Symulacja w Badaniach i Rozwoju
Vol. 1, No. 4/2010

Jerzy KRAWCZUK, Leon BOBROWSKI
Bialystok University of Technology
Faculty of Computer Science
Wiejska 45A Str., 15-351 Białystok
E-mail: j.krawczuk@pb.edu.pl, l.bobrowski@pb.edu.pl

Short term prediction of stock index changes based on linear classification

1 Introduction

Stock index forecasting, or more generally financial instrument price forecasting, is a difficult task, but an encouraging prospect for profit. Many methods of data mining are used for this task. First of all, time series analysis [1], using among others the Auto-Regressive Integrated Moving Average model (ARIMA) introduced by Box and Jenkins [2] and the autoregressive models with conditional heteroskedasticity (ARCH) introduced by Engel [3] and generalized by Bollerslev [4]. Many researchers claim to have good results using other methods, like neural networks [5], or supported vector machines SVM [6]. The approach proposed here, a linear classifier [7], is not commonly used for financial applications, but is widely used in pattern recognition.

2 Prediction model

The simplest form of this forecast is to determine whether the price of a stock index will rise or fall in a given time period, for example tomorrow. This kind of forecast can be treated as a problem of classification. We classify the current condition of the stock market into one of two classes:

ω^+ – in the next day the index will increase

ω^- – in the next day the index will fall

Classification is made based on the current value of an n-dimensional features vector $\mathbf{x}[n] = [x_1,...,x_n]^T$. The components of the $\mathbf{x}[n]$ features vector are real values ($x_i \in R^1$), with features such as:

- historical changes in the value of the index, such as one-day, weekly or monthly changes
- historical changes in the values of other financial instruments.

For example, consider 10 indexes, and for each of these indexes three historical price changes (daily, weekly and monthly). We will obtain a features vector $\mathbf{x}[n]$ of size $n = 30$.

The *linear classifier* $LC(\mathbf{w}[n_k],\theta)$ can be defined by the below decision (prediction) rule:

$$\begin{aligned} &\textit{if} \quad \mathbf{w}[n]^T\mathbf{x}[n] \geq \theta, \textit{then } \mathbf{x}[n] \text{ is located in class } \omega^+ \\ &\textit{if} \quad \mathbf{w}[n]^T\mathbf{x}[n] < \theta, \textit{then } \mathbf{x}[n] \text{ is located in class } \omega^- , \end{aligned} \tag{1}$$

where $\mathbf{w}[n] = [w_1,...,w_n]^T$ is a vector of weights w_i ($w_i \in R^1$) and θ is a threshold ($\theta \in R^1$).

The creation of predictive rules (1) requires the calculation of the parameter values $\mathbf{w}[n]$ and θ. The parameters $\mathbf{w}[n]$ and θ can be determined on the basis of learning sets G^+ and G^- containing examples of feature vectors $\mathbf{x}_j[n]$ from class ω^+ ($j \in J^+$) and from class ω^- ($j \in J^-$).

$$G^+ = \{\mathbf{x}_j[n]: j \in J^+\} \textbf{ and } G^- = \{\mathbf{x}_j[n]: j \in J^-\}. \tag{2}$$

Training set G^+ contains examples of such conditions for stock indexes $\mathbf{x}_j[n]$, which rose, while set G^- contains examples of such conditions $\mathbf{x}_j[n]$, where the value of the predicted index fell the next day.

Definition 1: The sets G^+ and G^- (1) are *linearly separable*, if and only if there exists such a weight vector $\mathbf{w}[n]$ ($\mathbf{w}[n] \in R^n$) and threshold θ ($\theta \in R$), that all the below inequalities for the inner products $\mathbf{w}[n]^T\mathbf{x}_j[n]$ are fulfilled:

$$\begin{aligned} &(\exists\, \mathbf{w}[n], \theta)\ (\forall \mathbf{x}_j[n] \in G^+) \quad \mathbf{w}[n]^T\mathbf{x}_j[n] > \theta \\ &\qquad \textbf{and} \quad (\forall \mathbf{x}_j[n] \in G^-) \quad \mathbf{w}[n]^T\mathbf{x}_j[n] < \theta. \end{aligned} \tag{3}$$

