TRAINING A SUPPORT VECTOR MACHINE CLASSIFIER

Fawzia M. Abujalala[1], Hajer R. Abulifa[2], Ahmed M. Abushaala[3]

[1,2,3]Faculty of Information Technology, Misurata University, Misurata, Libya
[1]f.abujalala@it.misuatau.edu.ly
[2]h.abulifa@it.misuatau.edu.ly
[3]a.mohamed@it.misuatau.edu.ly

ABSTRACT

Machine Learning (ML) is a very important field that enable researchers to develop many computer-based applications that can be used to facilitate employer's carrier in many scientific area. Support vector machine (SVM) is one of the most popular supervised learning algorithms in ML that are used to classify data. In this paper, researchers have trained a support vector machine classifier over linearly and nonlinearly separable data set. Over the nonlinearly separable dataset to types of kernels has been used, polynomial kernel and radial-basis function (RBF) kernel. The polynomial kernel reached the peak accuracy of 96% at degree of 10 and no additional capacity control. However, RBF kernel reached the peak accuracy of 96% at RBF sigma of 1 and misclassification error of 10. Moreover, cross validation technique has been used to improve and measure the performance of the nonlinear classifier.

KEYWORDS

Machine learning, SVM Classifier, Radial-basis Function, Classification.

1. INTRODUCTION

Researchers have been interested long time about building machines that are able to learn from experience. They have proved that machines can be trained to gain a significant level of learning. This make it possible to solve problems that cannot be solved by traditional programming techniques, as there are no available mathematical model, such as hand written recognition, finding genes in a DNA sequence, filtering mail, detecting or recognizing objects in machine vision, and so on. Machine learning algorithms provide the key solution to these problems [1]. For example, successful implementation of ML methods can help the integration of computer-based systems in the healthcare environment providing opportunities to facilitate and enhance the work of medical experts and ultimately to improve the efficiency and quality of medical care. So, machine learning improves the accuracy of medical diagnosis by analyzing data of patients.

The measurements in ML application are typically the results of certain medical tests (example blood pressure, temperature and various blood tests) or medical diagnostics (such as medical images), presence/absence/intensity of various symptoms and basic physical information about the patient(age, sex, weight .. etc.). On the basis of the results of these measurements, the doctors narrow down on the disease inflicting the patient.
Moreover, it is extremely important to be able to detect, identify and correct faults automatically in control systems and dangerous work environments. ML play a key role to deal with this problem and apply learning algorithms to detect and classify fault type. Then, machines should be able to automatically recover from these errors.

This paper addresses the problem of training a binary support vector machine classifier over a set of linear and nonlinear separable data. The performance measure of the proposed classifier has been measured using three measurement metrics, processing speed, correct rate, and error rate.

2. RELATED WORK

Deng, et al. (2017), have proposed a system for sensor fault detection and identification using improved support vector machine methods. They have applied online sparse least squares support vector machine (OS-LSSVM) to detect and predict sensor faults. Then, researchers classify these errors in order to identify them using error-correcting output codes support vector machine (ECOC-SVM). They have reached high accuracy for fault identification [2]. Ayyaz, Javed and Mahmood (2016) have proposed a system for offline handwritten character recognition system based on a hybrid feature extraction technique. The system consists of three main stages which are, pre-processing, feature extraction technique, and SVM based training/classification. Researchers have tested their system on handwritten characters taken from 30 different writers, who were allowed to write in their natural style. The system achieved 96.5% recognition accuracy on chosen digits data and 96% recognition accuracy on chosen alphabets data. Also, the trained support vector machine classifier has shown higher efficiency with respect to speed, memory, and classification accuracy as compared to other related approaches dealing with the handwritten character recognition problem [3]. Al-ayyoub and Al-zghool (2013) have considered the problem of detecting the existence of fraction and its type in long bones using X-ray images. They have addressed a binary classification problem of detecting whether a fraction exists or not, and 5-class classification problem to detect the type of the fraction. They have used SVM, DT, NB, and NN methods [4]. Mangia, Nayak and Kumar (2013), concerned classifiers of fundus images using NB, kNN, and SVM. Also, they used data and image pre-processing techniques to improve the performance of machine learning classifiers [5].

