

PCA BASED SUPPORT VECTOR MACHINE TECHNIQUE FOR VOLATILITY FORECASTING

R.Seethalakshmi¹, V.Saavithri², V. Badrinath³, C.Vijayabanu⁴

¹School of Humanities and Sciences, Department of Mathematics, SASTRA University, Thanjavur

²Department of Mathematics, Nehru Memorial College, Puthanampati, Tiruchirappalli

³Dean, School of Management, SASTRA University, Thanjavur

⁴Associate Professor, School of Management, SASTRA University, Thanjavur

Abstract

Conditional Volatility of stock market returns is one of the major problems in time series analysis. Support Vector Machine (SVM) has been applied for volatility estimation of stock market data with limited success, the limitation being in accurate volatility feature predictions due to general kernel functions. However, since Principal Component Analysis (PCA) technique yields good characteristics that describe of the volatility time series in terms of time varying risk estimates, this paper focuses designing PCA based SVM technique (PCASVM) to overcome the above limitations. Gaussian kernel trick has been invoked to build the learning mechanism of non-linear classifier. The applicability of this proposed combinatorial structure (PCASVM) for forecasting the volatility has been confirmed through experiments on real time bench mark data sets. The comparative study of PCASVM with SVM shows the superiority of the proposed technique.

Keywords: Conditional volatility, SVM, PCA, kernel function, stock market.

1. INTRODUCTION

As the future is often predictable from present and past events, financial time series prediction is identified as the recent demanding application of time series problems in modern times as it is possible to build software, which will be able to limitate the experienced investor's decision process. Such financial time series are naturally noisy, referring to the unavailability of entire information from the past behaviour of financial markets that fully catches the dependency between upcoming and past prices. It is also non-stationary and deterministically chaotic meaning that its short-term random but long-term deterministic.

This brings in the volatile nature of price of a financial instrument over time which states to the amount of ambiguity or risk about the size of changes in a security's value which can possibly be spread out over a large or short range of values depending on higher or lower volatility. Hence volatility often plays an extremely important role in measuring total sheik of financial assets, evaluating option prices and conducting hedging strategies as clever financial instruments make it possible for firms to hedge their risks. Hence it is of utmost important to search an effective method for volatility forecasting.

There are many machine learning techniques [1] available, of which, Neural networks an artificial intelligence technology have been identified to be successful in modelling financial time series.

Though Neural Networks (NN) are more supple, having the capacity to learn active learning mechanism through a reiterating process utilising a new data patterns, which speak

for themselves, they also has a wide capability which leads it to confine not just the useful source of information which augurs well with the training data but in additional unnecessary ill conditioned data. This results is only memorize the exercise data and leads to lack of generalizing capability to the outlier data.

Hence to enhance the generalization capability of NN, a modified NN algorithm has been explored, called Support Vector Machine (SVM) which employs the structural risk minimization principle which attempts to reduce an upper bound of the error due to generalization rather than minimizing the absolute error.

1.1 Background: (Literature Survey) and Recent Trends:-

Empirical research on security prices around 1950's supported "Random Walk" models which was only an approximate model to describe how security prices vary over time. As a logical extension of the random walk model an option pricing formula had been derived and Black Scholes [2] needed to model the movement of stock price in short intervals of time to find a trading strategy to rebalance a portfolio of an option with its underlying stock in the continuous manner. Though adequate extensions of [3-12] became the standard way to mould asset price behaviour both for derivative pricing and financial applications in general, empirical evidence showed the substantial difference of behaviour of asset returns in the real world [13] in terms of time variation in returns distribution, vast difference between transaction prices and equilibrium prices, un matching serial correlation, non-lognormal

observed price-changes, mean reversion in the volatility and mean reversion in price levels. This new degree of complexity induced by stochastic as well as time shifting nature of volatility namely two types of risk (risk due to price and risk due to volatility) affected the academic option value making Black-Scholes pricing no longer applicable.

