Support Vector Machines for Predicting The Electrical Faults

Dr. Tarik Rashid
Engineering College, University of Salahaddin/Erbil
Email: tarikrashid4@gmail.com
Salar J. Abdulhameed
Engineering College, University of Salahaddin/Erbil
salar_atroshi@yahoo.com

Received on: 31/10/2013 & Accepted on: 6/4/2014

Abstract

Support vector machines (SVMs) are a non-probabilistic binary linear classifier in machine learning techniques and are supervised learning algorithms that classify, predict, recognise and analyse patterns. This technique was developed in early 1990s.Training algorithms of support vector machines help build a model that assigns new examples into one class or the other when a set of training examples is recycled in the training process. This feature in SVM has attracted many of researchers to develop SVM methods and their applications. In this paper work support vector machines are used to tackle electrical faults in single phase circuits. Support vectors machines are evaluated against Simple Linear Regression techniques. Support vector machines outperformed Simple Linear Regression techniques.

Keywords: Support Vector Machines, A simple Linear Regression technique, Electrical Fault Perdition.

ماكينات دعم المتجه للتنبوء بالاعطال الكهربائية

الخلاصة:

تعتبر ماكينات دعم المتجه من المصنفات الثنائية غير الاحتمالية في تعليم الماكنة وتعتبر من انواع الخوارزميات المعتمدة على المشرف والتي تصنف وتتنبأ وتميز وتحلل الاصناف. طورت هذه التقنية في بدايات عام 1990. خوارزميات التدريب لهذه الماكينات تساعد في بناء نموذج يخصص أمثلة جديدة لصنف واحد أو أكثر عندما تتم أعادة الامثلة في مرحلة التدريب. هذه الخاصية تستقطب عدة باحثين لتطوير طرق ماكينات دعم المتجه وتطبيقاتها. في هذا البحث تم استخدام ماكينات دعم المتجه لتشخيص الاعطال الكهربائية في دوائر الطور الواحد. بعد تقييم أداء ماكينات دعم المتجه بالمقارنة مع تقنية الانحدار الخطي البسيط، تفوقت ماكينات دعم المتجه على تقنية الانحدار الخطي البسيط.

https://doi.org/10.30684/etj.32.8A4
2412-0758/University of Technology-Iraq, Baghdad, Iraq

Support Vector Machines

Support vector machine models are first coined by Boser, Guyon and Vapnik in 1992 [1]. They are learning algorithm which is used to solve a prediction task through constructing a model [2]. This type of algorithm has the power to anticipate unseen data in time series problems. SVM techniques can be applied to problems such as classification, recognition and regression analysis. Support vector machines received a lot of attention from many researchers and computational scientists once they revealed a better accuracy than standard neural networks when pixel maps used as inputs in handwriting recognition application [3, 4]. The key concept of this technique is to transform input variables into high dimensional feature space. Another feature of support vector machines is the regression surface which can be specified by a subset of points, these pointes are called support vectors which are considered to be vital and all other points which are not part of detecting the surface of regression are not important. Vapnik recognizes an ϵ-insensitive zone in the error loss function [1, 2]. Points that placed within the zone are believed to be correct, whereas other points that are outside the zone are deemed to be incorrect and provided to the error loss function. It is worth to mention that these incorrect points become a support vector in classification tasks, whereas these incorrect points which lie within ϵ-insensitive are not imperative in regression tasks as shown in Figure1.

