HW 4 : Learning Theory

(b) $P_d(x) = \sum_{i=0}^{d} a_i x^i = \langle a, \phi(x) \rangle$

$a = [a_0, a_1, \dots, a_d] \in \mathbb{R}^{d+1}$. So $\phi(x) = [1, x, x^2, \dots, x^d] \in \mathbb{R}^{d+1}$

will make $\langle a, \phi(x) \rangle = a_0 \cdot 1 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_d \cdot x^d$

$= \sum_{i=0}^{d} a_i x^i = P_d(x)$

(c) $P_d(x)$ is a degree "d" polynomial in x, and will have atmost d real roots according to fundamental theorem of algebra. We can note that it is only at these roots that a change of sign of $P_d(x)$ might occur, if at all.

Hence, if $P_d(x)$ has M roots, the hypothesis class it results in is $\mathcal{H}_M$. Since $P_d(x)$ can have $0 \leq M \leq d$ roots, the hypothesis class is:

$$\bigcup_{i=0}^{d} \mathcal{H}_i$$

Claim 1: If $\mathcal{H}_i$ shatters a set of points, each point should lie in a different interval.

Proof: If 2 points x_1, x_2 lie in the same interval, $h_i(x_1) = h(x_2)$ [property]

So we can't achieve opposite labels for them $\Rightarrow$ not shattered

$\Rightarrow$ contradiction

Claim 2: $\mathcal{H}_i$ can shatter atmost $i+1$ points (i.e) $VC(\mathcal{H}_i) \leq i+1$

Proof: $\mathcal{H}_i$ has $i+1$ intervals. If there are more $i+1$ points, by Pigeon hole principle, atleast 2 points will lie in same interval
$\Rightarrow$ points are not shattered by claim 1.

Hence, $P_d(x)$ with hypothesis class $\bigcup_{i=0}^{d} \mathcal{H}_i$ can shatter $\leq VC(\mathcal{H}_d) \leq d+1$ points.

Hence $VC(P_d(x)) \leq d+1$

→ Alternate solution Using Linear Algebra

Claim: A set with more than $d+1$ points cannot be shattered by f_d. — I

Proof: Suppose $\exists$ a set of "k" points $\{x_1, x_2, \ldots, x_d, \ldots, x_k\}$ where $k \geq d+2$. The set $\{\phi(x_1), \ldots, \phi(x_d), \ldots, \phi(x_k)\}$ is linearly dependent. → [Sub-proof: Suppose $M = \begin{bmatrix} | & & | \\ \phi(x_1) & \cdots & \phi(x_k) \\ | & & | \end{bmatrix}$; $\text{rank}(M) + \text{nullity}(M) = k$

by rank-nullity theorem. But $\text{rank}(M) \leq d+1$ since $\phi(x) \in \mathbb{R}^{d+1}$

$\Rightarrow \text{nullity}(M) \geq k-d-1 > 0$
↳ since $k > d+1$

Hence $\{\phi(x_1), \ldots, \phi(x_k)\}$ are linearly dependent]

Since $\{\phi(x_1), \dots, \phi(x_k)\}$ are linearly dependent, $\exists$ constant $c \in \mathbb{R}^k$ such that

$$c_1\phi(x_1) + c_2\phi(x_2) + \dots + c_k\phi(x_k) = 0$$

Suppose j is the largest index such that $c_j \neq 0$. Then we can write

$$c_j\phi(x_j) = -c_1\phi(x_1) - c_2\phi(x_2) - \dots - c_{j-1}\phi(x_{j-1})$$

$$\Rightarrow \phi(x_j) = \sum_{1\le i<j} \frac{-c_i}{c_j}\phi(x_i) = \sum_{1\le i<j} c_i'\phi(x_i) \quad \text{[for some constant } c_i'\text{'s]}$$

We prove the main claim by assuming the contrary. Suppose the k points $\{x_1, \dots, x_k\}$ are shattered by $f_d \Rightarrow$ For any set of labels $y \in \{-1, 1\}^k$, $\exists$ $a \in \mathbb{R}^{d+1}$ such that $\text{sign}(\langle a, \phi(x_i)\rangle) = y_i$.

