PCA-based Least Squares Support Vector Machines in Week-Ahead Load Forecasting

M. Afshin, *Member, IEEE*, and A. Sadeghian, *Member, IEEE*

***Abstract*— Week-Ahead load forecasting is essential in the planning activities of every electricity production and distribution company. This paper proposes the application of principal component analysis (PCA) to least squares support vector machines (LS-SVM) in a week-ahead load forecasting problem. New realistic features are added to better and more efficiently train the model. For instance, it was found that hours of daylight are influential in shaping the load profile. This is particularly important in case of cities that are situated in the northern hemisphere. To show the effectiveness, the introduced model is being trained and tested on the data of the historical load obtained from Ontario's Independent Electricity System Operator (IESO) for the Canadian metropolis, Toronto. Analysis of the experimental results proves that LS-SVM by feature extraction using PCA can achieve greater accuracy and faster speed than other models including LS-SVM without feature extraction and the popular feed forward back-propagation neural network (FFBP) model.**

***Index Terms*— Least squares support vector machines, load forecasting, principal component analysis**

I. Introduction

Load forecasting has always been a very important issue in economic and reliable power systems planning and operation [1], [2]. Specifically, short-term forecasting of daily electricity demand is crucial in unit commitment and maintenance, power interchange and task scheduling of both power generation and distribution facilities. Another area of application involves load flow studies, including contingency planning, load shedding, and load security strategies [3]. Economically, accuracy in load forecasting can allow utilities to operate at least cost which may contribute to millions of dollars in savings in major electric power companies.

Load forecasting for a week or less is categorized as short-term load prediction. When analyzing the reports and trends in the industry, it is evident that Week-Ahead daily peak load forecast plays an important role in the day-to-day activities of every power producing company.

A wide variety of forecasting techniques have been investigated and introduced so far. Traditionally, experts have applied conventional prediction methods for load forecasting including linear regression, stochastic processes, autoregressive moving average (ARMA) models and state space methods which are all based on time series or regression approaches [4], [5]. However, these techniques cannot properly express the complicated nonlinear relationship that exists between the load and the factors that influence it.

During the past decade, artificial neural networks have been emerged as a very successful approach to load forecasting [6]. However, this methodology has shortcomings when load patterns are not similar to those of weekdays, e.g., on weekends and public holidays. Also, the neural network requires many training samples and frequent retraining due to changing seasonal conditions. Therefore, the learning speed is comparatively slow [7], [8].

Recently, support vector machines (SVM) have attracted much attention in load forecasting [9]. They have been successfully employed to solve most nonlinear regression and time series problems. Typical advantages of SVMs include good generalization performance, the absence of local minima and accurate representation of solutions.

The theory of SVM is based on statistical learning theory [10]. Unlike most of the traditional methods which implement the empirical risk minimization principal, SVMs implement the structural risk minimization principal, which seeks to minimize an upper bound of the generalization error rather than minimizing the training error. Essentially, the SVM regressor maps the inputs into a higher dimensional feature space in which a linear regressor is constructed by minimizing an appropriate cost function. Using Mercer's theorem, the regressor is obtained by solving a finite dimensional quadratic programming problem in the dual space avoiding explicit knowledge of the high dimensional, mapping and using only the related kernel function. Therefore, the solution of SVM is always globally optimal. SVM is powerful for problems with nonlinearity and high dimension and it improves both training time and accuracy in comparison with other competitor forecasters [11], [12].

LS-SVM is a kind of SVM that uses equality constraints instead of the inequality constraints implemented in standard SVMs and a least squares error term to obtain a linear set of equations in the dual space [13].

The other aspect of load forecasting rather than the prediction algorithm is the feature selection and extraction. Principal Component Analysis (PCA) is used in order to

M. Afshin is with Computational Intelligence Initiative (CI2) lab, Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, M4S 2Y9 Canada (e-mail: mafshin@ryerson.ca).

A. Sadeghian is with Department of Computer Science, Ryerson University. He is the founder and director of the CI2 lab (e-mail: asadeghi@ryerson.ca).

