CSCI 6100 Machine Learning From Data
Fall 2018

HOMEWORK 3

Daniel Southwick
661542908
southd@rpi.edu

Exercise 1.13

(a)There will be two cases when $h(x)$ makes an error approximating $f(x)$. One is when $y = f(x)$ and $h(x)$ approximates correctly and one is when $y \neq f(x)$ and $h(x)$ approximates incorrectly. In both cases $h(x)$ will approximates $f(x)$. So the error it makes:
$P[error] = \mu * \lambda + (1 - \mu) * (1 - \lambda)$

(b)When the performance of $h(x)$ is independent of μ, we know that $P[error]$ must be independent of μ
$P[error] = \mu * \lambda + (1 - \mu) * (1 - \lambda) = 2 * \mu * \lambda - \mu - \lambda + 1$
So $2 * \mu * \lambda - \mu = 0$ Thus $\mu = 0.5$

Exercise 2.1

(a)Positive Rays: $m_H(N) = N + 1$. $m_H(1) = 1 + 1 = 2$, $m_H(2) = N + 2 = 3 < 2^2 = 4$, so 2 is a breakpoint for H

(b)Positive Intervals: $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$, $m_H(2) = \frac{1}{2} \times 2^2 + \frac{1}{2} \times 2 + 1 = 4 = 4$, $m_H(3) = \frac{1}{2} \times 3^2 + \frac{1}{2} \times 3 + 1 = 4 = 7 < 2^3 = 8$, so 3 is a breakpoint for H

(c)Convex Set: $m_H(N) = 2^N = 2^N$. There's no breakpoints since $m_H(N) = 2^N$

Exercise 2.2

(a)
(i)Positive Rays: From Exercise 2.1, $m_H(N) = N + 1$ with break point $k = 2$,

$$m_H(N) = 2 + 1 = 3 \leq \binom{N}{0} + \binom{N}{1} = N + 1 = 3$$

Thus the theorem holds

(ii)Positive Intervals: From Exercise 2.1, $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$ with break point $k = 3$,

$$m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1 = 7 \leq \binom{N}{0} + \binom{N}{1} + \binom{N}{2} = \frac{1}{2}N^2 + \frac{1}{2}N + 1 = 7$$

Thus the theorem holds

(iii)Convex Sets: From Exercise 2.1, $m_H(N) = 2^N$ with no break point. Thus no bound exists.

(b) $m_H(N) = N+2^{\lfloor N/2 \rfloor}$. There exists breakpoints k such that $m_H(k) < 2^k$. But from the theorem, for $N \leq k$, we can find a polynomial bound on $m_H(N)$ so $\lim_{x\to\infty} \frac{1}{N} log m_H(N) = 0$, but for actual $m_H(N), \lim_{x\to\infty} \frac{1}{N} log m_H(N) = \frac{1}{2} log 2$, which is inconsistence with the hypothesis. Thus, $m_H(N)$ cannot be bounded by any polynomial and there does not exist any hypothesis set.

Exercise 2.3

(a)Positive Rays: From Exercise 2.1, Smallest Breakpoint is 2, then $d_{vc} = 2 - 1 = 1$

(b)Positive Intervals: From Exercise 2.1, Smallest Breakpoint is 3, then $d_{vc} = 3 - 1 = 2$

(c)Convex Sets: From Exercise 2.1, Smallest Breakpoint is 2, then $d_{vc} = \infty - 1 = \infty$

Exercise 2.6

(a)$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$
For the testing error (E_{out}):
$M = 1, N = 200, \delta = 0.05$, so error bound $\sqrt{\frac{1}{2 \times 200} \ln \frac{2 \times 1}{\delta}} = 0.096$

For the training error (E_{in}):
$M = 1000, N = 400, \delta = 0.05$, so error bound $\sqrt{\frac{1}{2 \times 400} \ln \frac{2 \times 1000}{\delta}} = 0.115$
Thus E_{in} has a higher error bar.

(b) If we reserve more examples for testing, we can find a better fitted g, but then there are also less examples as training set, then the testing of g could be come useless with a large $E_{test}(g)$ and E_{out} could not be approximate to 0 since we did not find a good g to begin with.

Problem 1.11

(a)

Supermarket

h \ f	$+1$	-1
$+1$	0	1
-1	10	0

CIA

h \ f	$+1$	-1
$+1$	0	1000
-1	1	0

For the CIA application: the in-sample error:

$$E_{in} = \frac{1}{N}\sum_{i=1}^{N}(1 * C_{[h(x_i)=-1,f(x_i)=+1]} + 1000 * C_{[h(x_i)=+1,f(x_i)=-1]})$$

$(C_{[n]} = 1$ if n is true, else 0.)

For the Supermarket application: the in-sample error:

$$E_{in} = \frac{1}{N}\sum_{i=1}^{N}(10 * C_{[h(x_i)=-1,f(x_i)=+1]} + 1 * C_{[h(x_i)=+1,f(x_i)=-1]})$$

$(C_{[n]} = 1$ if n is true, else 0.)

Problem 1.12

(a)

$$E_{in}(h) = \sum_{i=1}^{N}(h - y_n)^2$$
$$= Nh^2 - 2h\sum_{i=1}^{N} y_n + \sum_{i=1}^{N}(y_n)^2$$
$$= N(h - \frac{1}{N}\sum_{i=1}^{N} y_n)^2 + \sum_{i=1}^{N} y_n^2 - \frac{1}{N}(\sum_{i=1}^{N} y_n)^2$$

Since both $\sum_{i=1}^{N} y_n^2$ and $\frac{1}{N}(\sum_{i=1}^{N} y_n)^2$ terms are constant, so $min(E_{in})$ reach its minimum when $h = \frac{1}{N}\sum_{i=1}^{N} y_n$.

(b)

$$E_{in}(h) = \sum_{n=1}^{N} |h - y_n|$$

We know that N data points are in the order of $y_1 \le y_2 \le y_3 \le ... \le y_N$, we suppose h lies between y_k and y_{k+1}. Now, if $N > 2k$, $y_1, y_2, ..., y_k, h, y_{k+1}, y_{k+2}, ..., y_{N-k+1}, ..., y_N$, then:

$$E_{in}(h) = \sum_{i=1}^{k}(y_{N-i+1} - y_n) + \sum_{i=k+1}^{N-k} |y_i - h|$$

If $N < 2k$, $y_1, y_2, ..., y_{N-k}, y_{N-k+1} ..., y_k, h, y_{k+1}, ..., y_N$, then:

$$E_{in}(h) = \sum_{i=1}^{N-k}(y_{N-i+1} - y_n) + \sum_{i=N-k+1}^{k} |y_i - h|$$

If h moves to the left direction, then the first term decrease and the second term stays the same. So when $E_{in}(h) achieves its minimum, h moves toward the middle position$ So if N is odd, h equals to the median of the N data points, $h = y_{\frac{N+1}{2}}$; If N is even, h is between $y_{\frac{N}{2}}$ and $y_{\frac{N}{2}+1}$. Either case, half the data points are at most h_{med} and half the data points are at least h_{med}, thus the estimate will be in h_{med}

(c) If y_N is perturbed to $y + \epsilon$, where $\epsilon \to \infty$, h_{mean} will grow to unbounded. However, h_{med} will stay the same. h_{med} will always lie between $y_{N/2}$ and $y_{N/2+1}$ when N is even, or equal to $y_{N/2+1}$ when N is odd.