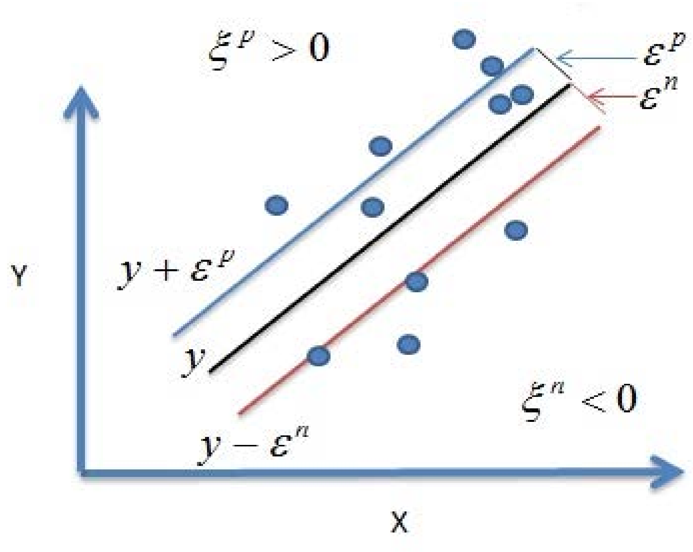

Figure (1) shows insentive error ε in regression tasks.

Mathematics for SVM Regression

A generalized version of SVMs method can be used for regression tasks. SVM attempts to craft a linear purpose such that the training data points lie inside a distance ϵ insensitive as can be seen in Figure 1.

Let our pilot data be (x_i, y_i) *where* $i = 1,...L$

$X \in \Re, y_i \in \Re,$ Where $\Re$, is a two dimension space. The aim is to predict a real value for the output y, and this can be represented in equation (1).

$$y_i = Wx_i + b..................(1)$$

W is the weight vector and the b is a scalar value.

In regression tasks, Support Vector Machine is effectively able to use a better penalty function, in other words, when the predicted value y_i is less than a distance (ε), away from the target, d_i regression will not assign a penalty, and this means $|d_i - y_i| < \varepsilon$ [5].

The regions guaranteed by $y_i + \varepsilon^p$ and $y_i + \varepsilon^n$, are called an insensitive tube; p and n stand for positive and negative respectively. The other amendmentto the penalty function is those output variables which are outside the tube are given one of two slack variable penalties ξ^p, which located above the tube and ξ^n, which located below the tube, where $\xi^p > 0$; $\xi^n > 0$ $\xi^n \forall i$ .These can be realized in equations (2) and (3) [5, 6].

$$d_i \leq y_i + \varepsilon + \xi^p ....(2)$$

And

$$d_i \geq y_i - \varepsilon - \xi^n ....(3)$$

In regression type of Support Vector, the function of error can be conveyed as in equation (4):

$$\frac{1}{2} \| W \|^2 + C \sum_{i=1}^{L} (\xi^p_{i_i} + \xi^n_{i_i})........(4)$$

In order to minimize the above expression, it has to be subject to the constraints where $\xi^p > 0$; $\xi^n < 0$ $\xi^n \forall i$ and both equations (2) and (3).

Using the Lagrange multipliers so that to minimize above

$\alpha^p \geq 0;\ \alpha^n \geq 0;\ \mu^p_{pi} \geq 0;\ \mu^n_i \geq 0\ \forall i$

The Lagrange can be represented as follows:

$$L = \frac{1}{2} \| W \|^2 + C \sum_{i=1}^{L} (\xi^p_{i_i} + \xi^n_{i_i}) - \sum_{i=1}^{L} (\mu^p \xi^p_i + \mu^n \xi^n_i) -$$
$$\sum_{i=1}^{L} \alpha^p (\varepsilon + \xi^p_{i_i} + y_i - d_i) - \sum_{i=1}^{L} \alpha^n (\varepsilon + \xi^n_i - y_i + d_i)........(5)$$

By using derivatives e.g. the derivatives are set to 0 and the substitution for y_i is done. The differentiation must be taken place with respect to $\xi^p_{i_i}$ *and* $\xi^n_{i_i}$ *and* to *w, b*:-

$$\frac{\partial L_p}{\partial w} = 0 \quad then \quad w = \sum_{i=1}^{L} (\alpha_i^p + \alpha_i^n) x_i ......(6)$$

$$\frac{\partial L_p}{\partial b} = 0 \quad then \quad C = \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) = 0.......(7)$$

and

$$\frac{\partial L_p}{\partial \varepsilon_i^p} = 0 \quad then \quad C = (\mu_i^p + \alpha_i^p)......(8)$$