Let us pick the labels such that $y_i = \text{sign}(c_i')$ for $1 \le i < j$, $y_j = -1$ and $y_i = +1$ for $j < i \le k$. Since the points are shattered, $\exists$ $a \in \mathbb{R}^{d+1}$ satisfying these.

Now, $\phi(x_j) = \sum_{1\le i<j} c_i'\phi(x_i) \Rightarrow \langle\phi(x_j), a\rangle = \langle \sum_{1\le i<j} c_i'\phi(x_i), a\rangle$

$$= \sum_{1\le i<j} c_i'\langle\phi(x_i), a\rangle \quad — \quad (1)$$

Since $\overbrace{\text{sign}(\langle\phi(x_i), a\rangle)}^{f_d(x_i)} = \text{sign}(c_i')$, $c_i'\langle\phi(x_i), a\rangle \ge 0$. Putting this in (1),

$$\underbrace{\langle\phi(x_j), a\rangle}_{<0} = \underbrace{\sum_{1\le i<j} \underbrace{c_i'\langle\phi(x_i), a\rangle}_{\ge 0}}_{\ge 0}.$$

because $y_j = -1 \Rightarrow f_d(x_j) < 0$

LHS is <0, RHS is ≥ 0

$\Rightarrow$ Contradiction!

$\Rightarrow$ Claim is proved!

claim (I) $\Rightarrow$ $VC(f_d) \leq d+1$

(d) Consider $d+1$ distinct points : $x_1, x_2, \dots, x_{d+1}$ ($x_i \neq 0 \ \forall \ 1 \leq i \leq d+1$)

(&(e)) Now consider the matrix :

$$V = \begin{bmatrix} | & | & & | \\ \phi(x_1) & \phi(x_2) & \cdots & \phi(x_{d+1}) \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 1 & & 1 \\ x_1 & x_2 & & x_{d+1} \\ x_1^2 & x_2^2 & \cdots & x_{d+1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^d & x_2^d & & x_{d+1}^d \end{bmatrix}$$

$\downarrow$

Vandermonde Matrix

<u>Claim</u>: If $\det(V) \neq 0$, $\{\phi(x_1), \dots, \phi(x_{d+1})\}$ are linearly independent. — II

<u>Proof</u>: $\det(V) \neq 0 \Rightarrow 0$ is not an eigenvalue of $V \Rightarrow$ Nullity $(V) = 0$

Then by rank-nullity theorem, $\text{rank}(V) = d+1 - \text{nullity}(V)$
$= d+1$ //

$\Rightarrow$ The $d+1$ column vectors are linearly independent.

So, now lets prove $\det(V_{d+1}) \neq 0$ by induction. (V_k is $k \times k$ Vandermonde matrix with distinct $x_1, \dots, x_k$; with none being 0)

$\downarrow$

so our $V = V_{d+1}$

- <u>Base-case</u>: $d=0 \Rightarrow V_1 = [1]$ $\det(V_1) = 1$ $(\neq 0)$.

- <u>Induction-hypothesis</u>: $\det(V_k) \neq 0$ for arbitrary $k \geq 1$

- <u>Induction Step</u>: Let's find $\det(V_{k+1})$.

$$V_{k+1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_k^k \end{bmatrix}.$$

Determinant is invariant when linear combination of other rows are added to some row.
(elementary row operations)

First for row 2, we do: $R_2 \leftarrow R_2 - x_1 . R_1$ to get:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_{k+1} - x_1 \\ x_1^2 & x_2^2 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_k^k \end{bmatrix}$$

Next we do $R_3 \leftarrow R_3 - x_1^2 . R_1 - x_1 . R_2$ to get:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_{k+1} - x_1 \\ 0 & x_2(x_2 - x_1) & \cdots & x_{k+1}(x_{k+1} - x_1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_k^k \end{bmatrix}$$

$\downarrow$

<u>Claim</u>:

If we continue the same way,

for $R_l \leftarrow R_l - x_1^{l-1} . R_1 - x_1^{l-2} . R_2 - \cdots - x_1 . R_{l-1}$,

we will get $x_c^{l-2} . (x_c - x_1)$ in the c^{th} column of R_l
(for $c \geq 1$, $l \geq 2$)

We can prove this by another induction:

- <u>Base case</u>: element in c^{th} column of R_2 after R_2's transformation is $x_c - x_1$ for $c \geq 1$
- <u>Induction step</u>: Assume the statement is true for R_n, for arbitrary $n \geq 2$.

