Homework 4 - LEARNABILITY
2015 Spring, Machine Learning
Choong-Wan Woo
March 2, 2015

Problem 2.3.

Concentric circles.

With concentric circles as concepts, we can define one annulus error area, E (see **Figure 1**), that has the probability of falling in this region, ϵ. The annulus area, E, can be defined as $E = \{(x, y) : e^2 \leq x^2 + y^2 \leq r^2\}$, with $e = \inf\{e : \Pr[\pi(r^2 - e^2)] \geq \epsilon\}$

By contraposition, if $R(\mathrm{R}_S) > \epsilon$ (here, $R(\mathrm{R}_S)$ denotes *the expected error of* R_S), then R_S must miss the error region E. As a result, we can write

$$\begin{aligned} \Pr[R(\mathrm{R}_S) > \epsilon] &\leq \Pr[\{\mathrm{R}_S \cap E = \emptyset\}] \\ &\leq (1 - \epsilon)^m \\ &\leq e^{-\epsilon m} \end{aligned}$$

For any $\delta > 0$, to ensure that $\Pr[R(\mathrm{R}_S) > \epsilon] \leq \delta$, we can impose

$$e^{-\epsilon m} \leq \delta$$

If we solve this in terms of m, we get

$$m \geq \frac{1}{\epsilon} \log \frac{1}{\delta}$$

Thus, for any $\epsilon > 0, \delta > 0$, if the sample size m is greater than $\frac{1}{\epsilon} \log \frac{1}{\delta}$, then $\Pr[R(\mathrm{R}_S) > \epsilon] \leq \delta$. Therefore, this class can be (ϵ, δ)-PAC-learnable from training data size $m \geq \frac{1}{\epsilon} \log \frac{1}{\delta}$

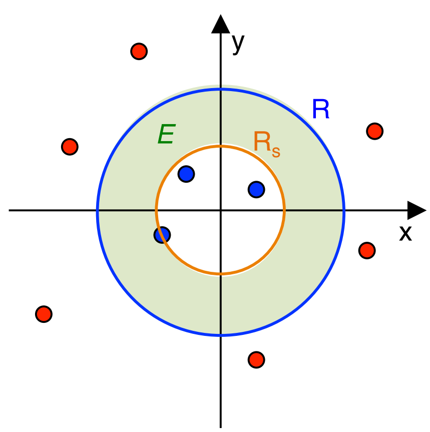

Figure 1. Illustration of the concentric circle case.

Problem 2.4.

Non-concentric circles. Can you tell Gertrude if her approach works?

Gertrude's approach will not work because her approach relies on wrong assumptions. Most importantly, with non-concentric circles, false positive errors can be made outside of a target concept, R, as demonstrated in **Figure 2**. Therefore, drawing three regions, r_1, r_2, r_3, inside R and defining the regions in terms of ϵ is not useful, and certainly, those three regions, r_1, r_2, r_3, do not have the equal probabilities of $\epsilon/3$.

In addition, we can think of a counterexample of Gertrude's approach. As **Figure 2** shows, even though training data do not miss r_1, r_2, r_3 regions, it can still make the generalization error. Thus the equation from her approach $\Pr[R(\mathrm{R}_S) > \epsilon] \leq \Pr[\bigcup_{i=1}^{3}\{\mathrm{R}_S \bigcap r_i = \emptyset\}]$, which is similar to the equation 2.5 of the textbook, should be wrong in this non-concentric circle case.

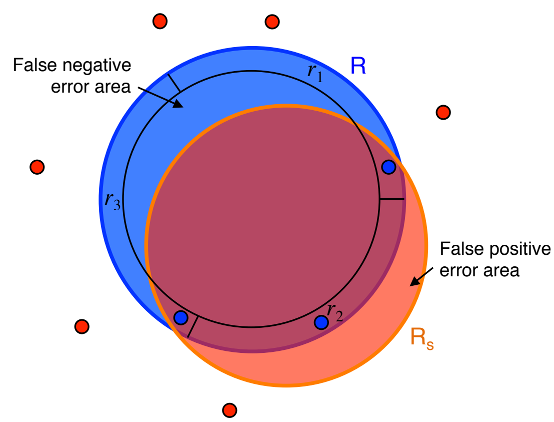

Figure 2. Illustration of the non-concentric circle case.

Problem 2.6.

Learning in the presence of noise-rectangles.