Other parametric models, formal models of time-varying volatility of ARCH family (Auto Regressive Conditional Heteroskedasticity), GARCH model and others suffer from serious limitations of their assumptions behind their parametric technique to forecast volatility of stock markets. After the advent of computer science, several non-parametric methods have been designed to meet the needs of

several Engineering applications and in particular time series analysis problems necessitate such type algorithms to resolve such type of complexities of non-linear relationship modelled by parametric techniques. Moreover several implementable non-parametric methods have been studied by researchers with machine learning techniques for majority of applications in finance varying from option pricing, stock index trading to exchange of currency. The recently established techniques, their features / characteristics, applications and limitations are summarized below. (Table-1).

Table-1: Summary of established techniques, features for characteristics, applications and limitations

S.No	Methods	Techniques developed	Features/ Characteristics	Applications	Limitations	Remarks
1.	SVM Approximation using kernel	Kernel PCA	Reduction of number of vectors, to a span of a fixed subspace	Hand written recognition	Fixed error bounds with limits $\pm \sqrt{\lambda k} / m$	Checked for two sample data sets
2.	PCA	Covariance matrix algorithm	Pattern recognition and image compression	Image reconstruction	Training and testing not considered	Limited improvement in accuracy
3.	Application of SVM in financial time series forecasting	Multilayer back propagation neural network	Method validated through four criteria NMSE, MAE, DS and WDS	Stock market forecasting	Features of PCA data not adequately described	Comparative study exhibits the superiority of the proposed scheme
4.	Financial forecasting using SVM	Back propagation algorithm	Compared with Arima prediction model, usage of five statistical metrics (NMSE, MAE, DS, CP and CD)	Sensitivity of SVM parameters described	Absence of feature extraction techniques	Limited number of support vectors
5.	SVM formulation to PCA analysis to its kernel version	Fisher discriminant analysis	Mapping on to a high dimensional feature space and implementation of kernel trick using Mercer theorem	Image Processing	Not validated with bench mark data sets.	Compared with other tools and the superiority of proposed method shown
6.	SVM with adaptive parameters in financial time series forecasting	RBF and BP network	Bench mark data sets considered	Financial time series problems	Sophisticated weight functions not considered	Exhaustive experimental validation

2. PCA PRELIMINARIES FOR DATA PROCESSING

PCA, used as a tool in exploratory data analysis, which describes most of the variance of data and which has been discovered by Karl Pearson in 1901, helps us to construct predictive models. After the development of computers, PCA has been widely used for pattern recognition problems due to its inherent capability of dimensionality reduction without losing information or capturing all features to generate a feature space. PCA works with the important concepts such as linear combination of original variables, Eigen values for each eigen vector of linear transformation and eigen vectors to construct factors and with the basic assumptions: (i) Linearity-Linearly frame for change of basis (ii) Large variances to have important and interesting structures (iii) The Principal components are orthogonal (solvable with linear algebraic decomposition techniques). PCA has been successfully combined Artificial Intelligence techniques such as Neural network and Fuzzy logic with limited success, the limitation being the usage of PCA as linear kernels to train the data. However use of SVM technique is expected to overcome the limitation as it maps the featured data on to high dimensional plane for classification and optimization. The following algorithm describes the essential steps of PCA as a feature extraction tool.

2.1 Essential Steps of PCA Algorithm:

Step:1 All sample data each consisting of five days closing price (five component vector) are converted into a row vector by concatenating each row consisting of five elements with first row in sequence.

Step:2 A 40 row vector matrix of 1200 data is constructed by combining together the row vectors of six sample vectors containing 30 closing price, x_i ; $i = 1, 2, \dots, 40$.

Step:3 A mean stock vector ψ of n row vectors ($n=40$) is obtained to extract the principle features of stock market data.

$$\psi = \frac{1}{n} \sum_{i=1}^n x_i, i = 1, 2, \dots, n.$$

Step:4 A matrix Y is obtained by subtracting mean stock vector ψ from each stock vector x_i of the training set.

$$Y_i = x_i - \psi.$$

Step:5 A data covariance matrix CV is constructed by multiplying matrix Y with its transpose matrix Y^T .

Step:6 40 highest valued eigen vectors are then picked to make a space of stock from the resultant covariance matrix CV .