3. SUPPORT VECTOR MACHINE TECHNIQUE

Support vector machine (SVM) is a useful technique for data classification and regression. It was developed in 1995 by Boser, Guyon, and Vapnic, and gained its popularity due to its high performance, accuracy, and ability to deal with high-dimensional data [6]. The basic concept of SVM is to find a linear separation (hyperplane) between classes and has gained the largest margin between the support vectors, which are a subnet of training samples that are the closest points to the hyperplane. Support vectors have a major influence on the hyperplane equation.

$$w.x_i + b > +1 \quad when \quad y_i = +1$$
$$w.x_i + b \le -1 \quad when \quad y_i = -1$$

Thus the equation of a hyperplane is of the form:

$$w^T x + b = 0$$

where, w is a weight vector, x is input vector, and b is bias.

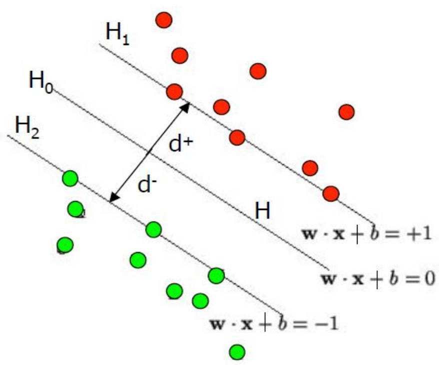

Figure 1. Hyperplane equation

From Figure 1, margin of separation (d), i.e. the total distance between H0 and H1 is given by:

$$\frac{|w.x+b|}{\|w\|}+\frac{|w.x+b|}{\|w\|}=\frac{2}{\|w\|}$$

In order to maximize the margin, w must be minimized, where:

$$w=\sum_{i=1}^{L}\propto_i y_i x_i$$

where α is a variable of the Lagrangian function.

Moreover, in some cases the dataset points cannot be separated by a linear hyperplane, so SVM algorithm transform the data to higher dimensional feature space. This is achieved by using kernel trick, which is a function turn the input data into another space, then calculates the hyperplane in that space.

$$K(x_i,x_j)=\Phi(x_i).\Phi(x_j).$$

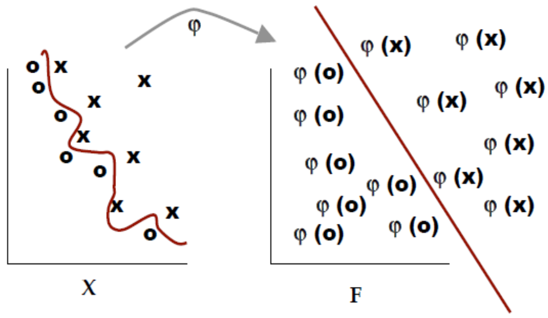

Figure 2. Transformation to separate

Among acceptable functions, the most popular kernels are polynomials, radial basis function, and sigmoid functions. The proposed classifier was trained on polynomials and radial basis function kernels, as showed in Table 1.

Table 1. Inner product kernels

Type of support vector machine	Inner product kernel $K(x,x_i), i=1,2,\ldots,N$	Usual inner product
Polynomial learning machine	$(x^T x_i+1)^P$	Power P is specified a priori by the user
Radial-basis function (RBF)	$Exp(1/(2\sigma^2)\|x-x_i\|^2)$	The width σ^2 is specified by a priori

3.1. Cross-Validation Measurement and Improve Performance

Cross-validation was used to estimate the percentage rates of correct and incorrect classifications that are needed for the analysis. Cross-validation is a way of estimating how the algorithm will perform on new, unknown data given a set of data with known classifications. The data set is randomly divided into two subsets – a training set and a testing set. The learning algorithm is then trained using the training set.

