2011 Third International Conference on Multimedia Information Networking and Security

Softmax Discriminant Classifier

Fei Zang
School of Science and the State Key Laboratory for Manufacturing Systems Engineering
Xi'an Jiaotong University
Xi'an, China
e-mail: fzang@163.com

Jiang-she Zhang
School of Science and the State Key Laboratory for Manufacturing Systems Engineering
Xi'an Jiaotong University
Xi'an, China
e-mail: jszhang@mail.xjtu.edu.cn

***Abstract*—A simple but effective classifier, which is called softmax discriminant classifier or SDC for short, is presented. Based on the softmax discriminant function, SDC assigns the label information to a new testing sample by nonlinear transformation of the distance between the testing sample and training samples. Experimental results on some well-known data sets demonstrate the feasibility and effectiveness of the proposed algorithm.**

Keywords-softmax discriminant function; classifier; nonlinear transformation

I. INTRODUCTION

The goal of classifier is to use labeled training samples from k distinct classes and characteristics of the testing sample to identify the class it belongs to. In practice, it is important to design an efficient classifier because of the significance its simplicity, classification speed and accuracy when it concerns implementation. As a successful classification method, nearest neighbor (NN) classifies the testing sample based on the best representation in terms of a single training sample. To seek the best representation for the testing sample x, NN classifier finds the nearest training sample. That is to say, it computes the minimal distance between x and all training samples. Linear regression classifier (LRC) [1] which falls into the category of nearest subspace (NS) classification formulates the identification task as a problem of linear regression. The inverse problem of LRC is solved through the least squares and the decision is ruled in favor of the class with the minimum reconstruction error. Sparse representation classifier (SRC) [2], which can be considered as a generalization of nearest neighbor and nearest subspace, exploits the discriminative nature of sparse representation to perform classification.

In this letter, a simple but effective classifier, which is called softmax discriminant classifier or SDC for short, is presented. Combined discriminant characteristic of softmax cost function with classification task, SDC assigns the label information to a new testing sample by nonlinear transformation of the distance between the testing sample and training samples. Experimental results on some well-known data sets demonstrate the feasibility and effectiveness of the proposed algorithm.

II. SOFTMAX DISCRIMINANT CLASSIFIER

The goal of softmax discriminant classifier is to correctly determine the class to which a testing sample belongs by weighting distance between the testing sample and the training samples from that class. Suppose that the training set $X=[X_1, X_2, \cdots X_k]\in R^{m\times n}$, comes from k distinct classes. $X_k=[X_1^k, X_2^k, \cdots, X_{n_k}^k]\in R^{m\times n_k}$ denotes n_k samples from the kth class, where $\sum_{i=1}^{k} n_i = n$. Let $x\in R^{m\times 1}$ be the testing sample. In SDC, we use the k class samples to represent the testing sample with the minimum reconstruction error. The aim of SDC is achieved by maximizing the nonlinear transformation value of distance between the testing sample and the k class samples. Now, we define the softmax discriminant classifier as following:

$$\begin{aligned} l(x) &= \arg\max_i d_x^i \\ &= \arg\max_i \log\Big(\sum_{j=1}^{n_i} \exp(-\lambda \left\| x - x_j^i \right\|_2\Big) \end{aligned} \tag{1}$$

where $d_x^i, l(x)$ denotes the distance between the testing sample and ith class, the identification of x, respectively. The parameter $\lambda > 0$ provides a relative penalty cost. If x belongs to ith class, that is to say, x and x_j^i have similar characteristics, and therefore, $\left\| x - x_j^i \right\|_2$ is close to zero and d_x^i can asymptotically achieves the maximum value. That is the reason why we maximize the d_x^i.

III. EXPERIMENTS

To evaluate classification performance of softmax discriminant classifier, some well-known data sets including UCI dataset [3], handwritten digit and alphabet [4], and IDA dataset [5] are used for our experiments. We first compare SDC with NN classifier which can be regards as "hard"

978-0-7695-4559-2/11 $26.00 © 2011 IEEE
DOI 10.1109/MINES.2011.123

TABLE I. SEVERAL DATA SETS IN THE UCI DATA SETS ARE USED IN OUR EXPERIMENTS

Data name	*Input data dimensionality*	*Number of classes*	*Number of training samples*	*Number of testing samples*
australian	14	2	413	227
banlance	4	3	373	252
cmc	9	3	885	588
glass	9	6	126	88
haberman	3	2	184	122
liver	6	2	207	138
pima	8	2	460	308
sonar	60	2	126	82
teaching	5	3	92	59
transfusion	4	2	449	299

classifier, and then with other four methods including principal component analysis (PCA) [6], linear discriminant analysis (LDA) [7], SRC and LRC. To PCA and LDA, NN classifier is used for performance evaluation in the feature space. Finally, we demonstrate the classification performance of SDC versus the variation of the parameter $\lambda > 0$. In all experiments, we run ten-fold cross-validation method to evaluate the optimal parameters in SDC and SRC.

