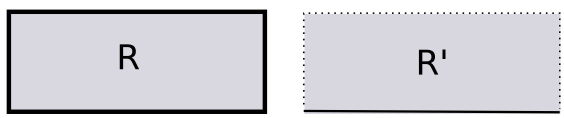
The following is a correction on the proof of PAC learnability of axis aligned rectangles given in [2]. Here an assumption on the continuity of the probability distribution is made, we drop it and proof the same learning bounds. The correct proof to this example is actually given in [1], but the details of the proof are rahter obscure and so we analyze them here. Before stating the example we give some simple notation and prove a lemma that will be useful for the final proof.

For any set $R \in \mathbb{R}^2$ we will denote by ∂R the boundary of R and by $\operatorname{int} R$ the interior of R.

Lemma 1. *For fixed $b > 0$, $\epsilon > 0$ and any probability measure P let $h = \inf\{h' | P([0,b] \times [0,h'] \geq \epsilon\}$. And call $R = [0,b] \times [0,h]$ if $R' = [0,b] \times \{0\} \cup \operatorname{int} R$, then $P(R') \leq \epsilon$.*

Before proceeding to the proof notice that we claim the existance of h this can only be asserted because probability measures are continuous from the right. The sets R and R' we are referring to are shown below



Proof. Define $R_k := [0,b] \times [0,k]$ and $R'_k = [0,b] \times \{0\} \cup \operatorname{int} R_k$ then the following statements are trivially true:

- $R'_k \subset R_k$
- If $k < h$ then $R_k \subset R_h$
- If $k_n \uparrow h$ then $\lim_{n\to\infty} R'_{k_n} = R'$

Notice that by definition of h, for any sequence $k_n \uparrow h$ we have

$$P(R'_{k_n}) \leq P(R_{k_n}) < \epsilon.$$

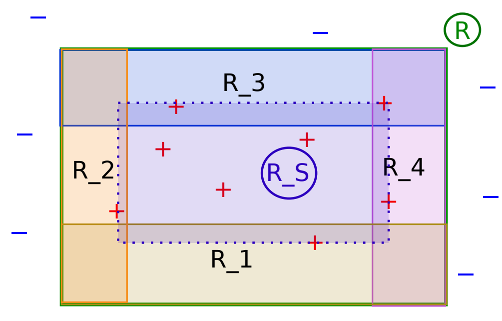
Thus $P(R') = \lim_{n\to\infty} P(R'_{k_n}) \leq \epsilon$ and we are done. □

We remind the reader of the problem, we are trying to learn the class of closed rectangles, i.e the indicator functions:

$$R(x) = \begin{cases} 1 & x \in R \\ 0 & x \notin R \end{cases}$$

Lemma 2. *The concept class of axis aligned rectangles is PAC learnable.*

Proof. Let $S = x_1, \cdots x_n$ be a sample and $\epsilon > 0$. And consider the hypothesis R_S which is the smallest rectangle that includes all positive examples in the sample. We bound the probability of error of this hypothesis. For this we let $R = [a, b] \times [c, d]$ be the real hypothesis. Notice that by construction $R_S \subset R$ and so the hypothesis can only give false negatives. Thus the region of error is precisely $R \backslash R_S$. If $P(R) < \epsilon$ then $P(R \backslash R_S) < \epsilon$ and we are done. So without loss of generality assume that $P(R) \geq \epsilon$. To bound the probability of error consider rectangles R_1, R_2, R_3, R_4 built as follows: Let $h = \inf\{h' | P([a, b] \times [c, c + h]) \geq \epsilon/4)$ and let $R_1 = [a, b] \times [c, c + h]$, in the same way construct a left, upper and right rectangles R_2, R_3, R_4. Assume that $R_S \cap R_i \neq \emptyset$ for all i, then we can see that $R \backslash R_S \subset \cup R'_i$ where $R'_1 = [a, b] \times \{0\} \cup \operatorname{int} R_1$ and R'_i is defined similarly for $i = 2, 3, 4$.



Using lemma 1 we know that $P(\cup R'_i) \leq \sum P(R'_i) \leq \epsilon$. Thus if R_S intersects all four rectangles $err(R_S) \leq \epsilon$. Or said it in another way:

$$
\begin{aligned}
P(err(R_S) > \epsilon) &\leq P(R_S \cap R_i = \emptyset \text{ for some } i) \\
&\leq \sum_{i=1}^{4} P(R_s \cap R_i = \emptyset) \\
&= \sum_{i=1}^{4} P(S \cap R_i = \emptyset) \\
&= \sum_{i=1}^{4} (1 - P(R_i))^n
\end{aligned}
$$

But by construction $P(R_i) \geq \epsilon/4$ and so $1 - P(R_i) \leq 1 - \epsilon/4$ and so:

$$P(err(R_s) > \epsilon) \leq 4\left(1 - \frac{\epsilon}{4}\right)^n \leq 4e^{-\frac{\epsilon}{4}}$$

Which implies that with probability at least $1 - \delta$ we can achieve an error of size ϵ if

$$n \geq \frac{4\log(4/\delta)}{\epsilon}$$

Which means that the concept class is PAC-learnable. □

References

[1] A. Blumer, A. Ehrenfeucht, D. Haussler, and M.K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965, 1989.

[2] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of Machine Learning*. Adaptive Computation and Machine Learning Series. Mit Press, 2012.