mld hw4

yasir Mohamed

September 2024

1 Exercise 2.4

1.1 A

Let's assume X is the following matrix:

$$\begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 1 & 1 & 0 & ... & 0 \\ . & . & . & ... & . \\ . & . & . & ... & . \\ . & . & . & ... & . \\ 1 & 0 & 0 & ... & 1 \end{bmatrix}$$

X is a non-singular matrix of (d+1)*(d+1) and is invertible therefore we can compute X^{-1} using the equation $y = w^T X$

$$w^T = yx^{-1}$$

$$w = y^T (X^{-1})^T$$

And since X's determinate is not zero and is invertible then for any y value representing any dichotomy. then there exists a solution w.therefore d+1 is shattered.

$$d_{VC} \geq d + 1$$

1.2 B

Lets assume X is the following (d+2)*(d+1): $\begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 1 & 1 & 0 & ... & 0 \\ . & . & . & ... & . \\ . & . & . & ... & . \\ . & . & . & ... & . \\ 1 & 0 & 0 & ... & 1 \\ 1 & ? & ? & ... & ? \end{bmatrix}$

this is the same as part a but with one more entry, x_{d+2}, that has n of its bits being 1s, with $n \geq 3$. As such as it can be represented as $x_{d+2} = x_{n_1} + x_{n_2} ... + x_{n_n} - (n-1) * x_1$ with x_i being the location first bit that is one in X_{d+2}. We then subtract $(n-1)*x_1$ since each addition from before adds $+1$ to the first bit yet the first bit should $= 1$.

As such the x_{d+2} data point value will fully depend on the value of the n points that are combined to produce it. Therefore we will have less than 2^N dichotomies so it is not shattered for $d_{VC} > d+1$ where i and j are the integers indicating the bits that are ones.

Since

$$d + 1 \leq d_{VC} \leq d + 1$$

then

$$d_{VC} = d + 1$$

2 Problem 2.3

2.1 A

For the case of N = 1:

α could either be less than or greater than the point and since $2^1 = 2$ then this case is shattered

$$m_H(1) = 2 = 2^1 = 2$$

for N = 2:

given two points A and B with A ¡ B then α can either be $a < \alpha < b, \alpha < a < b, a < b < \alpha$

the two edge cases where $\alpha < a < b$ and $a < b < \alpha$ will both provide the same two different dichotomies (+1={a,b} or -1 = {a,b}) therefore they both add a total of two dichotomies, however the center case $a < \alpha < b$ provides two dichotomies on its own, either (+1 =
a} -1 = {b}) when using the case of +1 on $(\infty, \alpha]$or (+1={b} -1={a}) when using the case of +1 on $[\alpha, \infty)$

so three dichotomies

$$m_H(2) = 4 <= 2^2 = 4$$

for N = N + 1: when adding an extra point, n, it can be seen as if adding a point right after the first point(since we started with A as our first point regardless of how many points we add A will always be first). In this case, all the locations in which α can be will provide the same amount of dichotomies except that there is a new possible location for $A\alpha$ which is $A < \alpha < n$ similar to the other center point it can either make all the points to the left of $alpha$ +1 or -1 meaning it adds two more dichotomies there for

$$m_H(N + 1) = m_H(N) + 2$$

since $m_H(1) = 2$ and $m_H(2) = 4$ and we grow by adding two then

$$m_H(N) = 2N$$

$$m_H(3) = 6 < 2^3 = 8$$

$$d_{VC} = 2$$

2.2 B

given N points:

if we start the interval from the first data point, $N_1 < N_2 < N3.... < N_n$ then we can decide the number of points included in the interval to be any integer in the range 1 to N giving us N dichotomies.

if we start at the second choice then we can contain any number of points in the range of 1 to N-1 giving us N - 1 dichotomies if we start at the n point then we have 1 - N-n dichotomies. we can rewrite this as dichotomies =

$$\sum_{i=1}^{N} i$$

For all the cases we can reverse the classification of all the points(everything that was a +1 is -1 and vice versa) meaning that they provide twice as many dichotomies

$$2\sum_{i=1}^{N} i$$

One problem with this is that for any selection starting with the first point $[n_1 - n_x]$ the possibilities are either all the values of $[n_1 - n_x]$, for all $n_x! = n_N$ are -1 with the rest being +1 or vice versa however, if we look at the vector starting after the last point

$$n_{x+1} - n_N$$

then the two possibilities generated by this are the same therefore we should ignore all the possibilities that are generated from the first point except the possibility in which all the points are included, $[n_1 - n_N]$ which generate two dichotomies hence we have

$$m_H(N) = 2 + 2\sum_{i=1}^{N-1} i = 2 + 2(\frac{n(n+1)}{2} - n) = n^2 - n + 2$$

for

$$m_H(4) = 16 - 4 + 2 = 10 < 2^4$$
$$d_{VC} = 3$$

2.3 C

regardless of the dimension d, when they are evaluated they are all assigned some inter value r through the formula, $\sqrt{x_1^2 + ... + x_d^2}$ therefore we can treat each point as some int value from $[0.\infty]$ and as such this become the same as the positive interval problem. From part B we excluded the combination including the first point as they could be replication in the combinations derived from other points by investing the sign however this is not possible in this case and as such we can represent all combinations as