The parameters $\mathbf{w}[n]$ and θ define the hyperplane $H(\mathbf{w}[n],\theta)$ in the feature space $F[n]$ ($\mathbf{x}[n] \in F[n]$):

$$H(\mathbf{w}[n],\theta) = \{\mathbf{x}[n]: \mathbf{w}[n]^T\mathbf{x}[n] = \theta\}. \tag{4}$$

If the inequalities (3) are fulfilled, then all the elements $\mathbf{x}_j[n]$ of the set G^+ are situated on the *positive side* of the hyperplane $H(\mathbf{w}[n],\theta)$ ($\mathbf{w}[n]^T\mathbf{x}_j[n] > \theta$) and all the elements of the set G^- are situated on the *negative side* of this hyperplane (3).

In practice it is not always possible to obtain the exact partition of learning sets described by the inequalities (3). This is not always desirable as well because of the danger of over-fitting to the data sets (2).

A quality of the classification rule (1) depends on the values of the parameters $\mathbf{w}[n]$ and θ. The optimal parameters $\mathbf{w}^*[n]$ and θ^* of the classification rule (1) can be determined through minimization of the convex and piecewise-linear (*CPL*) criterion functions defined on the learning sets G^+ and G^- (2).

3 Convex and piecewise-linear (CPL) criterion functions

Let us define the convex and piecewise-linear (*CPL*) penalty functions $\varphi_j^+(\mathbf{w}[n], \theta)$ and $\varphi_j^-(\mathbf{w}[n], \theta)$ in the below manner [8]:

$$(\forall \mathbf{x}_j[n] \in G^+): \quad \varphi_j^+(\mathbf{w}[n],\theta) = \begin{cases} \theta + 1 - \mathbf{w}[n]^T\mathbf{x}_j[n] & \textbf{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] < \theta + 1 \\ 0 & \textbf{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] \geq \theta + 1 \end{cases} \tag{5}$$

and

$$(\forall \mathbf{x}_j[n] \in G^-) \quad \varphi_j^-(\mathbf{w}[n], \theta) = \begin{cases} \theta - 1 + \mathbf{w}[n]^T\mathbf{x}_j[n] & \textbf{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] > \theta - 1 \\ 0 & \textbf{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] \leq \theta - 1 \end{cases} \tag{6}$$

The function $\varphi_j^+(\mathbf{w}[n], \theta)$ is equal to zero if the feature vector $\mathbf{x}_j[n]$ ($\mathbf{x}_j[n] \in G^+$) (2) is situated on the positive side of the hyperplane $H(\mathbf{w}[n],\theta)$ (4) and is not too near to it. Similarly, $\varphi_j^-(\mathbf{w}[n], \theta)$ is equal to zero if the vector $\mathbf{x}_j[n]$ ($\mathbf{x}_j[n] \in G^-$) is situated on the negative side of the hyperplane $H(\mathbf{w}[n], \theta)$ and is not too near to it.

The *perceptron criterion function* $\Phi(\mathbf{w}[n], \theta)$ is defined on feature vector $\mathbf{x}_j[n]$ from the sets G^+ and G^- (1) as the weighted sum of the penalty functions $\varphi_j^+(\mathbf{w}[n], \theta)$ (5) and $\varphi_j^-(\mathbf{w}[n], \theta)$ (6) [8]:

$$\Phi(\mathbf{w}[n], \theta) = \sum_{j \in J+} \alpha_j \varphi_j^+(\mathbf{w}[n], \theta) + \sum_{j \in J-} \alpha_j \varphi_j^-(\mathbf{w}[n], \theta), \tag{7}$$

where nonnegative parameters α_j represent *prices* linked to particular feature vectors $\mathbf{x}_j[n]$. The minimization of the criterion function $\Phi(\mathbf{w}[n], \theta)$ (7) allow to find the optimal parameters $\mathbf{w}[n]^*$ and θ^* of the prediction rule (1):

$$(\exists(\mathbf{w}[n]^*, \theta^*)) \ (\forall(\mathbf{w}[n], \theta)) \ \Phi(\mathbf{w}[n], \theta) \geq \Phi(\mathbf{w}[n]^*, \theta^*) = \Phi^* \geq 0. \tag{8}$$

It was proven that the nonnegative value Φ^* is equal to zero ($\Phi^* = 0$) if and only if the sets G^+ and G^- (2) are linearly separable (3) [8].