(All the above steps have been processed using Mat lab 7.10).

The advantage of PCA as a pre-processing tool is manifold and towards time series analysis it not only serves for dimensionality reduction favouring extraction of crisp features but also serves as a special case of factor analysis which is superior to many other bivariate statistical techniques. The existence of strong market factor is highly realized by PCA in a few of its principal components possessing large percentage of total variance. Though wavelet techniques do this work in obtaining wavelet features, it needs rigorous mathematical treatment such as Multi resolution analysis, Hilbert space assumptions, which are computationally expensive. The selection of principal components are such that the first to explain maximum proportion of total variance, the second to describe maximum of remaining variance etc., Due to this reason, principal component solution becomes a remarkable appropriate test for existence of strong market factor. [14]

3. THE ROLE OF SVM IN REGRESSION

Given a data set $D = \{(x_i, t_i)\}$ where x_i ; $i = 1, 2, \dots, n$ is the input vector and t_i are the target value which evolves from an unknown function. Regression approximation describes the method of estimating the unknown function based on the given data set. For a large set of data the estimation of these functions was taken by research community by minimizing the empirical risk with limited success. Of late, support vector machine, originally developed by Vapnik using the concepts of statistical learning theory has been proved to be a successful tool in non linear classification and estimation of unknown function governing a large data sets. The success of SVM lies in the principle of structural risk minimization (minimizing the upper bound of generalization error) [15] instead of empirical risk minimization (minimizing the upper bound of training error) by mapping the data onto higher dimensional space. With the given data SVM formulates a convex optimization problem (in most of the applications, it is Quadratic programming problems) for which we seek the solution for dual problem. Generally Mercer's theorem is used for the construction of positive definite kernel for SVM regressor. Various versions of kernels have been devised by researchers to propose suitable SVM classifier or Regressor for function estimation depending upon the nature of application.

3.1 Essential Steps of SVM Regression

(i) Given the feature vectors-training data set (from PCA) $F = \{x_i, t_i\}$ for $i = 1, \dots, N$ with input vectors $x_i = \{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)}\} \in \mathbb{R}^n$ and targets $t_i \in \{-1, 1\}$ for binary classifications (say risk and return).

Mapping of the data from input domain to a feature domain F with a non-linear function $\phi: X \mapsto F$ so as to define a discriminant function $f(x_i)$ such that $f(x_i) = \omega^T \phi(x_i) + b(1)$

(ii) Use of linear combination of training samples of volatility estimates

($\omega = \sum_{k=1}^n \alpha_k x_i^{(h)}$) for obtaining modified discriminant function

$$f(x_i) = \sum_{k=1}^n \alpha_k (x_i^k)^T x_i + b \text{-----}(2)$$

In feature space F,

$$f(x_i) = \sum_{k=1}^n \alpha_k \phi(x_i^k)^T \phi(x_i) + b \text{-----}(3)$$

(iii) Use of the kernel function

$$\ker(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \exp(-\nu \|x_j - x_i\|^2)$$

α_k - dual representation for decision boundary

$\nu > 0$ - parameter that controls Gaussian width,

(iv) Computation of $Q_{ij} = y_i y_j \ker(x_i, x_j) = \phi(x_i)^T \phi(x_j)$

(v) Use of penalty parameter for error term for formulation of Quadratic Programming Problem(QPP):

$$\underset{\alpha}{\text{Min}} Q(\alpha) = \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

subject to $y^T \alpha = 0$, $0 \leq \alpha_i \leq C$, for all α ranging from 1 to L.-----(4)

(vi) Solution of QPP for stock returns predictions, in terms of transformed variables.

