P-th Order twin support vector machine

ABSTRACT

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-th order of the L2-norm distances. A large number of experiments show that p-th order twin support vector machine (PTWSVM) can process the noise data and has a better accuracy.

1. Introduction

Support vector machine 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. In 2007,O.L. Mangasarian and E.W.Wild proposed a nonparallel plane classifier for binary data via generalized eigenvalue.

Different from PSVM and GEPSVM, a new nonparallel plane classifier termed as the Twin Support Vector Machine (TWSVM). It solves a pair of quadratic programming problems.

In this paper, we are absorbed in the problem of higher precision TWSVM on normal data set. In classical TWSVM, we are willing to minimize the distance with the squared distance. As we know, normal points account for a great proportion, the outliers just are very few points.

From this point，we hold the distance with a high orders, that to emphasis the percentage of normal points. A p-th order is used for the improvement to TWSVM that p ought to be higher than 2,e.g.,.

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

1．The modification of the TWSVM objective with p-th order l2-norm.

2．The formulation of proposed algorithm.

3. The proof of the algorithm convergence.

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 is about the extension on nonlinear kernel. Section 4 deals with the experiment and Section 5 summarize this paper.

1. P-Order twin support vector machine

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 m1\*n and m2\*n 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:

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 distances which it 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-th 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-th order does is to obtain a balance between the norm data and the outliers. Assuming squared distance is a benchmark, if , the data’s distance will be emphasized, if ，the data’s distance will be shortened. The paper hold the notion that the percentage of outliers decides the p value.

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

The Lagrangian function of the problem is:

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 :

Denote

,

the Lagrangian function can be written as:

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

(1)

(2)

(3)

(4)

(5)

(6)

(7)

Form ,, we have

We define

notice that can be signified as:

Combining (1) and (2) leads to

this can be written as:

i.e.:

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:

where and is an identity matrix of appropriate dimensions.

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

Similarly, the other one’s dual is:

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 p,C1,C2.

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.

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 the heart data set.

**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:

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

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

i.e.

Connect the lemma, it can be obtained that:

Combing the ???, we arrive at

Note that A is , and B is updated . Inequality ?? 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. The Nonlinear Kernel Classifier
2. Experimental Results

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 |

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 via linear kernel. Optimal values of the parameters were obtained by using a tuning set comprising of 10 percent of the data set.

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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2:Test Set Accuracy with a Linear Kernel | | | | | | | |
|  | Heart | Australian | Pima | Sonar | Spect | Germ | Monk1 |
| SVM | 0.8444 | 0.8551 | 0.7969 | 0.7117 | 0.7000 | 0.7200 | 0.6644 |
| PSVM | 0.8407 | 0.8580 | 0.7707 | 0.7840 | 0.7671 | **0.7660** | 0.6651 |
| GEPSVM | 0.8593 | 0.7362 | 0.6771 | 0.7019 | 0.6150 | 0.7040 | 0.7014 |
| TWSVM | 0.8519 | **0.8673** | **0.7970** | 0.8213 | 0.5027 | 0.7460 | 0.6597 |
| PTWSVM | **0.8741**  (p=3) | 0.8612  (p=2.1) | 0.7927  (p=3) | **0.8654**  (p=2.5) | **0.8127**  (p=1.3) | 0.7500  (p=3) | **0.7454**  (p=3) |
|  |  |  |  |  |  |  |  |

1. Conclusions

References