**[Machine Learning] hw5** B01902040

**鍾毓安**

1. (b)

The primal formulation of the linear soft-margin SVM problem is :



Therefore, there are (d + 1 + N) variables and 2N constraints.

1. (c)



By plotting on the figure, we can easily find the optimal separating hyperplane(hard-margin), which is



1. (b), (d)

Therefore, there’re 5 nonzero , , and .



1. (b)

The kernel function does the following transformation : 

I manually transform the original X by and then apply linear kernel instead of polynomial kernel. After deriving the trained model, we can then calculate and report w and b. The result turns out to have very minor numerical deviation, but it’s still easy to get (b) as the solution.



1. (c)

They use different kinds of kernels, and since different kernels implies that the definition of margin are accordingly different, the curves should be different in the X space.

1. (d)

Optimizing is equivalent to optimizing because :



(i) for any violating :(ii) for any feasible :



1. (a), (c), (d)

KKT conditions :

* primal feasible :



* dual feasible :



* primal-inner optimal :



* dual-inner optimal :



Therefore, from dual-inner optimal, we know (a), (d) are correct ; from primal-inner optimal, we know (b) is wrong and (c) is correct.

1. (e)



1. (e)



1. (a)

By primal-inner optimal from KKT conditions, we can pick some  with  to compute  as follows :



1. (a)

Let ,

we now try to solve  by setting , here we derive  as follows :

Now we set , our task is to figure out . We hope that by solving the hard-margin problem involving , we can equivalently solve the original soft-margin problem. Therefore,

 shall be equivalent to the original  for n = 1~N, we can then derive  as follows :



 is correct .

1. (a), (c)

To check whether  is a valid kernel function , we can apply Mercer’s Condition :  is a valid kernel function iff any  such that

 is finite, then . So now let’s check (a)~(d) one by one.

Note that since  are two valid kernels and  for all , we have two conditions in advance, which are  and . Now we can start checking (a)~(d) one by one.



1. (b), (d)

With condition  in hand, let’s go through and check if (a)~(d) are valid kernel functions.

 , both A and B are positive

semi-definite and symmetric.

 is not a valid kernel.



1. (c)

Our goal is to ensure , analysis is as follows :

, where b can be calculated when chosen a free support vector  as , so let’s rewrite and simplify  as 

Now let’s turn to  :



Since we hope both of them lead to an equivalent  classifier, we know that the support vectors are identical, and suppose we choose the same free support vector to compute b, we have :



**[Bonus]**

21. No

The objective value of the soft-margin SVM primal problem is . The parameter C controls the trade-off of large margin and margin violation ; when we set a relatively small C, it means that we allow the SVM classifier can have more margin violation, but with larger margin. As long as for all bounded support vectors, their corresponding ’s are smaller than 1, then we have , which indicates that the data is still linearly separable.

Experiment :

Suppose we have dataset 

Let’s train it with linear soft-margin SVM and set parameter C to be very small. For example, when , all of the four Lagrange multipliers  turns out to be equal to . This indicates that no free support vectors exist, but it’s trivial to see that the dataset is still linearly separable, some simple classification algorithms like PLA can easily separate point with different labels.

22. Yes

By complementary slackness, for all the bounded support vectors, their corresponding , thus . We can then easily compute their  by , since all bounded support vectors satisfy , we have . Simplify the inequality we can derive , which implies that those bounded support vectors are not linearly separable.