

CSE970 Homework #2

Implementation of Gaussian SVM

Chang Wang
Zhaohui Sun

March 25, 2003

Implementation issues

We implemented a non-linear classifier - Support Vector Machine using Gaussian kernel. The kernel function that we used is:

$$K(x, z) = \exp(- \|x - z\|^2 / c)$$

The update rule we used is the dual form of the large margin perceptron algorithm given in the lecture 4. We also used the pocket algorithm for training the classifiers. The training algorithm stops when the following ending algorithm is satisfied:

1. when the number of misclassification is smaller than certain value. (It is smaller than 5 in our experiments)
2. or when the iteration number reaches certain number.

For multiclass classification, we used ECOC. Four individual classifiers were used in our ECOC classifier.

Results of classifying 6 data sets

2-class classifiers

The 2-class classifiers were tested with data set 1 to 5. In each data set, the classifier was trained with 2 training sets and then tested with 2 corresponding test sets.

The error rate (error/total) of each test is reported:

data set	error rate of class 1	error rate of class 2	total error rate
data 1	3/3500 (0.086%)	7/1500 (0.46%)	10/5000 (0.2%)
data 2	2387/4000 (60%)	2/1000 (0.2%)	2389/5000 (47.8%)
data 3	57/2500 (2.3%)	60/2500 (2.4%)	117/5000 (2.3%)
data 4	0/3000 (0)	0/2000 (0)	0/5000 (0)
data 5	1/3000 (0.033%)	0/2000 (0)	1/5000 (0.02%)

$c = 1.0$, and the maximum numbers of passes on the training data = 500. This is generally the best results we got when we varied parameters. In next section, we also report other result we got with other setting of parameters.

4-class classifier

The 4-class classifier was tested with data set 6. The error rate (error/total) is reported:

class 1	class 2	class 3	class 4	total
840/2100 (40%)	297/350 (85%)	262/1050 (25%)	15/1500 (1%)	1414/5000 (28%)

$c = 1.0$ and the maximum number of passes on the training data = 3000

running time

We also report the time the program spent on training and testing for those 6 data sets:

data set	running time
data 1	69.97 s
data 2	637.99 s
data 3	146.60 s
data 4	10.70 s
data 5	32.47 s
data 6	2149.90 s

The maximum numbers of passes on training data are all 500 in data 1-6. The number of passes on training data is 1000 in data set 6.

The time spent in training is a very large portion of the total time, compared to the time spent in prediction.

Results on various parameter settings

We varied c value and the number of passes over training set when we ran our program. Different results were obtained.

100 passes, $c = 1.0$

data set	error rate of class 1	error rate of class 2	total error rate
data 1	3/3500 (0.086%)	7/1500 (0.46%)	10/5000 (0.2%)
data 2	3566/4000 (89%)	0/1000 (0.2%)	3566/5000 (71%)
data 3	57/2500 (2.3%)	60/2500 (2.4%)	117/5000 (2.3%)
data 4	0/3000 (0)	0/2000 (0)	0/5000 (0)
data 5	1/3000 (0.033%)	0/2000 (0)	1/5000 (0.02%)

250 passes, $c = 1.0$

data set	error rate of class 1	error rate of class 2	total error rate
data 1	3/3500 (0.086%)	7/1500 (0.46%)	10/5000 (0.2%)
data 2	3152/4000 (79%)	1/1000 (0.1%)	3153/5000 (63%)
data 3	57/2500 (2.3%)	60/2500 (2.4%)	117/5000 (2.3%)
data 4	0/3000 (0)	0/2000 (0)	0/5000 (0)
data 5	1/3000 (0.033%)	0/2000 (0)	1/5000 (0.02%)

20 passes, $c = 1.0$

data set	error rate of class 1	error rate of class 2	total error rate
data 1	11/3500 (0.31%)	5/1500 (0.33%)	16/5000 (0.32%)
data 3	962/2500 (38%)	13/2500 (0.52%)	975/5000 (19.5%)
data 4	0/3000 (0)	0/2000 (0)	0/5000 (0)
data 5	1/3000 (0.033%)	0/2000 (0)	1/5000 (0.02%)

500 passes, $c = 3.0$

data set	error rate of class 1	error rate of class 2	total error rate
data 1	2/3500 (0.06%)	11/1500 (0.73%)	13/5000 (0.26%)
data 2	3528/4000 (88%)	2/1000 (0.2%)	3530/5000 (70%)
data 3	88/2500 (3.5%)	41/2500 (1.6%)	129/5000 (2.6%)
data 4	1/3000 (0.03%)	0/2000 (0)	1/5000 (0.02%)
data 5	0/3000 (0)	22/2000 (1.1%)	22/5000 (0.44%)

500 passes, $c = 0.3$

data set	error rate of class 1	error rate of class 2	total error rate
data 1	0/3500 (0)	18/1500 (1.2%)	18/5000 (0.36%)
data 2	3062/4000 (77%)	8/1000 (0.8%)	3070/5000 (61%)
data 3	64/2500 (2.6%)	85/2500 (3.4%)	149/5000 (3.0%)
data 4	0/3000 (0)	0/2000 (0)	0/5000 (0)
data 5	1/3000 (0)	1/2000 (0.05%)	2/5000 (0.04%)

Testing 4-class classifier

Tested with data set 6: $c = 1.0$ and the maximum number of passes on the training data = 1000

class 1	class 2	class 3	class 4	total
1054/2100 (40%)	331/350 (94%)	504/1050 (48%)	39/1500 (2.6%)	1928/5000 (39%)