And

$$\frac{\partial L_p}{\partial \varepsilon_i^n} = 0 \quad then \quad C = (\mu_i^n + \alpha_i^n).....(9)$$

Both equations (6) and (7) are substituted, maximizing *Lp* with respect to α^p and α^n, in other words this can be explained as follows:-

$\alpha^p \geq 0; \ \alpha^n \geq 0 \ \forall i$

Where

$$L_p = \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) d_i - \varepsilon \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) - \frac{1}{2}$$

$$\sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n)(\alpha_i^p - \alpha_i^n) x_i x_j . \quad .....(10)$$

By means of $\mu_{pi}^p \geq 0$ and $\mu_i^n \geq 0 \ \forall i$ and through both equations (8) and (9), in other words, $\alpha_i^p \leq$ C; $\alpha_i^n \leq$ C $\forall i$

Then, this can be figured out:-

$$\max_{\alpha_i^p, \alpha_i^n} \begin{bmatrix} \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) d_i - \varepsilon \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) - \frac{1}{2} \\ \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n)(\alpha_i^p - \alpha_i^n) x_i x_j \end{bmatrix} . \quad ....(11)$$

In a way that:-

$\sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) = 0 \quad \forall i$ and $0 \leq \alpha_i^p \leq$ C , $0 \leq \alpha_i^n \leq$ C, If equation (6) is switched into equation (1),and y^{test} as a new prediction value is going to be used, as shown in equation (12) [5, 8].

$$y^{test} = \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) x_i x^{test} + b ......(12)$$

And now a set *S* of Support Vectors x_s can be created via the indices *i* where

$\xi_i^p = 0 \quad or \quad \xi_i^- = 0 \, and \quad 0 < \alpha < C$

b can be attained as in equation (13)

$$b = d_s - \varepsilon - \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) x_m x_s .....(13)$$

To take the average of equation (13) over all the indices *i* in *S*, as shown in equation (14):-

$$b = \frac{1}{N_s} \sum_{s \in S} \left[d_s - \varepsilon - \sum_{i=1}^{L} (\alpha_i^p - \alpha_i^n) x_m x_s \right] .....(14)$$

Support Vector Machines for nonlinearity

In this section, SVM is used to separate data linearly. This can be carried out by just forming a matrix *H* from the dot product of the input variables in the application [5, 9], and this can be expressed as:-

$$H_{ij} = y_i y_j K(x_i x_j) = x_i \times x_j = x_i^T \times x_j ....(15)$$

The term function $K(x_i x_j)$ in the above, is called a Kernel Function, the $K(x_i x_j) = x_i^T \times x_j$ is called a linear kernel function. Equation (15) is a set of kernel functions which is collected of alternatives and these alternatives stay all established on scheming the two vectors inner products [5, 10]. Obviously the trick of kernel function is beneficial and this is due to the fact that there are a lot of classification and regression problems that are not linearly separable or repressible in the space, this perhaps can be in high dimensional feature space when an appropriate mapping is given as in the following expression:-

$x \mapsto \phi(x)$

It is noticeable that occasionally a linear classifier is not a difficult task. SVM solution is aiming at mapping data into a richer feature space including nonlinear features, then building a hyper plane in that space [5] (see Figure 2), so all other equations are the same, the data can be managed with:-

$x \mapsto \phi(x)$ and then learn the map from $\phi(x)$ to y *or* $f(x)$:

