CSCI-4100 Assignment 4
Yichuan Wang
RIN:661414395

EXERCISES

2.4 (a)We can construct a input matrix as following:

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

Since the matrix is full rank, it is invertible.
Since we have $Xw = y$ and X is invertible, we have $w = X^{-1}y$ for all y where y is a d+1 length column of +1 and -1.
Since we can calculate w for each y using equation $w = X^{-1}y$, there must be at least one w for each y, and therefore these $d + 1$ data points can be shattered.
(b)Due to the fact that any $d + 2$ vectors of length $d + 1$ has to be linearly dependent, there must be one vector x_{d+2} such that $x_{d+2} = c_0x_0 + c_1x_1 + c_2x_2 + ... + c_{d+1}x_{d+1}$. Suppose there is a dichotomy A of first $d+1$ independent vectors such that $w^Tc_nx_n < 0$ for all coefficients c, then $sign(w^Tx_{d+2})$ must be -1. In this case the perceptron can only implement dichotomy $[A, -1]$; it cannot implement $[A, +1]$.

PROBLEMS

2.3
(a)
$m_H(1) = 2 = 2^1$ A single point can be classified as +1 or -1.
$m_H(2) = 4 = 2^2$ 2 points can have 4 dichotomies. They can be both 0 and 1, and have both 0,1 and 1,0 configuration since the ray can be positive or negative when placed in between these two points.
$m_H(3) = 6 < 2^3$ Since the data points are on a 1D array, the middle point must be the same as one of the side points, and therefore 0,1,0 and 1,0,1 cannot be achieved.

The VC dimension for positive or negative ray is 2.
(b)
According to Page 44 of LFD text book, a positive interval has a $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$. If H contains both positive and negative intervals, it would first cover all the results for positive intervals. It then should cover all the negative intervals. Since the negative interval that covers x_1 or x_n has its equivalence case in positive intervals, we just need to consider $\binom{N+1-2}{2}$ while setting both end points to be +1. There is no need to have +1 for negative interval since that is also a case covered in positive interval. In this case the total result would be $\binom{N+1}{2} + 1 + \binom{N-1}{2} = N^2 - N + 2$
$m_H(1) = 2 = 2^1$
$m_H(2) = 4 = 2^2$
$m_H(3) = 8 = 2^3$
$m_H(4) = 14 < 2^4$
The VC dimension is 3
(c)This case can be transformed into positive intervals. Each data point will be assigned a real number r based on their feature values $x_1 ... x_n$, and the concentric sphere is simply a positive interval among these real number values. The VC dimension in this case is 2.

2.8. There are only two possible cases for growth function $m_H(N)$: Power of 2 and polynomial. Also since there is theorem: "if $m_H(k) < 2^k$ some value k, then for all N, $m(N) < N^{k-1} + 1$ ". In this case $1 + N + \frac{N(N-1)(N-2)}{6}$ cannot be a answer: if we try $N = 2$ we get a result of 3, so we can conclude $k = 2$. However, it doesn't satisfy $m(N) < N^{k-1} + 1$ for $N = 2$ $(3 = 3)$.
The possible growth function are:
$1 + N$ This can be a positive positive ray
$1 + N + \frac{N(N-1)}{2}$ This can be a positive interval
2^N $d_{VC} = \infty$

2.10
The dichotomy of 2N points can be seen as a combination of two dichotomies of N points. The maximum number of combinations of two dichotomies of N points is at most all possible combinations of these two dichotomies, which is $m_H(N)^2$. So $m_H(2N) \leq m_H(N)^2$.
Generalization bound: $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4m_H(N)^2}{\sigma})}$

2.12

Iterative calculation starting with $N = 1000$

$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 1000)^{10} + 1}{0.05}) = 257251$$
$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 257251)^{10} + 1}{0.05}) = 434853$$
$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 434853)^{10} + 1}{0.05}) = 451652$$
$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 451652)^{10} + 1}{0.05}) = 452864$$
$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 452864)^{10} + 1}{0.05}) = 452950$$
$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 452950)^{10} + 1}{0.05}) = 452956$$
$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 452956)^{10} + 1}{0.05}) = 452957$$
$$N \geq \frac{8}{0.05^2} ln(\frac{4(2 \times 452957)^{10} + 1}{0.05}) = 452957$$

Answer: $N \geq 452957$