(i.e) c^{th} column value in row R_n after R_n transformation is $x_c^{n-2}(x_c - x_1)$ for $c \geq 1$

Now, performing $(n+1)^{th}$ row transformation:

$$R_{n+1} \leftarrow R_{n+1} - x_1^n \cdot R_1 - x_1^{n-1} \cdot R_2 - \cdots - x_1 \cdot R_n$$

For $c \geq 2$, c^{th} column value in R_{n+1} after transformation is:

$$x_c^n - x_1^n \cdot 1 - x_1^{n-1} \cdot [x_c - x_1] - x_1^{n-2} \cdot [x_c(x_c - x_1)] - \cdots - x_1 [x_c^{n-2}(x_c - x_1)]$$

$$= \left.\begin{aligned} & -x_1^n + x_c^n \\ & -x_c \cdot x_1^{n-1} + x_1^n \\ & -x_c^2 \cdot x_1^{n-2} + x_c \cdot x_1^{n-1} \\ & \quad \vdots \\ & -x_c^{n-2} \cdot x_1^2 + x_c^{n-3} \cdot x_1^3 \\ & -x_c^{n-1} \cdot x_1 + x_c^{n-2} \cdot x_1^2 \end{aligned}\right\} = x_c^n - x_c^{n-1} \cdot x_1 = x_c^{n-1}(x_c - x_1)$$

$$= x_c^{(n+1)-2} \cdot (x_c - x_1)$$

$\Rightarrow$ Statement is true for $n+1$.

Hence by induction, statement is true for all $n \geq 2$.

Also for $c=1$, all rows have zero value since $x_c = x_1$.

Hence, after sequential transformation of the rows $R_2, \ldots, R_k$ in V_{k+1}, we get $\tilde{V}_{k+1}$:

$$\tilde{V}_{k+1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & & x_{k+1} - x_1 \\ 0 & x_2(x_2 - x_1) & \cdots & x_{k+1}(x_{k+1} - x_1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_2^{k-1}(x_2 - x_1) & \cdots & x_{k+1}^{k-1}(x_{k+1} - x_1) \end{bmatrix}.$$

Now, $\det(V_{k+1}) = \det(\tilde{V}_{k+1})$

$$= \det\left(\begin{bmatrix} x_2 - x_1 & & x_{k+1} - x_1 \\ x_2(x_2 - x_1) & \cdots & x_{k+1}(x_{k+1} - x_1) \\ \vdots & \ddots & \vdots \\ x_2^{k-1}(x_2 - x_1) & \cdots & x_{k+1}^{k-1}(x_{k+1} - x_1) \end{bmatrix} \right)$$

(expanding by first column for det)

$$= \underbrace{(x_2 - x_1)(x_3 - x_1)\cdots(x_{k+1} - x_1)}_{\neq 0 \text{ since } x_1, \ldots, x_{k+1} \text{ are distinct}} \cdot \underbrace{\det\left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_2 & x_3 & & x_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_2^{k-1} & x_3^{k-1} & \cdots & x_{k+1}^{k-1} \end{bmatrix} \right)}_{\text{Vandermonde matrix } V_k \text{ for distinct } x_2, \ldots, x_{k+1}}$$

(since $x_1, x_2, \ldots, x_{k+1}$ itself are distinct)

$\Downarrow$

The determinant $\neq 0$ by induction hypothesis

$\neq 0 \quad \Rightarrow \det(V_{k+1}) \neq 0$.

