-based Knowledge Extraction routines, which may be more fruitfully applied to good Artificial Neural Network or SVM predictions than to raw data, but this will not be pursued here.

As a final step for the model, a Multinomial Logit regression is run (see Bolton and Champman (1986) and earlier discussion here) of Win (0-1) as a function of log-1+Odds and the Fundamental variable (i.e., the y's fitted from the SVM regression) using the training sample data. The resulting regression coefficients are 0.85 and 1.45, respectively, with respective T-Statistics of 6.4 and 4.0, suggesting clear marginal predictive value in each variable [the coefficients corresponding to the model of log-1+Odds alone are been 1.12 and 0 for the same dataset]. Also, the Likelihood R-squared (a measure of relative Likelihood Score enhancement over ignorance) is 19.2%, up from 17.3% for the model with bookmakers' odds alone, suggesting a reasonable chance of a betting advantage.

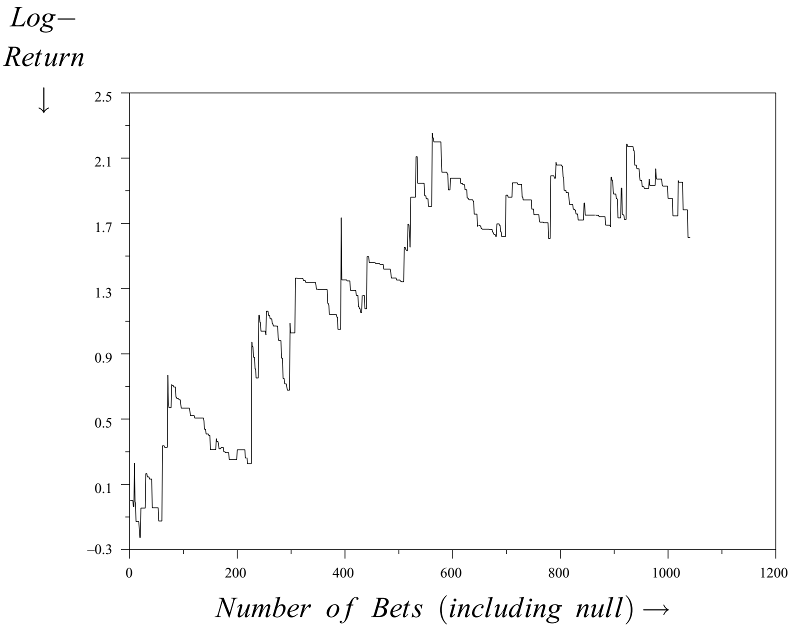
The full model is applied to the holdout set of 100 races, using Kelly betting (see Edelman (2001), Kelly (1956) and earlier discussion), where for instance betting amount b_i for a single bet in a race is given by

$$b_i = \frac{p_i(1 + o_i) - 1}{o_i},$$

where p_i is the forecast probability and o_i are the bookmakers' odds, and where a more involved formula (alluded to earlier) is required for multiple bets in the same race.

The resulting outlay without reinvestment is 9.9 units distributed among 221 bets (an average of two bets per race, with the average bet size comprising about 4% of current Bank level), with a gross return of 11.5 units, for a rate of return of approximately 16%. With reinvestment, this would have left an investor with a wealth growth factor of approximately 500% in the 100 races (approximately 1100 runners, most of which were associated with a bet of zero).

A graph of cumulative logarithmic return as a function of number of bets, including null (zero-stake) bets, is given below:



Cumulative Logarithmic Return (Holdout Sample)

A similar model was fit using Linear Regression methods in the first stage with the same input and output variables as for the SVM model, with the result that (full) Kelly betting yielded no bets at all. Hence, these results were not analysed further.

As a basis of comparison with a nonlinear model, a feedforward Neural Network model with sigmoidal (logit) activation functions and two hidden layers of 20 units each was fit on the same dataset. The resulting correlations between predicted and actual, within-sample and out-of-sample, were 46% and 34% respectively. The betting performance was disappointing, with the model breaking even, approximately, both within and out-of-sample for each of several betting methods, and betting occurring in only about 10% of the races.

3 Discussion

It appears that the use of Support Vector Machines is very successful in this application, as a method of fitting a nonparametric model with a large number of input variables relative to the number of training samples.

It is worth noting that the attempt to use ordinary regression or standard Neural Network methods on the same data failed to produce a profitable model, here resulting in little or no betting at all in the test sample.

