Homework 4

Runmin Lu

September 25, 2021

Exercise 2.4

(a)

Construct $d+1$ points as follows and put all of them into columns of a matrix called X where X has all 1's on the diagonal and the first row. The rest of the entries are 0's. It looks like

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

Clearly it's upper triangular so it must be singular, which means that it spans the entire $\mathbb{R}^{d+1}$. Therefore $\mathbf{w}X$ can span $\mathbb{R}^{d+1}$, so each entry can take positive or negative values, and hereby implements all 2^{d+1} dichotomies.

(b)

As the hint indicates, the span of any $d+1$ by $d+2$ matrix is at most $d+1$, which means that one entry of $\mathbf{w}X$ (which has dimension $d+2$) is linearly dependent on all other $d+1$ entries. For every combination of those $d+1$ entries, the one dependent entry must only take one value. Therefore not all 2^{d+2} dichotomies can be implemented.

Problem 2.3

(a)

N data points split the number line into $N-1$ inner intervals and 2 outer intervals.

For each of the $N-1$ intervals we have 2 dichotomies by shooting the ray in 2 directions.

For each of the 2 intervals, we have 2 dichotomies: all +1's and all −1's.

Therefore $m_{\mathcal{H}}(N) = 2(N-1) + 2 = \boxed{2N}$

$d_{VC} = \boxed{2}$ because for 3 data points we can't have +1's on two ends and −1's in the middle.

2 data points can be shattered because $2(2) = 2^2$

(b)

N data points split the number line into $N - 1$ inner intervals and 2 outer intervals.

The boundaries are either in the same interval or different.

If they're in the same interval, then there are only 2 dichotomies: all +1's or all −1's.

Otherwise, if the boundaries are in the $N-1$ intervals, then we have $2\binom{N-1}{2}$ dichotomies, where each section contains 1 or more +1's or −1's.

If we fix one boundary in an outer interval, then there are $2(N-1)$ dichotomies with 1 or more +1's followed by 1 or more −1's or the other way around.

In total,

$$\begin{aligned} m_{\mathcal{H}}(N) &= 2 + 2\binom{N-1}{2} + 2(N-1) \\ &= 2 + 2\frac{(N-1)(N-2)}{2} + 2N - 2 \\ &= (N-1)(N-2) + 2N \\ &= \boxed{N^2 - N + 2} \end{aligned}$$

$d_{VC} = \boxed{3}$ because we can't have 4 data points ordered as $-1, +1, -1, +1$ but $N = 3$ is shattered because $2^3 = 8 = 3^2 - 3 + 2$

(c)

Because the only property that matters to a data point is its distance to the origin, such $\mathcal{H}$ in $\mathbb{R}^d$ is simply equivalent to a positive interval in $\mathbb{R}^+$. Therefore, $m_{\mathcal{H}}(N) = \boxed{\frac{1}{2}N^2 + \frac{1}{2}N + 1}$ and $d_{VC} = \boxed{2}$ as shown in the example from last homework.

Problem 2.8

$1 + N$ is possible because it's for Positive Rays.

$1 + N + \frac{N(N-1)}{2}$ is possible because it's for Positive Intervals.

2^N is possible because it's for Convex Sets.

$2^{\lfloor\sqrt{N}\rfloor}$ is not possible.

Proof.

Assume for contradiction that it is possible. Then

$$\begin{aligned}
m_{\mathcal{H}}(1) &= 2^1 \\
m_{\mathcal{H}}(2) &= 2^1 < 2^2 \\
d_{VC} &= 1 \\
m_{\mathcal{H}}(N) &\leq N+1 \\
\text{However } m_{\mathcal{H}}(25) &= 2^5 = 32 > 25+1 \\
&\perp
\end{aligned}$$

□

$2^{\lfloor N/2\rfloor}$ is not possible. Proof is similar to the one above: $d_{VC} = 1$ but $m_{\mathcal{H}}(6) = 8 > 6+1$.

$1 + N + \frac{N(N-1)(N-2)}{6}$ is not possible. Proof is again similar: $d_{VC} = 1$ but $m_{\mathcal{H}}(3) = 5 > 3+1$.

Problem 2.10

Any $2N$ data points can be split into 2 sets of N data points, each of whose maximum number of dichotomies is $m_{\mathcal{H}}(N)$. Since $m_{\mathcal{H}}(2N)$ represents the maximum number of dichotomies of a combination of these 2 sets of N data points, it therefore can be at most $m_{\mathcal{H}}(N)^2$.

Problem 2.12

$$
\begin{aligned}
\varepsilon &= \sqrt{\frac{8}{N} \ln \frac{4(2N)^{d_{VC}}}{\delta}} \\
\varepsilon^2 &= \frac{8}{N} \ln \frac{4(2N)^{d_{VC}}}{\delta} \\
N &= \frac{8}{\varepsilon^2} \ln \frac{4(2N)^{d_{VC}}}{\delta} \\
&= \frac{8}{\varepsilon^2} d_{VC} \ln(2N) + \frac{8}{\varepsilon^2} \ln \frac{4}{\delta} \\
\text{Let } \ln(2N) &\approx 10 \\
N &\approx \frac{8}{0.05^2} 10(10) + \frac{8}{0.05^2} \ln \frac{4}{0.05} \\
&\approx \boxed{3.34 \times 10^5}
\end{aligned}
$$