Assignment 3

Alberto Mejia
RIN: 661514960

CSCI 4100 - Machine Learning from Data

September 22, 2019

1. (100) LFD Exercise 1.13

$$P(y \mid x) = \begin{cases} \lambda, & y = f(x) \\ 1 - \lambda, & y \neq f(x) \end{cases}$$

(a)

Hypothesis h makes an error with probability μ

$P(h(x) \neq f(x)) = \mu$

$P(y = f(x)) = \lambda$

$P(y \neq f(x)) = 1 - \lambda$

$P(h(x) = f(x)) = 1 - \mu$

Sum probabilities

$P[h(x) \neq y] = P(y = f(x))P(h(x) \neq f(x)) + P(y \neq f(x))P(h(x) = f(x))$

$P[h(x) \neq y] = \lambda\mu + (1 - \lambda)(1 - \mu)$

(b)

$P[h(x) \neq y] = \lambda\mu + (1 - \lambda)(1 - \mu)$

$= 2\lambda\mu - \mu - \lambda + 1$

$= \mu(2\lambda - 1) - \lambda + 1$

To make h independent of μ, let $2\lambda - 1 = 0$

$2\lambda - 1 = 0$

$2\lambda = 1$

$\lambda = 1/2 = 0.5$

2. (100) LFD Exercise 2.1

Positive rays :

Verify $m_H(k) \leq 2^k$ using $m_H(N) = N + 1$

Therefore, verify $2^k \leq N + 1$

The break point occurs at $k = 2$

For $N = 1$

$2^1 = 1 + 1$

$2 \leq 2$ ✓

For $N = 2$

$2^2 = 2 + 1$

$4 > 3$ Break point!

Positive intervals :

Verify $m_H(k) \leq 2^k$ using $m_H(N) = \binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$

Therefore, verify $2^k \leq \frac{1}{2}N^2 + \frac{1}{2}N + 1$

The break point occurs at k = 3

For N = 1

$2^1 = \frac{1}{2}(1)^2 + \frac{1}{2}(1) + 1$

$2 \leq 2.5$ ✓

For N = 2

$2^2 = \frac{1}{2}(2)^2 + \frac{1}{2}(2) + 1$

$4 \leq 4$ ✓

For N = 3

$2^2 = 2 + 1$

$4 > 3$ Break point!

Convex Sets :

Verify $m_H(k) \leq 2^k$ using $m_H(N) = 2^N$

$m_H(N) \leq 2^N$

Therefore, verify $2^N \leq 2^N$

$2^N = 2^N$

Which means

No break points! ✓

3. (100) LFD Exercise 2.2

(a)

Theorem 2.4. If $m_{\mathcal{H}}(k) < 2^k$ for some value k, then

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$

for all N. The RHS is polynomial in N of degree $k-1$.

(i) Positive rays: From the previous problem, positive rays have a break point at k = 2

$m_H(N) = \sum_{i=0}^{k-1}\binom{N}{i}$

$m_H(N) = N+1 \leq \sum_{i=0}^{k-1}\binom{N}{i}$

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

$N+1 \leq 1 + N$

The theorem holds

$m_H(2) = 2 + 1 \leq 1 + 2$

$= 3 \leq 3$ ✓

(ii) Positive intervals: From the previous problem, positive intervals have a break point at k = 3

$m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1 \leq \sum_{i=0}^{k-1}\binom{N}{i}$

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

$\frac{1}{2}N^2 + \frac{1}{2}N + 1 = 1 + N + \frac{N^2}{2} - N$

$\frac{1}{2}N^2 + \frac{1}{2}N + 1 = 1 + N + \frac{1}{2}N^2 - N$

$\frac{1}{2}N^2 + \frac{1}{2}N + 1 \equiv 1 + N + \frac{1}{2}\cdot N(N-1)$

These are equivalent

The theorem holds

$m_H(3) = \frac{1}{2}(3)^2 + \frac{1}{2}(3) + 1 \leq 1 + 3 + \frac{1}{2}\cdot 3(3-1)$

$= 7 \leq 7$ ✓

(iii) Convex sets: From the previous problem, convex sets have no break points. Therefore, $m_H(N)$ cannot be bounded by an expression.

(b) There does not exist a hypothesis set in which $m_H(N) = N + 2^{\lfloor N/2 \rfloor}$
This is because Theorem 2.4 describes a polynomial function, specifically a summation of polynomials and the expression proposed is an exponential. Exponentials will always grow faster than every polynomial (which can be proved using induction). Thus, it is impossible to find a polynomial to bound that exponential.

4. (100) LFD Exercise 2.3

$k = d_{vc} + 1$

$k - d_{vc} = 1$

$-d_{vc} = 1 - k$

$(-d_{vc})/(-1) = (1 - k)/(-1)$

$d_{vc} = -1 + k$

$d_{vc} = k - 1$

(i) Positive rays:

As seen, positive rays have a break point at k = 2

$d_{vc} = k - 1$

$d_{vc} = 2 - 1$

$d_{vc} = 1$

(ii) Positive intervals:

As seen, positive intervals have a break point at $k = 3$

$d_{vc} = k - 1$

$d_{vc} = 3 - 1$

$d_{vc} = 2$

(iii) Convex sets:

As seen, convex sets do not have a break point

$d_{vc} = k - 1$

$d_{vc} = \infty - 1$

$d_{vc} = \infty$

5. (100) LFD Exercise 2.6

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

(a)

In-sample error E_{in}:

$\delta = 0.5$

$N = 400$

$H = 1000$

Error bound $= \sqrt{\frac{1}{2(400)}\ln(\frac{2(1000)}{(0.05)})} = 0.11509$

Test error E_{test}:

$\delta = 0.5$

$N = 200$

$H = 1$

Error bound $= \sqrt{\frac{1}{2(200)}\ln(\frac{2(1)}{(0.05)})} = 0.09603$

Thus $E_{test} \leq E_{in}(g)$

E_{in} has a higher error bar.

