**[Machine Learning] hw2** B01902040

**鍾毓安**

1. (d)

From the question descriptions, we know that , thus the probability of error that h makes in approximating the noisy target y equals to , which turns out to be .

2. (d)

To make the performance of  be independent of , the probability of error that h makes in approximating the noisy target y must has no difference with the situation when only  matters, thus we have, by solving this equation, we get .

3. (c)

With given , we hope that , so we just plot 5 candidate N’s in the options and observe the result.

N = 420,000, the bound is about *697.753626145*

N = 440,000, the bound is about *2.14484271996*

N = 460,000, the bound is about *0.0064581256611*

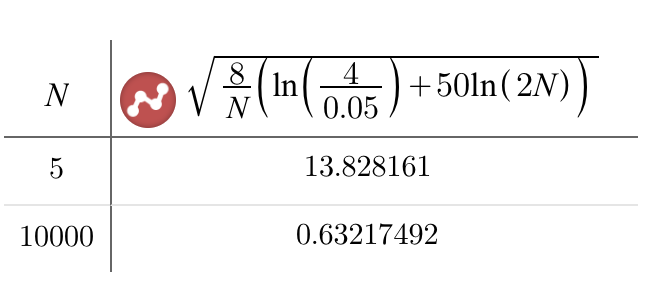
N = 480,000, the bound is about *1.90809518773e-05*

N = 500,000, the bound is about *5.54047747969e-08*

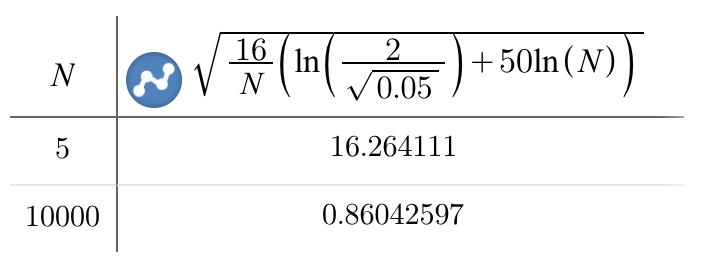
When N = 460,000, the VC bounds is closest to , which is 0.05.

4. (e)

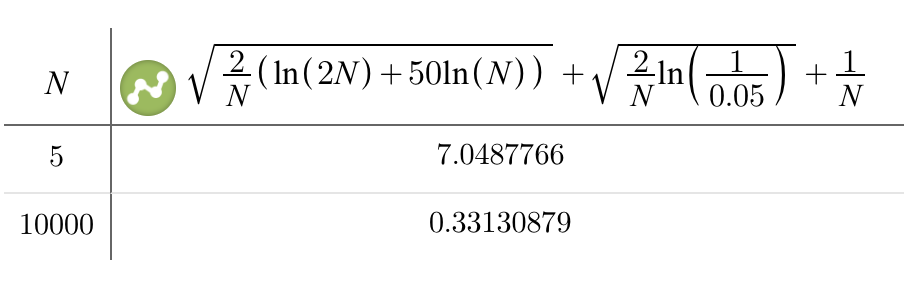
Original VC Bound :



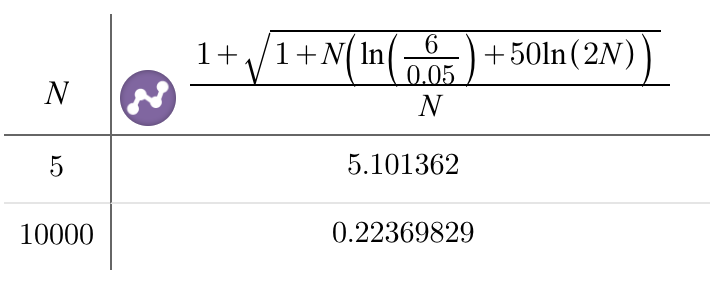
Variant VC Bound :



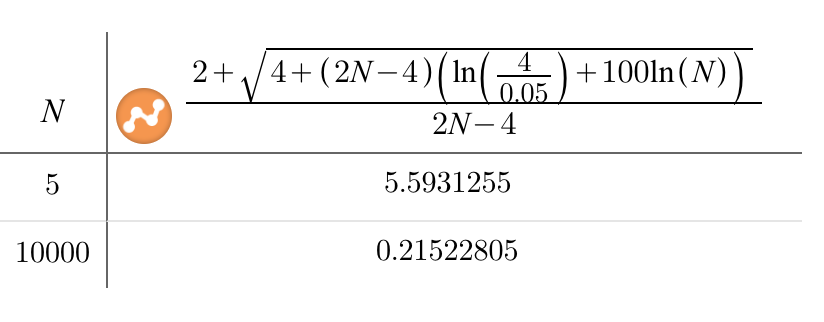
Rademacher Penalty Bound :



