P-Order twin support vector machine

Xu Ma, Qiaolin Ye

College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China

# ABSTRACT

Twin Support Vector Machine(TWSVM) [1] is an effective classifier especially for binary data, which is defined by squared distance in the objective. It is well known that distance is susceptible to outliers, which can lead to errors. It is desirable to develop a revised TWSVM. In this paper, a new robust twin support vector machine via p-order optimized algorithm was proposed. We improved the TWSVM algorithm by iterative method. Theoretical support shows that iterative method is effective in the solution to improve TWSVM via p-order instead of distances. A large number of experiments show that p-order Twin Support Vector machine (pTWSVM) can process the noise data effectively and has a better accuracy.

1. Introduction

Support vector machine [2] has been a vital method for pattern classification in the last decade. The standard Support vector machine devotes to get an optimal separating hyper plane that has the max margin between the two data sets. In 2001, G.Fungand and O.L.Mangasarian proposed a algorithm that two parallel planes are pushed apart as far as possible to classify points [3] . In 2007, O.L. Mangasarian and E.W.Wild proposed a nonparallel plane classifier for binary data via generalized eigenvalue [4].

Different from PSVM and GEPSVM, a new nonparallel plane classifier termed as the Twin Support Vector Machine (TWSVM) was proposed by Jayadeva in 2007. TWSVM obtains nonparallel planes around which the data points of the corresponding class get clustered. It solves a pair of quadratic programming problems. Each of the two quadratic programming problems has the formulation of a typical SVM, except that not all data points are used in the constraints of either problem at the same time.

In this paper, we are absorbed in the problem of robust TWSVM on data set with outlier data samples. In classical TWSVM, it is willing to minimize the distance with the squared distance. As we know, squared outliers distance will expand the error distance of samples. From this point，we hold the notion that distance with a lower orders can emphasis the percentage of normal points distance. A p-order is used for the improvement of TWSVM that p ought to be lower than 2, e.g. .

The p-order twin support vector machine (pTWSVM) method is focus on the following problems:

1．The modification of the TWSVM objective with p-order distance.

2．The formulation of proposed algorithm.

3. The proof of the algorithm convergence.

4.The nonlinear kernel classifier.

The paper is organized as follows: Section 2 dwells on our theoretical work for the new method in detail, including the improvement and related proof. Section 3 deals with the experiment. Section 4 is about the extension on nonlinear kernel and Section 5 summarize this paper.

1. P-Order Twin Support Vector Machine

2.1 Optimization Algorithm to the Proposed Method

Suppose we have data points of n-dimensional belongs to two classes represented by matrices A and B respectively. Assuming that A have m1 points and B have m2 points, so the sizes of matrices A and B are and respectively. The TWSVM devotes to obtaining two nonparallel hyper planes which each plan is as close as possible to one type points and as far as possible to the rest.

The TWSVM can be obtained by solving the following pairs of quadratic programming problems:

(1)

(2)

where are parameters and are vectors of ones of appropriate dimensions. The two nonparallel planes can be obtained by :

We can classify the point X by comparing the geometrical margin to the two planes respectively.

Form the TWSVM, it clearly shows that the squared distance in the formulas. It may be not satisfied the for the problem. The result we obtained could be affected by the outliers pronouncedly. That is, p-order is a good method for instead of squared distance. If we can find an appropriate p, the algorithm can emphasize norm data and overlook outliers best. Now, we can find that what the p-order value is to obtain a balance between the norm data and the outliers. Assuming squared distance is a benchmark, if ，the distance of data will be shortened and the outlier data samples will be alleviated. The paper holds the notion that the percentage of outliers decides the p value.

The improvement of TWSVM can be obtained by solving the following problem:

(3)

（4）

The Lagrange function of the problem is:

(5)

where are the vectors of Lagrange multipliers.

To solve the problem ,a good approach is splitting the distance to squared and (p-2)-th order :

(6)

Denote

(7)

the Lagrange function can be written as:

(8)

The derivative on every parameter, i.e., the Karush-Kuhn-Tucker(K.K.T) necessary and sufficient optimality conditions for the problem is:

(9)

(10)

(11)

(12)

(13)

(14)

(15)

(16)

Form ,, we have

(17)

We define

(18)

Notice that can be signified as:

(19)

Combining (9) and (10) leads to

(20)

this can be written as:

(21)

i.e.:

(22)

Although is always positive semidefinite , it is possible that it may not be well conditioned in some situations. So the problem can be regularized by introducing a regularization term as follows:

(23)

where and is an identity matrix of appropriate dimensions.

Using the Lagrange function and the K.K.T. conditions above, we obtain the Wolfe dual of p-th order TWSVM as follows:

(24)

Similarly, the other one’s dual is:

(25)

We can obtain the optimal via an iterative algorithm. In each iteration, are calculated with the current calculated . The iteration produce is repeat until converges. The iteration is started with a initialized . the are re-changed adaptively during each iteration.

The algorithm to solve the problem:

Input : Training data , parameter .

Give out .

Initialize .

While converge do

1. Calculate ;
2. Calculate via dual function;
3. Update , add regularization term if necessary;

End

Output .