1-4244-1291-9/07/$25.00 ©2007 IEEE

identify the most influential inputs in the context of the forecasting model [14]. It evaluates the input variables according to the projection of the largest eigenvector of the correlation matrix on the initial basis vector. This technique creates a new set of input variables which are orthogonal, so that they are uncorrelated with each other. The resulting orthogonal, principal components are ordered so that those with the largest variation take precedence, and components that contribute the least to the variation in the data set can be eliminated.

In this paper, the original inputs are first transformed into uncorrelated principal components using PCA. These new features are then used as inputs of LS-SVMs to solve time series load forecasting problems. The goal is to predict the daily peak load demand of the coming week. Evaluation of the algorithms is largely based on two error metrics, namely, mean absolute error (MAE) and mean absolute percentage error (MAPE). This paper relies on MAPE which is defined as:

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{L_i - \hat{L}_i}{L_i} \right|}{n} \times 100\,\%, \; n = 7 \qquad (1)$$

where L_i and $\hat{L}_i$ represent the actual and predicted peak daily loads and n is the number of the days of forecasting period.

The organization of this paper is as follows. In section II, we describe the process of feature selection and extraction. In addition, the analysis of load trend is also propounded. Sections III and IV will give a brief overview of principal component analysis followed by least squares support vector machines theory. In section V, the results of the proposed algorithms and other competitors are presented and compared. Finally, conclusions of this research are discussed in section VI.

II. Historical Data

A. Load Demand Analysis

An acceptable load forecast requires:

1) *Calendar information*
2) *Temperature profiles*
3) *Past load demands*

To begin with, it is clear that the load dynamics have multiple seasonal patterns, corresponding to a weekly and annual periodicity. Climatic conditions may vary depending on geographical location.

Fig. 1 shows that load consumption in summer is extensively higher than winter. In comparison with other Canadian cities, Toronto has a milder winter. During June to August, the load consumption reaches the maximum of the year.

Also, a load periodicity exists in every week. Load demand on weekends and holidays is usually lower than that of a weekday. Electricity demand on Saturday is a little higher than Sunday, because some businesses are open on Saturday. On the whole, it can be observed that broadly speaking, the peak load happens in the middle of the week, i.e. Wednesday, but there are always exceptions as discussed below.

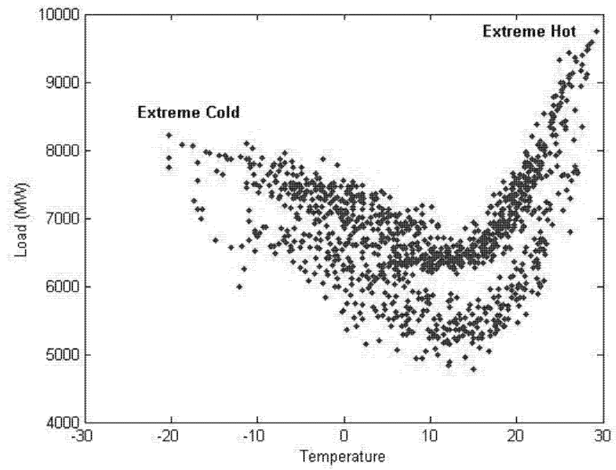

Fig. 1. Load vs. Temperature Profile, Toronto

1) Holiday Exceptions: As it was noted, the load demand usually decreases on holidays. Further study may show that load usage also depends on the (Christmas, New Year, etc.).

2) Extreme Temperatures: There is a complicated correlation between load demand and daily temperature [15]. As temperature increases above -20 °C, the load demand decreases until above 10 °C after which the trend reverses. Furthermore for temperatures above 30 °C (e.g., in Toronto) because of the Humidex effect, the load consumption will grow tremendously due to HVAC, cooling and ventilation Systems' usage.

B. Feature Selection

Historical load data was obtained from Ontario's IESO [16] for a Canadian major city, Toronto for the period May 2003 to July 2006 (hourly load).

