Assignment 3

Darwin Ding

September 22, 2016

Exercise 1.13

(a) h will fail to approximate y if $h(x) = f(x) \neq y$ or $h(x) \neq f(x) = y$. For the first case, $P[h(x) = f(x)] = 1 - \mu$ and $P[f(x) \neq y] = 1 - \lambda$, so $P[h(x) = f(x) \neq y] = (1 - \mu)(1 - \lambda)$.

For the second case, $P[h(x) \neq f(x)] = \mu$ and $P[f(x) = y] = \lambda$, so $P[h(x) \neq f(x) = y] = \mu\lambda$.

Therefore, $\boldsymbol{P[error] = (1 - \mu)(1 - \lambda) + \mu\lambda}$.

(b) When $\lambda = 0.5$, the $P[error]$ from the previous part $= (1-\mu)(1-0.5) + 0.5\mu = 0.5(1-\mu+\mu) = 0.5$. At this probability, μ is not even present in the formula for the error in approximation, and since $P[error] = 0.5$, the noisy target is completely random.

Problem 1.11

Assume for an input data set size of N:
a = <number of input data points where h(x) = 1 and f(x) = 1>
b = <number of points where h(x) = 1 and f(x) = -1>
c = <number of points where h(x) = -1 and f(x) = 1>
d = <number of points where h(x) = -1 and f(x) = -1>
By definition, N = a + b + c + d.

We want to create an E_{in} function where all the above categories are weighted properly according to the matrices given in the chapter. This E_{in}

function should also vary from 0 to 1.

The resultant $E_{in} = (a*w_a + b*w_b + c*w_c + d*w_d)/(N*max(w_a, w_b, w_c, w_d))$ should do all of the above. w_a, w_b, w_c, w_d are all the weights given in the matrix in the chapter.

For the supermarket, $\boldsymbol{E_{in} = (b + 10c)/(10N)}$
For the CIA, $\boldsymbol{E_{in} = (1000b + c)/(1000N)}$

Problem 1.12

(a)

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

Since we're trying to find the minimum of such E_{in}, we can take the derivative with respect to h and set that derivative to 0 to find which h gives the smallest E_{in}.

$$\frac{dE_{in}(h)}{dh} = 2Nh + \sum_{n=1}^{N}(-2y_n)$$

$$= 2Nh + N - 2\sum_{n=1}^{N} y_n$$

$$\frac{dE_{in}(h)}{dh} = 0 = 2Nh - 2\sum_{n=1}^{N} y_n$$

$$-2Nh = -2\sum_{n=1}^{N} y_n$$

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

(b)

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

$$= |h - y_1| + |h - y_2| + ... + |h - y_N|$$

Again, we find the minimum by taking a derivative with respect to h.

$$\frac{dE_{in}(h)}{dh} = \frac{d|h - y_1|}{dh} + \frac{d|h - y_2|}{dh} + ... + \frac{d|h - y_N|}{dh}$$

$d|x|/dx = |x|/x$ and $d(x - y_n)/dx = 1$ for all n, so we can use the chain rule to derive the individual derivatives of the absolute values in the summation above.

$$\frac{dE_{in}(h)}{dh} = \frac{|h - y_1|}{h - y_1} + \frac{|h - y_2|}{h - y_2} + \frac{|h - y_3|}{h - y_3} + ... + \frac{|h - y_N|}{h - y_N}$$

Each of the fractions above has value either +1 (if $x - y_n > 0$)or -1 (if $x - y_n < 0$). In order to get to zero, half of the values have to be above h and half the values need to be below h.

(c) As y_N approaches positive infinity, h_{mean} grows more and more as its sum increases, despite y_N being an outlier. However, h_{median} is not affected much at all due to the nature of medians naturally ignoring outliers. h_{median} may change, however, just due to the fact that if y_N used to be below the median and then became positive infinity, the median may increase by a point.

Exercise 2.1

(a) Positive rays get broken really early on. At N = 0, you can put the ray anywhere and it'll make sense. At N = 1, putting the ray is still trivial. If the one point is +1, start the ray to the point's left. Otherwise, start it at its right.

However, at N = 2 things get a little trickier. No matter where you put the two points on the number line, if the one on the left is +1 and the one on the right is -1, there is no possible way to positive ray the line. However, it is pretty trivial to ray the other 3 dichotomies. Since $3 < 4 = 2^2$, N = **2 is a break point for the positive rays**.

This coincides with the growth function derived in the question, as $m_H(N) = N + 1 < 2^N$ for $N >= 2$.

(b) Positive intervals also get broken pretty quickly. N = 0 is again trivial. N = 1 is also pretty trivial, you simply surround the point if it is +1, or avoid it if it is -1. For N = 2, it is still pretty easy to interval the points. For two -1s, simply avoid both points. For one +1 and one -1, you just interval tightly around the +1. For two +1s, the interval just needs to hold both points.

However, at N = 3, you can no longer interval the points consistently. You can interval all of the dichotomies pretty easily until you run into [+1, -1, +1]. Unfortunately, any interval that contains all the +1s here will by definition contain the -1 in the middle, which breaks the rule. Therefore N = **3 is a break point for positive intervals**.

This is in accordance with the growth function derived, as $m_H(2) =$

$\binom{2+1}{2} + 1 = 3 + 1 = 2^2$, but $m_H(3) = \binom{3+1}{2} + 1 = 7 < 8 = 2^3$.

(c) Convex sets are, as the growth formula implies, impossible to break. If you adhere to the strong strategy of placing all of the points equidistant around a circle, no two lines connecting any of those points can ever leave the set. Because this axiom about points on a circle never really ends, **there is no break point for convex sets**.

This doesn't really need to be verified because $2^N = 2^N$ for all N.

Exercise 2.2

(a) (a) Positive rays broke at k = 2, so we can plug this into the bound and compare to the calculated growth function:

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

Clearly, this holds.

(b) Positive intervals broke at k = 3, so plugging that into the bound and comparing to the growth function given:

$$\begin{aligned} m_H(N) = .5N^2 + .5N + 1 \leq \binom{N}{0} + \binom{N}{1} + \binom{N}{2} \\ = 1 + N + N(N-1)/2 \\ = 1 + N + .5N^2 - N/2 \\ = 1 + .5N + .5N^2 \end{aligned}$$

Clearly, this also holds.

(c) We cannot apply this bound here, because $m_H(N) = 2^N$ for all N. Convex sets did not break.

Exercise 2.3

Since the VC dimension is pretty much defined as the break point $k - 1$:

(a) Positive rays broke at k = 2, so the VC dimension is **1**

(b) Positive intervals broke at k = 3, so the VC dimension is **2**

(c) Convex sets did not break, so the VC dimension is $\boldsymbol{\infty}$

Exercise 2.6

(a) Our error bar function is $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N} log \frac{2|H|}{\delta}}$. With the same size hypothesis set (H = 1000 for both), and δ being the same for both (0.05), the lower N for E_{test} means that it has a lower error bar. However, the lower N also means that E_{test} might be more wild.

(b) By having a larger test set, you will have less examples used for your actual training set. As a result, $E_{test} \approx E_{out}$ but E_{test} has no guarantees of lowness.