CSCI567 2014 Homework Assignment 3

Student Name Yi Duan

USC ID 9966845562

I have collaborated with Qiang Wang

Yun Li

1. Kernelized Perceptron
2. Set ***w0*** = **0**. Since ***w*** is updated by ***or***  where and map to , by induction, we can get where .

Where

1. Initialize . For each , if the prediction is incorrect, update by .
2. Support Vector Machine without Bias Term

Eliminate and we get

1. The difference between the dual form in (e) and the dual form of original SVM(with bias term) is that dual form of original SVM contain an additional condition , which means the dual form with bias term has more constraint than that without bias term.
3. Sample questions for Quiz #1
4. Regularization

The objective function is

And we should maximize the objective function.

1. Logistic Regression

,

Since and is strictly positive, is positive definite. Thus is convex, with only one global optimum.

1. Programming
   1. Cross validation for linear SVM
2. Experiment 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 |
| Accuracy(%) | 52.10 | 78.30 | 79.70 | 81.50 | 80.20 | 79.80 | 79.60 | 79.90 | 79.90 |
| Training time(s) | 0.7920 | 0.5640 | 0.6280 | 0.6780 | 0.7740 | 0.9520 | 0.7680 | 0.9400 | 0.9580 |

Experiment 2

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 |
| Accuracy(%) | 51.80 | 79.10 | 80.50 | 80.40 | 80.20 | 79.70 | 80.10 | 80.20 | 80.20 |
| Training time(s) | 0.7180 | 0.5440 | 0.5560 | 0.6200 | 0.7700 | 0.8240 | 0.7620 | 0.7780 | 0.8980 |

When C becomes larger, the cross validation accuracy increases, while the average training shows a trend to become longer. The smaller C is, the bigger the margin of data is and the simpler the model is, thus, the shorter the training times is while the lower the accuracy is. C is a tradeoff between accuracy and training time.

1. As we do cross validation with the random training data every time, the experiment shows that C with the best accuracy change a little bit (as showed in the table in (a)). Sometimes the best C is 4-3, sometimes it is 4-4. So we choose both 4-3 and 4-4 as the best C and compute test accuracy respectively.

|  |  |  |
| --- | --- | --- |
|  | -4 | -3 |
| Test accuracy(%) | 85.43 | 84.64 |

* 1. Use linear SVM in LIBSVM

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 |
| Accuracy(%) | 51.7 | 78.8 | 80.5 | 79.7 | 79.6 | 79.4 | 79.5 | 79.3 | 79.2 |
| Training time(s) | 0.1220 | 0.1040 | 0.0820 | 0.0680 | 0.0700 | 0.0920 | 0.2000 | 0.4580 | 1.5120 |

1. The cross validation accuracy is almost the same as that in 4.3.
2. LIBSVM is about 10 times faster than my implementation.
   1. Use kernel SVM in LIBSVM
3. Polynomial kernel

Accuracy(%)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Degree | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 51.7 | 51.7 | 78.7 | 80.4 | 79.5 | 79.2 | 79.5 | 79.4 | 79.3 | 79.2 | 79.3 | 78.8 |
| 2 | 51.7 | 51.7 | 51.7 | 65.4 | 72.8 | 70.9 | 70.9 | 70.9 | 70.9 | 70.9 | 70.9 | 70.9 |
| 3 | 51.7 | 51.7 | 51.7 | 62.1 | 79.2 | 79 | 79.2 | 79.2 | 79.2 | 19.2 | 79.2 | 79.2 |

Training time(s)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Degree | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 0.174 | 0.118 | 0.106 | 0.090 | 0.096 | 0.076 | 0.148 | 0.218 | 0.546 | 1.972 | 7.148 | 13.62 |
| 2 | 0.112 | 0.110 | 0.112 | 0.142 | 0.106 | 0.120 | 0.126 | 0.128 | 0.118 | 0.126 | 0.120 | 0.120 |
| 3 | 0.114 | 0.130 | 0.116 | 0.142 | 0.120 | 0.136 | 0.126 | 0.124 | 0.120 | 0.140 | 0.130 | 0.118 |

1. RBF kernel

Accuracy(%)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| -7 | 51.7 | 51.7 | 51.7 | 51.7 | 51.7 | 77 | 79.5 | 79.8 | 79 | 79.4 | 81.8 | 85.3 |
| -6 | 51.7 | 51.7 | 51.7 | 51.7 | 76.1 | 79.7 | 79.7 | 79.6 | 81.7 | 85.4 | 87.8 | 87.4 |
| -5 | 51.7 | 51.7 | 51.7 | 73.1 | 79.9 | 80 | 81.1 | 85.7 | 88.1 | 87.4 | 87.4 | 87.4 |
| -4 | 51.7 | 51.7 | 54.5 | 80.3 | 81.7 | 84.7 | 88 | 87.2 | 87.2 | 87.2 | 87.2 | 87.2 |
| -3 | 51.7 | 51.7 | 61.2 | 83.5 | 87.2 | 88.6 | 88.2 | 88.2 | 88.2 | 88.2 | 88.2 | 88.2 |
| -2 | 51.7 | 51.7 | 51.7 | 51.7 | 76.6 | 77.9 | 77.9 | 77.9 | 77.9 | 77.9 | 77.9 | 77.9 |
| -1 | 51.7 | 51.7 | 51.7 | 51.7 | 57.3 | 57.9 | 57.9 | 57.9 | 57.9 | 57.9 | 57.9 | 57.9 |

Training time(s)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| -7 | 0.160 | 0.110 | 0.108 | 0.112 | 0.118 | 0.118 | 0.098 | 0.078 | 0.074 | 0.076 | 0.112 | 0.242 |
| -6 | 0.116 | 0.108 | 0.104 | 0.112 | 0.104 | 0.092 | 0.072 | 0.066 | 0.076 | 0.114 | 0.250 | 0.316 |
| -5 | 0.108 | 0.108 | 0.106 | 0.104 | 0.092 | 0.076 | 0.066 | 0.074 | 0.112 | 0.146 | 0.144 | 0.150 |
| -4 | 0.102 | 0.106 | 0.120 | 0.090 | 0.074 | 0.072 | 0.074 | 0.104 | 0.102 | 0.118 | 0.106 | 0.102 |
| -3 | 0.106 | 0.110 | 0.110 | 0.098 | 0.084 | 0.106 | 0.108 | 0.106 | 0.116 | 0.104 | 0.112 | 0.104 |
| -2 | 0.122 | 0.112 | 0.110 | 0.110 | 0.118 | 0.120 | 0.122 | 0.124 | 0.122 | 0.118 | 0.128 | 0.120 |
| -1 | 0.108 | 0.112 | 0.108 | 0.110 | 0.132 | 0.124 | 0.122 | 0.116 | 0.114 | 0.118 | 0.116 | 0.124 |

Based on the result, we choose RBF kernel with C = 41 and gamma = 4-3, and we get the test accuracy is 90.4368%