After the algorithm has been trained, it is then used to predict the classifications of the test data set. Since the data in the test set has known classifications, the known classifications are compared with the predictions made by the algorithm, and the percent classified correctly and incorrectly can be obtained.

4. Implementation of the Proposed SVM Classifier

In this work, the iris dataset has been used to construct a support vector machine (SVM) classifier. The dataset contains 3 classes of 50 instances each, where each class refers to a type of iris plantSetosa, Versicolour, and Virginica. One class is linearly separable from the other two; the latter is not linearly separable from each other. The dataset contain four attributes, sepal length, sepal width, petal length, and petal width. To visualize the problem, researchers restricted to two features that contain the most information about the class, namely the petal length and the petal width, as illustrated in the Figure 3.

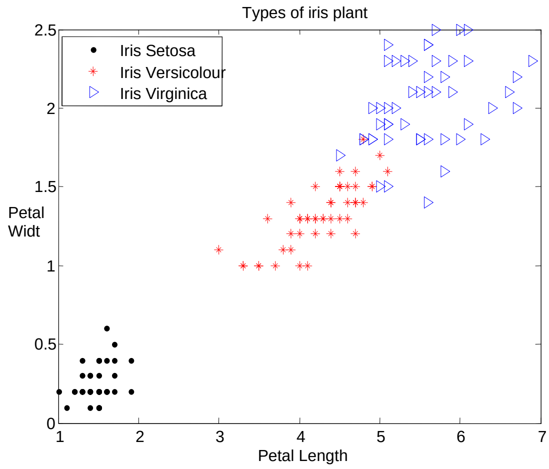

Figure 3. Types of iris plant

To analyze the performance of the proposed classifier, several experiments were conducted using a dataset of 100 samples. These samples were evenly divided between the two classes, 50 sample for each class. 80 samples were used for training, 40 samples from each class, and 20 samples were used for testing, 10 samples from each class.

The performance of the proposed classifier is also analyzed using three performance metrics, namely, speed of classification, correct rate, and error rate. The classifier was developed using MATLAB R2013a and all of the experiments were conducted on core i3 machine with 4GB RAM.

5. Results and Discussion

5.1. Linear Classification

Firstly, iris Setosa was easily separated from Versicolour and Virginica using linear hyperplane, with two support vectors circled as shown in the Figures 4, 5, 6 and 7.

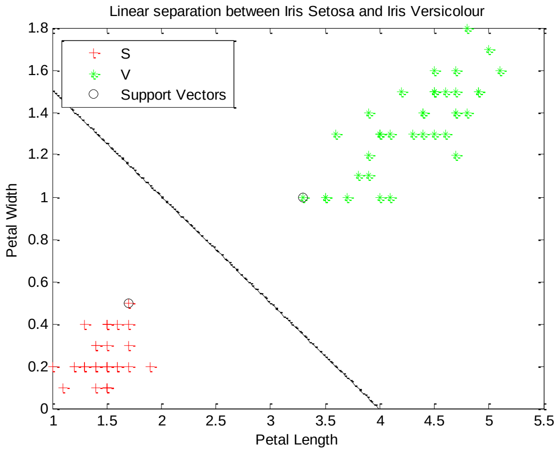

Figure 4. Linear separation between iris Setosa and iris Versicolour.

