Homework 3

Karan Sarkar
sarkak2@rpi.edu

September 23, 2019

Exercise 1.13.

(a) We are looking for $P[h(x) \neq y]$. We will consider two cases. First, suppose that $h(x) = f(x)$. This occurs with probability $1-\mu$. Now note that $y \neq f(x)$ with probability $1-\lambda$. Thus the case probability is $(1-\mu)(1-\lambda)$. For the second case, suppose that $h(x) \neq f(x)$. This occurs with probability μ. Now note that $y = f(x)$ with probability λ. Thus the case probability is $\mu\lambda$. The total probability is $1 - \lambda - \mu + 2\lambda\mu$.

(b) We are looking for the value of λ at which μ is irrelevant. The probability of error is $1 - \lambda - \mu + 2\lambda\mu = 1 - \lambda + \mu(2\lambda - 1)$. Thus, the cofficient for μ is zero when $\lambda = 0.5$.

Exercise 2.1.

1. A break point for positive rays occurs at $k = 2$. From the formula given, $m_{\mathcal{H}}(N) = N + 1$, thus $m_{\mathcal{H}}(2) = 3 < 2^2$.
2. A break point for positive intervals occurs at $k = 3$. From the formula given, $m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$, thus $m_{\mathcal{H}}(3) = 7 < 2^3$.
3. A break point for convex sets occurs at $k = \infty$. From the formula given, $m_{\mathcal{H}}(N) = 2^N$ so there is no finite break point.

Exercise 2.2.

(a)
1. For positive rays, the bound is tight. $m_{\mathcal{H}}(N) = N + 1 = \binom{N}{1} + \binom{N}{0}$.
2. The bound holds for positive intervals.

$$\begin{aligned}\binom{N}{2} + \binom{N}{1} + \binom{N}{0} &= \frac{1}{2}N(N-1) + N + 1 \\ &= \frac{1}{2}N^2 - \frac{1}{2}N + N + 1 \\ &= \frac{1}{2}N^2 + \frac{1}{2}N + 1 = m_{\mathcal{H}}(N)\end{aligned}$$

3. The bound holds for convex sets. The bound from Sauer's Lemma is $\sum_{i=0}^{N}\binom{N}{i}$. Note that this sums up the number of ways to pick from a set of size N a subset of size 0, 1, 2 all the way to N. In other words, we are looking for the number of ways to pick a subset from a set of size N. Because every element has two states in or out, we have $\sum_{i=0}^{N}\binom{N}{i} = 2^N = m_{\mathcal{H}}$.

(b) There exists no such hypothesis set. Suppose that $m_{\mathcal{H}} = N + 2^{\rfloor N/2 \lfloor}$. Note that $m_{\mathcal{H}}$ has a break point at $N = 3$. From Sauer's Lemma, we know that it is now less than $\binom{N}{3} + \binom{N}{2} + \binom{N}{1} + \binom{N}{0} < N^3 + 1$. Thus, the given growth function was impossible.

Exercise 2.3.

1. The smallest break point for positive rays occurs at $k = 2$. Thus, the VC-dimension is 1.

2. The smallest break point for positive intervals occurs at $k = 3$. Thus, the VC-dimension is 2.

3. The smallest break point for convex sets occurs at $k = \infty$. Thus, the VC-dimension is ∞.

Exercise 2.6.

(a) We get that the in-sample error bar is $\sqrt{\frac{1}{2N}\ln\left(\frac{2\mathcal{H}}{\delta}\right)} = \sqrt{\frac{1}{2\cdot 400}\ln\left(\frac{2\cdot 1000}{0.05}\right)} \approx 0.115$. We get that the test error bar is $\sqrt{\frac{1}{2N}\ln\left(\frac{2\mathcal{H}}{\delta}\right)} = \sqrt{\frac{1}{2\cdot 200}\ln\left(\frac{2\cdot 1000}{0.05}\right)} \approx 0.162$. The in-sample error is smaller than the test error.

(b) If you reserve more data for testing, you get a smaller error bar. However, you sacrifice approximation and E_{in} might not be as low even though the error bar is reduced.

Problem 1.11. We can calculate the error for the supermarket.

$$\begin{aligned}
E_{\text{in}}^S &= \frac{1}{N}\sum_{i=1}^{n} E[h(x_n), y_n] \\
&= \frac{1}{N}\left[\sum_{y_n=1}^{n} E[h(x_n), 1] + \sum_{y_n=-1}^{n} E[h(x_n), -1]\right] \\
&= \frac{1}{N}\left[\sum_{y_n=1}^{n} 10 \cdot [h(x_n) \neq 1] + \sum_{y_n=-1}^{n} [h(x_n) \neq -1]\right]
\end{aligned}$$

We can calculate the error for the CIA.

$$\begin{aligned}
E_{\text{in}}^S &= \frac{1}{N}\sum_{i=1}^{n} E[h(x_n), y_n] \\
&= \frac{1}{N}\left[\sum_{y_n=1}^{n} E[h(x_n), 1] + \sum_{y_n=-1}^{n} E[h(x_n), -1]\right] \\
&= \frac{1}{N}\left[\sum_{y_n=1}^{n} [h(x_n) \neq 1] + \sum_{y_n=-1}^{n} 1000 \cdot [h(x_n) \neq -1]\right]
\end{aligned}$$

Problem 1.12.

(a) To minimize $E_{\text{in}}(h)$ we have to find the point such that $E'_{\text{in}}(h) = 0$ with respect to h.

$$\begin{aligned}
E_{\text{in}}(h) &= \sum_{n=1}^{N} (h - y_n)^2 \\
E'_{\text{in}}(h) &= \sum_{n=1}^{N} 2(h - y_n) \\
0 &= \sum_{n=1}^{N} 2h - \sum_{n=1}^{N} 2y_n \\
Nh &= \sum_{n=1}^{N} y_n \\
h &= \frac{1}{N}\sum_{n=1}^{N} y_n = h_{\text{mean}}
\end{aligned}$$

(b) To minimize $E_{\text{in}}(h)$ we have to find the point such that $E'_{\text{in}}(h) = 0$ with respect to h. Let a be the number of y_n such that $y_n < h$. Let b be the number of y_n such that $y_n < h$.

$$
\begin{aligned}
E_{\text{in}}(h) &= \sum_{n=1}^{N} |h - y_n| \\
E_{\text{in}}(h) &= \sum_{y_n<h}^{N} |h - y_n| + \sum_{y_n>h} |h - y_n| \\
E_{\text{in}}(h) &= \sum_{y_n<h}^{N} (h - y_n) + \sum_{y_n>h} (-h + y_n) \\
E'_{\text{in}}(h) &= \sum_{y_n<h}^{N} 1 + \sum_{y_n>h} -1 \\
0 &= a - b \\
a &= b
\end{aligned}
$$

Thus, we have the quantity minimized, when the number of terms less than h is equal to the number of terms greater than h.

(c) When we add an outlier $y_N + \epsilon$ as $\epsilon \to \infty$, we see that $h_{\text{mean}} \to \infty$ whereas h_{med} at most shifts to the next highest value of y.