(a) The probability that R' misses a region r_j equals the sum of 1) the probability that no data in R' fall on a region r_j and 2) the probability that the positive training point that fall on a region r_j is flipped to negative with probability η'.

$$\begin{aligned}
\Pr[\mathrm{R}' \text{ misses } r_j] &= \Pr[\{\mathrm{R}' \cap r_j = \emptyset\}] + \eta' \Pr[\{\mathrm{R}' \cap r_j \neq \emptyset\}] \\
&= 1 - \frac{\epsilon}{4} + \eta' \times \frac{\epsilon}{4} \\
&= 1 + \frac{(\eta' - 1)\epsilon}{4}
\end{aligned}$$

(b) The upper bound on $\Pr[R(\mathrm{R}') > \epsilon]$ can be derived as follows:

$$\begin{aligned}
\Pr[R(\mathrm{R}') > \epsilon] &\le \Pr_{S\sim D^m}[\bigcup_{i=1}^{4}\{\mathrm{R}' \text{ misses } r_j\}] \\
&\le \sum_{i=1}^{4} \Pr_{S\sim D^m}[\{\mathrm{R}' \text{ misses } r_j\}] \\
&\le 4(1 + \frac{(\eta' - 1)\epsilon}{4})^m \\
&\le 4e^{(\eta'-1)m\epsilon/4}
\end{aligned}$$

For any $\delta > 0$, to ensure that $\Pr[R(\mathrm{R}') > \epsilon] \le \delta$, we can impose

$$4e^{(\eta'-1)m\epsilon/4} \le \delta$$

, solving for m:

$$m \ge \frac{4}{(1-\eta')\epsilon} \ln \frac{4}{\delta}$$

Problem 3.5.

Consider the really simple hypotheses that always provide 1 or -1. Then, VC-dimension of the hypothesis set is 1 because the functions cannot shatter two points. In this case, Prof. Jesetoo's new bound becomes

$$\mathfrak{R}_m(H) \le O(\frac{1}{m})$$

From eq. 3.31-3.32 on page 48 of the textbook, we know that

$$\hat{\mathfrak{R}}_m(H) \le O\Big(\sqrt{\frac{\log m}{m}}\Big)$$

From eq. 3.14 on page 36 of the textbook,

$$\begin{aligned}
\mathfrak{R}_m(H) &\le \hat{\mathfrak{R}}_m(H) + \sqrt{\frac{\log \frac{2}{\delta}}{2m}} \\
&\le O\Big(\sqrt{\frac{\log m}{m}}\Big)
\end{aligned}$$

With $m \ge 1$, $\sqrt{\frac{\log m}{m}}$ is always greater than $O(\frac{1}{m})$.

Problem 3.12.

VC-dimension of sine functions.

(a) To solve this problem, we need to think about the pattern of sine values of $x, 2x, 3x$, and $4x$. $\sin(2\omega x)$ has the twice larger frequency than $\sin(\omega x)$, $\sin(3\omega x)$ has the three times larger frequency than $\sin(\omega x)$, and $\sin(4\omega x)$

has the four-times larger frequency than $\sin(\omega x)$ and the twice larger frequency than $\sin(2\omega x)$. Therefore, the exact pattern of four sine values is repeated for x, as **Figure 3** shows (in **Figure 3**, *the repetitive pattern in red* points out the pattern). **Figure 3** shows twelve possible combinations of signs of four points that can be shattered by sine functions. Given that there are 16 possible combinations of signs in four points ($2^4 = 16$), there are four sign combinations that cannot be shattered by the sine functions, including $(+,+,-,+), (-,-,+,-), (+,-,-,-)$, and $(-,+,+,+)$. Therefore, the points, $x, 2x, 3x$, and $4x$ cannot be shattered by this family of sine functions.

