

Homework 4

Qinyun Song

1 Feature Transformation

1. For the instances x , consider the function

$$f = \text{sgn}(\|x - c_1\|_1 - k/2) \quad (1)$$

For this function, the resulting instances is still in 1 dimension. And it is linearly separable. For instance that is no further than $k/2$ to c_1 , it will return true. Otherwise it will return false.

2. Consider the function

$$f = \text{sgn}(\min(\|x - c_1\|_1, \|x - c_2\|_1) - k/2) \quad (2)$$

For this function, it is linearly separable. For instance that is no further than $k/2$ to c_1 and to c_2 , the function will return true. Otherwise it will return false.

3. Consider three points $x_1 = 4, x_2 = 5, x_3 = 6$. For these three points, we can see that for point that is close to 5, the function should return true. For the points that has distance 1 from 5 should return false. This will result into the condition in the first problem which means that this problem is not linearly separable.

4. Consider the function

$$f = \begin{cases} \text{true} & \text{if } \text{abs}(\cos(x/3 \times 2\pi + \pi/2) \times \cos(x/5 \times 2\pi + \pi/2)) = 0 \\ \text{false} & \text{otherwise} \end{cases} \quad (3)$$

For those instances who can be divided by 3 or 5, one of the cos function will equal 0 thus the function will return true. Otherwise the function will return false. And it is clear that this function is linear separable.

2 PAC Learning

1. The size of the hypothesis space is

$$5^{M \times W \times S \times C}$$

2. The size of the hypothesis space is now changed into

$$|H| = 5^{4 \times 2 \times 3 \times 1} = 5^{24}$$

To satisfy the requirement, the number of lattes needed needs to be larger than

$$\frac{1}{0.05} \times (\ln(|5^{24}|) + \ln(\frac{1}{0.3})) = 796.61 \approx 797 \quad (4)$$

3. There are $2^{|X|}$ ways to pick some elements from X . So there are $2^{|X|}$ functions in H . Thus we can see that

$$\ln(|H|) = \ln(2^{|X|}) = |X| \ln 2 \quad (5)$$

This is polynomial in the size of X , so this is PAC learnable. To find the true function, one needs to enumerate all possible subset of X and see given what subset the function returns true. So the computation complexity is also $2^{|X|}$. So as shown above,

$$\ln(2^{|X|}) = |X| \ln 2$$

it is polynomial in $|X|$, thus this is efficient PAC learnable.

3 VC Dimension

1. The VC-dimension of the hypothesis class is 2.
 - (a) If there is only one point, we can use one axis-aligned rectangle to enclose the point. If the point is true, set the area inside the rectangle to true. Otherwise set it to false. In this case, it is shattered.
 - (b) If there are two points, there are two different situations. First, if they have the same label, use one axis-aligned rectangle to enclose the two points. And assign the area inside the rectangle as the label of the two points. Secondly, if the two points have different labels, let one axis-aligned rectangle to enclose the true label. Set the area inside it to true and outside to false. In this case, they are all well-shattered.
 - (c) If there are three points. No matter how we place three points in the plane, there can always be a way to label them such that the rectangle cannot shatter them.

Thus the VC Dimension of one axis-aligned rectangle is 2.

2. If there is the union of two rectangles, it now can shatter three points. For three points forming a triangle, no matter how we label them, there can always exist a way to use the union to shatter them. So the VC-dimension of the union must be no smaller than 3. So the VC-dimension of the union is greater than that of a single rectangle.
3. Consider the condition where there are three points lying in one line. Then no matter how we label them, there can always be a way to draw a difference of two axis-aligned rectangles such that it can cover all the true labels and uncover all the false labels. So the VC-dimension of it must be no less than 3. Which shows that its VC-dimension is also greater than that of a single rectangle.
4. *Proof.* Suppose there is a finite concept class C with VC dimension more than $\log c$. Suppose it equals k . Then for k points, there are 2^k ways to label them. Thus the size of the concept class should be $2^k > 2^{\log C}$. But the size of the concept class should not be larger than itself. Thus the size of the VC dimension should be at most $\log(|C|)$. \square

4 Properties of VC Dimension

1. *Proof.* For two hypothesis class H, G . If $H \in G$, all the functions in H are in G . So all the situation that can be shattered by H can also shattered by G . So the size of the VC dimension of G should be no smaller than H . So we have

$$VCdim(H) \leq VCdim(G) \quad (6)$$

□

2. The VC-dimension of the hypothesis space is $|X| - k$. Suppose there are l samples needed to be partitioned. The number of true label can vary from 0 to l . Since the function we have can only give value of 1 to exactly k items, we need to choose $k - l$ items in the remaining $|X| - l$ samples. So for the remaining part, it should contains no less than k items. Otherwise we cannot have enough $k - l$ items to pick in the remaining part thus cannot shatter the l samples. Thus we have

$$|X| - l \geq k$$

which results in the final equation:

$$l \leq |X| - k \quad (7)$$

So the VC-dimension of the hypothesis space is $|X| - k$.

3. In this case, first we still need

$$l \leq |X| - k$$

But at the same time, we also need that for the right side, $k - l \geq 0$. Because when there are l true labels among the examples, we need that for the remaining part, $k - l$ is still a valid number. So after all, the VC-Dimension should be

$$\min(|X| - k, k) \quad (8)$$