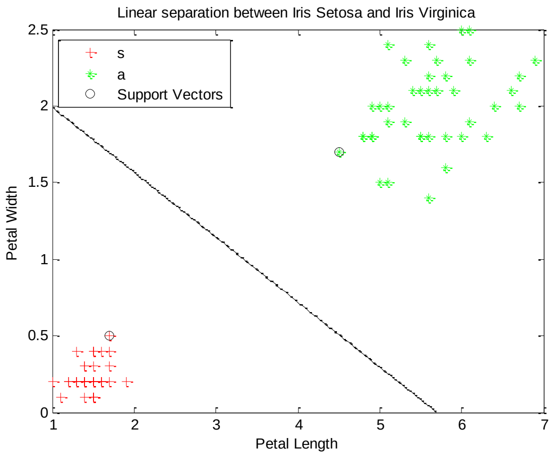

Figure 5. Linear separation between iris Setosa and iris Virginica.

Then, these classifiers were tested with 20 test samples and accuracy of 100% were obtained.

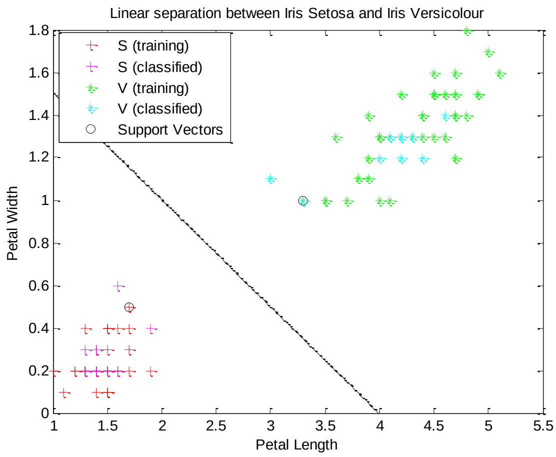

Figure 6. Linear separation between iris Setosa and iris Versicolour with test samples.

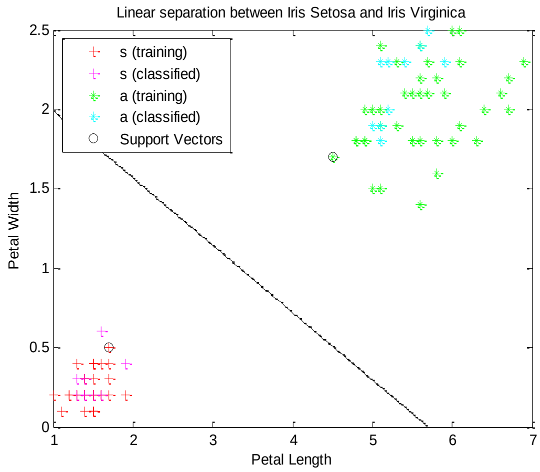

Figure 7. Linear separation between iris Setosa and iris Virginica with test samples.

5.2. Nonlinear classification using polynomial and Gaussian radial basis kernel

In order to separate Versicolour, and Virginica classes, polynomial kernel were used with degree of 10 and no additional capacity control, as in the Figure 8.

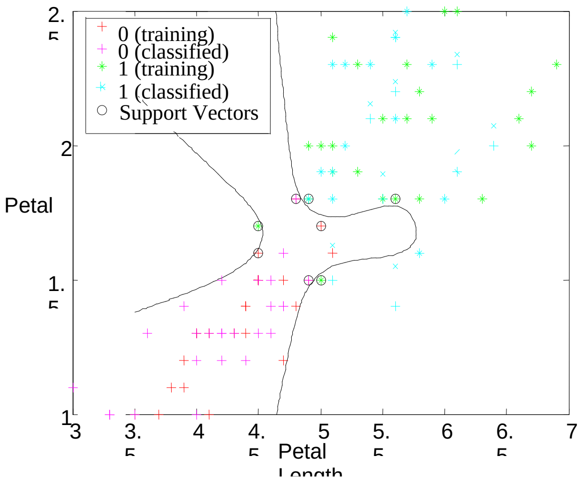

Figure 8. Nonlinear separation between Versicolour and Virginica using polynomial kernel.