$$f(x) = w \times \phi(x) + b ....(16)$$

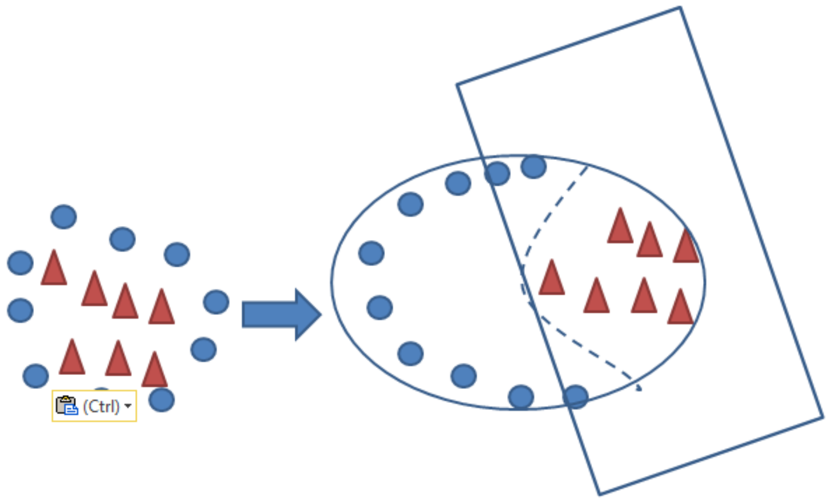

Figure (2) Mapping data into a richer feature space including nonlinear features.

In the regression task, it can be seen how large the insensitive loss region pans out via choosing appropriate values for the parameters C and ε and also it is required to see how meaningfully misclassifications would be dealt with:-
The key aim here is to find the below α_i^p *and* α_i^n in equation (11), this is maximized, subject to the constraints:-

$$\sum_{i=1}^{L}(\alpha_i^p - \alpha_i^n) = 0 \qquad \forall i$$

and $0 \le \alpha_i^p \le$ C , $0 \le \alpha_i^n \le$ C. This is done using quadratic problem. Then the following can be computed:-

$w = \sum_{i=1}^{L}((\alpha_i^p - \alpha_i^n)\phi(x_i)$. Determine the set of Support Vectors S by finding the indices such that

$\alpha_i < 0$
$\le$C and $\xi_i = 0 \forall i$, using equation (14).

Each new point x^{test} is determined by evaluating $y^{test} = w \times x^{test} + b$

Simple linear Regression

Regression is considered to be one of the most common statistical techniques that are usually used to build up models to find the relationship of output variable and other factors that are affecting the output variable [11]. The main disadvantage of these techniques is that the relationship between output variable and factors that are affecting the output is not stationary, but depends on temporal–spatial variations which cannot be addressed by these techniques [12].

Again let us have the same our pilot data as before (x_i, y_i) *where* $i = 1,...L$, the interested outputs are the values of estimated regression slope and estimated regression intercept. These will allow us to write the fitted regression line [13]. Recall that the equation of a line is:

$$y = a + bx...........(17)$$

Where *a*, is the intercept with *y* axis and *b*, is the slope.

The mean of *x* and *y* is expressed in the following equations:-

$$\overline{x} = \frac{\sum_{i=1}^{L} x_i}{L}...........(18)$$

$$\overline{y} = \frac{\sum_{i=1}^{L} y_i}{L}...........(19)$$

The square regression for x can be expressed by Sxx as in equation (20):-

$$Sxx = \sum_{i=1}^{L} x_i^2 - \frac{\left(\sum_{i=1}^{L} x_i\right)^2}{L}..........(20)$$

The square regression for y can be expressed by Syy as in equation (21):-

$$Syy = \sum_{i=1}^{L} y_i^2 - \frac{\left(\sum_{i=1}^{L} y_i\right)^2}{L}..........(21)$$

Where the variance is expressed as:

$$Sxy = \sum_{i=1}^{L} x_i y_i - \frac{\left(\sum_{i=1}^{L} x_i\right) \times \left(\sum_{i=1}^{L} y_i\right)}{L}..........(22)$$

The slope *b* can be expressed as in equation (23):-

$$b = \frac{Sxy}{Sxx}...........(23)$$

The intercept *a* can be expressed as in equation (24):-

$$a = \overline{y} - b\overline{x}...........(24)$$

SVMs and SLR Forecasting Models

SVM is introduced as a forecasting technique to tackle the problem of the prediction of electrical faults on a data set. The size of history data was relatively small, 278 samples, 58 samples are chosen for training and 20 samples are chosen for testing. Since it was a regression task, thus, the inputs and outputs are vectors of attributes of real values. The number of inputs was 11 variables and one output variable. The inputs namely are; the voltage, mega voltage ampere/short circuit capacity, sub transit current, transit peak current, and three different types of ampere current. The output was the length of the cable line.

