

EXERCISES

1.13

(a) There are two cases where $h(x) \neq y$. The first case is when $h(x) \neq f(x)$ and $f(x) = y$; the other case is the other mismatch. The probability of the first case is $u * \lambda$; the probability of the second case is $(1 - u) * (1 - \lambda)$. The overall probability of $h(x) \neq y$ is $1 + 2u\lambda - u - \lambda$

(b) $\lambda = 0.5$. In this case the noisy target will be completely random.

2.1

Positive Rays: The break point is $k = 2$.

$$2^1 = 1 + 1$$

,

$$2^2 > 2 + 1$$

Positive Intervals: The break point is $k = 3$.

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

$$2^3 > \frac{1}{2}(3^2 + 3) + 1$$

Convex Set: There is no break point.

2.2

(ai) Positive Rays: The break point is $k = 2$, the order of growth function $m_H(N) = N + 1$ is 1; consistent with Theorem 2.4.

(aii) Positive Intervals: The break point is $k = 3$, the order of growth function $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$ is 2; consistent with Theorem 2.4.

(aiii) Convex Set: Since there is no break point, Theorem 2.4 does not apply.

(b) No. According to Theorem 2.4, if $m_H(N) < 2^k$ for some break point k , it will turn into polynomial, so there shouldn't be any case such that

$$m_H(N) = N + 2^{\lfloor \frac{N}{2} \rfloor}$$

2.3

$$(i) d_{VC} = k - 1 = 1, N + 1 = 2 = 2^N$$

$$(ii) d_{VC} = k - 1 = 2, \frac{1}{2}(2^2 + 2) + 1 = 4 = 2^N$$

$$(iii) d_{VC} = k + 1 = \infty$$

2.6

(a)

$$\text{Error bar for } E_{out} : \sqrt{\frac{1}{2 \times 400} \ln\left(\frac{2 \times 1000}{0.005}\right)} = 0.1151$$

$$\text{Error bar for } E_{test} : \sqrt{\frac{1}{2 \times 200} \ln\left(\frac{2 \times 1}{0.005}\right)} = 0.0960$$

E_{out} has a higher error bar.

(b)

Yes. The test evaluates the training with a error bar E_{test} but higher E_{out} , indicating that our training is more likely to fail. So in this case we are trying to use a more accurate test to verify a training output g that is more likely to fail. If we reserve too many data for test we would simply be verifying a failure with high confidence, which won't help us a lot.

PROBLEMS

1.11

$$E_{in-CIA} = \frac{1}{N} \sum_{n=1}^N e(h(x_n), f(x_n))$$

where $e(1, -1) = 1000$, $e(1, -1) = 1$, $e(1, 1) = 0$ and $e(-1, -1) = 0$

$$E_{in-market} = \frac{1}{N} \sum_{n=1}^N e(h(x_n), f(x_n))$$

where $e(1, -1) = 1$, $e(1, -1) = 10$, $e(1, 1) = 0$ and $e(-1, -1) = 0$

1.12

(a)

Suppose we have $h = h_{mean} + x$, where y_{mean} is the mean of $y_1, y_2 \dots y_n$

Arrange the E_{in} equation to get:

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

Substitute h with h_{mean} and get

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

Since $\sum_{n=1}^N y_n^2$ is fixed, the goal is to minimize the rest of equation:

$$(h_{mean} + a) \sum_{n=1}^N ((h_{mean} + a) - 2y_n)$$

Since h_{mean} is the sample mean, we have $\sum_{n=1}^N y_n = N \times h_{mean}$ and hence

$$(h_{mean} + a) \sum_{n=1}^N ((h_{mean} + a) - 2y_n) = (h_{mean} + a) \times (N(h_{mean} + a) - 2N \times h_{mean}) = a^2 - h_{mean}^2$$

It is obvious that to minimize the result we need to have $a = 0$, in which case $h = h_{mean}$

(b)

Suppose there are X samples larger than h and Y samples smaller than h .

If we increase h by a without getting pass the next sample, there will be X sample deviations increasing by a and Y sample deviations decreasing by a .

The change of E_{in} is $a \times (X - Y)$

If we decrease h by a without getting pass the next sample, similarly, The change of E_{in} is $a \times (Y - X)$

We have three cases:

Case 1: $X < Y$, $h > h_{med}$.

Case 2: $X > Y$, $h > h_{med}$.

Case 3: $X = Y$, $h = h_{med}$.

These cases hold for both even and odd number of samples.

Consider case 1, if we move h closer to h_{med} by a without passing the next sample, that is, we increase h by a , since $X < Y$, $a \times (X - Y) < 0$, E_{in} decreases. Same thing happens for Case 2: moving h closer to h_{med} decreases E_{in} .

For case 3, when $h = h_{med}$, for odd number of samples, h is the middle y_n and moving h in either positive or negative direction would result in case 1 or 2 and thus cause E_{in} to increase. For even number of samples, as long as h is within the middle two samples, E_{in} won't change since the shift would cause equal amount of increase and decrease in smaller-sample deviations and larger-sample deviations ($X = Y$). In conclusion, when $h \neq h_{med}$, moving h closer to h_{med} can decrease E_{in} , thus $h = h_{med}$ is when E_{in} is at minimum.

(c) An infinitely large outlier would ruin the least square error sum algorithm since it will pull the mean to infinity.

An infinitely large outlier wouldn't affect the absolute error sum algorithm since the median is determined by the order of samples. Since y_n is already the maximum of all samples, making it infinity won't change the sample order, and thus won't affect the algorithm.