Homework 4 - LEARNABILITY
2015 Spring, Machine Learning
Choong-Wan Woo
February 20, 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 $m \geq \frac{1}{\epsilon} \log \frac{1}{\delta}$

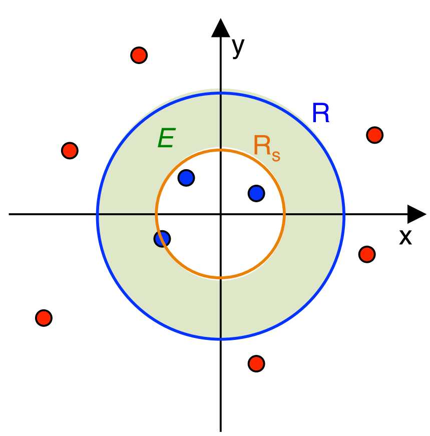

Figure 1. Illustration of the concentric circle case.

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.

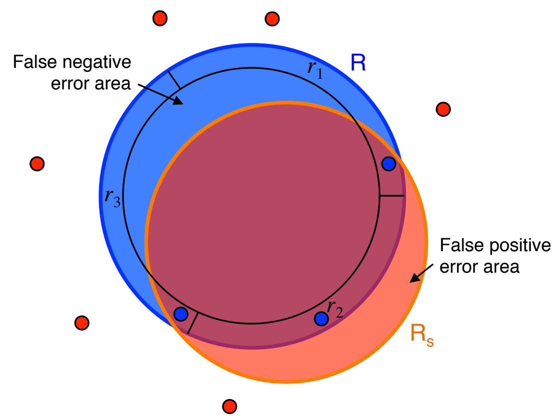

Figure 2. Illustration of the non-concentric circle case.

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:

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

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

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

, solving for m:

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