SVM

By Weiyi He 2022.10.27

1.Theory of Supported Vector Machine

Considering soft margin, we use two method to solve the problem (6.35) instead of (6.6).

- SMO

  The SMO algorithm.

  Briefly, that is to choose a pair of Lagrange Multipliers each time and to renew them while fixing other parameters.

  In detail, first we select the variable that violates the KKT condition to the greatest extent, and then choose the second variable so that it can maximize the interval between the selected two sample.

- Gradient decreasing

  The gradient decreasing algorithm.

  Choosing hinge loss as the loss function, calculate the gradient of the loss function and renew the parameters and we can also get the separation hyperplane.

  What's more, by formula derivation, we can easily find that hinge loss is the traditional SVM in a geometric sense, thus gradient decreasing is also the way to solve (6.35) in book.

2.The result

The result of my two method are all contained in the following part.

3.The comparison of the methods

- The header:

  Method 1: SMO; Method 2: Gradient decreasing; Method 3: sklearn

  Dim: the dimension of the sample

  Num: the number of the sample

  Time: the training time (seconds)

  Accuracy: the test accuracy

  Iteration: the iteration times

- The comparison table is as follow:

Method	Dim	Num	mislabel	Time	Accuracy	Iteration
1	10	1000	0.0290	422.1244	0.7400	100

Method	Dim	Num	mislabel	Time	Accuracy	Iteration
2	10	1000	0.0290	44.6557	0.9440	1000
3	10	1000	0.0290	0.0998	0.9720	/
1	10	800	0.0400	274.8571	0.8400	100
2	10	800	0.0400	54.7152	0.9750	1000
3	10	800	0.0400	0.0898	0.9550	/
1	5	1000	0.0300	326.7760	0.8120	100
2	5	1000	0.0300	33.6579	0.9680	1000
3	5	1000	0.0300	0.0419	0.9640	/

- **Note**

    1. I do not choose the dimension and sample numbers suggested on gitee considering my CPU.
    2. I use hold-out method by setting 75% of the data set as the train set.
    3. For the SMO, theoretically it should be faster and more accurate, but my SMO run more slowly and it's not that accurate compared with gradient decreasing method. (Also it seems not stable) I write totally the same as the text book, and I also set breakpoints to track and modify the way I write the cycle but still don't know what is wrong...