

**My notes while reading:**  
**An Introduction To Computational Learning**  
**Theory**

Mohammed D. Belgoumri

April 23, 2022

# Contents

Title page	1
List of figures	2
List of code snippets	3
1 The Probably Approximately Correct Learning Model	4
1.1 A Rectangle Learning Game . . . . .	4
1.2 A General Model . . . . .	7

# List of Figures

1.1	The target rectangle $\mathcal{R}$ along with a labeled sample of points . . . .	5
1.2	The tightest fit rectangle for the example of Figure 1.1 . . . . .	6

# List of Code Snippets

1.1	A python implementation of the described strategy . . . . .	5
-----	---	---

# Chapter 1

## The Probably Approximately Correct Learning Model

### 1.1 A Rectangle Learning Game

The objective of this game is to learn an unknown *target* (axis-aligned) rectangle  $\mathcal{R} = [a, b] \times [c, d] \subset \mathbb{R}^2$ .

The player can gain information about  $\mathcal{R}$  only by choosing random points according to some distribution  $\mathcal{D}$ , and asking the game whether they are inside  $\mathcal{R}$ . By convention, points inside  $\mathcal{R}$  are considered positive.

Figure 1.1 shows an example of a possible target rectangle  $\mathcal{R}$  along with some points labeled using it.

The player's goal is to find a hypothesis rectangle  $\mathcal{R}'$  that “approximates”  $\mathcal{R}$  “as closely as possible”. To measure the quality of this approximation, we will consider the region  $\mathcal{R} \Delta \mathcal{R}'$  of points that  $\mathcal{R}$  and  $\mathcal{R}'$  label differently. More precisely, we will consider the probability  $\mathbb{P}(\mathcal{R} \Delta \mathcal{R}')$  of falling with this region according to  $\mathcal{D}$  and try to minimize this quantity.

Since  $\mathcal{R}$  is unknown in practice,  $\mathcal{R}'$  is evaluated by sampling a number of points from  $\mathcal{D}$ , labeling them using  $\mathcal{R}'$  and comparing to their true labels by asking the game. We then take the number of falsely labeled points divided by the total of chosen points as an estimation of  $\mathbb{P}(\mathcal{R} \Delta \mathcal{R}')$ . Note that this score is the same as  $1 - \text{accuracy}$ .

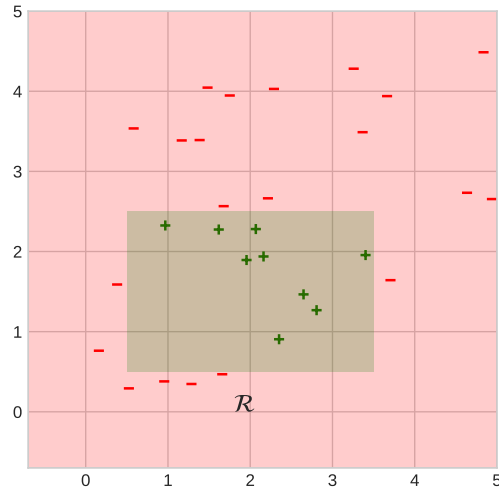


Figure 1.1: The target rectangle  $\mathcal{R}$  along with a labeled sample of points

A simple strategy for the player is to sample a sufficiently large number  $m$  of points from  $\mathcal{D}$  and request their labels, then chose as hypothesis  $\mathcal{R}'$  the smallest rectangle containing all positive examples and no negative ones<sup>1</sup>. If all  $m$  points are negative, we chose  $\mathcal{R}' = \emptyset$ . Figure 1.2 illustrates this strategy for one example, and Code Snippet 1.1 shows a python implementation.

```

1 def fit_rectangle(X, y, ec='black', alpha=.1) -> Rectangle:
2     """
3     Fit a rectangle to the given data.
4     """
5     h_min = X[y].min(axis=0)
6     h_max = X[y].max(axis=0)
7     return Rectangle(h_min, *(h_max - h_min), edgecolor=ec, alpha=
alpha)

```

Code Snippet 1.1: A python implementation of the described strategy

We will now prove that this strategy works, more specifically, we will show the following theorem:

**Theorem 1.1.**

Let  $\mathcal{D}$  be a distribution over  $\mathbb{R}^2$ ,  $\mathcal{R}$  a rectangle, and  $\varepsilon, \delta > 0$  be positive real

---

<sup>1</sup>Note that such at least one rectangle with this property is garenteed to exist, because  $\mathcal{R}$  is one such a rectangle

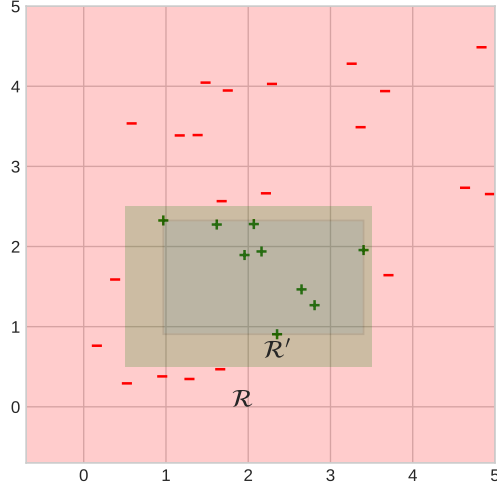


Figure 1.2: The tightest fit rectangle for the example of Figure 1.1

numbers. There exists an integer  $m \in \mathbb{N}$ , such that the hypothesis rectangle  $\mathcal{R}'$  generated by  $m$  sampled points has with probability  $\geq 1 - \delta$  an error  $\leq \varepsilon$ .

*Proof.*

Consider a distribution  $\mathcal{D}$ , a rectangle  $\mathcal{R} = [a, b] \times [c, d]$ , a hypothesis rectangle  $\mathcal{R}' = [a', b'] \times [c', d']$  constructed using the strategy we described, and positive real numbers  $\varepsilon, \delta > 0$ .

Note that  $\mathcal{R}'$  cannot have false positives. In fact,  $\mathcal{R}' \subset \mathcal{R}$  and therefore  $\mathcal{R}' \Delta \mathcal{R} = \mathcal{R} \setminus \mathcal{R}'$ . This region can be expressed as the union of four rectangles:

$$\begin{aligned} \mathcal{R} \setminus \mathcal{R}' = & [a, a'] \times [c', d'] \cup \\ & [a, b] \times [c, c'] \cup \\ & [b', b] \times [c', d'] \cup \\ & [a, b] \times [d, d'] \end{aligned}$$

We denote the union of four rectangles respectively by  $L, B, R, T$ . It immediately follows that:

$$\mathbb{P}(\mathcal{R} \setminus \mathcal{R}') = \mathbb{P}(L \cup B \cup R \cup T) \leq \mathbb{P}(L) + \mathbb{P}(B) + \mathbb{P}(R) + \mathbb{P}(T)$$

Consequently, to show that  $\mathbb{P}(\mathcal{R} \setminus \mathcal{R}') \leq \varepsilon$ , it suffices to show that  $\mathbb{P}(X) \leq \frac{\varepsilon}{4}$  for all  $X \in \{L, B, R, T\}$ .

□

## 1.2 A General Model