

Mihaela GHEORGHE

Faculty of Economic Cybernetics, Statistics and Informatics
Bucharest University of Economic Studies, Romania

A SUPPORT VECTOR MACHINE APPROACH FOR DEVELOPING TELEMEDICINE SOLUTIONS: MEDICAL DIAGNOSIS

Case study

Keywords

Support vector machine

Medical diagnosis

Classification

Artificial neural network

Kernel function

JEL Classification

C45, I10

Abstract

Support vector machine represents an important tool for artificial neural networks techniques including classification and prediction. It offers a solution for a wide range of different issues in which cases the traditional optimization algorithms and methods cannot be applied directly due to different constraints, including memory restrictions, hidden relationships between variables, very high volume of computations that needs to be handled. One of these issues relates to medical diagnosis, a subset of the medical field. In this paper, the SVM learning algorithm is tested on a diabetes dataset and the results obtained for training with different kernel functions are presented and analyzed in order to determine a good approach from a telemedicine perspective.

1. Introduction

Neural networks represent one of the recent artificial intelligence technologies that have gained an important significance due to their processing methods used to solve many real world issues where it is difficult to elaborate and define a conventional algorithm.

Medical diagnosis has different difficulties related to the large amount of medical data and also to a large number of similarities between symptoms associated with different diseases. Neural networks can help in this situation by offering data mining techniques and therefore providing a tremendous opportunity to collect all the medical data needed for an investigation, find and interpret hidden and complex connections between symptoms.

Therefore, neural networks are becoming one of the most important classifier technologies that can be used for developing informatics solutions for solving different tasks, especially related to medical diagnosis process. In this study, the results of the investigation taken in what concerns some of the potential benefits of using a support vector machine-based approach for medical diagnosis of patients that may present diabetes disease, are presented.

2. Support vector machine: mathematical framework

Support vector machine (SVM) represents an example of a two-class linear classifier which can be generalized to an n-class classifier. For exemplification, within this paper, the SVM algorithm will be presented from a two-class perspective. Therefore, the investigation starts from a set of instances which can be labeled with one of the values from the domain class. Assuming the labels are +1, for the positive class and -1, for the negative one and let X be a vector with x_i , as components, y_i the class associated with x_i , object from X is named a pattern.

Based on the fact that SVM is a kernel method and therefore, an algorithm which depends on the data only through their scalar product, it has the ability to generate non-linear decision boundaries [2]. SVM is mainly based on a linear discriminant function represented by (1).

$$f(x) = w^T x + b \quad (1)$$

where:

w - represents the weight vector
 x - represents the input vector
 b - represents a bias or a threshold

If the threshold has a value equal to zero ($b=0$), the scalar product of the vectors w and x will generate a set of points x which are all perpendicular on the weight vector and form a hyperplane. If this is not applicable ($b \neq 0$) this will translate the hyperplane

away from the origin. The decision boundary of SVM represents a new boundary region, which divides the space into two individual classes. Each instance that needs to be classified will be added within one of the two regions classified as positive and negative based on the sign of the discriminant function illustrated by (1). An example of a support vector machine classifier is represented in figure 1 as shown in [6].

By replacing the input vector x with a non-linear function $\phi(x)$ the equation illustrated in (1) becomes (2) as follows:

$$f(x) = w^T \phi(x) + b \quad (2)$$

Furthermore, this substitution of a linear function with a kernel one generates those decision boundaries mentioned previously. In order to analyze the maximum margin of these, the following equations (3) are used:

$$\begin{cases} w^T x + b \geq 1 \\ w^T x + b \leq -1 \end{cases} \quad (3)$$

These equations can be combined within (4):

$$y_i(w^T x + b) - 1 \geq 0, \quad i \quad (4)$$

Based on (4) a separating hyperplane with a large margin is defined by (5).

$$M = \frac{w}{\|w\|} (x_2 - x_1) = \frac{2}{\|w\|} \quad (5)$$

where:

x_1, x_2 - represents two support vectors perpendicular on w .