Hence, by induction, $\det(V_k) \neq 0$ is true for all $k \geq 1 \Rightarrow \det(V) \neq 0$

$[V = V_{d+1}]$

$\Rightarrow$ From (II), $\{\phi(x_1), \dots, \phi(x_{d+1})\}$ are linearly independent for distinct $x_1, \dots, x_{d+1}$

To take a specific example $x_1 = 1$, $x_2 = 2$, $\dots$, $x_{d+1} = d+1$ form $d+1$ independent $\{\phi(x_1), \dots, \phi(x_{d+1})\}$

(f) Take the $d+1$ independent vectors from part (d) $\Rightarrow \{\phi(1), \phi(2), \dots, \phi(d+1)\}$

Consider $M = \begin{bmatrix} | & | & & | \\ \phi(1) & \phi(2) & \cdots & \phi(d+1) \\ | & | & & | \end{bmatrix}$; arbitrary $a \in \mathbb{R}^{d+1}$ and labels $y \in \{-1, 1\}^{d+1}$ for the $d+1$ points

For the given set of $d+1$ points to have labels y under a:

$$\text{sign}(\langle \phi(i), a \rangle) = y_i \quad \forall \ 1 \le i \le d+1$$

(i.e) $\text{sign}(Ma) = y$

Let's put an even harder condition: $Ma = y$ (this $\Rightarrow \text{sign}(Ma) = y$ since $\text{sign}(y) = y$)

Since M has $d+1$ independent $(d+1)$-dimensional vectors, it has <u>full rank</u>.

and <u>nullity = 0</u> [$\text{nullity}(M) = d+1 - \text{rank}(M) = d+1 - (d+1) = 0$, by rank-nullity theorem]

$\left.\begin{array}{l} \text{null space} = \phi \Rightarrow \text{one-to-one linear map} \\ \text{full rank} \Rightarrow \text{onto linear map} \end{array}\right\} \Rightarrow M$ is invertible

Let the unique inverse be M^{-1}

For any set of labels $y \in \{-1, 1\}^{d+1}$,

$$Ma = y \implies M^{-1}Ma = M^{-1}y \implies a = M^{-1}y$$

(i.e) a has a unique solution for $Ma = y$.

$\implies a$ has a solution for $\text{sign}(Ma) = y$

$\implies \exists$ a hypothesis in f_d - hypothesis class for any set of labels y_i's $\in \{-1, 1\}$, for the $d+1$ points $\{1, 2, \dots, d+1\} \implies \{1, 2, \dots, d+1\}$ is shattered by f_d-classifier.

$$\implies \boxed{VC(f_d) \geq d+1}$$

For a given label set $y \in \{-1, 1\}^{d+1}$, the f_d-classifier will have $a = M^{-1}y$, where

$$M = \begin{bmatrix} | & & | \\ \phi(1) & \dots & \phi(d+1) \\ | & & | \end{bmatrix}$$

so the f_d-classifier is:

$$f_d(x) = \begin{cases} +1 & \text{if } \langle M^{-1}y, \phi(x) \rangle \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

$$= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 3 & & d \\ 1 & 4 & 9 & & d^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2^d & 3^d & & d^d \end{bmatrix}$$

for the vectors in part (d)

(g) From part (c), we got VC (d^{th} polynomial classifier f_d) $\leq d+1$ — ①

From part (f), we got VC (d^{th} polynomial classifier f_d) $\geq d+1$ — ②

① & ② $\Rightarrow$ $\boxed{\text{VC}(d^{th} \text{ polynomial classifier } f_d) = d+1}$

(h) Family of all possible polynomial classifiers will be infinity.

Proof: Assume the contrary. Suppose VC (all possible polynomial classiers) is finite, $= k$, say.

We know VC (k^{th} degree polynomial classifier) $= k+1$ (extrapolating from part (g))

But All possible polynomial classifiers $\supset$ k^{th} degree polynomial classifier

$\Rightarrow$ VC(All possible polynomial classifiers) $\geq$ VC(k^{th} degree polynomial classifier) $\geq k+1$ $\Rightarrow$ Contradiction!

Hence, proved.