CSCI 4100 Assignment 3

Boliang Yang 661541863

1. Exercise 1.13 in LFD

(a) When $h(x) = f(x) \neq y$ or $h(x) \neq f(x) = y$, then h makes error.

Then we have $P[error] = P[h(x) = f(x) \neq y] + P[h(x) \neq f(x) = y] = (1-\mu)(1-\lambda) + \mu\lambda = 1 - \mu - \lambda + 2\mu\lambda$

(b) $P[error] = 1 - \mu - \lambda + 2\mu\lambda = 1 - \lambda + \mu(2\lambda - 1)$, and we want it to be independent of μ, then we have $(2\lambda - 1) = 0$, which gives us $\lambda = \frac{1}{2}$, and $P[error] = 1 - \frac{1}{2} = \frac{1}{2}$.

2. Exercise 2.1 in LFD

(a) Positive rays:

$k = 2$ is the break point since for $k = 2$ dichotomy cannot shatter all possibilities of points. For example, +1 on the left and -1 on the right.

From formula, $m_H(N) = N + 1$, then $m_H(2) = 2 + 1 = 3 < 2^2 = 4$.

(b) Positive intervals:

$k = 3$ is the break point since for $k = 3$ dichotomy cannot shatter all possibilities of points. For example, [+1, -1, +1].

From formula, $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$, then $m_H(3) = \frac{1}{2}\times 3^2 + \frac{1}{2}\times 3 + 1 = 7 < 2^3 = 8$.

(c) Convex sets:

Break point for convex sets does not exist since dichotomy will always shatter all possibilities of points.

From formula, $m_H(N) = 2^N$, then $m_H(N) = 2^N$ for all N.

3. Exercise 2.2 in LFD

(a) (i) Positive rays:

The break point is $k = 2$, the polynomial degree is 1. Since $m_H(N) = 1 + N = \binom{N}{0} + \binom{N}{1} = \sum_{i=0}^{1}\binom{N}{i} \leq \sum_{i=0}^{2-1}\binom{N}{i}$, then $m_H(N) \leq \sum_{i=0}^{k-1}\binom{N}{i}$ holds where $k = 2$.

(ii) Positive intervals:

The break point is $k = 3$, the polynomial degree is 2. Since $m_H(N) = 1 + \frac{1}{2}N + \frac{1}{2}N^2 = 1 + N + \frac{1}{2}N(N-1) = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} = \sum_{i=0}^{2}\binom{N}{i} \leq \sum_{i=0}^{3-1}\binom{N}{i}$, then $m_H(N) \leq \sum_{i=0}^{k-1}\binom{N}{i}$ holds where $k = 3$.

(iii) Convex sets:

Since convex sets don't have a break point, the Theorem 2.4 cannot be applied here.

(b) From class we know there are only two types of hypothesis sets: either be of the form 2^N or polynomial. Since $m_H(N) = N + 2^{\lfloor N/2 \rfloor}$ is neither of those types, then there does not exist such hypothesis set.

Prove by contradiction: $k = 3$ is a break point since $m_H(3) = 3 + 2^{\lfloor 3/2 \rfloor} = 3 + 2 = 5 < 2^3 = 8$. By Theorem 2.4, $m_H(N) \leq \sum_{i=0}^{k-1}\binom{N}{i} = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} = 1 + N + \frac{1}{2}N(N-1)$. But when $N = 14$, $m_H(14) = 14 + 2^{\lfloor 14/2 \rfloor} = 142 > 1 + 14 + \frac{1}{2} \times 14 \times 13 = 106$, which contradicts the Theorem 2.4.

4. Exercise 2.3 in LFD

If k is the smallest break point for H, then $d_{vc} = k - 1$.

(i) Positive rays: since the break point is $k = 2$, then $d_{vc} = 2 - 1 = 1$.

(ii) Positive intervals: since the break point is $k = 3$, then $d_{vc} = 3 - 1 = 2$.