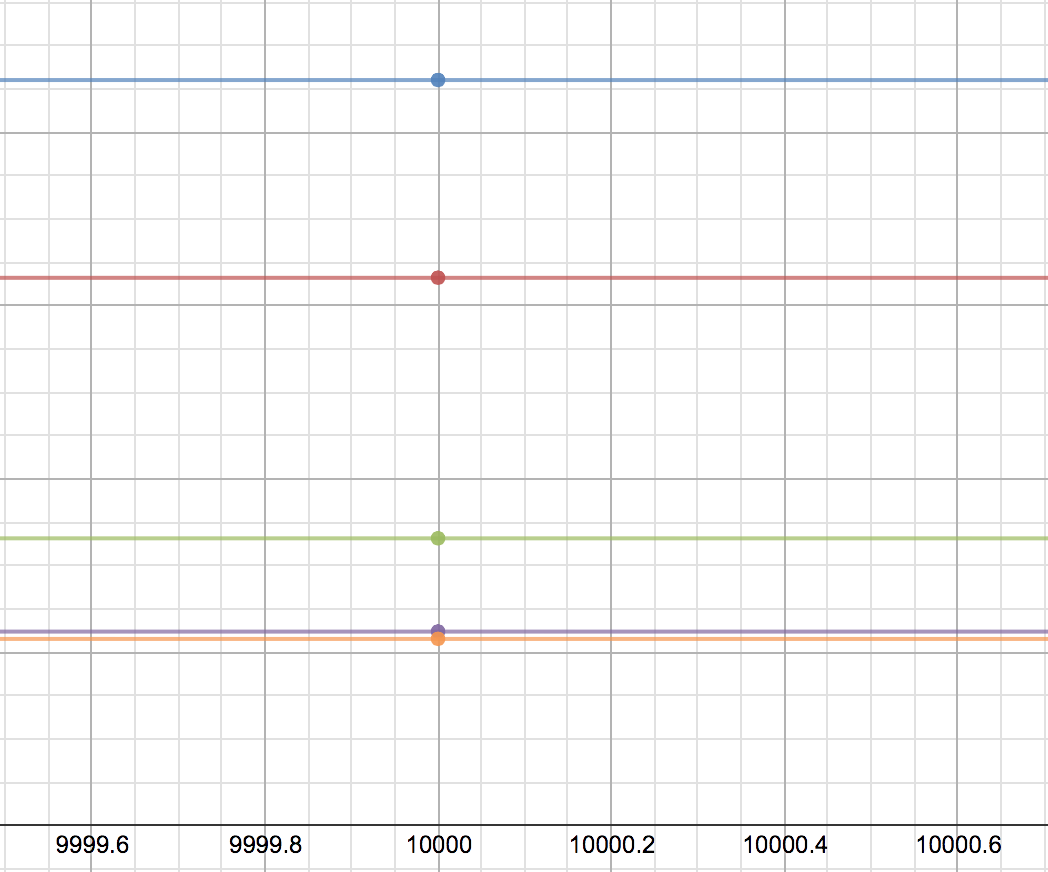
Parrondo and Van den Broek Bound :



Devroye Bound :