For solving a classification problem, a maximization process of $\lim_{\|w\|} \frac{1}{\|w\|}$ is needed and this can be obtained with the Lagranges multiplier as described in (6).

$$\begin{aligned} L_P &= \frac{1}{2} \|w\|^2 - \alpha [y_i(x_i \cdot w + b) - 1 \quad \forall i] \equiv \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^L \alpha_i [y_i(x_i \cdot w + b) - 1] \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^L \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^L \alpha_i \quad (6) \end{aligned}$$

This will be solved with the partially differentiate equation represented by (7).

$$\begin{cases} \frac{\partial L_P}{\partial w} = 0 & w = \sum_{i=1}^L \alpha_i y_i x_i \\ \frac{\partial L_P}{\partial b} = 0 & -\sum_{i=1}^L \alpha_i y_i = 0 \end{cases} \quad (7)$$

The maximization of the margin which is equivalent with the $\|w\|^2$ minimization, is augmented with a regularization parameter $C > 0$ used within a term $C \sum_i \xi_i$ needed in order to

penalize misclassification and margin errors [1]. The optimization parameter becomes (8).

$$\min\left(\frac{1}{2} \|w\|^2 + C \sum_i \varepsilon_i\right) \quad (8)$$

This C parameter is also known as the soft margin of the SVM classifier. If this is set to a large value, an associated large penalty will be assigned to the margin error. If this is lowered, each point that needs to be classified may become a margin error. In order to map each x_i component of the X vector into a higher dimensional space, a kernel function is introduced [3]. Its general form is represented by (9).

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (9)$$

There are many kernel functions which can be used with a support vector machine algorithm. In this paper, the results obtained by using some of these functions are also presented and compared between each other in order to determine the most precise of them. Some of the equations used for this study are presented as follows: polynomial kernel (10), Gaussian kernel (11) and sigmoid kernel (12).

$$K(x,y) = (\gamma x^T y + r)^d \quad (10)$$

$$K(x,y) = \tanh(\gamma x^T y + r) \quad (11)$$

$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (12)$$

3. Case study: SVM for medical diagnosis

Within this paper a case study is being developed and presented for different categories of the SVM kernel functions which are used to train and validate a database that consists in some clinical instances taken from National Institute of Diabetes and Digestive and Kidney Diseases, from India. This dataset was obtained from UCI Repository of Machine Learning Database [4].

3.1. Dataset description

The Diabetes dataset used for this study contains 768 instances representing patient's associated values for each symptom taken into consideration for confirming the diagnosis. It has a number of 8 attributes and a class, all described in table 1.

Regarding the class distribution, this can take one of the following two labels: "tested positive", in case the patient's symptoms will confirm this disease and "tested negative", otherwise.

Before running the SVM algorithm, a preprocessing step needs to be followed in order to normalize the dataset and improve the overall performance. This consists in scaling each input attribute dimension to the same interval. In this paper, a scaling from the initial interval to [0,1] is being applied, and thus, it will prevent that some

input dimension will completely influence the others. The distribution of each attribute values after applying the normalization process is presented in figure 2.

3.2. Experimental results and performance evaluation

Within this section, the results obtained by training and validating the neural network associated with the described database with the SVM classifiers, are described. The classification phase consists in two main sub-phases represented by training the dataset and afterwards, testing it with the same dataset used initially.

In what concerns the evaluation metrics [5], three main indicators are used for each type of SVM classifier to analyze the overall performance, as follows:

- Accuracy which represents the proportion of the total number of predictions that were correctly classified for their class. It is determined by (13).

$$Ac = \frac{A+D}{A+B+C+D} \quad (13)$$

where:

A - is the number of correct prediction that an instance is negative

B - is the number of incorrect prediction that an instance is positive

C - is the number of incorrect prediction that an instance is negative

D - is the number of correct prediction that an instance is negative

- Sensitivity or recall which represents the percent of all the negative cases correctly classified, out of all possible instances labeled as negative. This is calculated based on (14).

$$S = \frac{A}{A+B} \quad (14)$$

- Root mean square error (RMS) measures the difference between the predicted values and the actually observed values, It represents the standard deviation between these category of values and is calculated based on (15).

$$RMS = \sqrt{\frac{\sum_i e_i^2}{n-1}} \quad (15)$$

where:

e_i - represents the error calculated as a difference between each observed and calculated value

n - represents the total number of instances taken into consideration for the experiment

The results obtained after training and validating the neural network by applying SVM classifier are presented in table 2 for each kernel function described previously and in each case, for different combination of parameters. For the polynomial kernel, these parameters are represented by gamma and the degree of the polynomial function and in what concerns the sigmoid and Gaussian ones, the variables are represented by the gamma parameter only. In all cases, the regularization parameter was set to $C=100$.

The performance of the classifier is determined by the computation of the total accuracy, sensitivity and RMS. As illustrated in table 2, the best results are obtained with a polynomial kernel function that has a degree of 5 and a gamma parameter of 0.9.

As shown, for these values, an accuracy of 88.9323 was obtained.