A. UCI data sets

We select ten popular data sets from the UCI dataset. Those data sets have not the missing attribute values. Table I depicts the detail information of those data sets, including the dimensionality of data set, number of class from each data set, number of training and testing samples, which are used in our experiments. In this experiment, we show the classification performance of SDC and NN. We randomly select about 60% samples from each class for training and the remaining samples for testing. We repeat these trials independently ten times and compute the average classification accuracy. Table II gives the best results of SDC and NN classifier on those data sets. Meanwhile, we display the standard deviation corresponding to the best classification accuracy. From Table II, we can see that the classification performance of SDC obviously outperforms NN classifier, and the improvements are very significant. For example, the classification accuracy of NN classifier and SDC on balance dataset are $61.8\%, 90.4\%$, respectively. The small standard deviation of SDC shows that SDC is more robust than NN classifier.

TABLE II. THE CLASSIFICATION ACCURACIES(%) AND THE CORRESPONDING STANDARD DEVIATIONS OF SDC AND NN CLASSIFIER ON THE UCI DATA SETS

Data name	*NN Classifier*	*SDC*
australian	61.8±0.28	**66.5**±0.26
banlance	70.3±0.28	**90.4**±0.07
cmc	44.7±0.26	**56.0**±0.11
glass	65.7±0.30	**73.2**±0.36
haberman	64.8±0.27	**75.6**±0.14
liver	57.0±0.33	**69.0**±0.15
pima	65.0±0.34	**75.3**±0.11
sonar	78.8±0.30	**85.6**±0.37
teaching	49.5±0.32	**56.8**±0.54
transfusion	70.6±0.26	**77.8**±0.15

B. Binary Alphadigits data sets

The Binary Alphadigits (BA) data set is collected from a binary 20×16 digits database of "0" through "9" and capital "A" through "Z". In this experiment, we display the classification performance of SDC on three data sets including handwritten digits, capital and BA database. For each subject, p images are randomly selected as the training set and the rest are considered as the testing set. For each given p, we average the results over ten random splits. Fig. 1, 2, 3 shows the classification performance of SDC and NN classifier versus the number of training samples. As shown in Fig. 1, 2, 3, the classification accuracy of SDC obviously outperforms NN classifier. That also validates the feasibility and effectiveness of our proposed classifier.

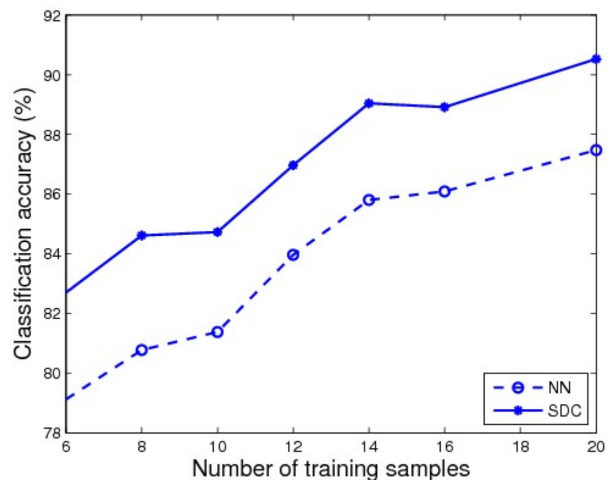

Figure 1. The classification accuracy of SDC and NN classifier versus the variation of number of training samples on Handwritten digits.

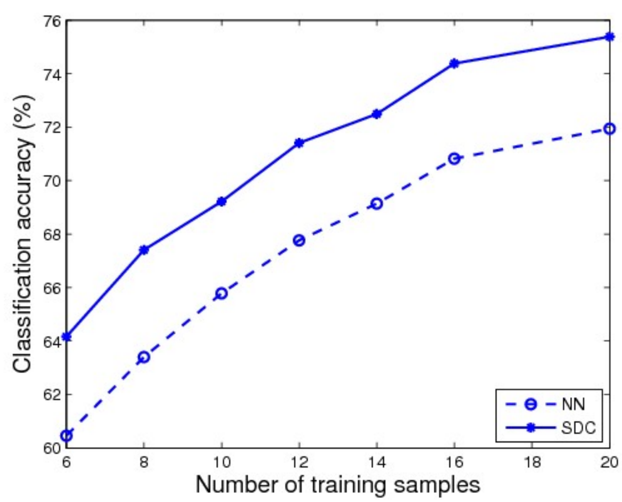

Figure 2. The classification accuracy of SDC and NN classifier versus the variation of number of training samples on Handwritten capitals.

