Assignment 4

Alberto Mejia
RIN: 661514960

CSCI 4100 - Machine Learning from Data

September 29, 2019

1. (50) LFD Exercise 2.4

(a)

$X = \{1\} \times \mathrm{R}^{d}$

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_{d+1}^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 1 & 0 & \cdots & 0 & 1 \end{bmatrix},\ \mathrm{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix},\ \mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{d+1} \end{bmatrix} = \begin{bmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix} = [\pm 1\ \pm 1 \cdots \pm 1]^{\mathrm{T}}$$

X has all rows and columns linearly independent, giving it full rank and thus, is an invertible and nonsingular matrix.

For any y, finding a w such that sign(Xw) = y, then the perceptron can shatter (d+1) points. This is because Xw = y and X is invertible, meaning $\mathrm{w} = \mathrm{X}^{-1}\mathrm{y}$.

(b)

$\mathrm{X} = [x_1, x_2, \ldots, x_d]$

So, for there to be any new vector, it must be a linear combination of the d+1 vector. That means any d+2 vectors that have length d+1 (more points than dimensions) must be linearly dependent, or

$$x_j = \sum_{i \neq j} a_i x_i$$

Where not all of the $a_i's$ are equal to 0.

Now, we create a dichotomy that the perceptron learning algorithm cannot shatter

$$\mathrm{y_i} = \begin{cases} \mathrm{sign}(\mathrm{a_i}), & i \neq j \\ -1, & i = j \end{cases}$$

Using the PLA, we have

$$x_j = \sum_{i \neq j} a_i x_i -> \mathrm{w}^{\mathrm{T}} x_i = \sum_{i \neq j} a_i \mathrm{w}^{\mathrm{T}} x_i$$

We know that $a_i x_i = \mathrm{w}^{\mathrm{T}} x_i$, so if $\mathrm{y_i} = \mathrm{sign}(\mathrm{w}^{\mathrm{T}}\mathrm{x_i})$ and sign($\mathrm{a_i}$), then $a_i \mathrm{w}^{\mathrm{T}} x_i > 0$.

Thus,

$\mathrm{w}^{\mathrm{T}} x_i = \sum_{i \neq j} a_i \mathrm{w}^{\mathrm{T}} x_i > 0$

Must be true since each of our classifications will be > 0. This means having a correct classification of $\mathrm{y_i}$ is impossible because $\mathrm{y_i} = \mathrm{sign}(\mathrm{w}^{\mathrm{T}}\mathrm{x_i}) = +1$

Thus, the PLA cannot implement this dichotomy.

$\therefore\ \nexists$ {d+2 points} that can be shattered by the PLA

2. (300) LFD Problem 2.3

(a) Positive or negative ray:

As previously computed in exercise 2.1, we know that the growth function for positive rays is equal to N + 1.
For positive rays, we have m_H(N) = N + 1
For negative rays, we have m_H(N) = N – 1
The sum of these two rays gives us
m_H(N) = (N + 1) + (N – 1) = 2N.
The largest value of N we can use where m_H(N) = 2^N is 2.
m_H(N) = 2^N
$m_H(2) = 2(2) = 4$ & $2^2 = 4$, -> $4 = 2^2$
$m_H(3) = 2(3) = 6$ & $2^3 = 6$, -> $6 < 2^3$

$\therefore d_{vc} = 2$

(b) Positive or negative interval:

As previously computed in exercise 2.1, we know that the growth function for positive intervals is equal to $\binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$.
For positive intervals, we have m_H(N) = $\frac{1}{2}N^2 + \frac{1}{2}N + 1$
For negative intervals, we have $\binom{N-1}{2}$ The +1 is included in the positive intervals

-We get N-2 new intervals giving us N+1-2 = N-1

The sum of these two intervals give us
$m_H(N) = \binom{N+1}{2} + \binom{N-1}{2} = N^2 - N + 2$
The largest value of N we can use where m_H(N) = 2^N is 3
m_H(N) = 2^N
$m_H(2) = (2)^2 - 2 + 2 = 4$ & $2^2 = 4$, -> $4 = 2^2$
$m_H(3) = (3)^2 - 3 + 2 = 8$ & $2^3 = 8$, -> $8 = 2^3$
$m_H(4) = (4)^2 - 4 + 2 = 10$ & $2^4 = 16$, -> $14 < 2^4$

$\therefore d_{vc} = 3$

(c) Two concentric spheres in R^d:

This problem can be viewed as solving the problem of positive intervals. This can be done by mapping a point in R^d into a point y ∈ R giving us

$$\text{Map: } \{x_1,\ldots,x_d\} \to y = \sqrt{x_1^2+\ldots+x_d^2} \text{ in } 0 \le y \le \infty$$

