CSCI 4100 Assignment 4

Boliang Yang 661541863

1. Exercise 2.4 in LFD

(a) We can construct a nonsingular $(d+1)\times(d+1)$ matrix whose rows represent the $(d+1)$ points, that is, put $(d+1)$ input vectors $x_n \in R^{d+1}$ to be $X \in R^{(d+1)\times(d+1)}$ whose rows represent x_n, then we define the target vector $y \in R^{d+1}$. Suppose $G = \begin{bmatrix} w^T x_1 \\ \vdots \\ w^T x_{d+1} \end{bmatrix} = Xw$, where $w \in R^{d+1}$ since there's hypothesis $g(x_i) = \text{sign}(w^T x_i)$. To show that $(d+1)$ points in X that the perceptron can shatter we can show that $\text{sign}(G) = \begin{bmatrix} \text{sign}(w^T x_1) \\ \vdots \\ \text{sign}(w^T x_{d+1}) \end{bmatrix} = \text{sign}(Xw) = y$ has solution for w, that is, we show $G = Xw$ has solution for w where $\text{sign}(G) = y$. Since there exists a X such that X is nonsingular, and $w = X^{-1}G$, we can then construct a linearly independent matrix: $\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$. Thus we show $(d+1)$ points in X that the perceptron can shatter and therefore, $d_{vc} \geq d+1$.

(b) Suppose input space X is $(d+1)$ dimension, $x_1, x_2, \ldots, x_{d+1}$ are linearly independent and form a basis for X. Now there is a x_{d+2}, by the fact that any $d+2$ vectors of length $d+1$ have to be linearly dependent. Then there exists $a_1, a_2, \ldots, a_{d+1} \in R$ such that $x_{d+2} = a_1 x_1 + \cdots + a_{d+1} x_{d+1}$. Then $g(x_{d+2}) = \text{sign}(w^T x_{d+2}) = \text{sign}(\mathrm{w}^{\mathrm{T}} a_1 x_1 + \cdots + w^T a_{d+1} x_{d+1}) = \text{sign}\left(\sum_{i=1}^{d+1} a_i w^T x_i\right)$. Now we assign $g(x_i)$ to +1 if $a_i > 0$ and $g(x_i)$ to -1 if $a_i < 0$. Then when $a_i > 0$, $g(x_i) = +1$, $w^T x_i > 0$, and $a_i w^T x_i > 0$; when $a_i < 0$, $g(x_i) = -1$, $w^T x_i < 0$, and $a_i w^T x_i > 0$. Now we have $\sum_{i=1}^{d+1} a_i w^T x_i > 0$, $g(x_{d+2}) = +1$. When we choose the class of other vectors carefully like this way, then the classification of x_{d+2} is dictated: here $g(x_{d+2}) = +1$ and cannot be -1. Thus, $d+2$ points cannot be shattered, and $d_{vc} \leq d+1$.

2. Problem 2.3 in LFD

   (a) Positive or negative rays:

   For N points, there are N+1 regions and the rays have two endpoints which can be put into N+1 regions. To compute $m_H(N)$, we consider to add both ends are in same point, then $m_H(N) = \binom{N+1}{2} + 1 = \frac{N(N+1)}{2} + 1$. Since $\frac{N(N+1)}{2} + 1 \leq \sum_{i=0}^{2}\binom{N}{i}$, then by Theorem 2.4, $d_{vc} = k - 1 = 2$.

   (b) Positive or negative interval:

   To compute $m_H(N)$, we have $\binom{N+1}{2}$ choices with two sets of values (+1, -1). Thus, $m_H(N) = 2\left(\binom{N+1}{2} + 1\right) = N^2 + N + 2$. Since $N^2 + N + 2 \leq \sum_{i=0}^{3}\binom{N}{i}$, then by Theorem 2.4, $d_{vc} = k - 1 = 3$.

   (c) Two concentric spheres in R^d:

   This problem is same as positive intervals, because we can rewrite the problem as $r = \sqrt{x_1^2 + \cdots + x_d^2}$, where $a \leq r \leq b$. Then, $m_H(N) = \binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$. Since $\frac{1}{2}N^2 + \frac{1}{2}N + 1 \leq \sum_{i=0}^{2}\binom{N}{i}$, then by Theorem 2.4, $d_{vc} = k - 1 = 2$.

3. Problem 2.8 in LFD

   We can verify if the function is possible growth function by Theorem 2.4.

   $1 + N$: Possible, $m_H(N) \leq \sum_{i=0}^{1}\binom{N}{i} = 1 + N$.

   $1 + N + \frac{N(N-1)}{2}$: Possible, $m_H(N) \leq \sum_{i=0}^{2}\binom{N}{i} = 1 + N + \frac{N(N-1)}{2}$.

   2^N: Possible, if no break point exists, then $m_H(N) = 2^N$.

   $2^{\lfloor\sqrt{N}\rfloor}$: Impossible, break point is 2 since $m_H(2) = 2^{\lfloor\sqrt{2}\rfloor} = 2 < 2^2$. Then by Theorem 2.4 we have $m_H(N) \leq \sum_{i=0}^{2-1}\binom{N}{i} = 1 + N$, but $m_H(25) = 2^{\lfloor\sqrt{25}\rfloor} = 2^5 = 32 > 1 + 25$. Thus, there is a contradiction.

   $2^{\lfloor N/2\rfloor}$: Impossible, break point is 2 since $m_H(2) = 2^{\lfloor 2/2\rfloor} = 2 < 2^2$. Then by Theorem 2.4 we have $m_H(N) \leq \sum_{i=0}^{2-1}\binom{N}{i} = 1 + N$, but $m_H(6) = 2^{\lfloor 6/2\rfloor} = 2^3 = 8 > 1 + 6$. Thus, there is a contradiction.

   $1 + N + \frac{N(N-1)(N-2)}{6}$: Impossible, break point is 2 since $m_H(2) = 1 + 2 = 3 < 2^2$.

Then by Theorem 2.4 we have $m_H(N) \leq \sum_{i=0}^{2-1}\binom{N}{i} = 1 + N$, but $m_H(3) = 1 + 3 + 1 = 5 > 1 + 3$. Thus, there is a contradiction.

4. Problem 2.10 in LFD

Suppose $m_H(N) = k > 0$ where k is the maximum number of dichotomies that H can implement on N points. Then for any 2N points, it can be first N points plus a second N points, each of which have at most k dichotomies. Thus, for 2N points we at most can implement $k \times k$ combinations, then $m_H(2N) \leq k^2 = m_H(N)^2$.

Then from generalization bound, we have $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N}\ln\frac{4m_H(2N)}{\delta}} \leq E_{in}(g) + \sqrt{\frac{8}{N}\ln\frac{4m_H(N)^2}{\delta}}$.

5. Problem 2.12 in LFD

From generalization bound, we have $N \geq \frac{8}{\epsilon^2}\ln\frac{4\left((2N)^{d_{vc}}+1\right)}{\delta}$, where $d_{vc} = 10, \delta = 1 - 0.95 = 0.05$, $\epsilon = 0.05$. Then $N \geq \frac{8}{0.05^2}\ln\frac{4\left((2N)^{10}+1\right)}{0.05}$.

For $N = 10000$, then $N \geq 330934$.

For $N = 330934$, then $N \geq 442912$.

For $N = 442912$, then $N \geq 452239$.

For $N = 452239$, then $N \geq 452906$.

For $N = 452906$, then $N \geq 452953$.

For $N = 452953$, then $N \geq 452956$.

For $N = 452956$, then $N \geq 452956$.

Sample size N converges to 452956. Thus, $N \geq 452957$.