Temperature information was acquired from Environment Canada [17] for the same period. The minimum, maximum and average daily temperatures, precipitation and the snow level on the ground are the most important parameters influencing the load usage in a typical day. Looking at other sources [18], it was observed that occasionally cloud cover is being considered in load forecasting, but this parameter alters instantaneously in real time. For this reason, it is not known to have a reliable impact on prediction accuracy. Moreover, wind speed and humidity are being used exceptionally, but in this paper, we have ignored them, due to rapid changes in their values in Toronto. Finally, official holidays are determined for the relevant years.

Another very important parameter to be taken into account in every load forecasting issue is the number of daylight hours. As illustrated in Fig. 2 the daylight fluctuates significantly from summer to winter (depending on the geographical location), and this has to be factored into data selection. To do so, the time between sunrise and sunset is readily determined as light hours during daytime.

To represent temperature, we use 3 numerical attributes: the minimum, the maximum and the average daily temperatures of the target day followed by temperatures from the previous two weeks. To represent load, we employ 25 real values for the past hourly loads and daily peak load demand. In order to forecast the maximum load of target day, we use the load information of the last two weeks only for the day of interest and the day before it. To represent calendar, we use two different features:

- An integer for each day of the week from Monday to Sunday (1 to 7),
- Three binary digits to distinguish a day type which includes weekdays, weekends and public holidays.

Different data encodings affect the selection of modeling schemes. The above selection of data was obtained through empirical results and in almost all cases it led to the least error (MAPE) in forecasts.

C. Feature Extraction

In developing any forecasting problem, the first step is feature selection (new features are selected from the original inputs) and then feature extraction (new features are transformed from the original inputs) [14].

In other words, all available information can be used as inputs, but irrelevant or strongly correlated features could unfavorably impact the generalization performance due to the dimensionality problem [19].

On this basis, there are changes to the original load data received from Ontario's IESO. A quick look at the historical load data shows that starting 14*th* August 2003 at 16:10, due to a huge blackout in southern Ontario and northeastern US, the load in transmission grid decreased tremendously and it remained low for the next day, until full recovery from the widespread power outage. Thus, to improve the accuracy and be more realistic, the hourly loads of those two days are substituted with the average of the hourly loads of the days before and after the Blackout occurred.

Eventually, the data considered for training, validation and testing have to be separately clarified.

III. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a well-known method for feature extraction [20]. By calculating the eigenvectors of the sample covariance matrix, PCA linearly transforms the original inputs into uncorrelated new features (called principal components), which are the orthogonal transformation of the original inputs based on the eigenvectors. The obtained principal components in PCA have second-order correlations with the original inputs.

Given a set of centered input vectors $x_t (t = 1,..., l$ and $\sum_{t=1}^{l} x_t = 0)$, each of which is of m dimension

$$x_t = (x_t(1), x_t(2),..., x_t(m))^T$$

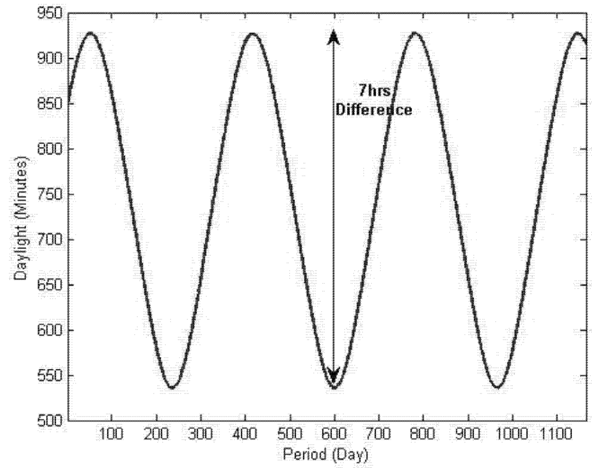

Fig. 2. Daylight Time Trend, Toronto

(usually $m < l$) , PCA linearly transforms each vector x_t , into a new one s_t , by

$$s_t = U^T x_t \tag{2}$$

where U is the $m \times m$ orthogonal matrix whose i-th column u_i is the i-th eigenvector of the sample covariance matrix:

$$C = \frac{1}{l}\sum_{t=1}^{l} x_t x_t^T$$

Basically, PCA firstly solves the eigenvalue problem (3):

$$\lambda_i u_i = C u_i, i = 1,..., l \tag{3}$$

where λ_i is one of the eigenvalues of C and u_i is the corresponding eigenvector. Based on the estimated u_i , the components of s_t , are then calculated as the orthogonal transformations of x_t .

