Assignment 3

Darwin Ding

September 21, 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 due to the nature of medians naturally ignoring outliers.