Homework 4

Karan Sarkar
sarkak2@rpi.edu

October 1, 2019

Exercise 2.4.

(a) We will show that there exists $d+1$ points that can be shattered. Suppose we arrange these points into the $(d+1) \times (d+1)$ matrix $\mathbf{X}$ such that each row corresponds to a data point. We will choose $\mathbf{X}$ to be a lower-diagonal matrix filled with ones. Thus, all entries below and on the diagonal are 1 and all else is 0. Note that this is invertible because the rows are all linearly independent.

Let the weight vector of the perceptron be $\mathbf{w}$. Now let the classification results vector be $\mathbf{y}$ where $y_i = \mathbf{x_i} \cdot \mathbf{w}$. Thus, we have $\mathbf{Xw} = \mathbf{y}$. Now to shatter the points in $\mathbf{X}$, we need to show that for every possible vector $\mathbf{y} \in \{-1, 1\}^{d+1}$, there exists $\mathbf{w}$ such that $\mathbf{Xw} = \mathbf{y}$. Because $\mathbf{w} = \mathbf{X}^{-1}\mathbf{y}$, we can shatter the points if $\mathbf{X}$ is invertible. Because $\mathbf{X}$ was constructed to be non-singular, $d+1$ points can be shattered.

(b) Assume for the sake of contradiction that $d+2$ points can be shattered. Let those points be $\mathbf{x_i}$ where $1 \leq i \leq d+2$. Because these points are $d+1$ dimensional, they are linearly dependent. Thus, $\mathbf{x_{d+2}} = \alpha_1 \mathbf{x}_i + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_{d+1}\mathbf{x}_{d+1}$ where α_i is a constant for $1 \leq i \leq d+1$. Let the weight vector of the perceptron be $\mathbf{w}$. Now let the classification results vector be $\mathbf{y}$ where $y_i = \mathbf{x_i} \cdot \mathbf{w}$. It follows that:

$$\begin{aligned}
\mathbf{x_{d+2}} &= \alpha_1 \mathbf{x}_i + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_{d+1}\mathbf{x}_{d+1} \\
\mathbf{w} \cdot \mathbf{x_{d+2}} &= \alpha_1 \mathbf{w} \cdot \mathbf{x}_i + \alpha_2 \mathbf{w} \cdot \mathbf{x}_2 + \cdots + \alpha_{d+1}\mathbf{w} \cdot \mathbf{x}_{d+1} \\
y_{d+2} &= \alpha_1 y_i + \alpha_2 y_2 + \cdots + \alpha_{d+1} y_{d+1}
\end{aligned}$$

Because we assumed that $d+2$ can be shattered, given our $\mathbf{x}_i$ and α_i constants, the sign vector $\text{sign}(\mathbf{y})$ can equal any element of $\{-1, 1\}^{d+2}$. Suppose we choose that $\text{sign}(y_i) = \text{sign}(\alpha_i)$ for all $1 \leq i \leq d+1$. Therefore, $\alpha_i y_i > 0$. Thus $\text{sign}(y_{d+2}) = +1$. This means the dichotomy $(\text{sign}(\alpha_1), \text{sign}(\alpha_2), \ldots, \text{sign}(\alpha_{d+1}), +1)$ can be implemented. However, the dichotomy, $(\text{sign}(\alpha_1), \text{sign}(\alpha_2), \ldots, \text{sign}(\alpha_{d+1}), -1)$ cannot be implemented. Because not all dichotomies can be implemented, for $N \geq d+2, m_{\mathcal{H}}(N) > 2^N$. Thus, $d+2$ points cannot be shattered.

Problem 2.3.

(a) Note that the positive ray can implement $N+1$ dichotomies. This reflects the fact that there are $N+1$ intervals in which to place the ray's origin point. Thus, the negative ray can also implement $N+1$ dichotomies. Note that the all-positive and all-negative are the only two dichotomies that can be implemented by either positive or negative ray. Thus, a positive or negative ray has growth function $m_{\mathcal{H}}(N) = N+1+N+1-2 = \boxed{2N}$. The largest value of N for which $m_{\mathcal{H}}(N) = 2^N$ is 2 as $m_{\mathcal{H}}(3) = 6$. Thus, $\boxed{d_{VC} = 2}$.

(b) We will first count the growth function for a positive interval. There are $N+1$ intervals to put the endpoints of the interval. If the endpoints are chosen from different intervals, we have $\binom{N+1}{2}$ dichotomies implemented. If both endpoints are from the same interval, we have 1 additional dichotomy, the all negative dichotomy. Thus, the growth function for a positive interval is $\binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$.