$$s_t(i) = u_i^T x_t, i = 1,..., m \tag{4}$$

The new components are called principal components. By using only the first several eigenvectors sorted in descending order of the eigenvalues, the number of principal components in s_t can be reduced. This is the dimensional reduction characteristic of PCA. The size of the input vectors will be reduced by retaining only those components which contribute more than a specified fraction of the total variation in the data set. That means a new variable, minimum fraction variance component, should be defined. The comparison of different values of this parameter is explained in section V.

IV. Least Squares Support Vector Machines

In general, any ideal forecasting algorithm must satisfy the following criteria [15]:

1) Non stationarity of load series: When modeling the load series, it is important to consider the dynamic, nonlinear and complex input-output relationships that exist in the load trend.

2) Adaptiveness of the forecasting model: Previous researches have proved that the characteristics of load series between regular workdays and anomalous days (weekends and public holidays) are different.

3) Robustness of the forecasting model: A universal model is the top priority.

Kernel based estimation techniques, such as support vector machines (SVMs) especially least squares SVMs (LS-SVM) have proven to be powerful nonlinear classification and regression techniques and it has turned out that they can fulfill all of the above conditions perfectly.

With the help of a kernel function, LS-SVMs perform the linear regression in the transformed space by nonlinearly mapping the input data into a high dimensional feature space [13]. We will use x and z to denote the input vector and the feature space vector respectively, and $z = \phi(x)$.

Let the training set, x_i and y_i, consist of N data points, where x_i is the i-th input vector and y_i is the corresponding target value. The goal of LS-SVMs regression is to estimate a function that is as "close" as possible to the target values y_i for every x_i and at the same time, is as "flat" as possible for good generalization. The function f is represented using a linear function in the feature space:

$$y = f(x) = w \cdot \phi(x) + b \tag{5}$$

where $\phi(x)$ is a function that maps the input space into a higher dimensional feature space. Also, b denotes the bias, as in all SVM designs, we define the kernel function, where "." represents inner product in the space.

$$k(x, \hat{x}) = \phi(x) \cdot \hat{\phi}(x)$$

This will result in the optimization problem in primal weight space:

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^{N} e_k^2 \tag{6}$$

subject to

$$y_k = w^T \phi(x_k) + b + e_k, k = 1,..., N$$

where w is weight vector in primal weight space and e_k is the error variable. The cost function J includes a sum squared error and a regularization term. γ is a positive real constant that determines the penalties to estimation errors.

Next, the model in (6) can be computed in dual space instead of the primal space, which results in Lagrangian with Lagrange multipliers $\alpha_k \in R$, called support values.

$$L(w,b,e;\alpha) = J(w,e) - \sum_{k=1}^{N} \alpha_k \{ w^T \varphi(x_k) + b + e_k - y_k \} \tag{7}$$

The circumstances for optimality are as follows:

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^{N} \alpha_k \varphi(x_k) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^{N} \alpha_k = 0 \\ \dfrac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k, k = 1,..., N \\ \dfrac{\partial L}{\partial \alpha_k} \rightarrow w^T \varphi(x_k) + b + e_k - y_k = 0, k = 1,..., N \end{cases} \tag{8}$$

Based on [21], with the application of Mercer's theorem, $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$, with a positive definite kernel function K, we can eliminate w and e_i, obtaining

$$y_j = \sum_{i=1}^{N} \alpha_i K(x_i, x_j) + b + \frac{\alpha_j}{\gamma}$$

Building the kernel matrix $\Omega_{i,j} = K(x_i, x_j)$ and writing the equations in matrix notation, we express the final system in dual form,