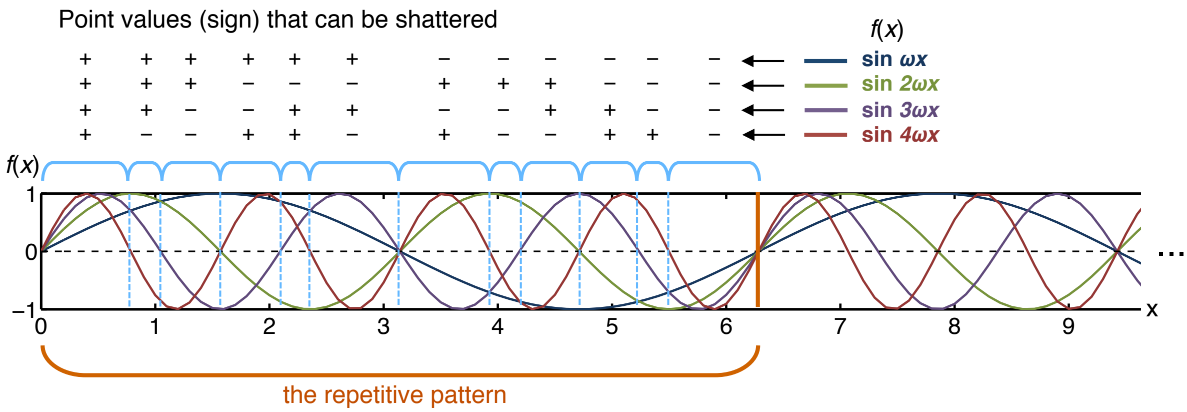

Figure 3. Point values (signs) that can be shattered by four sine functions $(\sin(\omega x), \sin(2\omega x), \sin(3\omega x), \sin(4\omega x))$. The maximum number of possible combinations of signs of four points is $2^4 = 16$, and the number of cases that can be shattered by sine functions are twelve. Therefore, there are four combinations that cannot be shattered by the family of sine functions.

(b)
According to the hint, consider $x_i = 2^{-i}$ for $i \leq m$. If we can find the correct parameter ω that is the function of y_i, such that ωx_i yields the values between $-\pi < \omega x_i < 0$ or $\pi < \omega x_i < 2\pi$ when $y_i = -1$, and the values between $0 < \omega x_i < \pi$ when $y_i = +1$. This was my approach to this problem, but it was difficult to find the right ω value.

Problem 3.13. (or 3.6 in the new version of textbook)
VC-dimension of the union of k intervals.

The answer is $2k$. The union of k intervals can shatter $2k$ points in one the real line, as **Figure 4a** suggests. In **Figure 4a**, I showed the most complicated case using $2k$ points (where $+1$ and -1 alternates) and how the union of k intervals can shatter it. Then, we can add one additional $+1$ point after the last point of $2k$ points. The $(2k+1)$-th point cannot be shattered with k intervals, as **Figure 4b** suggested. Therefore, the VC-dimension of the union of k intervals is $2k$.

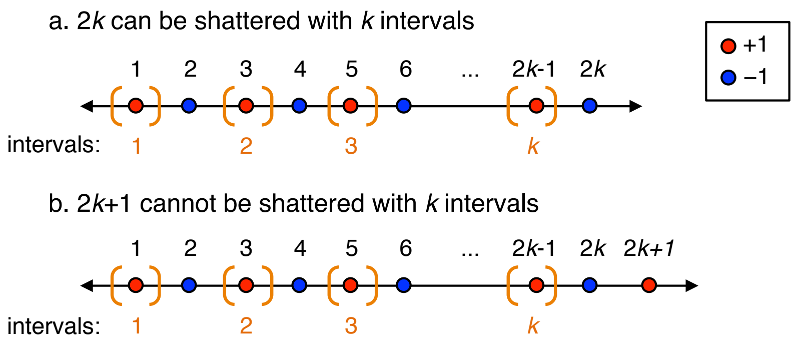

Figure 4. VC-dimensions of the union of k intervals.

Problem 3.19.

Biased coins.

(a)

$$\begin{aligned} error(f_o) &= \Pr[f_o(S) \neq x] \\ &= \Pr[f_o(S) = x_A | x = x_B] \Pr[x = x_B] + \Pr[f_o(S) = x_B | x = x_A] \Pr[x = x_A] \end{aligned}$$

where $\Pr[x = x_A] = \Pr[x = x_B] = 1/2$. Therefore,

$$= \frac{1}{2} \Pr[f_o(S) = x_A | x = x_B] + \frac{1}{2} \Pr[f_o(S) = x_B | x = x_A]$$

where $f_o(S) = x_A$ is same with $N(S) < m/2$, and $f_o(S) = x_B$ is same with $N(S) \geq m/2$ by definition. Therefore,

$$\begin{aligned} &= \frac{1}{2} \Pr[N(S) < \frac{m}{2} | x = x_B] + \frac{1}{2} \Pr[N(S) \geq \frac{m}{2} | x = x_A] \\ &\geq \frac{1}{2} \Pr[N(S) \geq \frac{m}{2} | x = x_A] \end{aligned}$$