(vii) Classification of data into two classes: Return & Risk

3.2 Choice of Kernel Function

The quadratic complexity induced in the quadratic programming in the feature space F is feasible in the lower dimensional data. Though we used PCA for dimensionality reduction while pre-processing data, handling the data in terms of thousands of dimensions with quadratic complexity (even with sparse matrices) will show the limited efficiency of the vector machine. Introduction of kernel function in SVM solves this problem by expressing the weights as a linear combination of training samples expressed as in (ii) and in feature space as in (iii). Though any function satisfying the condition due to Mercer [16] can be utilized as the kernel, widely used kernel function includes the algebraic kernel: $\ker(x, y) = (xy+1)^t$, the kernel function

Gauss is $\exp(-\nu \|x_i - x_j\|^2)$ where t is the degree of the algebraic kernel and ν is the band width of Gauss. As parameter of kernel describes the structure of high dimensional feature space $\phi(x)$ that maintaining the difficulty under control and there is no predetermined way in choosing all open parameters of SVM, much care is exercised by researchers for choosing an appropriate value of these parameters. In this paper Gaussian kernel is chosen as it generally provides good results under general assumptions that include requirements of no additional information about the general data etc.,

4. THE NEED FOR THE COMBINATORIAL STRUCTURE OF PCA & SVM(PCASVM)

Some of the specific limitations of PCA such as “no realized probability distribution of data sets, inability to discover the true economic interpretations of PCs using the transformed variables, drawback in standardization of data need other intelligent computational tool such as Neural networks, SVM” to overcome the limitations so that the features produced by PCA can be used as essential vector data to create an initial model which can further be fine-tuned by SVM to propose an appropriate model for better prediction

Furthermore there are a number of benefits in this conditional structure. (1) Dimensionality reduction and feature extraction reduce the extra computational burden of SVM, (2) The range length in which the free and user parameters of SVM lie gets reduced. (3) And hence validation of tool will not be a time consuming process.

5. DATA SETS AND ALGORITHM PERFORMANCE:

The S&P 500 data is one of the most commonly used equity indices that include the index of leading indicators. Hence on the wide stock market movements and performance this focus is naturally a good choice. Another noteworthy characteristic of this data set is that it includes the data corresponding to time-duration of unforeseen financial crisis (2008-2010). The market returns that includes dividends, have continuously been compounded several data sets on daily basis and monthly basis have been taken up for study which have been taken from Amit Goyal's web page [17], that have been considered as bench mark data by a wide Research fraternity and the analysis have been carried out in each case.

The volatility represents the range between maximum and least prices in a day, known as a measure of market risk. As there are four advantages in applying transformation: Relative difference in percentage (RDP) to raw data as interpreted in [18], the raw data has been converted to a five day RDP.

At first, PCA as explained in section 2.(a) has been applied to the RDP data. A sample of 600 data has been processed and 11 clusters have formed out of this process. The

principal components are labelled and cluster centres have been identified using K-means algorithm. After this dimensionality reduction by PCA the crisp data has been fed to SVM (section 3-1) for construction of high dimensional feature space.

In this paper especially for S&P 500 data which includes the data corresponding to the period (2008-2010) also, three kernel functions namely linear, polynomial or radial basis function, Gaussian have been taken up for experimental study and prediction performance of the algorithm has been evaluated in terms of Normalized Mean Square Error (NMSE) and Normalized Mean Absolute Error (NMAE). It is observed that Gaussian kernel function is best suited for this hybrid, that minimises both NMSE and NMAE which are shown in table 5.1. The statistical metrics used in this study are as described in [15],

$$NMSE = \left[\frac{\sum_{i=1}^N (s_i^2 - g_i^2)}{\sum_{i=1}^N (g_{i-1}^2 - g_i^2)^2} \right]^{\frac{1}{2}}$$

$$NMAE = \frac{\sum_{i=1}^N |s_i^2 - g_i^2|}{\sum_{i=1}^N |g_{i-1}^2 - g_i^2|}$$

Where N is the sample size of test data, s^2 is the conditional variance predicted, g_i is the predicted return. Here NMSE is for all majority of data, whereas NMAE is robust against the outliers. Hence smaller NMSE and NMAE ensure the prediction accuracy.