With regard to serious handicapping, there are many other variables which should be included, including information from more than a single previous race, race class information, time standard information, jockey information, horses' ages, as well as 'videoform' (quantifiable characteristics extracted from race replays). Benter (1994, 2003) considers such information (albeit for arguably less efficient Hong Kong racing markets), with much greater success than was achieved here, using a mere Linear model, citing a better than 30% return on investment using (full) Kelly betting.

As it is not the purpose of this paper to demonstrate statistical significance of betting efficacy, but rather to suggest that the method may be worthy of further investigation on a larger-scale, no statistical tests regarding the betting returns achieved in the previous section were undertaken. [Indeed, any horserace handicapper with practical betting experience knows that profitability out-of-sample over 100 races, no matter how significant, can only be regarded as an interesting start].

It remains to be seen what the SVM approach, when applied to a more thorough set of input variables and trained on a larger dataset might produce.

As a final note, it is perhaps worth speculating as to why the SVM approach has apparently outperformed the linear and Neural Network approaches so significantly. While investigation into this question has thus far been of an exploratory nature only, one point appears to be fairly clear, which is that the improvement in performance does not appear to be particularly related to the change of loss-function from the squared-error to the ε-insensitive loss function. Firstly, merely fitting the linear model with this modification appears to change very little in the character of the resulting model. Further, there does not appear to be evidence of the kind of heavy-tailed behaviour in the fitted residuals of the SVM model which one might associate with a large discrepancy with Least-Squares methods. It would appear, then, that the added value of the SVM approach must lie in the ability of the method to capture subtle nonlinear and interactive effects of input variables, which is not surprising, as such types of subtleties are likely to be the last to be discovered and exploited in any Market tending towards Efficiency. The failure of the Neural Network model in this case would appear to relate more to the nature of the sigmoidal activation functions than anything else, as a Radial Basis Neural Network which is suitably designed and trained should be expected to be able to perform in a manner similar to the Support Vector Machine model, though such a Network might be expected to be much more difficult to design and more delicate to train. Comparison of the methods presented here with other Kernel methods could prove enlightening, though the relative ease of implementation of the SVM approach would always need to be taken into account when assessing potential marginal gains in performance which might be achieved from other methods.

References

Benter, W. (1994)."Computer Based Horserace Handicapping and Wagering Systems: A Report." In L. Hausch and Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. San Diego, CA: Academic Press, pp. 183–198.

Benter, W. (2003). "Advances in the Mathematical Modeling of Horse Race Outcomes." *12th International Conference on Gambling and Risk-Taking*, Vancouver, BC, Canada (May 2003) (Proceedings).

Bolton, R., and R.G. Chapman. (1986). "Searching for Positive Returns at the Track: A Multinomial Logit Model for Handicapping Horseraces." *Management Science*, 32(8), 1040–1060.

Breiman, L. (1961)."Optimal Gambling Systems for Favorable Games." *Proceedings of the 4th Berkeley Symposium on Mathematics Statistics and Probability*, **1**, 63–68.

Edelman, D. (2000). "On the Financial Value of Information," *Annals of Operations Research*, 100, 123–132.

Edelman, D. (2001). *The Compleat Horseplayer.* Sydney: De Mare Consultants.

Hausch, et al. (1994). In L. Hausch and Ziemba (eds.), *Efficiency of Racetrack Betting Markets*, Ch. 4. San Diego, CA: Academic Press.

Kelly, J. (1956). "A New Interpretation of the Information Rate." *Bell System Technology Journal*, 35, 917–926.

Maclean, L.C., W.T. Ziemba, and G. Blazenko. (1994). "Growth versus Security in Dynamic Investment Analysis." L. Hausch and Ziemba (eds.), *Efficiency of Racetrack Betting Markets*. San Diego, CA: Academic Press, pp. 127–150.

Markowitz, H.M. (1952). "Portfolio Selection." *Journal of Finance*, 7, 77–91.

Platt, J. (1998). "Sequential Minimal Optimisation: A Fast Algorithm for Training Support Vector Machines." Microsoft Research Technical Report MSR-TR-98-14.

Platt, J. (1999). "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods." In A. Smola, P. Bartlett, B. Schöelkopf, and D. Schuurmans (eds.), *Advances in Large Margin Classifiers.* MIT Press.

Scott, D. (1985). *Winning More*. Sydney: Horwitz/Grahame.

Sharpe, W. (1994). "The Sharpe Ratio." *Journal of Portfolio Management*, 21(1), 49–58.

Shin, H.S. (1991). "Optimal Betting Odds Against Insider Traders." *Economic Journal*, 101, 1179–1185.

Shin, H.S. (1993). "Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims." *Economic Journal*, 103, 1141–1153.

Vapnik, V. (2001). *The Nature of Statistical Learning Theory*, 2nd ed. New York: Springer.