Experimental Results

The epsilon parameter, which is called loss function regression for all kernels in general, was set to a default value 0.1. SVMs were trained with the training set and tested with the testing set. A validation set is chosen from the dataset in such a way that the number of input patterns of the validation set is equal to the number of the input patterns of the testing set. A linear search is performed to obtain the values of sigma *(s)* in such a way that these *(s)* values correspond to the minimum error on the validation set. These *(s)* values are used on the training set to retrain the SVMs and then used on the testing set. Figure 3 and 4 show the results of SVM on training set and testing set respectively.

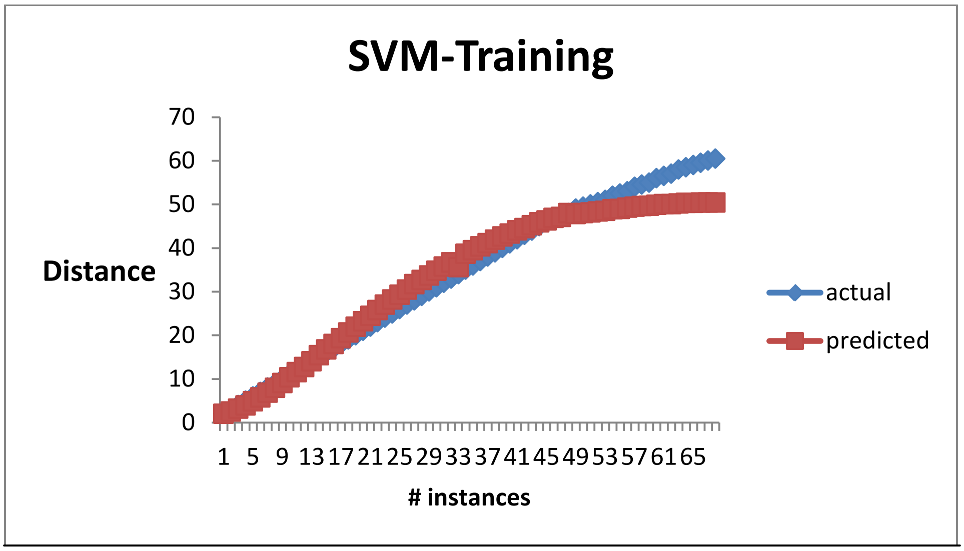

Figure (3) shows the results of SVM on training data.

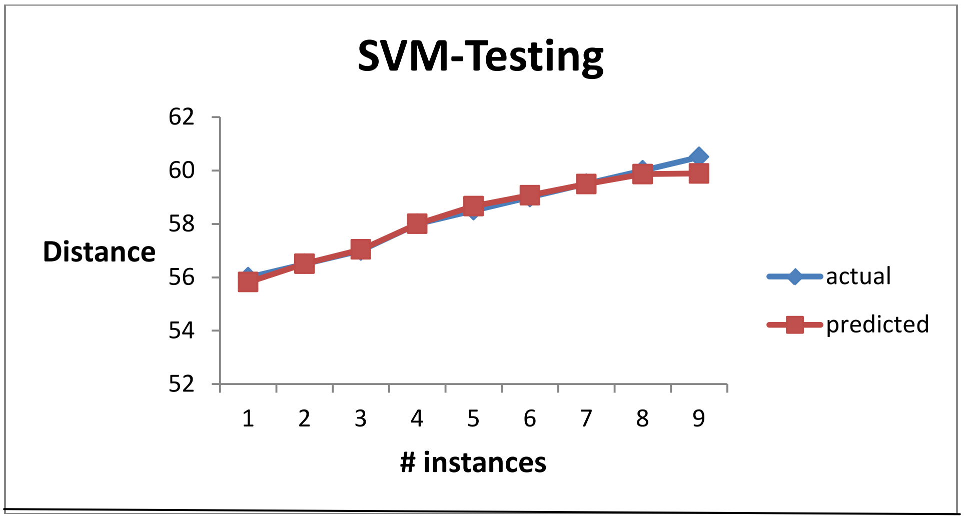

Figure (4) shows the results of SVM on testing data.