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x_j) + b \tag{9}$$

The following choices for Kernel function $K(x_i, x_j)$ are available:

- Linear kernel: $K(x_i, x_j) = x_i^T x_j$
- Polynomial kernel with degree d and tuning parameter c: $K(x_i, x_j) = (x_i^T x_j + 1)^d$
- Radial basis function (RBF), where δ is a tuning parameter: $K(x_i, x_j) = \exp(-\frac{\| x_i - x_j \|^2}{2\delta^2})$

Usually the training of the LS-SVM model involves an optimal selection of the tuning (kernel) parameter δ and regularization parameter γ, which can be done using cross-validation or bayesian inference techniques [22].

V. Implementation and Evaluation

Here we mention the steps taken in the PCA-based LS-SVM to forecast the week-ahead peak load demand:

Step 1: Preprocess the historical load data sets (e.g., removing the abnormal samples of 13*th* and 14*th*, August 2003 Blackout) and then normalize all sample sets to zero mean and unit variance.

Step 2: Implement PCA on the input data and based on trial and error, determine the appropriate minimum fraction variance component i.e. number of features to be entered into the LS-SVM model.

Step 3: Build the target equation of (9) and use the test data sets to predict the next seven day's maximum load demands.

When training a LS-SVM model, there are some parameters to be selected which influence the performance of the model considerably:

- Regularization parameter (γ)
- Kernel bandwidth parameter (δ)
- The kernel function $k(x, \hat{x})$, RBF is used in our LS-SVM model.
- The size of training data sets, i.e. how many previous days are included for one training data.

Step 4: Apply the Bayesian evidence framework [22] to the LS-SVM regression algorithm and keep estimating the optimal regularization parameter (γ) and kernel parameter (δ) until enough accuracy is reached. Optimizing the PCA would be a hard task, and we use trial and error to get the most accurate forecasts. For this matter, part of the data set is dedicated to validation purposes (Usually 25% of the data).

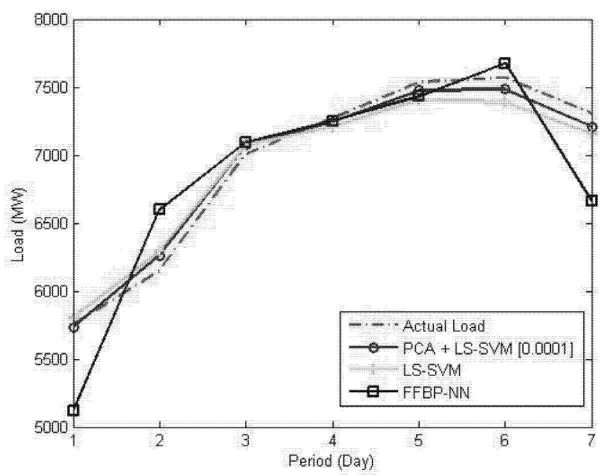

Fig. 3. Actual vs. Predicted Loads

TABLE I
FORECASTING ERRORS FOR A TYPICAL WEEK

Estimation Technique	No. features	MAPE (%)	MAE (MW)
Single LS-SVM	85	1.5755	110.3582
FFBP Neural Network	85	2.9231	203.1979
PCA + LS-SVM [0.0001]	34	1.1454	80.0693
PCA + LS-SVM [0.00001]	58	0.8082	56.1873

The comparison of different forecasting models is shown in Table. 1. For instance, the 0.01 for PCA + LS-SVM means those components that contribute more than one percent to the variance in the data set are kept and the remaining are eliminated. LS-SVM with PCA and minimum fraction variance of 0.01% obtained better results than LS-SVM without feature extraction and feed forward back propagation neural networks. Using smaller minimum fraction variance will improve MAPE and the model performance both in time and accuracy.

The corresponding program for this forecasting model was developed from LS-SVMlab (MATLAB/C toolbox for Least Squares Support Vector Machines) [23], with major changes to fulfill the application.