(b) If we reserve even more examples for testing, our result cold be worse. We may find a better hypothesis g and better our in-sample error but then it could increase our out-of-sample error making our results actually worse. This is known as bias variance tradeoff.

6. (200) LFD Problem 1.11

		f	
		+1	−1
h	+1	no error	false accept
	−1	false reject	no error

		f	
		+1	−1
h	+1	0	1
	−1	10	0

Supermarket

		f	
		+1	−1
h	+1	0	1000
	−1	1	0

CIA

Error = E(h,f)
error measure = e(h(x), f(x))
Supposing that there are N learning examples

$$E_{in}(\text{h}) = \frac{1}{N}\sum_{n=1}^{N} e(h(x), f(x))$$

Supermarket:

$$E_{in}(\text{h}) = \sum_{n=1}^{N} e(h(x), f(x))$$

Based on the Supermarket matrix above, $E_{in}(\text{h})$ = the above formula where
e(1,1) = 0
e(1,-1) = 1
e(-1, 1) = 10
e(-1,-1) = 0

Now we compute the sum of the errors, ignoring 0 errors, and we get

$$E_{in}(\text{h}) = \sum_{n=1}^{N} (1 * [h(x_n) = +1, f(x_n) = -1]) + (10 * [h(x_n) = -1, f(x_n) = +1])$$

CIA:

$$E_{in}(\text{h}) = \sum_{n=1}^{N} e(h(x), f(x))$$

Based on the CIA matrix above, $E_{in}(\text{h})$ = the above formula where
e(1,1) = 0
e(1,-1) = 1000
e(-1, 1) = 1
e(-1,-1) = 0

Now we compute the sum of the errors, ignoring 0 errors, and we get

$$E_{in}(\mathrm{h}) = \sum_{n=1}^{N}(1000 \; * \; [h(x_n) = +1, f(x_n) = -1]) + (1 \; * \; [h(x_n) = \mathrm{v} \; -1, f(x_n) = +1])$$

7. (300) LFD Problem 1.12

(a)

$$E_{in}(\mathrm{h}) = \sum_{n=1}^{N}(h - y_n)^2$$

$$h_{mean} = \frac{1}{N}\sum_{n=1}^{N} y_n$$

$E_{in}(\mathrm{h}) = \sum_{n=1}^{N}(h - y_n) \times (h - y_n)$ Expand out exponents

$= \sum_{n=1}^{N}(hh - hy_n - hy_n + y_n y_n)$ Multiply binomials

$= \sum_{n=1}^{N} h^2 - 2hy_n + (y_n)^2$ Simplify algebraic expression

$= \sum_{n=1}^{N} h^2 - \sum_{n=1}^{N} -2hy_n \sum_{n=1}^{N} {y_n}^2$ Apply finite sum definition

$= \sum_{n=1}^{N} h^2 - 2h\sum_{n=1}^{N} -y_n + \sum_{n=1}^{N} {y_n}^2$ Pull out constant terms

$= Nh^2 - 2h\sum_{n=1}^{N} y_n + \sum_{n=1}^{N} {y_n}^2$ Summation of a constant property

$E_{in}(\mathrm{h})$ is a function of this number (variable) h, one way we can minimize this is too take the derivate and set to zero.

$E_{in}(\mathrm{h})' = 2\mathrm{Nh} - 2\sum_{n=1}^{N} y_n$ Compute Derivative

$2\mathrm{Nh} - 2\sum_{n=1}^{N} y_n = 0$ Set to zero

$2\mathrm{Nh} = +2\sum_{n=1}^{N} y_n$ Solve for h

$\mathrm{h} = \frac{2\sum_{n=1}^{N} y_n}{2N}$

$\mathrm{h} = \frac{\sum_{n=1}^{N} y_n}{N}$

$\mathrm{h} = \frac{1}{N}\sum_{n=1}^{N} y_n$

Which is equal to the in-sample mean, or h_{mean}

(b)

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

Again, we'll use Calculus to determine the h that minimizes the in-sample median

$$E_{in}(\mathrm{h})' = (\sum_{n=1}^{N} |\mathrm{h} - y_n|)' \; = \; 0$$

However, because of the absolute value, we will be getting positive terms and thus, the only way to get the derivative to equal zero is if we have the same number of negative terms to cancel out our positive terms. Thus, our hypothesis h has to be the median for this in-sample sum.

(c) If y_N is perturbed to $y_N + \varepsilon$, where $\varepsilon \to \infty$

> h_{mean} will approach infinity as it will follow y_N to positive infinity. The mean is highly sensitive to outliers.
>
> h_{med} will not change as the concept of median is not going to change. The median is basically insensitive to outliers.