Robt 407 Zarema Balgabekova HW2

0.1 VC Dimension

1. Exercise 2.2 (b)

Does there exist a hypothesis set for which $m_H(N) = N + 2^{\lfloor N/2 \rfloor}$ where $\lfloor N/2 \rfloor$ is the largest integer $\leq N/2$?

From Theorem 2.4, we know that if $m_H(k) < 2^k$ for some value k, then $m_H(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$ for all N. The RHS is polynomial in N of degree $k-1$.

$$\Rightarrow \quad m_H(N) = N + 2^{\lfloor N/2 \rfloor} \overset{?}{\leq} \sum_{i=0}^{k-1} \binom{N}{i} \quad \text{for all } N$$

In this case, $k = 3$ $(m_H(3) = 5 < 2^3)$ $\Rightarrow$ $m_H(N) \overset{?}{\leq} \frac{N^2}{2} + \frac{N}{2} + 1$

The inequality won't hold for all N because LHS experiences exponential growth while RHS experiences polynomial growth (2nd order) $\Rightarrow$ the hypothesis set does not exist

2. Exercise 2.6

400 training examples 1000 hypotheses

200 test examples

$\delta = 0.05$

a) $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$

$$\sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} = \sqrt{\frac{1}{2 \cdot 400} \ln \frac{2 \cdot 1000}{0,05}} \approx 0,115$$

$\Rightarrow E_{out}(g) \leq E_{in}(g) + 0,115$

For test set, we have only one hypothesis (final hypothesis produced by training)

$$E_{out}(g) \leq E_{test}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

$$\sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} = \sqrt{\frac{1}{2 \cdot 200} \ln \frac{2 \cdot 1}{0,05}} \approx 0,096$$

$\Rightarrow E_{out}(g) \leq E_{test}(g) + 0,096$

$\Rightarrow E_{in}(g)$ will have the higher 'error bar'

b) Yes, if we reserve even more examples for testing, we will use fewer examples for training that are important to find a good hypothesis

3. Problem 2.3(b)

From example 2.2, we know that for positive intervals the maximum number of dichotomies $m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$

For negative intervals, we have $\binom{N-1}{2}$ additional dichotomies

$$\binom{N-1}{2} = \frac{(N-1)!}{2!\,(N-3)!} = \frac{(N-2)(N-1)}{2} = \frac{N^2-3N+2}{2}$$

=> Total number of max dichotomies ; $m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1 + \frac{1}{2}N^2 \; \frac{3}{2}N + 1$

$= N^2 - N + 2$

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

4. Problem 2.16

$\mathcal{X} = \mathbb{R}$

For $\mathcal{H} = \{ h_c \mid h_c(x) = \text{sign}\left(\sum_{i=0}^{D} c_i x^i \right) \}$

prove that the VC dimension of $\mathcal{H}$ is exactly $(D+1)$

a) Construct a square matrix with $(D+1) \times (D+1)$ dimensions:

$$X = \begin{bmatrix} 1 & x_0^1 & \cdots & x_0^D \\ 1 & x_1^1 & \cdots & x_1^D \\ \vdots & \vdots & & \vdots \\ 1 & x_D^1 & \cdots & x_D^D \end{bmatrix} \qquad (x_k^0 = 1)$$

where $x_0, x_1, \ldots, x_D$ are $D+1$ points in $\mathbb{R}$

$\det X \neq 0$ since x_k are all different

Arbitrary dichotomy: $y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_D \end{bmatrix} \in \{-1, +1\}^{D+1}$

Let $c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_D \end{bmatrix} = X^{-1} y \Rightarrow Xc = y$

$\Rightarrow h_c(x_k) = \text{sign}\left(\sum_{i=0}^{D} c_i x_k^i \right) = y_k$ for all $k = 0, \ldots, D$

$\Rightarrow$ $D+1$ points $x_0, x_1, \ldots, x_D$ can be shattered by the hypothesis set $\mathcal{H}$

b) If we consider $D+2$ points in $\mathbb{R}$ $x_0, x_1, \dots, x_D, x_{D+1}$

and $D+2$ vectors in the form $\underbrace{[x_k^0, x_k^1, \dots, x_k^D]}_{D+1 \text{ elem.}}$ for $k = 0, \dots, D+1$,

these vectors will be lin dependent

$\Rightarrow$ some vector is a linear combination of all other vectors:

$$[x_l^0, x_l^1, \dots, x_l^D] = \sum_{k \neq l} a_k \, [x_k^0, x_k^1, \dots, x_k^D]$$

($D+1$ coefficients a_k not all equal to zero)