The choice of the kernel is the biggest drawback of the support vector approach [14, 15, 16]. The size and the speed are considered to be the drawbacks both in training and testing stages, the data must be discrete is also a problem. The selections of parameters such as sigma in Guassian function and epsilon insensitive loss function are not easy [2]. Feasibly the most thoughtful problem with SVMs is the high algorithmic complexity [17, 18]. Figure 5 and 6 show results of SLR on both training and testing sets.

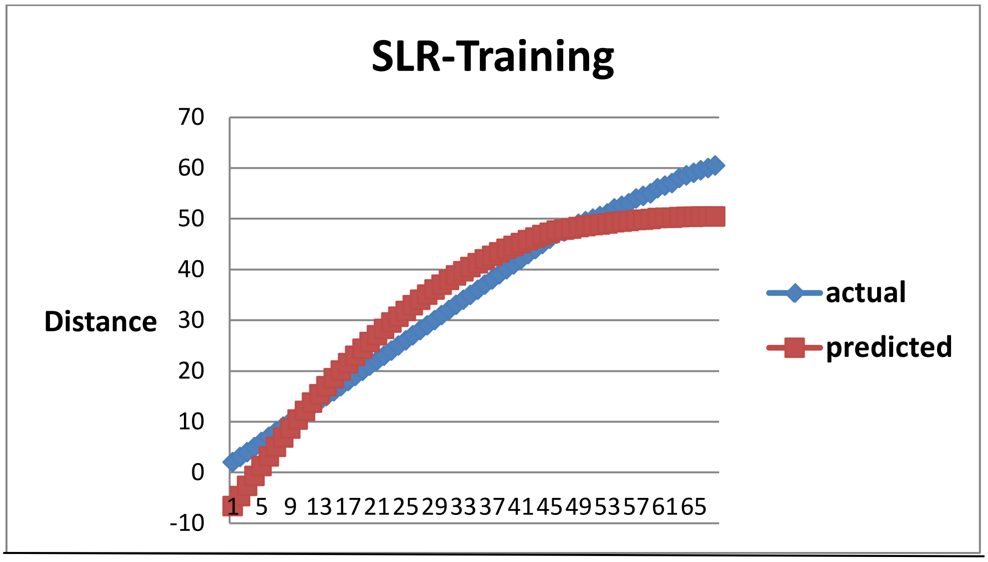

Figure (5) shows the results of SLR on training data.