Table 2. The result when using different polynomial degrees and misclassification errors

Degree	Box Constraint	Time	Correct Rate	Error Rate
15	1	0.825 Sec	0.92	0.08
	10	0.852 Sec	0.92	0.08
	INF	0.848 Sec	0.96	0.04
8	1	0.841 sec	0.90	0.1
	10	0.851 sec	0.92	0.08
	INF	0.948 sec	0.94	0.06
10	1	0.814 sec	0.90	0.1
	10	0.827 sec	0.90	0.1
	INF	0.843 sec	0.96	0.04

As illustrated in Table 2, additional testing has been done using different polynomial degrees and misclassification errors to obtain the best combination of these parameters. It was found that polynomial kernel with degree of 10 and no additional capacity control gave the best accuracy rate of 96%, and processing speed of 0.843 seconds.

Furthermore, Versicolour, and Virginica classes were separated using RBF kernel with RBF sigma of 0.2 and misclassification error of 10, as in the Figure 9.

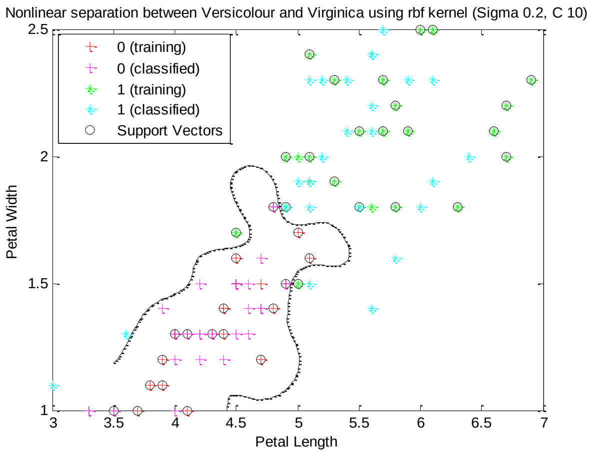

Figure 9. Nonlinear separation between Versicolour and Virginica using RBF kernel-sigma 0.2

The Figure 10 shows the use of RBF kernel with RBF sigma of 1 and misclassification error of 10, to separate the second and third classes.

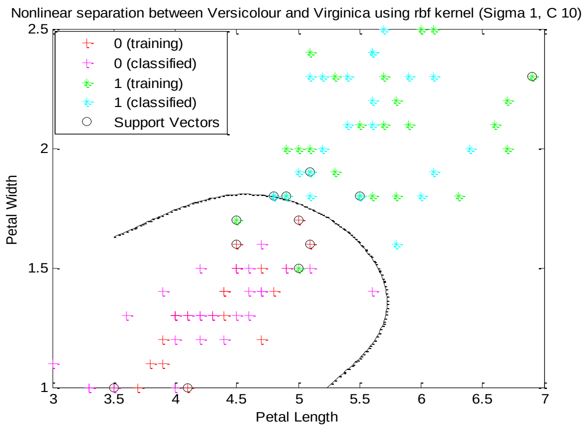

Figure 10. Nonlinear separation between Versicolour and Virginica using RBF kernel-sigma 1

Then, additional testing has been done using different RBF sigma values and misclassification errors to obtain the best combination of these parameters as shown in Table 3. It was found that RBF kernel with RBF sigma value of 0.2 and 1 misclassification error of 10 gave the best accuracy rate of 92% and 96%, and processing speed of 0.881 and 0.861 seconds respectively.

Table 3. The result when using different RBF kernel with RBF sigma values and misclassification errors

Sigma Value	Box Constraint	Time	Correct Rate	Error Rate
0.2	1	0.789 sec	0.90	0.10
	10	0.881 Sec	0.92	0.08
	INF	0.901 Sec	0.92	0.08
1	1	0.872sec	0.96	0.04
	10	0.861 sec	0.96	0.04
	INF	1.191 sec	0.88	0.12
2	1	0.878 sec	0.96	0.04
	10	0.854 sec	0.96	0.04
	INF	2.952 sec	No Convergence	

Next, k-fold cross validation has been implemented over 10 subsets randomly chosen from the training data to improve the performance of the proposed RBF kernel classifier. It was found that the correct rate was75% in the first iteration, as it can be seen in Table 4. Then it continues to increase until it reaches 93.75% in the sixth iteration, and then starts to decrease. This means the cross validation should stop on the sixth iteration to obtain best performance.