We have successfully transformed the problem from a R^d to a R problem where y ∈ [0,..., ∞). This means that m_H(N) no longer depends on d.
Thus, m_H(N) = $\frac{1}{2}N^2 + \frac{1}{2}N + 1$
The largest value of N we can use where m_H(N) = 2^N is 2

m_H(N) = 2^N

$m_H(2) = (1/2)(2)^2 + (2/2) + 1 = 4$ & $2^2 = 4$, -> $4 = 2^2$
$m_H(3) = (1/2)(3)^2 + (3/2) + 1 = 7$ & $2^3 = 8$, -> $7 < 2^3$

$\therefore d_{vc} = 2$

3. (200) LFD Problem 2.8

We know since $m_H(N) \leq 2^N$, there can only exist two cases. The first is that if $m_H(N) = 2^N$ for all N, then $d_{vc}(H) = \infty$. The second is $d_{vc}(H)$ is a finite value, meaning that $m_H(k) \leq 2^k$ for some k, and $m_H(N)$ is bounded, such that
$m_H(N) \leq N^{d_{vc}} + 1$ Def. 2.5 & (2.10) | pg.50 in LFD

$m_H(N) = 1 + N$: ✓

$m_H(1) = 2$ & $2^1 = 2$ -> $2 = 2^1$
$m_H(2) = 3$ & $2^2 = 4$ -> $3 < 2^2$ [$\therefore d_{vc} = 1$]
This means that it must be bounded by 1 + N, ∀N. This is a possible growth function

$m_H(N) = 1 + N + \frac{N(N-1)}{2}$: ✓

$m_H(1) = 2$ & $2^1 = 2$ -> $2 = 2^1$
$m_H(2) = 4$ & $2^2 = 4$ -> $4 = 2^2$
$m_H(3) = 7$ & $2^3 = 8$ -> $7 < 2^3$ [$\therefore d_{vc} = 2$]
This means that it must be bounded by 1 + N, ∀N. This is a possible growth function

$m_H(N) = 2^N$: ✓

This one is trivially a possible growth function as $2^N \leq 2^N$

$m_H(N) = 2^{\lfloor\sqrt{N}\rfloor}$: ×

$m_H(1) = 2^1$ & $2^1 = 2$ -> $2 = 2^1$
$m_H(2) = 2^1$ & $2^1 = 2$ -> $2 = 2^1$[$\therefore d_{vc} = 1$]

$m_H(N) \leq N^{d_{vc}} + 1$
$m_H(N) \leq N + 1$
Thus, this must be bounded by N+1 ∀N but that is not true.
Ex: When N = 16 [$m_H(16) = 2^4 = 16 > 5$]
This is not a possible growth function.

$m_H(N) = 2^{\lfloor\sqrt{N/2}\rfloor}$: ×

$m_H(1) = 2^0$ & $2^1 = 2$ -> $1 < 2^1$[$\therefore d_{vc} = 0$]
$m_H(N) \leq N^{d_{vc}} + 1$
$m_H(N) \leq N^0 + 1 = 1+1 = 2$
Thus, this must be bounded by 1+1=2 ∀N but that is not true.
Ex: When N = 32 [$m_H(32) = 2^4 = 16 > 2$]
This is not a possible growth function.

$m_H(N) = 1 + N + \frac{N(N-1)(N-2)}{6}$: ×

$m_H(1) = 2$ & $2^1 = 2$ -> $2 = 2^1$
$m_H(2) = 3$ & $2^2 = 4$ -> $3 < 2^2$ [$\therefore d_{vc} = 1$]
$m_H(N) \leq N^{d_{vc}} + 1$
$m_H(N) \leq N^1 + 1 = N+1$
Thus, this must be bounded by N+1 ∀N but that is not true.
Ex: When N = 3 [$m_H(3) = 5$ & N+1 = 4 | 5 > 2]
This is not a possible growth function.

4. (100) LFD Problem 2.10

Show $m_H(2N) \leq m_H(N)^2$

Suppose that there exist 2N points. Partition the 2N points into two sets of N points. Based on the definition of the growth function, we know that each set of N points cannot generate greater than $m_H(N)$ dichotomies. If each group of N points can only produce up to $m_H(N)$ dichotomies, then combining them gives us one set of 2N points that can produce a max number $2\times m_H(N)$ dichotomies. Thus, 2N points produce at most $m_H(N) \times m_H(N) = m_H(N)^2$.
$\therefore m_H(2N) \leq m_H(N)^2$

VC Generalization bound:

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4m_H(2N)}{\delta})}$$

With our results, combining it with the VC generalization bound gives us:

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4m_H(2N)}{\delta})} \leq E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4m_H(N)^2}{\delta})}$$

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4m_H(N)^2}{\delta})}$$

5. (100) LFD Problem 2.12

VC Generalization bound:

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4m_H(2N)}{\delta})}$$

$d_{vc} = 10$
Need 95% confidence
$\delta = 5\% = 0.05$
VC Generalization bound:

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} ln(\frac{4((2N)^{\wedge}10 + 1)}{0.05})} \leq 0.05$$

With 10,000 training examples,
$N = 4.52957 \times 10^5$
Thus, a sample size of 4.52957×10^5 or greater is needed to receive such confidence.