Choose dichotomy y: $y_k = \text{sign}(a_k)$ if $a_k \neq 0$
and $y_l = -1$

Let $c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_D \end{bmatrix} \Rightarrow$

$$[x_l^0, x_l^1, \dots, x_l^D] \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_D \end{bmatrix} = \sum_{k \neq l} a_k [x_k^0, x_k^1, \dots, x_k^D] \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_D \end{bmatrix} = \sum_{k \neq l} \sum_{i=0}^{D} c_i a_k x_k^i$$

Assume that there exists $c \in \mathbb{R}^{D+1}$ s.t

$$y_k = h_c(x_k) = \text{sign}\left(\sum_{i=0}^{D} c_i x_k^i\right) \quad \text{for all } k \ (a_k \neq 0)$$

$$y_\ell = h_c(x_\ell) = \text{sign}\left(\sum_{i=0}^{D} c_i x_\ell^i\right)$$

$$\Rightarrow \text{sign}(a_k) = y_k = \text{sign}\left(\sum_{i=0}^{D} c_i x_k^i\right)$$

$$\Rightarrow \sum_{i=0}^{D} c_i a_k x_k^i > 0 \quad \text{for any } k \ (a_k \neq 0)$$

Because of multiplication with the same signs

$$\Rightarrow \sum_{k \neq \ell} \sum_{i=0}^{D} c_i a_k x_k^i = [x_\ell^0, x_\ell^1, \ldots, x_\ell^D] \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_D \end{bmatrix} = \sum_{i=0}^{D} c_i x_\ell^i > 0$$

$$\Rightarrow y_\ell = \text{sign}\left(\sum_{i=0}^{D} c_i x_\ell^i\right) = +1$$

But we chose $y_\ell = -1$ $\Rightarrow$ there is a dichotomy that cannot be implemented $\Rightarrow$ there are no $D+2$ points which are shattered by $\mathcal{H}$

From a and b, VC dimension of $\mathcal{H}$ is exactly $D+1$.

5. Problem 2.18

Prove that the following hypothesis set for $x \in \mathbb{R}$ has an infinite VC dimension:

$$\mathcal{H} = \{ h_\alpha \mid h_\alpha(x) = (-1)^{\lfloor \alpha x \rfloor}, \text{ where } \alpha \in \mathbb{R} \}$$

where $\lfloor A \rfloor$ is the biggest integer $\leq A$

Consider N points $x_1, \dots, x_N$ where $x_n = 10^n$

Arbitrary dichotomy: $y_1, \dots, y_N \in \{-1, +1\}^N$

If we choose $\alpha = 0.a_1 a_2 \dots a_N$ with $a_i = 1$ if $y_i = -1$ (odd)

$a_i = 2$ if $y_i = +1$ (even)

$$\Rightarrow h_\alpha(x_n) = (-1)^{\lfloor \alpha \cdot 10^n \rfloor} = y_n \quad \text{for all } n = 1, \dots, N$$

$$\Rightarrow \mathcal{H}(x_1, \dots, x_N) = \{-1, +1\}^N \quad \Rightarrow m_{\mathcal{H}}(N) = 2^N \text{ for all } N$$

$\Rightarrow$ The hypothesis set has an infinite VC dimension

D.2 Perceptron Dimension versus VC Dimension

1. Exercise 2.4 (a)

Input space $\mathcal{X} = \{1\} \times \mathbb{R}^d$ (including $x_0 = 1$)

Show that the VC dimension of the perceptron (with $d+1$ parameters counting w_0) is exactly $d+1$

a) Construct nonsingular $(d+1) \times (d+1)$ matrix:

$$X = \begin{bmatrix} 1 & x_{01} & \cdots & x_{0d} \\ 1 & x_{11} & \cdots & x_{1d} \\ \vdots & \vdots & & \\ 1 & x_{d1} & \cdots & x_{dd} \end{bmatrix}$$

$(x_{k0} = 1)$

where $x_0, x_1, \dots, x_d$ are $d+1$ distinct points in $\mathbb{R}^d$

$\det X \neq 0$ since x_k are all different

Dichotomy: $y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_d \end{bmatrix} \in \{-1, 1\}^{d+1}$

let $w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} = X^{-1} y \Rightarrow Xw = y$

$$\Rightarrow h_w(x_k) = \text{sign}\left(\sum_{i=0}^{d} w_i x_{ki}\right) = y_k \quad \text{for all } k = 0, \dots, d$$

$\Rightarrow$ Perceptron can shatter $x_0, \dots, x_d$ $\Rightarrow$ $m_{\mathcal{H}}(d+1) = 2^{d+1}$