$$sum_{i=0}^{N} i$$

Then we need to account for the edge case where no points are within the interval so all points are -1

$$m_H(N) = 1 + \sum_{i=0}^{N} i = \frac{n^2}{2} + \frac{n}{2} + 1$$

$$
\begin{aligned}
m_H(1) &= 2 = 2^1 \\
m_H(2) &= 4 = 2^2 \\
m_H(3) &= 7 < 2^3
\end{aligned}
$$

therefore

$$d_{VC} = 2$$

3 Problem 2.8

3.1 N+1

Here the $d_{VC} = 1$ and for all N the formula $N + 1 \leq \sum_{i=0}^{1} \binom{N}{i}$ is true therefore it is possible

3.2 $1 + N + \frac{N(N-1)}{2}$

$$1 + N + \frac{N(N-1)}{2} = \frac{N^2}{2} + \frac{N}{2} + 1$$
$$m_H 1 = 2 = 2^1$$
$$m_H 2 = 4 = 2^2$$
$$m_H 3 = 7 < 2^1$$
$$d_{VC} = 2$$

3.3 2^N

for all N $m_H(N) = 2^N$ and $d_{VC} = \infty$ therefore it is possible

3.4 $2^{\lfloor \sqrt{N} \rfloor}$

$m_H(N) = 2^{\lfloor \sqrt{N} \rfloor} < 2^N$ should have $d_{VC} = 1$
and

$$m_H(N) \leq \sum_{i=0}^{d_{VC}} \binom{N}{i}$$

however this is not true for N = 5:

$$m_H(25) = 2^{\lfloor \sqrt{25} \rfloor} = 2^5 = 32 > \sum_{i=0}^{1} \binom{N}{i} = 1 + 25 = 26$$

therefore it is not a possible growth function

3.5 $2^{\lfloor \frac{N}{2} \rfloor}$

the $d_{VC} = 0$ as such it must be bounded by

$$\sum_{i=0}^{0} \binom{N}{i}$$

but for N = 2 $m_H(2) = 2^{\lfloor \frac{2}{2} \rfloor} = 2^1 = 2 > \sum_{i=0}^{0} \binom{N}{i} = 1$ therefore it is not a possible growth function

3.6 $1 + N + \frac{N(N-1)(N-2)}{6}$

$$1+N+\frac{N(N-1)(N-2)}{6} = 1+N+\frac{N^3-2N^2-N+2}{6} = \frac{N^3-2N^2+5N+8}{6}$$

$$m_H(1) = 2 = 2^1$$

$$m_H(2) = 3 < 2^2$$

$$d_{VC} = 1$$

as such it must be bounded by

$$\sum_{i=0}^{1} \binom{N}{i}$$

however for n = 3 we get

$$m_H(3) = 5 > \sum_{i=0}^{1} \binom{3}{i} = 4$$

therefore it is not possible

4 Problem 2.10

We have a few cases: First if $d_{VC} = \infty$ or$d_{VC} \geq 2N$:

$$m_H(2N) = 2^{2N} = (2^N)^2$$

$$m_H(N)^2 = (2^N)^2 = m_H(2N)$$

The case where $d_{VC} \leq N$: Here according to the bound theorem:

$$m_H(N) \leq \sum_{i=0}^{d_{VC}} \binom{N}{i} \leq N^{d_{VC}} + 1$$

Hence:

$$m_H(2N) \leq (2N)^{d_{VC}} + 1$$

$$m_H(N)^2 \leq (N^{d_{VC}} + 1)^2 = N^{2d_{VC}} + 2N^{d_{VC}} + 1 > (2N)^{d_{VC}} + 1$$

The third case if if $N < d_{VC} < 2N$

$$m_H(N) = 2^N$$

$$m_H(N)^2 = 2^{2N}$$

by definition of the d_{VC}

$$m_H(2N) < 2^{2N}$$

then

$$m_H(N)^2 = 2^{2N} \geq m_H(2N) \leq 2^{2N}$$

the generalization bound:

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln \frac{4(m_H(N))^2}{\delta}}$$

5 Problem 2.12

We use te equation

$$N \geq \frac{8}{\epsilon^2} ln(\frac{4m_H(2N)}{\delta}$$

$$N \geq \frac{8}{\epsilon^2} ln(\frac{4((2N)^{d_{VC}} + 1)}{\delta})$$

From the question we have $\epsilon = 0.05, d_{VC} = 10, \delta = 0.05$

$$N \geq \frac{8}{0.05^2} ln(\frac{4((2N)^{10} + 1)}{0.05})$$

now selecting an n value and plugging it in the equation then repeating it multiple times we find that N converges to ≈ 452956.86 see code segment 2.12 iterative.py