This proves the fact that support vector machine and furthermore, the neural artificial networks, can represent a powerful tool for medical diagnosis. By using a polynomial SVM, training the proposed model achieves higher accuracy in comparison with other kernels and represents a good approach within this field, for classification purposes.

4. Conclusions and future research

Within this paper, a comparison of the performance obtained by training and validating a diabetes neural network was made for different parameters of the support vector machine classifier algorithm, using standard statistical metrics. The results of this study show that the SVM algorithm implemented for a polynomial function gave the best performance. The classifier accuracy and sensitivity have been found to be high and thus, making it a good solution for medical diagnosis field. Therefore, it can be concluded that the artificial neural network model described in this paper represents an efficient quantitative and qualitative instrument for the classification process involving diagnosis. This study will be further enhanced and expended by applying the algorithm on other medical dataset and also by incorporating other classification algorithms.

Acknowledgment

This work was financially supported through the project "Routes of academic excellence in doctoral and post-doctoral research - READ" co-financed through the European Social Fund, by Sectoral Operational Programme Human Resources Development 2007-2013, contract no POSDRU/159/1.5/S/137926.

5. Reference list

- [1] Battista, B. (2011). Support Vector Machines Under Adversarial Label Noise, *Workshop and Conference Proceedings 20 of Asian Conference on Machine Learning*, 97-112
- [2] Cheng, S.O., & Soren, S., Support Vector Machines and Kernels for Computational Biology, available at:
<http://svmcompbio.tuebingen.mpg.de/plos-svm.pdf>
- [3] Hofmann, T., & Bernhard, S. (2008). Kernel methods in machine learning, *The Annals of Statistics*, Vol. 36, No. 3, 1171–1220
- [4] <https://archive.ics.uci.edu/ml/datasets/Diabetes>
- [5] <http://www-users.cs.umn.edu/~kumar/dmbook/ch4.pdf>
- [6] http://www.cise.ufl.edu/class/cis4930sp11dtm/notes/intro_svm_new.pdf

Appendices

Appendix A

Table No. 1
Diabetes database attributes

| Attribute name | Unit |
|------------------------------|---------------------------------|
| Number of times pregnant | Numeric |
| Plasma glucose concentration | mg/dl |
| Diastolic blood pressure | mmHg |
| Triceps skin fold thickness | mm |
| 2-Hour serum insulin | µU/ml |
| Body mass index | ((weight in kg/(height in m)^2) |
| Diabetes pedigree function | Numeric |
| Age | Numeric (years) |
| Class variable | Numeric (0 or 1) |

Source: prepared by the author based on [4]

Table No. 2
Results obtained after training and validating the neural network

| Kernel function | | RMS | Accuracy (%) | Sensitivity (%) |
|--------------------|------------|--------|--------------|-----------------|
| Polinomyal | | | | |
| Gamma (γ) | Degree (d) | | | |
| 0.00001 | 3 | 0.5907 | 65.1042 | 42.4 |
| 0.0003 | 3 | 0.5449 | 70.3125 | 74.5 |
| 0.0009 | 5 | 0.4677 | 78.1250 | 77.9 |
| 0.005 | 3 | 0.4774 | 77.2135 | 76.8 |
| 0.009 | 4 | 0.4732 | 77.6042 | 77.1 |
| 0.009 | 1 | 0.476 | 77.3438 | 77.0 |
| 0.01 | 3 | 0.4719 | 77.7344 | 77.3 |
| 0.1 | 3 | 0.4593 | 78.9063 | 78.3 |
| 0.1 | 4 | 0.455 | 79.2969 | 79.1 |
| 0.3 | 4 | 0.433 | 81.25 | 81.1 |
| 0.5 | 4 | 0.4177 | 82.5521 | 82.4 |
| 0.7 | 4 | 0.4082 | 83.3333 | 83.2 |
| 0.7 | 5 | 0.3698 | 86.3281 | 86.4 |
| 0.9 | 4 | 0.3886 | 84.8958 | 84.9 |
| 0.9 | 5 | 0.3327 | 88.9323 | 89.0 |
| 0.006 | 5 | 0.4732 | 77.6042 | 77.1 |
| 0.007 | 4 | 0.4705 | 77.8646 | 77.5 |
| Gaussian - gamma | | | | |
| 0.00001 | | 0.5907 | 65.1042 | 42.4 |
| 0.0009 | | 0.4841 | 76.5625 | 76.8 |
| 0.005 | | 0.4746 | 77.474 | 77.1 |
| 0.009 | | 0.4732 | 77.6042 | 77.2 |
| 0.09 | | 0.4621 | 78.6458 | 78.3 |
| 0.01 | | 0.4719 | 77.7344 | 77.3 |
| 0.1 | | 0.4593 | 78.9063 | 78.6 |
| 0.3 | | 0.455 | 79.2969 | 79.0 |
| 0.7 | | 0.4285 | 81.6406 | 81.4 |
| 0.9 | | 0.4114 | 83.0729 | 82.9 |

