Assignment 4

Name: Mohaiminul Al Nahian
RIN: 662026703
Course no.: CSCI 6100

1 Exercise 2.4

(a) Let X be a non-singular matrix of dimension (d+1)x(d+1), where each row corresponds to a datapoint and each column is the elements of a datapoint. and Y is a vector of (d+1) outputs and W is a vector of weights with (d+1) elements.

$$X = \begin{bmatrix} x0_1 & x1_1 & . & . & xd_1 \\ x0_2 & x1_2 & . & . & xd_2 \\ . & . & . & . & . \\ x0_{(d+1)} & x1_{(d+1)} & . & . & xd_{(d+1)} \end{bmatrix} W = \begin{bmatrix} w0 \\ w1 \\ . \\ w_d \end{bmatrix} Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ y_{d+1} \end{bmatrix}$$

If XW=Y has a solution W, then perceptron can shatter (d+1) points. since, X is non-singular, so it is invertible and we get $W = X^{-1}Y$, that gives a solution of W. So, $d_{vc} \geq d+1$

(b) We know that, d+2 vectors of length d+1 is linearly dependent. So, we may assume that, $x_{d+2} = a_0x_0 + a_1x_1 + ... + a_dx_d$.
If there is a dichotomy D of first d+1 datapoints such that $w^Ta_ix_i < 0$ for all a_i then the perceptron can only implement [D,-1] and not [D,+1]. So, in conlusion, perceptron cannot shatter d+2 datapoints , so $d_{vc} < d+2$ or $d_{vc} \leq d+1$.

From (a) $d_{vc} \geq d+1$ and from (b) $d_{vc} \leq d+1$, so it is evident that VC dimension of the perceptron $d_{vc} = d+1$

2 Problem 2.3

(a) Positive or Negative rays: For N points, for positive rays $m_H(N) = N+1$ and also for negative rays $m_H(N) = N+1$. subtracting the two cases where all points are +1 or -1, we have $m_H(N) = (N+1) + (N+1) - 2 = 2N$. Here, $m_H(3) = 6 < 2^3$, so $d_{vc} = 2$

(b) Positive or Negative interval: Positive interval has $\binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$ dichotomies. For negative intervals we need to add $\binom{N+1-2}{2} = \binom{N-1}{2} = \frac{1}{2}N^2 - \frac{3}{2}N + 1$ more dichotomies, while setting the end points to +1.
So, maximum number of dichotomies= $\frac{1}{2}N^2 + \frac{1}{2}N + 1 + \frac{1}{2}N^2 - \frac{3}{2}N + 1 = N^2 - N + 2$
$m_H(3) = 8 = 2^3$, $m_H(4) = 14 < 2^4$, so, $d_{vc} = 3$

(c) Two concentric spheres in R^d: This is equivalent to the problem of positive intervals. We map a point in R^d into a point in $y\epsilon R$, where $y = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$. So, for N points in d dimensions has corresponding $y\epsilon R$ and splits into N+1 regions. So,

$m_H(N) = \binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1.$
$m_H(2) = 4 = 2^2$, $m_H(3) = 7 < 2^3$, so, $d_{vc} = 2$

3 Problem 2.8

For growth functions, the choices are $d_{vc} = \infty$ and $m_H(N) = 2^N$ or d_{vc} is finite and $m_H(N)$ is bounded by $\sum_{i=0}^{d_{vc}} \binom{N}{i}$

(i) for $m_H(N) = 1 + N$: $m_H(2) = 3 < 2^2$, so $d_{vc} = 1$ and $m_H(N)$ must be bounded by $\sum_{i=0}^{1} \binom{N}{i} = 1 + N$ for all N in this case, which is true. So, it is a possible growth function.

(ii) for $m_H(N) = 1 + N + N(N-1)/2$: $m_H(3) = 7 < 2^3$, so $d_{vc} = 2$ and $m_H(N)$ must be bounded by $\sum_{i=0}^{3-1} \binom{N}{i} = 1 + N + N(N-1)/2$ for all N, which is true in this case as $N >= 1$. So, it is a possible growth function.

(iii) $m_H(N) = 2^N$ is a possible growth function with $d_{vc} = +\infty$

(iv) For $m_H(N) = 2^{\lfloor \sqrt{N} \rfloor}$: $m_H(2) = 2^1 < 2^2$, so $d_{vc} = 1$ and $m_H(N)$ must be bounded by $\sum_{i=0}^{1} \binom{N}{i} = 1 + N$ for all N. But this is not true for all cases, for example, $m_H(25) = 2^{\lfloor \sqrt{25} \rfloor} = 32 > 1 + 25$. So, it cannot be a growth function.

(v) For $m_H(N) = 2^{\lfloor N/2 \rfloor}$: $m_H(1) = 2^0 < 2^1$, so $d_{vc} = 0$ and $m_H(N)$ must be bounded by $\binom{N}{0} = 1$ for all N. But this is not true for all cases, for example, $m_H(2) = 2^{\lfloor 2/2 \rfloor} = 2 > 1$. So, it cannot be a growth function.

(vi)For $m_H(N) = 1 + N + \frac{N(N-1)(N-2)}{6}$: $m_H(2) = 3 < 2^2$, so $d_{vc} = 1$ and $m_H(N)$ must be bounded by $\sum_{i=0}^{1} \binom{N}{i} = 1 + N$ for all N. But this is not true for all cases, for example, $m_H(3) = 5 > 1 + 3$. So, it cannot be a growth function.

4 Problem 2.10

The dichotomy of 2N points can be separated into two dichotomies of N points.Their individual maximum number of dichotomies is $m_H(N)$.So, for 2N points the maximum number of dichotomies $m_H(2N)$ is at most the total combination $m_H(N) \times m_H(N) = m_H(N)^2$
So, $m_H(2N) \leq m_H(N)^2$

So, the generalized bound is: $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4 \times m_H(2N)}{\delta})} \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4 \times m_H(N)^2}{\delta})}$

5 Problem 2.12

We know that $\epsilon \geq \sqrt{\frac{8}{N} ln(\frac{4 m_H(2N)}{\delta})}$
Taking the upperbound of $m_H(2N) \leq (2N)^{d_{vc}} + 1$ we get $\epsilon \geq \sqrt{\frac{8}{N} ln(\frac{4((2N)^{dvc}+1)}{\delta})}$
So, $N \geq \frac{8}{\epsilon^2} ln(\frac{4((2N)^{dvc}+1)}{\delta})$

We have, $d_{vc} = 10, \epsilon = 0.05, \delta = 0.05$

So, we start with N=100 and then iterate for convergence

$$N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 100)^{10} + 1)}{0.05}) = 183568$$
$$=> N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 183568)^{10} + 1)}{0.05}) = 424054$$
$$=> N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 424054)^{10} + 1)}{0.05}) = 450847$$
$$=> N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 450847)^{10} + 1)}{0.05}) = 452807$$
$$=> N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 452807)^{10} + 1)}{0.05}) = 452946$$
$$=> N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 452946)^{10} + 1)}{0.05}) = 452956$$
$$=> N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 452956)^{10} + 1)}{0.05}) = 452957$$
$$=> N \geq \frac{8}{0.05^2} ln(\frac{4((2 \times 452957)^{10} + 1)}{0.05}) = 452957$$

So, $N \geq 452957$