For the purpose of the feature selection, a modified criterion function $\Phi_\lambda(\mathbf{w}[n], \theta)$, was introduced [8]. $\Phi_\lambda(\mathbf{w}[n], \theta)$ includes additional *CPL* penalty functions in the form of the absolute values $|w_i|$ and the *costs* γ_i ($\gamma_i > 0$) related to particular features x_i:

$$\Psi_\lambda(\mathbf{w}[n], \theta) = \Phi(\mathbf{w}[n], \theta) + \lambda \sum_{i \in I} \gamma_i |w_i|, \tag{9}$$

where λ ($\lambda \geq 0$) is the *cost level*, and $I = \{1,\ldots\ldots,n\}$.

Similarly to the function $\Phi(\mathbf{w}[n], \theta)$ (7) The criterion function $\Psi_\lambda(\mathbf{w}[n],\theta)$ (9), similarly to the function $\Phi(\mathbf{w}[n], \theta)$ (8) is convex and piecewise-linear (*CPL*). The basis exchange algorithms allow to efficiently find the optimal vector of parameters (*vertex*) $\mathbf{w}_\lambda[n]$ and the optimal threshold θ_λ constituting the minimum of the *CPL* function $\Psi_\lambda(\mathbf{w}[n], \theta)$, even in the case of large data sets G^+ and G^- (1) [8]:

$$(\exists(\mathbf{w}_\lambda[n], \theta_\lambda)) \ (\forall \mathbf{w}[n], \theta) \ \ \Psi_\lambda(\mathbf{w}[n],\theta) \geq \Psi_\lambda(\mathbf{w}_\lambda[n], \theta_\lambda) = \Psi_\lambda^{\wedge} \tag{10}$$

The criterion function $\Psi_\lambda(\mathbf{w}[n],\theta)$ (9) is used in the *relaxed linear separability* (*RLS*) method of feature selection [9]. The below feature reduction rule is used in the *RLS* method:

$$(w_{\lambda i} = 0) \Rightarrow (\text{the feature } x_i \text{ is reduced}), \tag{11}$$

where $\mathbf{w}_\lambda[n] = [w_{\lambda 1},...,w_{\lambda n}]^T$ is the vector constituting the minimum (10) of the function $\Psi_\lambda(\mathbf{w}[n], \theta)$ (9).

In accordance with the *RLS* method, a gradual increase of the cost level λ value in the criterion function $\Psi_\lambda(\mathbf{w}[n], \theta)$ (9) allows the reduction of (11) successive features x_i [9]. In result a reduction (11) of successive features x_i, the descended sequence of feature subspaces F_k $[n_k]$ can be generated:

$$F[n] \supset F_1[n_1] \supset F_2[n_2] \supset \ldots \supset F_{k'}[n_{k'}], \tag{12}$$

where $n_k > n_{k+1}$.

The manner of feature subspaces $F_k[n_k]$ evaluation is to define in order to determine the stop criterion of the feature reduction in the above sequence.

4 Evaluation of feature subspaces $F_k[n_k]$

In accordance with the *RLS* method of feature selection, a quality of a given feature subspace $F_k[n_k]$ is evaluated on the basis of evaluation of the optimal linear classifiers (1) designed in this subspace.

The decision rule of the optimal linear classifier $LC(\mathbf{w}^*[n_k],\theta^*)$ in the feature subspace $F_k[n_k]$ ($F_k[n_k] \subset F[n]$) is defined in a similar manner to (1). The optimal parameters $\mathbf{w}^*[n_k]$ and θ^* of the linear classifier $LC(\mathbf{w}^*[n_k],\theta^*)$ can be determined through the minimization (8) of the criterion function $\Phi_k(\mathbf{w}[n_k],\theta)$ (7) defined on the feature vectors $\mathbf{x}_j[n_k]$ ($\mathbf{x}_j[n_k] \in F_k[n_k]$). The better optimal linear classifier $LC(\mathbf{w}^*[n_k],\theta^*)$ in the feature subspace $F_k[n_k]$ means a higher quality of this subspace.