Table 4. Performance measurement using cross-validation

Iteration	1	2	3	4	5	6	7	8	9	10
Correct Rate	0.75	0.8125	0.8750	0.9063	0.9250	0.9375	0.9107	0.9063	0.8889	0.9000

Finally, additional performance measure was measured with respect to the size of the training set and testing set.

Table 5. Performance measure with respect to the size for training and testing sets

Training Set	Testing Set	Correct Rate * 100
10	70	75.71%
20	60	83.33%
30	50	94.00%
40	40	92.50%
50	30	83.33%
60	20	95.00%
70	10	90.00%

As illustrated in Table 5, It was found that the correct rate was 75.71% with 10 samples for training and 70 samples for testing. Then it continued to increase until it reaches 94% with 30 samples for training and 50 samples for testing. Then it starts to decrease and reaches a peak of 95% with 60 samples for training and 20 samples for testing. According to the results from table 5, it can be seen that the correct rate decreases when the training set become larger than 75% of total dataset. This is due to the problem of over-fitting, where the classifier becomes narrowly surrounded around the training set.

3. CONCLUSIONS

In this work, we presented a SVM classifier whose training on the iris dataset, Firstly, iris Setosa was easily separated from Versicolour and Virginica using linear hyperplane. Then, these classifiers were tested with 20 test samples and the accuracy was 100%. Secondly, nonlinear classification using polynomial and Gaussian radial basis kernel. It was found that polynomial kernel with degree of 10 and no additional capacity control give the best accuracy rate of 96%, and processing speed of 0.843 seconds. Furthermore, Versicolour, and Virginica classes were separated using RBF kernel with RBF sigma of 0.2 and 1 with misclassification error of 10 to obtain the best accuracy rate. The values of 92% and 96%, and processing speed of 0.881 and 0.861 seconds respectively were the best accuracy. The k-fold cross validation has been implemented over 10 subsets randomly chosen from the training data to improve the performance of the proposed RBF kernel classifier. It was found that the correct rate was 75% in the first iteration. Then it continue to increase until it reaches 93.75% in the sixth iteration, and then starts to decrease. This indicates the cross validation should stop on the sixth iteration to obtain best performance. Finally, according to the results from table 5, it can be seen that the correct rate decreases when the training set become larger than 75% of total dataset. This is due to the problem of over-fitting, where the classifier becomes narrowly surrounded around the training set.

REFERENCES

1. Russell, S.J. and P. Norvig, *Artificial intelligence: a modern approach*. 2016: Malaysia; Pearson Education Limited.
2. Deng, F., et al., *Sensor multifault diagnosis with improved support vector machines.* IEEE Transactions on Automation Science and Engineering, 2017. **14**(2): p. 1053-1063.
3. Ayyaz, M.N., I. Javed, and W. Mahmood, *Handwritten character recognition using multiclass SVM classification with hybrid feature extraction.* Pakistan Journal of Engineering and Applied Sciences, 2016.
4. Al-Ayyoub, M. and D. Al-Zghool, *Determining the type of long bone fractures in x-ray images.* WSEAS Transactions on Information Science and Applications, 2013. **10**(8): p. 261-270.
5. Mangai, J.A., J. Nayak, and V.S. Kumar, *A novel approach for classifying medical images using data mining techniques.* International Journal of Computer Science and Electronics Engineering (IJCSEE), 2013. **1**(2).
6. *Ray, S. "Understanding Support Vector Machine algorithm from examples (along with code)", 2017, available at: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/, accessed at: [20/12/2017]*