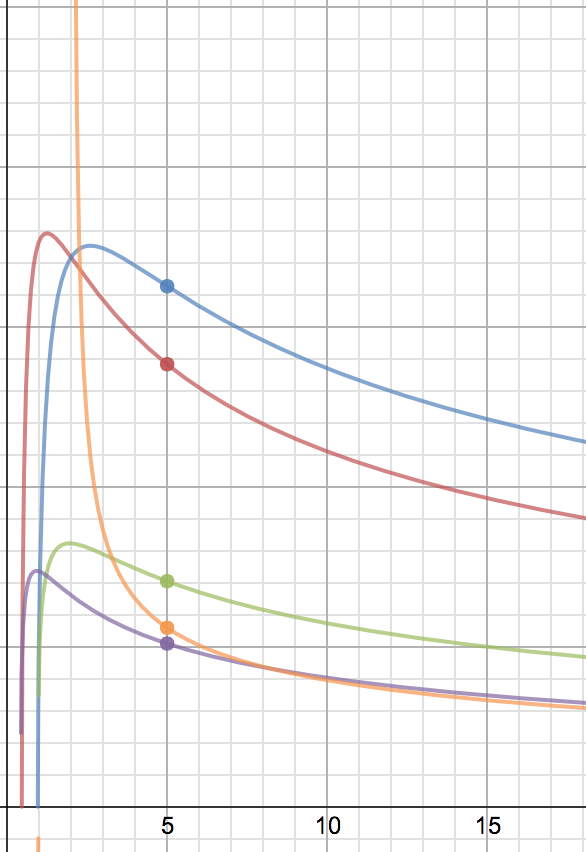


N = 10000



5. (d)

N = 5



6. (a)

We can’t simply conclude that , where  is the growth function of positive intervals in . The reason is that, say when , the positive intervals hypothesis would generate some dichotomies like oxxx, ooxx, ooox, oooo, xxxx, xxxo, xxoo, xooo, which the negative intervals hypothesis would also generate. So if we conclude that , some dichotomies will be repeating counted.

To derive the correct result, all dichotomies like ox should be counted only once, and the two dichotomies o…o、x…x will be added at the end. Therefore, .

7. (b)

N = 3, 

N = 4, 

k = 4 is the minimum break point, therefore .

8. (c)

The hypothesis “positive donuts in ” works just the same as the positive intervals hypothesis in . Therefore, there are at most  (add 1 for all negative ) dichotomies when there’re N inputs. .

9. (b)

A polynomial equation  has at most D roots, which means that there can be at most  intervals to place  inputs and shatter them all. Therefore, .

10. (a)

For any given threshold vector  of -dim, the -dim hyperspace can be partitioned into  sub-spaces. We can then place  inputs in every different sub-spaces with each input represents a different vector .

Then, by simply selecting which vectors to be put in , the hypothesis can shatter the  inputs. On the other hand, if we put more than  inputs in the hyperspace, since there are only  sub-spaces, at least two of the inputs will be place in the same sub-space, which belongs to the same vector. Therefore, this hypothesis can at most shatter  inputs, .

11. (d)

Our goal is to show that for any , there exists N points that can be shattered by this hypothesis. Let , for any of the  dichotomies , set the parameter , where  are defined as , .

By this setting, we have 



Since any dichotomy can be realized by , .

12. (a), (c), (d)

(a) When ; when ,  decreases a lot from

. The extent of decreasing becomes larger and lager, therefore,  bounds .

(b) is a constant once the hypothesis is decided

(c) Same reason with (a)

(d) 

(e) Here’s a counterexample : When , we have



 can’t be bound by 

13. (a)

(a) Ex. the convex sets taught in the lecture.

If , there must exist some break points for this hypothesis, and if the minimum break point  is found, the following inequality must hold : .

Therefore, for options (b), (c), (d), since they are all exponential functions, even with ceil & floor functions , they can never be bound by any polynomial functions. (b), (c), (d) are illegal growth functions.

As for option (e), the minimum break point , which means that the growth function should be bound by , but it’s obvious that cannot be bound by , thus (e) is also an illegal growth function.

14. (b)

By intersection operator, we obtain a new hypothesis that is less or equally powerless than the weakest hypothesis in .

Therefore, we have .

15. (e)

First, we have \_\_(1) since at least we can choose the most powerful hypothesis from  to find dichotomies.

Next, we try to prove that :

Consider there are only two hypothesis  and , and let ,  and  denotes the growth function of the union hypothesis, break point of , break point of , respectively. Then, we have the following analysis : (The ‘=’ in inequality  happens only when  and  are mutually exclusive and mutually exhaustive.)

When , .

Such result indicates that .

Now let’s turn back to the case of K hypothesis. By applying the above result to K hypothesis, we derive \_\_(2).

With (1) and (2), we have 

 is not tight enough. Consider there are two hypothesis 

the former one always answers yes while the latter one always answers no, therefore, both  and  equals to 0. But after the union operator, the union hypothesis can at most shatter N = 1, so its VC dimension is 1, which is larger than .

16. (c)