The quality of the linear classifier $LC(\mathbf{w}^*[n_k],\theta^*)$ (1) defined by the parameters $\mathbf{w}^*[n_k]$ and θ^* can be evaluated by using the error estimator (*apparent error rate*) $e_a(\mathbf{w}^*[n_k],\theta^*)$. The apparent error rate is defined as the fraction of wrongly classified elements $\mathbf{x}_j[n_k]$ of the sets G_k^+ and G_k^- (1) in the feature subspace $F_k[n_k]$ [10]:

$$e_a(\mathbf{w}^*[n_k],\theta^*) = m_a(\mathbf{w}^*[n_k],\theta^*) \ / \tag{13}$$

where m is the number of all elements $\mathbf{x}_j[n_k]$ of the sets G_k^+ and G_k^- (1), and $m_a(\mathbf{w}^*[n_k],\theta^*)$ is the number of elements $\mathbf{x}_j[n_k]$ from these sets wrongly allocated by the rule (1).

The parameters $\mathbf{w}^*[n_k]$ and θ^* of the optimal linear classifier in the feature subspaces $F_k[n_k]$ were estimated from the learning sets G_k^+ and G_k^- (2) through minimization of the perceptron criterion function $\Phi_k(\mathbf{w}[n_k],\theta)$ (7) defined on elements $\mathbf{x}_j[n_k]$ of these sets. Because the same data $\mathbf{x}_j[n_k]$ are used for classifier designing and for classifier evaluation, the evaluation result (13) is too optimistic (*biased*) [10]. For example, if the sets G_k^+ and G_k^- (2) are linearly separable (3), then all the elements $\mathbf{x}_j[n_k]$ of the learning sets are correctly classified by the optimal classifier $LC(\mathbf{w}^*[n_k],\theta^*)$ (1), and the apparent error (13) is equal to zero ($e_a(\mathbf{w}^*[n_k],\theta^*) = 0$). But it is typically found in practical applications that the error rate of the classifier $LC(\mathbf{w}^*[n_k],\theta^*)$ (1) evaluated on vectors $\mathbf{x}[n_k]$ that do not belong to the learning sets G_k^+ and G_k^- (2) is higher than zero.

For the purpose of the classifier bias diminishing, the cross validation procedures can be applied [10]. The term *p-fold cross validation* means that data sets G_k^+ and G_k^- (2) have been randomly divided into p parts P_i, where $i = 1,\ldots, p$. The vectors $\mathbf{x}_j[n_k]$ contained in $p - 1$ parts P_i were used for the definition of criterion function $\Phi_k(\mathbf{w}[n_k],\theta)$ (8) and in the computation of optimal parameters $\mathbf{w}^*[n_k]$ and θ^*. The remaining vectors $\mathbf{x}_j[n]$ were used as a *test set* (one p-part $P_{i'}$) for the evaluation of error rate $e_{i'}(\mathbf{w}^*[n_k],\theta^*)$ (13). This evaluation is repeated p times, and during each time a different p-part $P_{i'}$ is used as the test set. After this, the mean value $e_c(\mathbf{w}^*[n_k],\theta^*)$ of the errors rates $e_{i'}(\mathbf{w}^*[n_k],\theta^*)$ (13) on the elements of the test sets $P_{i'}$ is computed. The cross validation procedure allows us to use different vectors $\mathbf{x}_j[n_k]$ (1) for designing of the classifier (1), its evaluation, and as a result,

to reduce the bias of the error rate estimation (13). The error rate $e_c(\mathbf{w}^*[n_k],\theta^*)$ (13) estimated during the *cross validation* procedure will be called the *cross-validation error* (*CVE*). A special case of the *p-fold cross validation* method is the *leave-one out* procedure. In the case of the *leave-one out* procedure, the number p of the parts P_i is equal to the number m of elements $\mathbf{x}_j[n_k]$ in the sets G_k^+ and G_k^- (2).