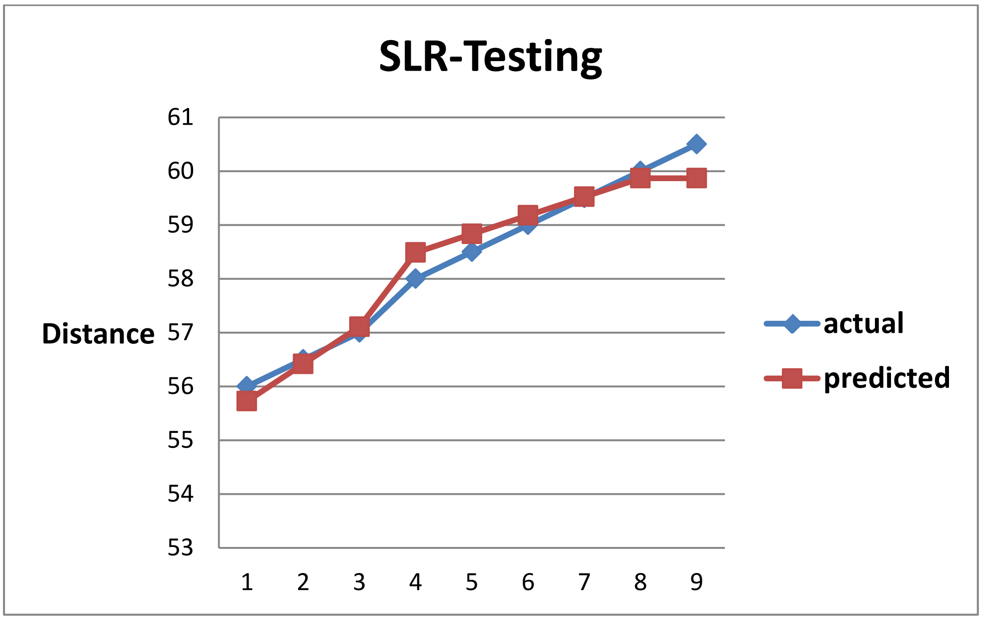

Figure (6) shows the results of SLR on testing data.

Figure 7, shows different error performance measurements, error performance measurements for both techniques SVM and SLR on training data show that SVM machine outperformed SLR; this means that SVM is preferred over SLR. However, in the correlation coefficient performance measurement, both SVM and SLR scored almost exactly the same rate, this means that both predicated value and the desired value are closely related in both techniques.

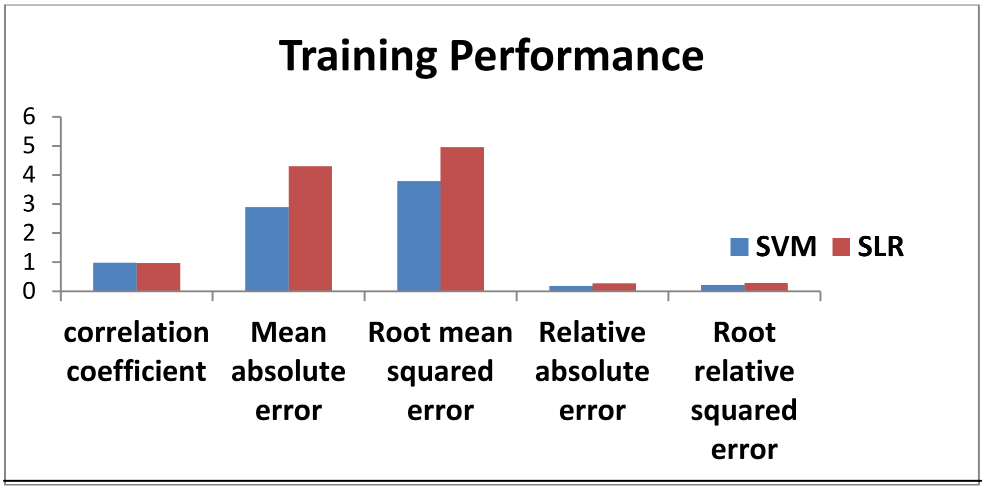

Figure (7) shows the performance results on SVM and SLR on training samples.

Figure 8, shows different error performance measurements, error performance measurements for both techniques SVM and SLR on testing data show that SVM machine outperformed SLR; this means that SVM is preferred over SLR. It is also worth mentioning that the correlation coefficient performance measurement in SVM is seen higher than SLR, this means that, SVM responded better than SLR. This would prove the predicated value and desired value on testing data are very close in SVM than in the case of SLR.