| Sigmoid - gamma | | | |
|-----------------|--------|---------|------|
| 0.00001 | 0.5907 | 65.1042 | 42.4 |
| 0.0003 | 0.5907 | 65.1042 | 42.4 |
| 0.009 | 0.4691 | 77.9948 | 77.8 |
| 0.9 | 0.5013 | 74.8698 | 748 |
| 0.09 | 0.4814 | 76.8229 | 76.3 |
| 0.01 | 0.4663 | 78.2552 | 78.1 |
| 0.1 | 0.4828 | 76.6927 | 76.1 |

Source: prepared by the author after running SVM algorithm

Appendix B

Figure 1
SVM Classifier [6]

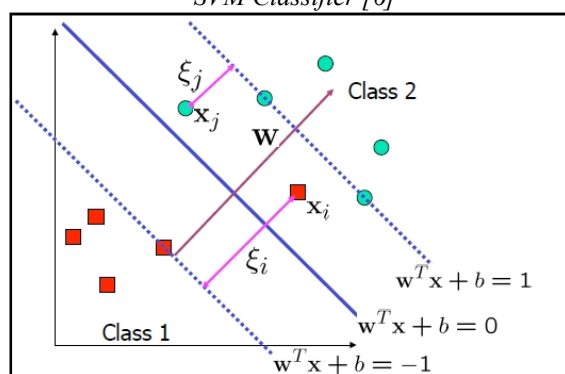
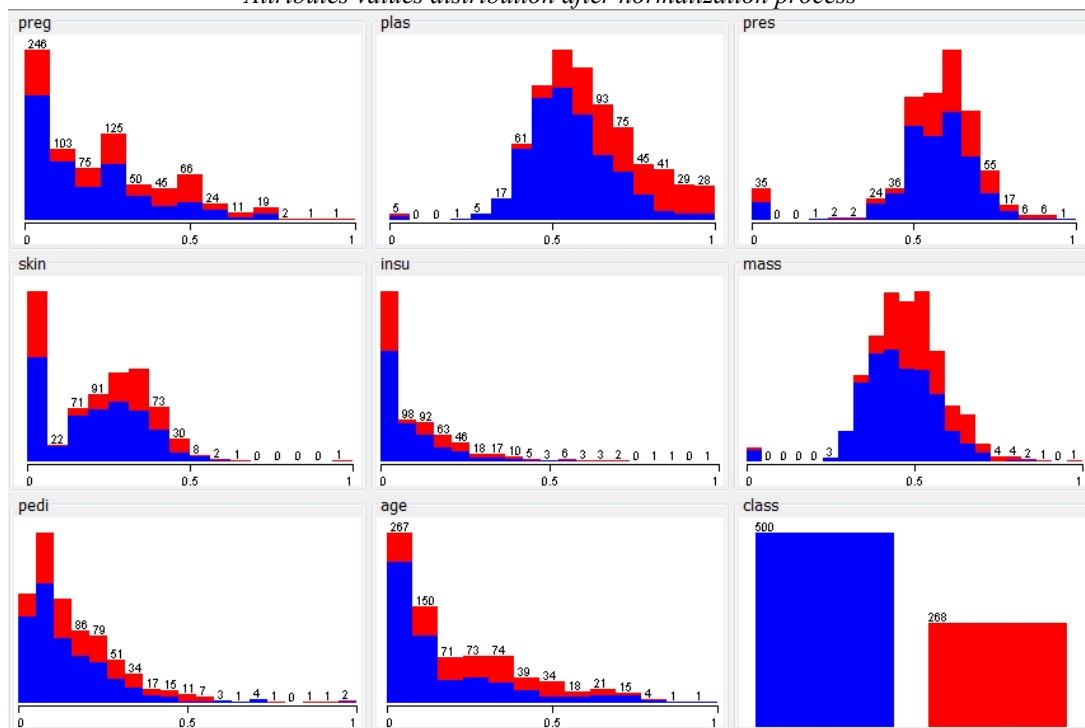


Figure 2

Attributes values distribution after normalization process



Source: prepared by author after running the normalization function on all attributes present in the dataset