$\Rightarrow$ $d_{vc} \geq d+1$

2. Exercise 2.4 b)

b) If we consider $d+2$ points represented by a vector of length $d+1$, these vectors will be linearly dependent

$\Rightarrow$ some vector is a lin. combination of all other vectors:

$$[x_{l0}, x_{l1}, \dots x_{ld}] = \sum_{k \neq l} a_k (x_{k0}, x_{k1}, \dots, x_{kd}]$$

($d+1$ coefficients a_k not all equal to zero)

Choose dichotomy y: $y_k = \text{sign}(a_k)$ if $a_k \neq 0$
and $y_l = -1$

Let $w = [w_0, w_1, \dots, w_d]^T$ $\Rightarrow$

$$[x_{l0}, x_{l1}, \dots, x_{ld}] \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} = \sum_{k \neq l} a_k [x_{k0}, x_{k1}, \dots, x_{kd}] \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} =$$

$$\sum_{k \neq l} \sum_{i=0}^{d} w_i a_k x_{ki}$$

Assume that there exists $w \in \mathbb{R}^{d+1}$ s.t.

$$y_k = h(x_k) = \text{sign}\left(\sum_{i=0}^{d} w_i x_{ki}\right) \quad \text{for all } k \quad (a_k \neq 0)$$

$$y_l = h(x_l) = \text{sign}\left(\sum_{i=0}^{d} w_i x_{li}\right)$$

$$\Rightarrow \quad \text{sign}(a_k) = y_k = \text{sign}\left(\sum_{i=0}^{d} w_i x_{ki}\right)$$

$$\Rightarrow \quad \sum_{i=0}^{d} w_i a_k x_{ki} > 0 \quad \text{for any } k \ (a_k \neq 0)$$

Because of multiplication with the same signs

$$\Rightarrow \sum_{k \neq l} \sum_{i=0}^{d} w_i a_k x_{ki} = [x_{l0}, x_{l1}, \dots, x_{ld}] \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} = \sum_{i=0}^{d} w_i x_{li} > 0$$

$$\Rightarrow \quad y_l = \text{sign}\left(\sum_{i=0}^{d} w_i x_{li}\right) = +1$$

But we chose $y_l = -1$ $\Rightarrow$ there is a dichotomy that cannot be implemented $\Rightarrow$ for $N \geq d+2$ $m_H(N) < 2^N$

$\Rightarrow d_{VC} \leq d+1$

0.3 The Upper Bound

1. Exercise 2.7 (a)

For Binary target functions, show that $\mathbb{P}[h(x) \neq f(x)]$ can be written as an expected value of a mean squared error measure if the convention used for the Binary function is 0 or 1.

There are 4 cases:

$h(x)$	$f(x)$
0	0
0	1
1	0
1	1

probability of each case is $\frac{1}{4}$

Expected value of a mean squared error measure:

$$\mathbb{E}[(h(x) - f(x))^2] = \frac{1}{4} \cdot (0-0)^2 + \frac{1}{4}(0-1)^2 + \frac{1}{4}(1-0)^2 +$$

$$\frac{1}{4}(1-1)^2 = \frac{1}{2}$$

$$\mathbb{P}[h(x) \neq f(x)] = \mathbb{P}(\text{'}h(x)=0, f(x)=1\text{'} \cup \text{'}h(x)=1, f(x)=0\text{'}) =$$

$$\mathbb{P}(\text{'}h(x)=0, f(x)=1\text{'}) + \mathbb{P}(\text{'}h(x)=1, f(x)=0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$\Rightarrow \mathbb{P}[h(x) \neq f(x)] = \mathbb{E}[(h(x) - f(x))^2]$$

Also, algebraically:

$$\mathbb{P}[h(x) \neq f(x)] = \mathbb{P}[h(x) \neq f(x)] \cdot 1 + \mathbb{P}[h(x) = f(x)] \cdot 0 =$$

$$\mathbb{P}[h(x) \neq f(x)]\,(h(x) - f(x))^2 + \mathbb{P}[h(x) = f(x)]\,(h(x) - f(x))^2 =$$

$$\mathbb{E}[(h(x) - f(x))^2]$$

2. Exercise 2.7(b)

The convention used for the binary function is ± 1.