The another one is similarly like the process above.

2.2 Convergence Analysis

To prove the convergence of the new algorithm, we need the following lemmas that proved by Hua Wang in his paper:

Lemma: For any nonzero vectors A, B, when , the following inequality holds:

(26)

In order to use this lemma in this paper, we can transform it into the following form:

(27)

Considering the objective function, suppose , and the updated is B. According the objective function, we know that

(28)

i.e.

(29)

Connect the lemma, it can be obtained that:

(30)

(31)

(32)

Combing the (31) and (32), we arrive at

(33)

Note that A is , and B is updated . Inequality (33) indicates that will converge after each iteration. According , so the is convergent in each iteration. is a positive number, this leads to it will converge to an optimal value via an iterative approach. The convergence of also means that ’s convergence.

1. Experimental Results

3.1 Binary data

In order to directly compare the differences between TWSVM and pTWSVM, we did experiments on an artificial data set. A simple data set was constructed, with ten points distributed over and respectively.



(TWSVM) (pTWSVM)

fig1 binary data experiments pictures

The above diagram shows that the two algorithms have good classification effect on binary data sets and the classification surfaces are almost the same. Next we add a little noise to the data set to test the robustness of the two algorithms.



（TWSVM） （PTWSVM） （TWO）

fig2 binary data with outliers experiments pictures

Form the picture1 and picture2 we can find that the classification surfaces are similar in terms of structure. Picture3 shows that pTWSVM provides a better classification. This proves that pTWSVM is much less susceptible to noise than TWSVM and has good robustness.

3.2 Study the p value of the new proposed method

The new method leads to a problem that what the value of p is. Considering the objective function, we hold the notion that the p’s value is under the influence of outliers. In order to get a higher accuracy, the greater proportion of noise, p value is smaller，and vice versa. Formula 11？perspicuously indicate that p value directly affect the result of the formula. Splitting the formula into two parts: the outliers functional margin and the normal data’s functional margin. The role of p value is to emphasize the proportion of the two.

We experiment with several benchmark data set as examples. Considering the objective function, we hold the notion that the parameter p value can directly affect experiment accuracy. Our goal is to obtain the best parameter value to get the best classification accuracy. We vary p of the proposed objective in the range of 0.1 to 2 to study its impacts to the classification performance. Through the experimental data, we simulate the corresponding correct rate curve.



fig3 accuracy with different p value

Pictures show that the accuracy of pTWSVM on different data sets will increase with increasing p value. When p is about 1.3, the accuracy reaches the maximum. When , the accuracy will be slightly decreased. This can be attributed that, when p is big, the distance measurement will be compromised. This conclusion implies p should not be too big to invalidate the distance measurement in the Euclidean space. Upon the result in fig3, empirically, we set in all our subsequent experiments.

## 3.3 Comparison of accuracy

In this section, we compare the pTWSVM algorithm with other algorithm by experiments. We evaluate the proposed method on several widely used benchmark datasets in machine learning studies. The descriptions of the datasets are given in table 1.

|  |  |  |
| --- | --- | --- |
| Table1: Data sets descriptions | | |
| Data sets | Number | Dimension |
| Heart | 270 | 13 |
| Australian | 690 | 14 |
| Pima | 768 | 8 |
| Sonar | 208 | 60 |
| Spect | 267 | 44 |
| germ | 1000 | 24 |
| Monk1 | 561 | 6 |
| cancer | 683 | 9 |
| Ionodata | 351 | 34 |
| splice | 1519 | 60 |
| cmc | 642 | 9 |
| blood | 748 | 4 |
| haberman | 306 | 3 |
| liver | 327 | 6 |
| diabetes | 768 | 8 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 2:Test Set Accuracy with a Linear Kernel | | | | | |
|  | SVM | PSVM | GEPSVM | TWSVM | PTWSVM |
| Heart | 0.8444 | 0.8407 | 0.8593 | 0.8519 | 0.8741 |
| Australian | 0.8551 | 0.8580 | 0.7362 | 0.8041 | 0.8653 |
| pima | 0.7969 | 0.7707 | 0.6771 | 0.7073 | 0.8013 |
| sonar | 0.7117 | 0.7840 | 0.7019 | 0.8068 | 0.9038 |
| spect | 0.7000 | 0.7671 | 0.6150 | 0.6578 | 0.8289 |
| germ | 0.7200 | 0.7660 | 0.7040 | 0.7020 | 0.7660 |
| monk1 | 0.6644 | 0.6651 | 0.7014 | 0.6667 | 0.6829 |
| cancer | 0.9795 | 0.9604 | 0.7251 | 0.9591 | 0.9708 |
| ionodata | 0.9261 | 0.8603 | 0.8239 | 0.8750 | 0.8750 |
| splice | 0.7320 | 0.8360 | 0.7620 | 0.5981 | 0.5981 |
| CMC | 0.6916 | 0.7511 | 0.6822 | 0.5794 | 0.7259 |
| blood | 0.7941 | 0.7733 | 0.6123 | 0.8182 | 0.8476 |
| haberman | 0.2745 | 0.7351 | 0.7582 | 0.7255 | 0.6601 |
| liver | 0.6646 | 0.6970 | 0.4512 | 0.6098 | 0.7073 |
| Diabetes | 0.7969 | 0.7695 | 0.6771 | 0.7292 | 0.8047 |

The table 2 compares the performance of the pTWSVM classifier with that of some other SVMs. The experiments of each algorithm were implemented by using MATLAB R2014b. All classifiers are trained with a linear kernel. Optimal values of the parameters were obtained by using a tuning set comprising of 10 percent of the data set.