This PCA based SVM algorithm has been implemented in mat lab and the program has been for various values of C that range from 25 to 700 for all the sample data mentioned in the table 5.1. Since PCA based feature extraction technique provides predominant features representing all data, principle components have been extracted as the input vector to SVM. This hybrid technique has been applied to run a few samples of SP500 data and the classification (risk and return features) has been shown in Figures (1) and (2)

Table 5.1

	SVM				SVM + PCA			
	NMSE		NMAE		NMSE		NMAE	
	C=100	C=650	C=100	C=650	C=100	C=650	C=100	C=650
S&P 500 data (200 data)	1.7362	1.4362	0.8362	0.5622	0.9862	0.6562	0.6322	0.4312
(300 data)	1.7382	1.4389	0.8491	0.5738	0.9899	0.6539	0.6438	0.4414
All data	1.7402	1.4391	0.8501	0.5750	0.9893	0.6543	0.6457	0.4423

Hence C has been chosen and trial and error basis and the optimum value of C have alone been furnished in the table 5.1. The main advantage of this hybrid is that C need not be found in any structured way. Selected choice of C themselves ensure optimum classification.

6. OBSERVATION AND EXPERIMENTAL RESULTS

In order to examine the stability of a proposed technique three year S&P 500 data has been divided into four groups (200 data, 300 data, all data). The capability of SVM classifier is depending on two main things: the choice of kernel and the parameter C. In this study Gaussian kernel is found to be apt. Though the power of other kernels has been examined, they are not tabulated. Smaller choice of C causes under fitting while larger choice of C cause over fitting. Though several statistics (such as arithmetic mean, Geometric mean and Root mean square value of various C ranging from 25 to 700) have been tried to find the optimum C, the value of C for which SVMPCA outperforms SVM alone is quoted in the table. The SVMPCA program is executed for these groups and the corresponding NMSE and NMAE has been calculated for various values of C and for the optimum values of C, the values of these metrics are tabulated [5.1]. Once again SVM technique alone is also employed and the results are furnished in the same table 5.1. From the table it is clear that PCA based SVM performs better than ordinary SVM in classifying the data into two classes namely return and risk and SVMPCA shows 96% classification accuracy while SVM alone shows 88% classification accuracy only.

7. CONCLUSIONS

In this paper we have designed a PCASVM combined computation model to forecast the deviations of stock market indices in terms of risk and return components. For classification of data by PCASVM we have used 1(return) and -1(risk) as targets. In the proposed model, principal component features were extracted by processing the covariance matrices and SVM have been used to process the PCA feature vectors for support vector generations. Volatility of this data has been classified in to risk and return groups. The superiority of this computational model driven by Gaussian kernel function has been explained with the optimized parameter C and the quality metrics NMSE & NMAE. Though S&P 500, DowJones, Nifty Index three data sets were considered for testing, S&P 500 data analysis have been shown in this paper. As a future study the structured model for obtaining the optimum parameter C can be studied. A complete computational model can also be developed if we associate appropriate geometry to the distribution of stock data. This in turn be useful in obtaining the optimum statistical and geometrical parameters also. This work is presently on

Fig-1

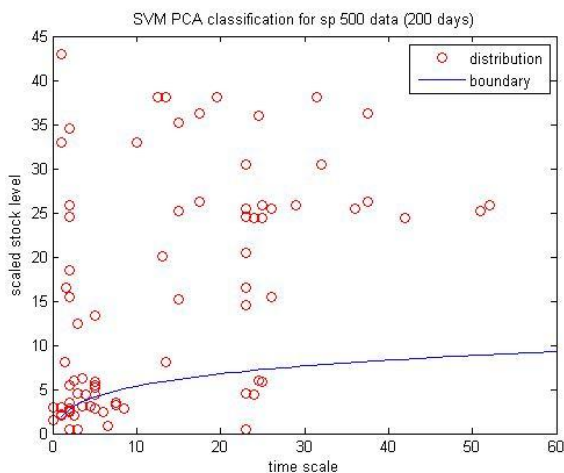
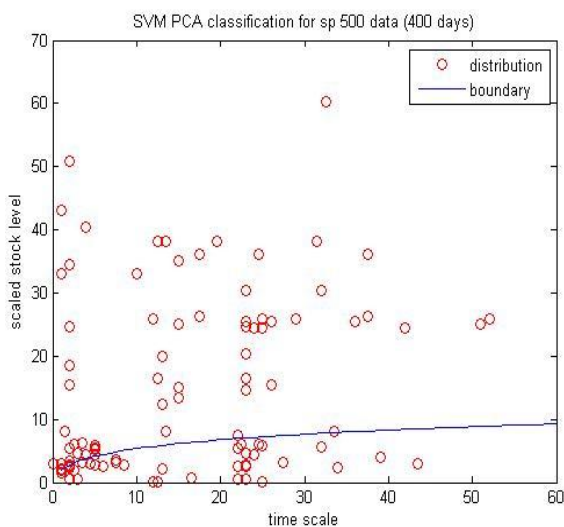