There are 4 cases:

$h(x)$	$f(x)$
$+1$	$+1$
$+1$	-1
-1	$+1$
-1	-1

probability of each case is $\frac{1}{4}$

Expected value of a mean squared measure:

$$\mathbb{E}[((h(x) - f(x))^2] = \frac{1}{4}(1-1)^2 + \frac{1}{4}(1-(-1))^2 + \frac{1}{4}(-1-1)^2 +$$

$$\frac{1}{4}((-1) - (-1))^2 = 2$$

$$\mathbb{P}[h(x) \neq f(x)] = \mathbb{P}('h(x) = +1, f(x) = -1' \cup 'h(x) = -1, f(x) = +1') =$$

$$\mathbb{P}('h(x) = +1, f(x) = -1') + \mathbb{P}('h(x) = -1, f(x) = +1') = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$\Rightarrow \mathbb{P}[h(x) \neq f(x)] = \frac{1}{4} \mathbb{E}[(h(x)-f(x))^2]$$

Also, algebraically:

$$\mathbb{P}[h(x) \neq f(x)] = \frac{1}{4} \mathbb{P}[h(x) \neq f(x)] \cdot 4 + \frac{1}{4} \mathbb{P}[h(x) = f(x)] \cdot 0 =$$

$$\frac{1}{4} \mathbb{P}[h(x) \neq f(x)] (h(x)-f(x))^2 + \frac{1}{4} \mathbb{P}[h(x) = f(x)] (h(x)-f(x))^2 =$$

$$\frac{1}{4} \mathbb{E}[(h(x)-f(x))^2]$$

3. Problem 2.8

Which of the following are possible growth functions $m_{\mathcal{H}}(N)$ for some hypothesis set:

$$1+N, \quad 1+N+\frac{N(N-1)}{2}, \quad 2^N, \quad 2^{\lfloor\sqrt{N}\rfloor}, \quad 2^{\lfloor N/2 \rfloor},$$

$$1+N+\frac{N(N-1)(N-2)}{6}$$

$m_{\mathcal{H}}(N)$ is upper-bounded by 2^N

If d_{vc} is finite, $m_{\mathcal{H}}(N) \leqslant N^{d_{vc}} + 1$ for all N

If $d_{vc} = \infty$, $m_{\mathcal{H}}(N) = 2^N$ for all N

1. If $m_H(N) = 1 + N$, $m_H(1) = 2 = 2^1$ and $m_H(2) = 3 < 2^2$

$\Rightarrow d_{vc} = 1 \Rightarrow m_H(N) \leq N^1 + 1$ for all N (this is true $\Rightarrow$

$m_H(N) = 1 + N$ is a possible growth function.)

2. If $m_H(N) = 1 + N + \frac{N(N-1)}{2} = 1 + \frac{N}{2} + \frac{N^2}{2}$,

$m_H(2) = 4 = 2^2$ and $m_H(3) = 7 < 2^3 \Rightarrow d_{vc} = 2 \Rightarrow$

$m_H(N) \leq N^2 + 1$ for all N (this is true $\Rightarrow$

$m_H(N) = 1 + N + \frac{N(N-1)}{2}$ is a possible growth function)

3. If $m_H(N) = 2^N$, $d_{vc} = \infty$ and $m_H(N) = 2^N$ for all N

(this is true $\Rightarrow$ $m_H(N) = 2^N$ is a possible growth function)

4. If $m_H(N) = 2^{\lfloor \sqrt{N} \rfloor}$, $m_H(1) = 2 = 2^1$ and

$m_H(2) = 2^{\lfloor \sqrt{2} \rfloor} = 2 < 2^2 \Rightarrow d_{vc} = 1 \Rightarrow$

$m_H(N) \leq N^1 + 1$ for all N (this is not true for all N (e.g. if $N = 25$, $2^{\lfloor \sqrt{25} \rfloor} = 32 \not\leq 26$) $\Rightarrow$

$m_H(N) = 2^{\lfloor \sqrt{N} \rfloor}$ is not a possible growth function)

5. If $m_H(N) = 2^{\lfloor N/2 \rfloor}$, $m_H(0) = 1 = 2^0$ and

$m_H(1) = 2^{\lfloor 1/2 \rfloor} = 1 < 2^1 \Rightarrow d_{vc} = 0 \Rightarrow$

$m_H(N) \leq N^0 + 1 = 2$ for all N (this is not true

for all N (starting from $N=4$) $\Rightarrow m_H(N) = 2^{\lfloor N/2 \rfloor}$ is

not a possible growth function)

6. If $m_H(N) = 1 + N + \frac{N(N-1)(N-2)}{6} = \frac{N^3}{6} - \frac{N^2}{2} + \frac{4N}{3} + 1$,

$m_H(1) = 2 = 2^1$ and $m_H(2) = 3 < 2^2 \Rightarrow d_{vc} = 1 \Rightarrow$

$m_H(N) \leq N^1 + 1$ for all N (this is not true for all N

(starting from $N=3$) $\Rightarrow m_H(N) = 1 + N + \frac{N(N-1)(N-2)}{6}$ is

not a possible growth function