VI. CONCLUSION

In this paper, a PCA-based least squares support vector machine was presented and its performance was evaluated through a simulation study. PCA was used to reduce the input variable dimension. A wide range of the minimum fraction variance was tested on the model and the results were satisfactory. The LS-SVM technique shows satisfactory performance, such as powerful regression ability, acceptable predicting accuracy and perfect foundations in theory. Also, a crucial and effective feature was added to the data collection namely daylight time. Depending on the region, this feature could vary tremendously and must be taken into account.

LS-SVM by feature extraction using PCA outperformed other techniques in week-ahead load forecast including LSSVM without feature extraction and the well-known feed forward back propagation neural networks. It obtained better accuracy, faster speed and superb generalization.

VII. ACKNOWLEDGMENT

The authors would like to thank Mr. Douglas Barnard, Customer Communications, Ontario's IESO for valuable guidance and providing information during this research.

VIII. REFERENCES

[1] G. Gross and F. Galiana, "Short term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573, 1987.
[2] R. Sadownik and E. Barbosa, "Short-term forecasting of industrial electricity consumption in brazil," *Int. J. Forecast.*, vol. 18, pp. 215–224, 1999.
[3] K. Kim, H. S. Youn, and Y. C. Kang, "Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 559–565, 2000.
[4] S. Tzafestas and E. Tzafestas, "Computational intelligence techniques for short-term electric load forecasting," *Journal of Intelligent and Robotic Systems*, vol. 31, p. 7–68, 2001.
[5] A. Lotufo and C. Minussi, "Electric power systems load forecasting: A survey," *Int. J. Forecast.*, vol. 18, pp. 215–224, 1999.
[6] H. S. Hippert, C. E. Pedreira, and R. Castro, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, 2001.
[7] D. Srinivasan, A. Liew, and C. Chang, "Forecasting daily load curves using a hybrid fuzzy-neural approach," *IEE Proc.*, vol. 141, no. 6, 1994.
[8] D. Srinivasan, S. S. Tan, C. S. Chang, and E. K. Chan, "Practical implementation of a hybrid fuzzy neural network for one-day-a head load forecasting," *IEE Proceedings*, vol. 145, no. 6, 1998.
[9] B. Chen, M. Chang, and C. Lin, "Load forecasting using support vector machines: A study on eunite competition 2001," *IEEE Trans. PowerSystem*, vol. 19, no. 4, pp. 1821–1830, 2004.
[10] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel Based Learning Methods*. Cambridge University Press, 2000.
[12] B. Scholkopf, C. Burges, and A. Smola, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1999.
[13] Y. Li, T. Fang, and E. Yu, "Short-term electrical load forecasting using least square support vector machines," *Proceedings. PowerCon*, vol. 1, pp. 230–233, 2002.
[14] J. Cao, K. Chua, and L. Guan, "Combining kpca with support vector machine for time series forecasting," *CIFEr Hong Kong*, 2003.
[15] S. Fan and L. Chen, "Short-term load forecasting based on an adaptive hybrid method," *IEEE Trans. PowerSystem*, vol. 21, no. 1, pp. 392–401, 2006.

[16] Independent electricity system operator. [Online]. Available: http://www.ieso.ca
[17] Environment canada. [Online]. Available: http://www.ec.gc.ca/
[18] “Methodology to perform long term assessments,” Public, IESO, 2006.
[19] F. Tay and L. Cao, “Saliency analysis of support vector machines for feature selection,” *Neural Network World*, vol. 2, no. 1, pp. 153–166, 2001.
[20] L. Cao and W. Chong, “Feature extraction in support vector machine: A comparison of pca, kpca and ica,” *Proceedings of the 9th International Conference on Neural Information Processing*, vol. 2, pp. 1001–1005, 2002.
[21] V. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
[22] X. WEN, Y. ZHANG, W. YAN, and X. XU, “Nonlinear decoupling controller design based on least squares support vector regression,” *Journal of Zhejiang University SCIENCE*, 2005.
[23] J. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle. (2002) Ls-svmlab: a matlab/c toolbox for least squares support vector machines. [Online]. Available: http://www.esat.kuleuven.ac.be/sista/lssvmlab