(b)

Tossing a coin (x_A) follows a binomial distribution with $p = (1 - \epsilon)/2$. Slud's inequality (D.3.1) on page 374 of the textbook says, if $p \leq 1/2$ and $mp \leq k \leq m(1-p)$,

$$\Pr[B \geq k] \geq \Pr[N \geq \frac{k - mp}{\sqrt{mp(1-p)}}]$$

$\frac{1}{2}\Pr[N(S) \geq \frac{m}{2}|x = x_A]$ in (a) can be formulated as $B(m,p) \geq k$, where $k = m/2$ and $p = (1-\epsilon)/2$. Therefore,

$$\begin{aligned}
error(f_o) &\geq \frac{1}{2}\Pr[N(S) \geq \frac{m}{2}|x = x_A] \\
&\geq \frac{1}{2}\Pr[N \geq \frac{\frac{m}{2} - m(\frac{1-\epsilon}{2})}{\sqrt{m(\frac{1-\epsilon}{2})(\frac{1+\epsilon}{2})}}] \\
&= \frac{1}{2}\Pr[N \geq \frac{m\epsilon/2}{\sqrt{m(1-\epsilon^2)/4}}] \\
&= \frac{1}{2}\Pr[N \geq \frac{\sqrt{m}\epsilon}{\sqrt{1-\epsilon^2}}]
\end{aligned}$$

According D.3.2. on page 374 of the textbook, if N is a random variable following the normal distribution, then for $u > 0$,

$$\Pr[N \geq u] \geq \frac{1}{2}(1 - \sqrt{1 - e^{-u^2}})$$

From the previous equation, we can consider $u = \frac{\sqrt{m}\epsilon}{\sqrt{1-\epsilon^2}}$. Then,

$$error(f_o) \geq \frac{1}{4}(1 - \sqrt{1 - e^{-\frac{m\epsilon^2}{1-\epsilon^2}}})$$

(c)
If m is odd, we can use $k = (m+1)/2$ instead of $k = m/2$ in the probability expression. Therefore, $\Pr[N(S) \geq \frac{m}{2}|x = x_A]$ becomes $\Pr[N(S) \geq \frac{m+1}{2}|x = x_A]$. For both cases, we can use the following expression,

$$\Pr[N(S) \geq \lceil\frac{m}{2}\rceil|x = x_A]$$

Therefore, we can replace $m/2$ in previous equation with $\lceil m/2 \rceil$:

$$error(f_o) \geq \frac{1}{4}(1 - \sqrt{1 - e^{-\frac{2\lceil m/2\rceil\epsilon^2}{1-\epsilon^2}}})$$

(d)
Oskar's error is at most δ,

$$\frac{1}{4}(1 - \sqrt{1 - e^{-\frac{2\lceil m/2\rceil\epsilon^2}{1-\epsilon^2}}}) < \delta$$

$$\Longrightarrow 1 - \sqrt{1 - e^{-\frac{2\lceil m/2\rceil\epsilon^2}{1-\epsilon^2}}} < 4\delta$$

$$\Longrightarrow \sqrt{1 - e^{-\frac{2\lceil m/2\rceil\epsilon^2}{1-\epsilon^2}}} > 1 - 4\delta$$

$$\Longrightarrow 1 - e^{-\frac{2\lceil m/2\rceil\epsilon^2}{1-\epsilon^2}} > (1 - 4\delta)^2$$

$$
\begin{aligned}
\implies e^{-\frac{2\lceil m/2\rceil \epsilon^2}{1-\epsilon^2}} &< 1-(1-4\delta)^2 \\
&= 1-1+8\delta-16\delta^2 \\
&= 8\delta(1-2\delta)
\end{aligned}
$$

$$
\implies \frac{-2\lceil m/2\rceil \epsilon^2}{1-\epsilon^2} < \ln(8\delta(1-2\delta))
$$

$$
\implies \lceil m/2\rceil > \frac{1-\epsilon^2}{2\epsilon^2}\ln(8\delta(1-2\delta))
$$

$$
\implies m > 2\lceil \frac{1-\epsilon^2}{2\epsilon^2}\rceil \ln(8\delta(1-2\delta))
$$

(e)
Couldn't solve this.