We now want to find the growth function of a positive or negative interval. Note that both implement $\frac{1}{2}N^2 + \frac{1}{2}N + 1$ dichotomies individually. Now consider the number of dichotomies that can be implemented by either a positive or negative interval. A dichotomy can be implemented by both if there is only one sign switch i.e. positive points followed by negative or negative points followed by positive. The number of dichotomies with only one sign switch is the number of dichotomies that can be implemented by a positve or negative ray $2N$. Thus, $m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1 + \frac{1}{2}N^2 + \frac{1}{2}N + 1 - 2N = \boxed{N^2 - N + 2}$. The largest value of N for which $m_{\mathcal{H}}(N) = 2^N$ is 3 as $m_{\mathcal{H}}(4) = 14$. Thus, $\boxed{d_{VC} = 3}$.

(c) In order to determine the growth function for concentric circles we use the transform $\phi : \mathbb{R}^d \to \mathbb{R}$. We define $\phi(x_1, x_2, \dots x_d) = \sqrt{x_1^2 + x_2^2 + \cdots x_d^2}$. Thus, the problem of concentric circles in $\mathbb{R}$ is equivalent to positive intervals in $\mathbb{R}$. Thus, from (b), we have $\boxed{m_{\mathcal{H}}(N) = \dfrac{1}{2}N^2 + \dfrac{1}{2}N + 1}$. The largest value of N for which $m_{\mathcal{H}}(N) = 2^N$ is 2 as $m_{\mathcal{H}}(3) = 7$. Thus, $\boxed{d_{VC} = 2}$.

Problem 2.8. A growth function $m_{\mathcal{H}}(N)$ has two possibilies. First, $d_{VC} = \infty$ and $m_{\mathcal{H}}(N) = 2^N$ for all N or d_{VC} is finite and $m_{\mathcal{H}}(N) \leq N^{d_{VC}} + 1$. Note that d_{VC} is the largest value such that $m_{\mathcal{H}}(N) = 2^N$.

1. When $m_{\mathcal{H}}(N) = 1 + N$, we have $d_{VC} = 1$ as $m_{\mathcal{H}}(2) = 3$. Thus, it must be bounded by $N + 1$ for all N which is true. Thus, $\boxed{m_{\mathcal{H}}(N) = N + 1 \text{ is a growth function}}$

2. When $m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)}{2}$, we have $d_{VC} = 2$ as $m_{\mathcal{H}}(3) = 7$. Thus, it must be bounded by $N^2 + 1$ for all N which is true. Thus, $\boxed{m_{\mathcal{H}}(N) = 1 + N + \dfrac{N(N-1)}{2} \text{ is a growth function}}$

3. $\boxed{m_{\mathcal{H}}(N) = 2^N \text{ is clearly a growth function}}$ as $d_{VC} = \infty$.

4. When $m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor}$, we have $d_{VC} = 1$ as $m_{\mathcal{H}}(2) = 2$. Thus, it must be bounded by $N + 1$ for all N which is false. When $N = 6$, $36 + 1 < 2^6$. Thus, $\boxed{m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor} \text{ is not a growth function}}$

5. When $m_{\mathcal{H}}(N) = 2^{\lfloor N/2 \rfloor}$, we have $d_{VC} = 0$ as $m_{\mathcal{H}}(1) = 1$. Thus, it must be bounded by $N^0 + 1 = 2$ for all N which is false. When $N = 6$, $2 < 2^3$. Thus, $\boxed{m_{\mathcal{H}}(N) = 2^{\lfloor N/2 \rfloor} \text{ is not a growth function}}$

6. When $m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)(N-2)}{6}$, we have $d_{VC} = 1$ as $m_{\mathcal{H}}(2) = 3$. Thus, it must be bounded by $N + 1$ for all N which is false. When $N = 4$, $m_{\mathcal{H}}(4) = 9 < 5$. Thus, $\boxed{m_{\mathcal{H}}(N) = 1 + N + \dfrac{N(N-1)(N-2)}{6} \text{ is not a growth function}}$

Problem 2.10. We are trying to bound $m_{\mathcal{H}}(2N)$. Suppose we partition our $2N$ points into two groups of N points. Each group of N has at most $m_{\mathcal{H}}(N)$ dichotomies. Thus, the number of dichotomies for $2N$ points can be upper bounded by choosing one dichotomy for each set of N points. Thus, $m_{\mathcal{H}}(2N) < m_{\mathcal{H}}(N)m_{\mathcal{H}}(N) = m_{\mathcal{H}}(N)^2$. We can combine this with the VC generalization bound to get:

$$E_{\text{out}} \leq E_{\text{in}} + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$$
$$E_{\text{out}} \leq E_{\text{in}} + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(N)^2}{\delta}}$$

Problem 2.12. With $d_{VC} = 10, \epsilon = 0.05, \delta = 0.05$, we can achieve the following bound.

$$\begin{aligned}
\epsilon &\leq \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}} \\
\epsilon^2 &\leq \frac{8}{N} \ln \frac{4((2N)^{d_{VC}} + 1)}{\delta} \\
0.05^2 &\leq \frac{8}{N} \ln \frac{4((2N)^{10} + 1)}{0.05}
\end{aligned}$$

We can solve this equation numerically. We get $N \approx 4.53 \times 10^5$.