The *CVE error rate* $e_c(\mathbf{w}^*[n_k],\theta^*)$ (13) of the optimal linear classifier $LC(\mathbf{w}^*[n_k],\theta^*)$ (1) is used in the relaxed linear separability method as the criterion for evaluation of particular feature subspaces $F_k[n_k]$ in the sequence (12) [10]. Feature subspace $F_{k'}[n_{k'}]$ that is linked to the linear classifier $LC(\mathbf{w}^*[n_{k'},\theta^*)$ (1) with the lowest *CVE error rate* $e_c(\mathbf{w}^*[n_{k'}],\theta^*)$ (13) is considered as the optimal one in accordance with the *RLS* method of feature selection.

5 Experiment

In this experiment historical market data were used starting from Jan 2nd 2008 and ending October 30th 2010. During that time period there were 712 working days, which give us exactly same number of feature vectors $\mathbf{x}_j[n]$, one feature vector per day. Each vector describes the condition of the stock indexes for each day. There were 294 features used to describe the market environment per day. These features were derived from historical prices of 42 selected financial instruments. For each of the 42 instruments, the change in price was taken from 1,2,3,4,5,10 and 22 days back. The full list of instruments are presented in table 3, with short descriptions and their exchange symbols. All of these instruments are traded on stock exchanges in the United States, but they reflect the situation in different countries and different markets. All of them are ETF's (exchange-traded funds)[9] **.**

An exchange-traded fund (ETF) is an investment fund traded on stock exchanges, much like stocks. An ETF holds assets such as stocks, commodities, or bonds and trades at approximately the same price as the net asset value of its underlying assets over the course of the trading day.

www.wikipedia.org

24 of them track the major indexes in 24 countries all over the world, 4 are for world regions. For example, GML is an Emerging Latin American ETF. 8 of them are for different currencies against the US dollar, 2 are for the commodities oil (USO) and natural gas (UNG), 2 are for the precious metals gold (GLD) and silver (SLV). One (DBA) is tracking agricultural commodity prices like corn, wheat, soy beans and sugar.

The forecast is done for the one-day price change of a given index. Each of the feature vectors $\mathbf{x}_j[n]$, is classified in class ω_0 or ω_1. Class ω_0 was assigned if in the next day there was a decline in the value of the forecasted instrument, class ω_1 if there was an increase. There were 42 assignments done for same vectors x[n], one assignment for each of the 42 predicted instruments used in the experiment. In the result there were 42 predictive models (1) built, one for each instrument. For each model the classification accuracy was calculated (13).

6 Experiment results

All 712 features vectors $\mathbf{x}_j[n]$ were divided into 2 sets, a learning set and a test set. The division was done in a two ways. The first way was to use data from years 2008 and 2009 as a learning set, and data from 2010 as the test set. The second was to divide randomly all

data into 3 sets, and use each set as a test set, and the other 2 as learning sets (otherwise known as 3 folds cross-validation). In the second case statistics were calculated as an average of results from the 3 runs.

Table 1. Classification accuracy for linear classifier on all features

	1. Time split		2. Random 3 folds cross-validation	
	Training set, 2008-2009	Test set 2010	Training set random	Test set random
	Apparent accuracy	Accuracy	Apparent accuracy	Accuracy
Average	***100***	***51.9***	***100***	***51.9***
symbol				
DBA	100	45.5	100	50.4
EWO	100	59.7	100	53.4
GAF	100	48.3	100	50.6
RSX	100	48.8	100	51.6
EWZ	100	55.5	100	52.3
EWW	100	47.4	100	49.1
GML	100	52.1	100	52.9

In the first case, the learning set had a size of 505 days, and test set size of 207 days. All 505 vectors $\mathbf{x}_j[n]$ for every predicted symbol were linearly separated. That means that apparent accuracy (accuracy of classification measured on same data that the model was built with) is equal to 1 for all symbols. In table 1 results for a few selected symbols are presented, and also the minimum, average and maximum for each statistic. The models built with data from 2008-2009 were then used to predict price changes in 2010. The classification accuracy on the test set with data from 2010 in average is equal to 51.9%. That means that on average the model correctly predicted the direction of market moves only 51.9% of the time. This result is close to random or 50%. The best result was achieved for EWO (Austria) at 59.7%, the worst for DBA (Agriculture) at 45.5%.