(iii) Convex sets: since $m_H(N) = 2^N$ for all N, then $d_{vc} = \infty$.

5. Exercise 2.6 in LFD

(a) Error bar for E_{in}:

$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N}\ln\frac{2|H|}{\delta}}$ where $N = 400, |H| = 1000, \delta = 0.05$, then

$E_{out}(g) \leq E_{in}(g) + 0.11509$

Error bar for E_{test}:

We have one hypothesis then we can use Hoeffding' inequality for a single fixed hypotheisis. $P[|v-\mu| \geq \epsilon] = 0.05 \leq 2e^{-2\epsilon^2 N}$ where $N = 200$. Then we have $\epsilon \leq 0.09603$.

Thus, the error bar for E_{in} is higher.

(b) If we have a larger test set, then we will have less data for training set. There is a trade-off between samples for E_{in} and E_{test}. In this case, we will not have enough samples for training, and we will have a good E_{in} but worse and wild E_{test} and E_{out}.

6. Problem 1.11 in LFD

For CIA:

$$E_{in} = \frac{1000P[h(x) = +1 \text{ and } f(x) = -1] + P[h(x) = -1 \text{ and } f(x) = +1]}{1001}$$

For supermarket:

$$E_{in} = \frac{10P[h(x) = -1 \text{ and } f(x) = +1] + P[h(x) = +1 \text{ and } f(x) = -1]}{11}$$

7. Problem 1.12 in LFD

(a) $E_{in}(h) = \sum_{n=1}^{N}(h-y_n)^2 = \sum_{n=1}^{N}(h^2 - 2hy_n + y_n^2) = Nh^2 - 2h\sum_{n=1}^{N} y_n + \sum_{n=1}^{N} y_n^2$

If we want to minimizes the E_{in}, we need $\frac{dE_{in}(h)}{dh} = 0$. Then $\frac{dE_{in}(h)}{dh} = 2Nh - 2\sum_{n=1}^{N} y_n = 0$, which gives us $h = \frac{1}{N}\sum_{n=1}^{N} y_n = h_{mean}$.

(b) $E_{in}(h) = \sum_{n=1}^{N} |h - y_n|$, again, we need $\frac{dE_{in}(h)}{dh} = 0$ in order to minimize the E_{in}.

Then $\frac{dE_{in}(h)}{dh} = \sum_{n=1}^{N} \frac{d|h-y_n|}{dh} = \sum_{n=1}^{N} \frac{d|h-y_n|}{d(h-y_n)} \times \frac{d(h-y_n)}{dh} = \sum_{n=1}^{N} \frac{d|h-y_n|}{d(h-y_n)} = \frac{d|h-y_1|}{d(h-y_1)} + \frac{d|h-y_2|}{d(h-y_2)} + \dots + \frac{d|h-y_N|}{d(h-y_N)}$. Each of the fractions is either +1 or -1 since $d(h-y_n)$ can be positive or negative. And since $\frac{dE_{in}(h)}{dh} = \frac{d|h-y_1|}{d(h-y_1)} + \frac{d|h-y_2|}{d(h-y_2)} + \cdots + \frac{d|h-y_N|}{d(h-y_N)} = 0$, then half of the data points are at most h, and half of the data points are at least h, here h

should be the median h_{med}.

(c) Since $h_{mean} = \frac{1}{N}\sum_{n=1}^{N} y_n$ is the average of sum of y_n, then when y_N is perturbed to $y_N + \epsilon$ where $\epsilon \to \infty$, h_{mean} will grow more and more and $h_{mean} \to \infty$.

But since h_{med} is the median of all the data points, when y_N is perturbed to $y_N + \epsilon$ where $\epsilon \to \infty$, h_{med} will shift at most by one. If original y_N is below h_{med}, then h_{med} will increase by one point; if original y_N is above h_{med}, then h_{med} will not change.