Machine Learning from Data CSCI 4100
Assignment 4
Jae Park
661994900

Exercise 2.4

(a) X can be represented as

$$X = \begin{bmatrix} x_{1\,0} & x_{1\,1} & & x_{1\,d} \\ x_{2\,0} & x_{2\,1} & \dots & x_{2\,d} \\ \vdots & \vdots & & \vdots \\ x_{d+1\;0} & x_{d+1\;1} & & x_{d+1\;d} \end{bmatrix}$$

and w & y are given as

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}, y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_d \end{bmatrix}$$

X is an invertible, nonsingular matrix of dimension $(d+1) \times (d+1)$ where each column is an individual $\vec{x}$.
y is given as $y = [y_0, y_1, ..., y_d]^T$ where each element of y is composed of $+1's$ and $-1's$

Given $X \cdot w = y$, if w exists that can satisfy $sign(Xw) = y$, then the perceptron shatters the $(d+1)$ points. In such cases, now $w = X^{-1}y$ holds true because X is an invertible matrix, as stated above.
Because a valid solution matrix of w exists, it concludes the inequality $d_{vc} \geq d+1$.

(b)Let each point of X as $[\vec{x}_0, \vec{x}_1, ..., \vec{x}_d]$ like the construction earlier in (a). Then new vector $\vec{x}_{d+2}$ should be the linear combination of the previous $(d+1)$ vectors in such form

$$\vec{x}_{d+2} = c_0\vec{x}_0 + c_1\vec{x}_1 + c_2\vec{x}_2 + ... + c_{d+1}\vec{x}_{d+1}$$

Now, we assume that $w^T c_i x_i < 0$ for all c_i. If this is true, then for some dichotomy, it is always true that $sign(w^T \cdot x_{d+2}) = -1$. This means that there is some dichotomy, more specifically $[+1]$, that cannot be implemented.

Problem 2.3

(a) With N data points, you can make $N+1$ regions. And for each positive and negative ray, you get $m_H(N)$ is $2N+2$. From this, you subtract two cases where points are either all +1's or -1's. So $m_H(N) = 2N$.
$m_{\mathcal{H}}(1) = 2 = 2^1$
$m_{\mathcal{H}}(2) = 4 = 2^2$
$m_{\mathcal{H}}(3) = 6 < 2^3$
So $d_{vc} = 2$

(b) (referring to the graph in LFD pg.44) In the positive intervals, the maximum number of dichotomies is $\frac{1}{2}N^2 + \frac{1}{2}N + 1$.
Now, the negative intervals has the same case in positive intervals, and the only additional case arises when both end points are set to be +1. In such case, the number of dichotomies is $\binom{N+1-2}{2}$, which is $\frac{(N-1)!}{(2)!(N-3)!} = \frac{1}{2}N^2 - \frac{3}{2}N + 1$.

$$m_{\mathcal{H}}(N) = (\tfrac{1}{2}N^2 + \tfrac{1}{2}N + 1) + (\tfrac{1}{2}N^2 - \tfrac{3}{2}N + 1) = N^2 - N + 2$$

$m_{\mathcal{H}}(1) = 2 = 2^1$
$m_{\mathcal{H}}(2) = 4 = 2^2$
$m_{\mathcal{H}}(3) = 8 = 2^3$
$m_{\mathcal{H}}(3) = 14 < 2^4$
So $d_{vc} = 3$

(c)For the values $x_1, x_2, ...x_m$, y is given as $y = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$. This is essentially same as finding the interval, where x_i mapped onto a line, and the interval is split by the points into at most $N+1$ regions. According to LFD pg 44, $m_{\mathcal{H}}(N) = \begin{pmatrix} N+1 \\ 2 \end{pmatrix} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$. $m_{\mathcal{H}}(2) = 2 = 2^1$
$m_{\mathcal{H}}(2) = 4 = 2^2$
$m_{\mathcal{H}}(3) = 7 < 2^3$
So $d_{vc} = 2$

Problem 2.8

A growth function is either bounded by some polynomial function of N or the function of 2^N. If the growth function has a break point, then it is bounded by a polynomial function; otherwise, 2^N.

$m_{\mathcal{H}}(N) = 1 + N$: $m_{\mathcal{H}}(1) = 2 = 2^1$ $m_{\mathcal{H}}(2) = 3 < 2^2$
$d_{vc} = 1$ so a possible growth function.

$m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)}{2}$: $m_{\mathcal{H}}(2) = 4 = 2^2$ $m_{\mathcal{H}}(3) = 7 < 2^3$
$d_{vc} = 2$ so a possible growth function.

$m_{\mathcal{H}}(N) = 2^N$: $d_{vc} = \infty$ so a possible growth function.

$m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor}$: $m_{\mathcal{H}}(0) = 1 = 2^0$ $m_{\mathcal{H}}(1) = 2 = 2^1$ $m_{\mathcal{H}}(2) = 2 < 2^2$
$d_{vc} = 1$ and this does not satisfy the inequality $m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{vc}} \binom{N}{i}$

$m_{\mathcal{H}}(N) = 2^{\lfloor N/2 \rfloor}$: $m_{\mathcal{H}}(0) = 1 = 2^0$ $m_{\mathcal{H}}(1) = 1 < 2^1$
$d_{vc} = 0$ and this does not satisfy the inequality $m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{vc}} \binom{N}{i}$

$m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)(N-2)}{6}$: $m_{\mathcal{H}}(0) = 1 = 2^0$ $m_{\mathcal{H}}(1) = 2 = 2^1$ $m_{\mathcal{H}}(2) = 3 < 2^2$
$d_{vc} = 1$ and this does not satisfy the inequality $m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{vc}} \binom{N}{i}$

Therefore, $m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor}, 2^{\lfloor N/2 \rfloor}, 1 + N + \frac{N(N-1)(N-2)}{6}$ cannot be a possible growth function.

Problem 2.10

$m_{\mathcal{H}}(2N)$ means the greatest number of dichotomies obtainable by two independent groups of N points, or in other words, maximum possible combination of two different groups of dichotomies. The case for this maximum possible combination is given as
$m_{\mathcal{H}}(N) \times m_{\mathcal{H}}(N) = m_{\mathcal{H}}(N)^2$. Therefore, $m_{\mathcal{H}}(2N) \leq m_{\mathcal{H}}(N)^2$

A generalization bound which involves only $m_{\mathcal{H}}(N)$ is given as

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4 m_{\mathcal{H}}(N)^2}{\delta}} \quad \ldots \text{LFD (2.12)}$$

which is the most conservative bound possible with probability $\geq 1 - \delta$.

Problem 2.12

$$N \geq \frac{8}{\epsilon^2} \ln \left(\frac{4((2N)^{d_{vc}} + 1)}{\delta} \right)$$

...LFD (2.13) on pg. 57

where, $d_{vc} = 10$, $\epsilon = 0.05$, $\delta = 0.05$.

$$N \geq \frac{8}{0.05^2} \ln \left(\frac{4((2N)^{10} + 1)}{0.05} \right)$$

Trying an initial guess of $N = 1,000$ in the RHS, we get

$$N \geq \frac{8}{0.05^2} \ln \frac{4((2 \times 1000)^{10} + 1)}{0.05} = 257251$$

After an iterative process, N rapidly converges to an estimate of $N \approx 452957$
Therefore, the sample size needs to be greater than or equal to 452957.