Since market data are probably time dependent, there could be an argument that the 2010 market environment was different than in 2008 and 2009, so the classifier learnt trading rules that weren't appropriate for 2010. To verify this hypothesis a second test was done, which randomly split all data into training and test sets. 3 folds cross-validation was applied. The learning set had a size of 474-475 days, and the test set 237-238 days. The right part of table 1 shows the detailed results. Average accuracy for the test set was exactly the same as in the first case, 51.9%.

7 Feature selection

A complex model with 294 features that produces only slightly better than random forecasts is, in practice, not very useful. To build a better model that can discover more general rules in the data, we reduced the number of features.

Features were selected using the RLS (Relax-Linear-Separability[10]) method. This method consists of subsequent rejections of one or more features. Following each rejection, the quality of classification is evaluated by leave one out cross-validation.

The resulting quality of classification for the instrument GAF (Emerging Middle East & Africa), depending on the number of features, is shown in graph 1.

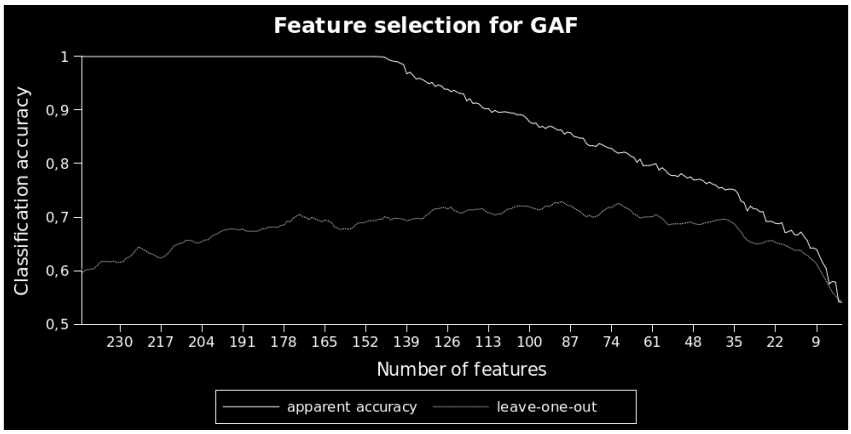

Fig. 1. Feature selection for predicting the GAF

There are 2 measures. First is apparent accuracy, which is equal to 1 until reducing the number of features to 150, at which point we still have linear separability. For less than 150 features apparent accuracy goes down linearly until around 15 features then it drops very quickly.

The second accuracy measure is calculated using the leave-one-out (loo) method. We use one day as validation and the remaining days as training data to build the model. This is repeated so that each observation in the sample is used once as the validation data. In this case data that are used to build the model are not used to calculate accuracy. The RLS feature selection method chooses those feature subsets that result in the highest leave-one-out accuracy. For GAF the maximum accuracy (74.5%) was achieved with 74 features. This was the highest accuracy found for any instrument. Loo accuracy grows with feature reduction, with local accuracy maximums at 175 and 130 features and the global maximum at 74 features. Then accuracy goes down, more quickly for less features.

8 Feature selection results

Detailed results for all symbols are presented in table 2. The methodology of testing was the same as before. The experiment was done by dividing all data into training and test sets. In one case the test set was chosen as Jan-Oct 2010, in another case as a 1/3 random part of all data. Results are presented only for the best subset of features for each instrument.

