

1: PAC Learning**1.**

(a). As described in the question, we can consider that from an overall N parts, for each part present, we can keep that part or neglect the same. Thus it implies that for every part of the overall N parts, we have two cases. In terms of number of combinations we have 2^N combinations that can be made. But an important point to note here is that combination cannot be considered where any of the part is not selected. Thus it implies that the required hypothesis space is $2^N - 1$.

(b). According to the 2nd rule, we can now take 4 cases for each part from the overall total of N parts. We can describe the 4 cases as below:

a. Consider the particular part.

b. Don't consider the part.

c. Split the part. Now we have two parts. Consider the 1st split part and neglect the 2nd split.

d. Split the part. Now we have 2 parts again. Only consider the 2nd part in this case.

Here one more case is possible in which we take both the split parts after breaking a particular part, But this will not make logical connections as it will be the same case in which we didn't split that particular part and used it directly.

Thus, considering the above 4 cases, we can conclude that 4 cases for every part out of total N part is, $4^N - 1$, in terms of combinations. Thus, this is the required size of hypothesis space now.

(c). The required hypothesis size = $4^N - 1$.

Accordingly, here $N = 6$, thus the size of hypothesis class will become 4095.

Given : $\epsilon = 0.01$ and $\delta = 1 - 0.99 = 0.01$.

Here the Hypothesis space is containing the true concept class, so we are using the following equation:

Now using these values, in $m > \frac{1}{\epsilon} (\ln(|H|) + \ln(1/\delta))$, we get the required value as 1292.2

Thus, to conclude we can say that greater than or equal to 1293 examples will be needed by the robot in the given scenario.

2. Reference: Vivek's Office hour discussions on 1st November.

Our goal is to find the no. of examples enough to guarantee that every hypothesis $h \in H$ will be having,

$$err_D(h) < (1 + \epsilon)err_s(h)$$

We know by the definition of Chernoff Bounds:

$$err_D(h) = p, err_s(h) = S/m$$

So by using the above 2 equations:

$$p < (1 + \epsilon)S/m \text{ (Equation A)}$$

As we want to compute $err_D(h) < (1 + \epsilon)err_s(h)$, thus our task is to minimize the probability of $err_D(h) < (1 + \epsilon)err_s(h)$.

2nd inequality of Chernoff Bound will give us probability as:

$$p > ((1 + \epsilon)S/m).$$

Thus we can see that by using this inequality of Chernoff Bounds:

$$\frac{S}{m} < (1 - \gamma)p$$

$$p > \frac{\left(\frac{1}{1-\gamma}\right) * S}{m}$$

Thus by comparing the above equations $(1 + \epsilon)$ is replaceable by $\frac{1}{(1-\gamma)}$ as the given inequality is true for the 2 values. Thus we can conclude that:

$$\gamma = \frac{\epsilon}{(1+\epsilon)}$$

Thus we can see using the second equation above and equation A, we get

$$Prob\left(\frac{S}{m} < (1 - \gamma)p\right) \leq e^{-\frac{mp\gamma^2}{2}}$$

$$Prob[err_D > (1 + \epsilon)err_s(h)] \leq e^{-\frac{mp\gamma^2}{2}}$$

Thus for each hypothesis in H:

$$Prob[err_D(h) > (1 + \epsilon)err_s(h)] \leq |H| * e^{-\frac{mp\gamma^2}{2}}$$

For minimizing the above probability value or lets consider it less than a particular small δ :

$$|H| * e^{-\frac{mp\gamma^2}{2}} < \delta$$

$$\log(|H|) - \frac{mp\gamma^2}{2} < \log(\delta)$$

$$\log(|H|) - \log(|\delta|) < \frac{mp\gamma^2}{2}$$

$$\log(|H|) + \log(1/(|\delta|)) < \frac{mp\gamma^2}{2}$$

$$m > \frac{2}{p\gamma^2} * [\log(|H|) + \log(1/(|\delta|))]$$

Thus, replacing p with $err_D(h)$

$$m > \frac{2(\epsilon+1)^2}{\epsilon^2 * err_D(h)} (\ln|H|) + (\ln(1/\delta))$$

Thus, the above equation explains that m is sufficient to guarantee that every hypothesis $h \in H$ will be having error,

$$err_D(h) < (1 + \epsilon)err_s(h)$$

2: VC Dimensions

1. Let us assume that the VC Dimension of C is d that implies that $VC(C) = d$. We will be needing atleast 2^d examples to shatter the d examples as given. The above statement means that $2^d \leq |C|$ which further shows that $VC(C) \leq \log_2 |C|$.

2.

Reference Vivek's Office Hour Discussion

(a) Let us assume that the hypothesis class here is incorporating the hypotheses $h_1, h_2, h_3, \dots, h_n$. As given in the question, the no. of examples whose labels is one is perfectly k , and also given that $k \leq |\mathcal{X}|$. We already know by the definition that a particular mentioned hypothesis class contains all possible different arrangements of k examples, whose label is 1.

If we take any one particular point, then the hypothesis class can shatter it.

Similarly if we take any two points, the possible arrangements for those two points will be 00, 01, 10, 11.

A function must be present in the given hypothesis class such that it satisfies this labeling. Thus, we can say that these two points are also shatterable.

Thus, in a similar fashion we observe that K points are also shatterable, as a function must be existing in the hypothesis class for every arrangement of labeling possible.

Further extending this to $K+1$ points, a function won't exist (in hypothesis class) for labeling all one's.

Therefore the hypothesis class won't be able to shatter $K + 1$ points, and thus we can say that the VC Dimension of this particular class is K .