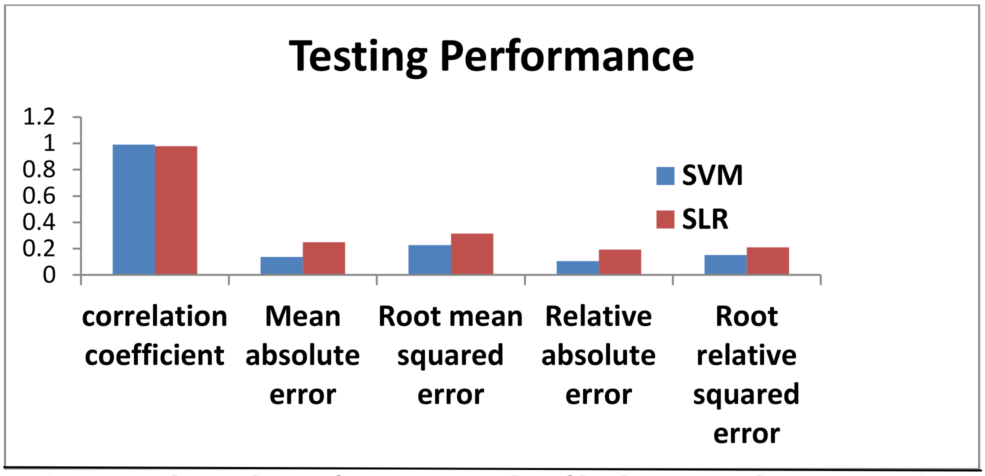

Figure (8) shows the performance results of both SVM and SLR on testing samples.

Conclusions

This paper presented Support Vector Machines for a regression purpose. SVM is introduced and guided mathematically and explained in detail for predicting electrical faults. The paper also introduced Simple Linear Regression in detail. SVMs are also compared to Simple linear regressions. The obtained results from SVM validate this technique and compare favorably with the result of simple regression methods.

Bibliography

[1] E. B. Boser, M. G. Isabelle, and V. N. Vapnik, “A training algorithm for optimal margin classifiers”, In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, Pennsylvania, 1992, ACM Press.

[2] Christopher J.C. Burges, “A tutorial on support vector machines for pattern recognition”, Data Mining and Knowledge Discovery, 2(2):121–167, 1998.

[3] W. Hua Chen and J. Ying Shih,” Comparison of support vector machines and back propagation neural networks in forecasting”, the six major Asian stock markets, Int. J. Electronic Finance, 1(1), 2006.

[4] C. Cortes and V. Vapnik ,” Support-vector networks”, Machine Learning, 20(3):273–297, 1995.

[5] Tristan Fletcher, “Support Vector Machines Explained” www.cs.ucl.ac.uk/sta_/T.Fletcher/, UCL, March 1, 2009

[6] N.Cristianini, J.Shawe-Taylor,“An introduction to support Vector Machines:and other kernel-based learning methods”,Cambridge University Press,New York,NY,USA,2000.

[7] J. Shawe-Taylor, N. Cristianini, “Kernel Methods for Pattern Analysis”, Cambridge University Press, New York, NY, USA, 2004.

[8] C. J. C. Burges, Data Mining and Knowledge Discovery 2, 121, 1998.

[9] C. M. Bishop, “Pattern Recognition and Machine Learning (Information Science and Statistics)”, Springer, 2006.

[10] Jason Weston, Support Vector Machine (and Statistical Learning Theory) Tutorial, NEC Labs America 4, Independence Way, Princeton, USA. jasonw@nec-labs.com.

[11] Tarik Rashid and M-T. Kechadi, “Study of artificial neural networks for daily peak electricity load forecasting”, The 2nd International Conference on Information Technology (ICIT 2005) Amman, Jordan, 2005.

[12] Tarik Rashid, M-T.Kechadi, and Brain Gleeson,”Neural network based methods for electric load forecasting”,The 3rd International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2006) in Orlando, USA,2006.

[13] Simple Linear regression Simple, Linear Regression Documents Prepared For Use In Course B01.1305, New York University, Stern School Of Business, nd.

[14] Rasmussen C.E., Williams C.K.I., “Gaussian Processes for Machine Learning”, MIT Press, 2006.

[15] Sch¨olkopf B., Smola A., “Learning with Kernels”, MIT Press, 2002.

[16] Shawe-Taylor J., Cristianini N., “Kernel Methods for Pattern Analysis”, Cambridge University Press, 2004.

[17] Suykens J.A.K., Van Gestel T., De Brabanter J., De Moor B., VandewalleJ., “Least Squares Support Vector Machines”, World Scientific, Singapore, 2002.

[18] Suykens J.A.K., Horvath G., Basu S., Micchelli C., Vandewal le J. (Eds.), “Advances in Learning Theory: Methods, Models and Applications”, vol. 190 NATO-ASI Series III: Computer and Systems Sciences, IOS Press, 2003.