Table 2. Classification accuracy for linear classifier on subset of features selected by RLS method

	Training set, 2008-2009			Test set 2010	Training set, random 3 folds cross-validation			Test set random
	1.size	2.app	3. loo	4. acc	1.size	2.app	3. loo	4. acc
Average	108,74	86.0	71.1	***51.8***	106,58	87.0	71.5	***52.7***
Symbol								
DBA	98	84.1	72.7	47.4	117,67	89.4	72.6	54.1
EWO	120	88.5	72.4	57.8	102,67	87.1	72.5	55.9
GAF	74	81.5	74.5	52.1	101	87.1	73.5	54.0
RSX	60	77.6	70.4	52.6	88	85.1	72.1	55.1
EWZ	154	90.5	69.0	52.1	111	87.1	72.0	51.5
EWW	80	79.2	69.3	46.0	104,67	85.9	70.0	48.8
GML	77	79.4	69.0	58.8	105,33	86.6	71.5	53.1

Columns description:
1. size – optimum number of features selected by RLS method
2. app – apparent accuracy on training set
3. loo – leave-one-out accuracy on training set
4. acc – accuracy measured on test set, for model built on training set

The optimal number of features selected by the RLS method was different for different symbols. The lowest number was 60 features for RSX (Russia), the highest was 154 features for EWZ (Brazil). On average 108.74 features were selected. Classification results on the training set measured by leave-one-out were pretty high at 71.1%, but on the test set they are rather low at 51.8%. The best accuracy for 2010 test set was achieved for GML (Emerging Latin America) at 58.8% and worst for EWW (Mexico) at 46.0%. This result is similar to the previous experiment that used all features. Slightly better performance of 52.7% was achieved on random 3-folds cross-validation.

9 Summary

All results on test sets are in average similar at 51.9% for linear classification on all features, and 51.8% for the best subset of features in 2010. A little better results were observed for a subset of features on randomly split data, 52.7%, which is better by 0.9%. This may suggest that there is time structure in the data, meaning that rules learnt from the training data from 2008-2009 are not working well in 2010, or at least the first few months. More research needs to be done to verify this hypothesis. One approach could to be use training and test sets moved over time. For example, 6 months of training data then 1 month of test data.

Classification accuracy does not directly translate into profits when using the model for investing decisions. For example if we used some model to trade over two days, where one day we made 2% and one day we lost lost 0.1%, accuracy would only be 50%, but we have had a high profit of 1.9%. To address this effect models could be built and used to simulate actual buy/sell transactions on the market, and portfolio histories could then be analyzed.

References

1. Mills T.C. *Time Series Techniques for Economists.* Cambridge University Press, 1990
2. Box G. and Jenkins G., *Time series analysis: Forecasting and control*, San Francisco: Holden-Day, 1970
3. Engle R.F., Autoregressive Conditional Heteroscedasticity with Eshtimates of Variance of United Kingdom Inflation, *Econometrica* 50:987-1008, 1982
4. Bollerslev T., Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31:307-327, 1986
5. Haefke C., Helmenstein C., Neural Networks in the Capital Markets: An Application to Index Forecasting, *Computational Economics*, pp. 37-50, 1996
6. Tay F.E.H., Cao L., Application of support vector machines in financial time series forecasting, *Omega* 29 (2001) 309–317
7. Duda O. R., Hart P. E., Stork D. G.: *Pattern Classification,* J. Wiley, New York, 2001
8. Bobrowski L.: *Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych*, Wydawnictwa Politechniki Białostockiej, Białystok, 2005
9. Bobrowski L., Łukaszuk T: Feature selection based on relaxed linear separabilty, *Biocybernetics and Biomedcal Engineering*, 2009, Volume 29, Number 2, pp. 43-59
10. Lachenbruch P.A., *Discriminant Analysis*, Hafner Press, New York, 1975
11. Liu H., Motoda H. (Eds.) *Computational Methods of Feature Selection*, Chapmann&Hall. CRC, New York 2008
12. Carrell L., *ETFs for the Long Run: What They Are, How They Work, a and Simple Strategies for Successful Long-Term Investing*. JW Wiley, 2008 ISBN 978-0-470-13894-6

Abstract

This article describe the linear classifier based on convex and piecewise-linear function (CPL) and it application to market prediction. In an experiment we use CPL linear classifier to predict direction of one day change in stock index price. We use classification approach to predict only direction of change (grow or decline) of the index, not it quantity as in regression approach. Total number of instruments used in experiment including currencies is 42. Prediction of one index is based on historical prices of all 42 indexes. Using 7 historical values for each index it produce 294 attributes. Such high dimensional feature space was reduced by feature selection method - relaxed linear separability (RLS). Details of this methodology are also presented. Features was selected and model was build on training data. Test data (holdout data) was used for checking model accuracy. Model in average correctly classify (predict) 51.9% direction of daily index changes.