(b)

Reference Vivek's Office Hour Discussion

The VC Dimension for this class is minimum of $2k+1$ and total number of examples, i.e. X . This can be proved as, when we have $2k$ examples, there would be at max k 1s or k 0s, or atleast either if 1s or 0s will be less than k , if other one occurs more than k times. When we consider $2k+1$ examples, again there would be atmost k occurrences of either 1 or 0. For example, if $x=7$ and $k=3$, when considering 7 examples, it could be that 1s occur 4 times and 0 occurs 3 times. Either of the two will occur at most k times for $2k+1$ examples. This hence is shatterable by this class of functions. When we take $2k+2$ examples, for a combination where 1s occur $k+1$ times and 0s occur $k+1$ times, there will be no hypothesis in H which will be able to shatter it. Hence we say that VC dimension is less than $2k+2$ and greater than equal to $2k+1$.

Consider edge case where total examples are less than $2k+1$, our VC dimension in that case would be atmost X , i.e. atmost the total number of examples.

3. Here we can observe that the VC Dimension of \mathcal{H} is 4.

To arrive at the result these steps were followed:

We can tell that for a set consisting of 5 points, there can exist an arrangement like

$[+, -, +, -, +]$ which can't be shattered, and we can tell that according to the definition of VC dimension, the VC dimension is not five.

In other case if we take arrangements of 4 points, it is possible to shatter that by using the hypothesis space \mathcal{H} . Therefore, we can conclude that an instance space containing real numbers can be shattered using a hypothesis space \mathcal{H} , in which \mathcal{H} has two disjoint intervals, explained by $[a, b]$ and $[c, d]$, as given in the question.

4. Here we can observe that the VC Dimension of this particular class is 2. We can show the proof of the above statement as below: For any three points there must exist an arrangement where this hypothesis class is not able to shatter

Let us consider 2 equivalence classes:

(1) 3 points are placed in a collinear fashion.

(2) 3 points are placed in a non-collinear fashion.

Case 1: If the three points are collinear, there exist an arrangement $[+, -, +]$ such that the hypothesis class cannot shatter these labelings. Thus, in this particular case we can conclude that the equivalence class isn't shatterable.

Case 2: If the 3 points are not collinear then there will be an arrangement $[+, -, +]$ where '-' will fall in the +ve region and thus this particular class is also not shatterable.

Now a case where we consider two points. In this case the given hypothesis class can shatter all the arrangements. Thus, VC Dimension of this class is 2.

5. Let us assume, $VC(H_1) = d$. Thus we know that there exist d points that can be shattered by H_1 that implies that for each and every existing labelling of the d points, a hypothesis will exist $h \in H_1$ that satisfies that labelling.

Now, as we see H_2 has all the Hypothesis in H_1 , then we can say that H_2 shatters the similar set,

and therefore we can say the required proof $VC(H_2) \geq d = VC(H_1)$.

3: AdaBoost

The required values will be following:

Here h_a is used for round 1, h_b is used for round 2, h_d is used for round 3. Round 4 is having error more than $1/2$ so we stop computation here.

Values $h_a(\mathbf{x}) = \text{sgn}(\mathbf{x}_1)$, $\epsilon_1 = 1/4$, $\alpha_1 = \frac{\ln 3}{2}$, $\mathbf{Z}_1 = \frac{\sqrt{3}}{2}$

$x = [x_1, x_2]$	y_i	$h_a(x)$	D_1	$D_1(i)y_i h_i(x_i)$	D_2
[1,1]	-1	1	1/4	-1/4	1/2
[1,-1]	1	1	1/4	1/4	1/6
[-1,-1]	-1	-1	1/4	1/4	1/6
[-1,1]	-1	-1	1/4	1/4	1/6

Values $h_b(\mathbf{x}) = \text{sgn}(\mathbf{x}_1 - 2)$, $\epsilon_2 = 1/6$, $\alpha_2 = \frac{\ln 5}{2}$, $\mathbf{Z}_2 = \frac{\sqrt{5}}{3}$

$x = [x_1, x_2]$	y_i	$h_b(x)$	D_2	$D_2(i)y_i h_i(x_i)$	D_3
[1,1]	-1	-1	1/2	1/2	3/10
[1,-1]	1	-1	1/6	-1/6	1/2
[-1,-1]	-1	-1	1/6	1/6	1/10
[-1,1]	-1	-1	1/6	1/6	1/10

Values $h_d(\mathbf{x}) = -\text{sgn}(\mathbf{x}_2)$, $\epsilon_3 = 1/10$, $\alpha_3 = \frac{\ln 9}{2}$, $\mathbf{Z}_3 = \frac{3}{5}$

$x = [x_1, x_2]$	y_i	$h_d(x)$	D_3	$D_3(i)y_i h_i(x_i)$	D_4
[1,1]	-1	-1	3/10	3/10	1/6
[1,-1]	1	1	1/2	1/2	5/18
[-1,-1]	-1	1	1/10	-1/10	1/2
[-1,1]	-1	-1	1/10	1/10	1/18

So now we can observe that $h_c(\mathbf{x})$ is leftover, we can also observe that error we got is greater than $1/2$.

Thus we will stop computations here.

(Reference: Vivek's office hour discussions on November 1st)

Therefore, $H_{final}(\mathbf{x}) = \text{sgn}(\alpha_1 h_a + \alpha_2 h_b + \alpha_3 h_d)$

$$\text{that implies} = \text{sgn}\left(\frac{\ln 3}{2} \times \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix} + \frac{\ln 5}{2} \times \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} + \frac{\ln 9}{2} \times \begin{bmatrix} -1 \\ +1 \\ +1 \\ -1 \end{bmatrix}\right)$$

Finally after solving the above we get,
[−1 + 1 − 1 − 1].