Fig-2



ACKNOWLEDGEMENTS

Two of the authors wish to thank Dr.K.Kannan, Dean, Humanities and Sciences, for his kind permission to audit the Machine Learning Techniques course handled by him, who in turn wishes to thank Dr.S.VaithyaSubramaniam, Dean, Planning and Development, SASTRA University who suggested Gaussian Kernel for SVMPCA implementation.

REFERENCES

- [1]. Option pricing using Machine Learning techniques, Amit Deoda, Dual degree dissertation, Indian Institute of Technology, Bombay, June 2011.
- [2]. Machine Learning for Financial Market Prediction PhD Thesis, Tristan Fletche, Computer Science University College London
- [3]. P. Amster, C.G. Averbuj, C.M. Marirani and D. Rial, A Black-Scholes option pricing model with transaction costs, Journal of Mathematical Analysis and Applications, 303(2005), 688-695.
- [4]. G. Barles and H.M. Soner, Option pricing with transaction costs and a nonlinear Black-Scholes equation, Finance and Stochastics, 2(1998), 369-397.
- [5]. F.E. Benth, K.H. Karlsen and K. Reikvam, Optimal portfolio selection with consumption and nonlinear integro-differential equations with gradient constraint: A viscosity solution approach, Finance and Stochastics, 5(2001), 275-303.
- [6]. T. Bjork and H. Hult, A note on Wick products and the fractional Black-Scholes model, Finance and Stochastics, 9(2005), 197-209.
- [7]. I. Bouhouev and V. Isakov, Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets, Inverse problems, 15(1999), R95-R116.
- [8]. C.K. Cho, T. Kim and Y.H. Kwon, Estimation of local volatilities in a generalized Black-Scholes model, Applied Mathematics and Computation, 162(2005), 1135-1149.
- [9]. R. Company, L. Jodar, G. Rubio and R.J. Villanueva, explicit solution of Black-Scholes option pricing mathematical models with an impulsive Payoff function, Mathematical and Computer Modelling, 45(2007), 80-92.
- [10]. Z.C. Deng, J.N. Yu and L. Yang, An inverse problem of determining the implied volatility in option pricing, Journal of Mathematical Analysis and Applications, 340(2008), 16-31.
- [11]. L. Jodar, P. Sevilla-Peris, J.C. Cortes and R. Sala, A new direct method for solving the Black-Scholes equation, Applied Mathematical Letters, 18(2005), 29-32.
- [12]. Q.L. Luo and W.C. Sheng, A unique solution to a semilinear Black-Scholes partial differential equation for valuing multi-assets of American options, Journal of Shanghai University, 11(2007), 344-350.
- [13]. Huseyin Ince, Theodore B. Trafalis, Kernel principal component analysis and support vector machines for stock price prediction, Neural Networks 2004, IEEE International joint conference, 2053-2058 (volume 3)
- [14]. Tuan Anh, Sander Magi (2009), National Cheng Kung University, "Principal component Analysis", Final paper in Financial Pricing.

- [15].Lijuan Cao and Francis E.H. Tay, Financial ForecastingusingSupport Vector Machines, Neural Computing &Applications(2001)10: 184-192.
- [16]. Francis E. H. Tay, LijuanCao, Applications of support vectormachines in financial time series forecasting, Omega, TheInternational journal of Management Science(2001), 309-317.
- [17]. <http://www.hec.unil.ch/agoyal/>.
- [18]. Charles Yuan, Predicting S&P 500 Returns using Support VectorMachines: Theory and Empirics (2011).