Jerzy KRAWCZUK, Leon BOBROWSKI

Krótkoterminowe prognozowanie indeksów giełdowych w oparciu o klasyfikator liniowy

Streszczenie

W artykule opisano klasyfikator liniowy oparty o wypukłe i odcinkowo-liniowe funkcje kary (CPL) i jego zastosowanie w prognozowaniu giełdy. W przeprowadzonym eksperymencie klasyfikator liniowy CPL został użyty do prognozy kierunku jednodniowej zmiany indeksów giełdowych. W zastosowanym podejściu klasyfikacyjnym prognozowano jedynie kierunek zmian (wzrost lub spadek), a nie dokładną wartość indeksu (podejście regresyjne). W eksperymencie użyto 42 instrumentów finansowych w tym m.in. kursów walut. Jednodniowa prognoza wybranego instrumentu budowana jest w oparciu o wartości historyczne wszystkich 42 instrumentów. Używając 7 danych historycznych dla każdego instrumentu, uzyskano w sumie 294 atrybuty. Tak wielowymiarowa przestrzeń została zredukowana metodą selekcji cech opartą o relaksację liniowej separowalności. Metoda ta została opisana szczegółowo. Selekcja cech i budowa modelu w wybranej podprzestrzeni została przeprowadzona na zbiorze uczącym (treningowym). Natomiast ocena modelu została przeprowadzona na zbiorze testowym. Otrzymany wynik to średnio 51.9% prawidłowo sklasyfikowanych (prognozowanych) dziennych zmian indeksów giełdowych.

Praca wspierana przez projekt S/WI/2/08 Politechniki Białostockiej.

Table 3: Financial instruments ETF's used in experiment

Region Type	Symbol	Description
Europe	FEU	SPDR STOXX Europe 50
	GUR	SPDR S&P Emerging Europe
	RSX	Market Vectors Russia ETF
	EWU	iShares MSCI United Kingdom Index
	EWL	iShares MSCI Switzerland Index
	EWI	iShares MSCI Italy Index
	EWD	iShares MSCI Sweden Index
	EWG	iShares MSCI Germany Index
	EWP	iShares MSCI Spain Index
	EWQ	iShares MSCI France Index
	EWO	iShares MSCI Austria Investable Mkt Idx
	EWK	iShares MSCI Belgium Investable Mkt Idx
	EWN	iShares MSCI Netherlands Invstbl Mkt Idx
Asia	EWJ	iShares MSCI Japan Index
	EWM	iShares MSCI Malaysia Index
	EWT	iShares MSCI Taiwan Index
	EWY	iShares MSCI South Korea Index
	EWH	iShares MSCI Hong Kong Index
	FXI	iShares FTSE/Xinhua China 25 Index
	EWA	iShares MSCI Australia Index
	EWS	iShares MSCI Singapore Index
North America	SPY	SPDR S&P 500 USA
	EWC	iShares MSCI Canada Index
Latin America	GML	Emerging Latin America
	EWW	iShares MSCI Mexico Index Fund
	EWZ	iShares MSCI Brazil Index Fund
Africa	EZA	iShares MSCI South Africa Index
	GAF	SPDR S&P Emerging Middle East & Africa
Commodities	USO	United States Oil Fund
	UNG	United States Natural Gas
	DBA	PowerShares DB Agriculture
	GLD	SPDR Gold Shares
	SLV	iShares Silver Trust
Currencies	FXA	CurrencyShares Australian Dollar Trust
	FXB	CurrencyShares British Pound Sterling Trust
	FXC	CurrencyShares Canadian Dollar Trust
	FXE	CurrencyShares Euro Trust
	FXY	CurrencyShares Japanese Yen Trust
	UDN	PowerShares DB US Dollar Index Bearish
	UUP	PowerShares DB US Dollar Index Bullish
	DBV	PowerShares DB G10 Currency Harvest