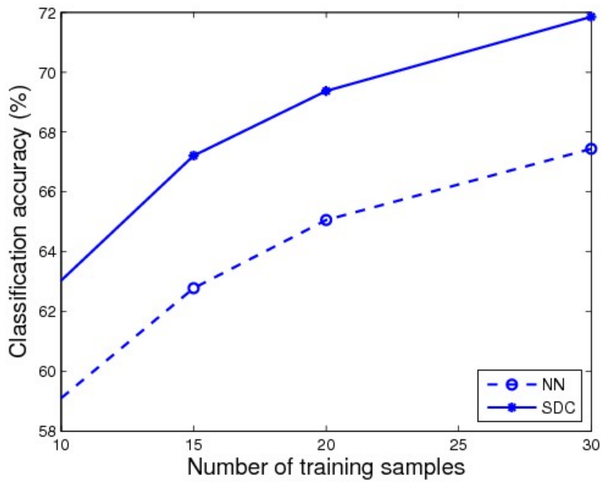

Figure 3. The classification accuracy of SDC and NN classifier versus the variation of number of training samples on Binary alphadigits.

C. IDA data sets

To further evaluate the performance of SDC, we employ the IDA dataset and compare SDC with other five methods, such as SRC and LRC. Table Ⅲ depicts the IDA dataset specifications which are standard binary classification data sets. For each object, we repeat these trials independently twenty or a hundred times and compute the average classification accuracy. Table Ⅳ presents an overall comparison of the six methods. From Table Ⅳ, we can see that SDC consistently outperforms other methods and the improvements are significant.

Lastly, Fig. 4 demonstrates the classification accuracy versus the variation of parameter $\lambda > 0$ on glass, digit 6 and banana data sets. From above experiments, the performances of the SDC based on classical softmax discriminant function are superior than other methods, such as NN classifier and LRC. That is to say, as a generalization of nearest neighbor and nearest subspace, SDC can well reflects the intrinsic structures of those data sets.

TABLE III. IDA DATA SPECIFICATIONS, WHICH ARE STANDARD BINARY CLASSIFICATION DATA SETS.

Data name	***Input data dimensionality***	***Number of training samples***	***Number of testing samples***	***Number of realizations***
banana	2	400	4900	100
breast	9	200	77	100
diabetes	8	468	300	100
flare	9	666	400	100
german	20	700	300	100
heart	13	170	100	100
splice	60	1000	2175	20
twonorm	20	400	7000	100
	21	1000	1000	100
waveform				

TABLE IV. AVERAGE CLASSIFICATION ACCURACIES (%) OF SX METHODS ON IDA DATA SETS, WHICH ARE STANDARD BINARY CLASSIFICATION DATA SETS.

Data name	***PCA***	***LDA***	***SRC***	***LRC***	***NN Classifier***	***SDC***
banana	86.4	61.7	56.8	47.1	86.4	**89.3**
breast	67.4	64.9	67.2	39.4	67.4	**76.8**
diabetes	69.9	68.7	64.8	47.0	69.9	**74.9**
flare	60.9	61.1	60.8	44.7	60.9	**66.3**
german	70.9	67.8	70.5	38.5	70.5	**75.1**
heart	78.0	76.9	70.5	47.8	76.8	**84.1**
splice	77.5	79.1	77.1	50.7	71.1	**81.7**
twonorm	96.5	96.5	50.4	50.0	93.3	**97.5**
waveform	88.2	81.4	61.8	41.7	84.2	**89.5**

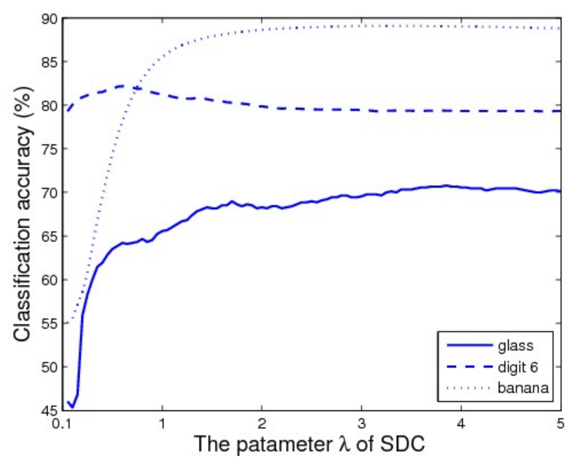

Figure 4. The classification accuracy of SDC versus the variation of parameter on three different data sets.

IV. CONCLUSION

We present a simple but effective classifier based on classical softmax discriminant function, called softmax discriminant classifier. Experimental results on some well-known data sets demonstrate the feasibility and effectiveness of our proposed classifier.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the anonymous referees as well as the Editor and Associate Editor for their valuable comments which lead to substantial improvements of the paper. This work was supported by the National Basic Research Program of China (973 Program) (Grant No. 2007CB311002), the National Natural Science Foundation of China (Grant Nos. 60675013, 61075006) and the Research Fund for the Doctoral Program of Higher Education of China (No. 20100201120048).

REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Recognition and Machine Learning, vol. 31, Feb. 2009, pp. 210-227.

[2] http://www.ics.uci.edu/mlearn/MLRepository.html, 1998.

[3] http://cs.nyu.edu/~roweis/data.html, 2008.

[4] http://ida.first.fraunhofer.de/projects/bench/benchmarks.html, 2001.

[5] I. T. Jolliffe, Principal component analysis, 2rd ed., New York, Springer, 2002.

[6] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2rd ed., Boston, Academic Press, 1990.