Form the table we can find that pTWSVM performs best on the vast majority of data sets compared to several other algorithms. And a detail is that pTWSVM behavers better than TWSVM on any dataset, or the same.

The experimental results indicate that pTWSVM is not only effective, but also can be a better choice on most data sets. In addition, the accuracy of pTWSVM is very close to the best.

## 3.4 Robustness Against Outliers Samples

Since the main advantage of the new proposed pTWSVM algorithm dedicated to process noisy samples, we will focus on the processing of the data sets with outliers in the following experiments.

First, we construct a noise matrix . The noise matrix will be involved in the benchmark data set and they will constitute the data sets with outliers , where is the noise factor that can determine the degree of data contamination. The differences in accuracy of each algorithm will be obtained by comparing the performances of the origin data and contaminated data. To get a deep association, we take different value in experiment. The following pictures summarizes the performance of different algorithms on some benchmark datasets with different values of .





fig4 accuracy with different noise factor value

Form the pictures above, we can get the following points:

Fist, the proposed pTWSVM method is consistently better then TWSVM method on the experimental data sets, which demonstrate that the proposed new methods is able to effectively improve the clustering accuracy on noisy data with outlier data samples. This also shows that the new pTWSVM method in the practical application will achieve better results.

Second, no matter what the noise factor value is, the accuracy of pTWSVM always be higher then the accuracy of TWSVM. Although the improvements by pTWSVM method over the comparing methods on the original benchmark data sets without noise are mediocre as shown in Table??，the improvements by our new method on the contaminated data with outlier data samples are considerably large. For example, on the heart data set with outliers, the average pTWSVM accuracy of different value is 0.7481, and TWSVM accuracy is 0.6633. So our proposed method improves the clustering accuracy over the TWSVM method by . In contrast, the improvement of

clustering accuracy on the same data set under the noiseless condition is about 4.47% =(0.8667-0.8296)/0.8296 . The same observations can be seen on all the other experimental data sets, which show that the proposed method has better capability to cluster on contaminated data.

Third, the pictures show that the change in accuracy of pTWSVM is flat and does not change much. This clearly indicate that the new proposed pTWSVM method is faster and easier to stabilize than original TWSVM method. The feature confirms pTWSVM method’s robustness against outlier data samples.

1. The Nonlinear Kernel Classifier

In order to extend our new method to nonlinear classifiers, we have modified the new algorithm by using the kernel method.

As we know, kernel-generated surfaces for TWSVM

,and (34)

where

and K is an appropriately chosen kernel. Note that if the K is a linear kernel like , it will degenerate into an ordinary plane.

We construct an optimization problem KPTWSVM as follows:

(35)

where is a parameter. Next, we define a Lagrange function L by the above formula:

(36)

To solve the problem, we split the distance into two parts:

(37)

In this formula, can be represented by . The Lagrange function is updated as follows:

(38)

We obtain the K.K.T. conditions for KPTWSVM as follows:

(39)

(40)

(41)

(42)

(43)

(44)

(45)

(46)

Combing (39) and (40), we obtain

(47)

Let

, (48)

and the augmented vector . Then the formula can be solved as:

(49)

i.e.:

(50)

The Wolfe dual of KPTWSVM is given by

(51)

In a similar manner, the another KPTWSVM kernel-generated surface can be obtained by solving a new dual function.

Once the two KPTWSVM problems are solved to obtain the surfaces, a new data can be classified in a manner similar to the linear case.

In the actual experiments, if the number of patterns is large, then the rectangular kernel technique can be used to reduce the dimensionality of KPTWSVM. In the linear case, a regularization term always be useful.

1. Conclusions

We have proposed a robust TWSVM based on the -th order of distance, which formulated a non-smooth non-convex minimization problem. Compare to the squared distance，the -th order TWSVM has better accuracy and it is very robust against outlier data samples. The new proposed method takes much more challenging optimization problem than that in the traditional TWSVM. To solve the problem, we introduced an efficient iterative algorithm and provided the rigorous theoretical analysis on the convergence of our algorithm.

# References

[1] Khemchandani R, Jayadeva, Chandra S. Fuzzy Twin Support Vector Machines for Pattern Classification[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(5):905-910.

[2] Cortes C, Vapnik V. Support-Vector Networks[J]. Machine Learning, 1995, 20(3):273-297.

[3] Fung G, Mangasarian O L. Proximal support vector machine classifiers[C]// ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2001:77--86.

[4] Mangasarian O L, Wild E W. Multisurface proximal support vector machine classification via generalized eigenvalues[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(1):69-74.