SOLUTIONS TO CHP 2 EXERCISES

DANIEL ROY

Chapter 2. A Gentle Start

Exercise 2.1.

This solution is due to Alon Gonen and Dana Rubinstein.

Given $S = ((x_i, y_i))_{i=1}^m$ define the polynomial

$$p(x) = -\prod_{i \in [m] \text{ s.t. } y_i = 1} \|x - x_i\|^2 \tag{1}$$

Since this is (the negation of) a product of nonnegative terms, the polynomial is nonpositive. Further, it is clearly zero at any input that was labeled 1.

Exercise 2.2.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{S|_x \sim \mathcal{D}^m} [L_S(h)] && && (2)\\
&= \mathop{\mathbb{E}}_{S|_x \sim \mathcal{D}^m} \left[\frac{1}{m} \sum_{i=1}^m \mathbf{1}[h(x_i) \neq f(x_i)] \right] && \text{by definition} && (3)\\
&= \frac{1}{m} \sum_{i=1}^m \mathop{\mathbb{E}}_{S|_x \sim \mathcal{D}^m} [\mathbf{1}[h(x_i) \neq f(x_i)]] && \text{by linearity of exp.} && (4)\\
&= \frac{1}{m} \sum_{i=1}^m \mathop{\mathbb{E}}_{x_i \sim \mathcal{D}} [\mathbf{1}[h(x_i) \neq f(x_i)]] && \text{expression only depends on } x_i && (5)\\
&= \frac{1}{m} \sum_{i=1}^m L_{(\mathcal{D},f)}(h) && \text{by definition} && (6)\\
&= L_{(\mathcal{D},f)}(h). && && (7)
\end{aligned}$$

Exercise 2.3. **Rectangular Hypotheses**

Credit for this problem's solution goes to Jeffrey Negrea.

2.3.1. By construction, $A = A(S)$ is the smallest rectangle enclosing the positive examples, and so the labels of positive examples are correctly classified. It suffices to show that A does not mislabel any negative example.

By realizability, there is a rectangle $A' = (a'_1, b'_1, a'_2, b'_2)$ such that

$$\mathcal{D}(\{x \in \mathcal{X} : f(x) = \mathbf{1}[x \in A']\}) = 1. \tag{8}$$

It suffices to show that $A \subseteq A'$, because otherwise, A and A' will agree on all negative examples. Assume otherwise, i.e., $A \setminus A'$ is nonempty. Since A is the smallest rectangle enclosing the positive examples, this would imply that there was a positive example outside A', a contradiction.

2.3.2. The solution proposed via the hint obviously assumes that the measure $\mathcal{D}$ is continuous (can you explain why? because it assumes that we can select subsets of $\mathcal{X}$ of precise probability mesaure). We make this assumption explicit here.
Step 1: Showing that $R(S) \subseteq R^\star$ is equivalent to the argument used above that $A(S) \subseteq A'$.
Step 2: Suppose that each of the rectangles $R_i : i \in [4]$ contain a training point. Then, clearly, $R^\star \setminus \bigcup_{i\in[4]} R_i \subseteq R(S) \subseteq R^\star$. The probability of a training error is exactly $\mathcal{D}(R^\star \setminus R(S))$. Therefore we have that

$$\begin{aligned}
L_{(\mathcal{D},f)}\left(A(S)\right) &= \mathcal{D}(R^\star \setminus R(S)) && (9)\\
&\le \mathcal{D}\left(R^\star \setminus \left(R^\star \setminus \bigcup_{i\in[4]} R_i\right)\right) && (10)\\
&= \mathcal{D}\left(\bigcup_{i\in[4]} R_i\right) && (11)\\
&\le \sum_{i\in[4]} \mathcal{D}(R_i) && (12)\\
&= \epsilon && (13)
\end{aligned}$$

Step 3: The probability that S contains no examples from R_i is exactly $(1-\epsilon/4)^m$ for each $i \in [4]$ because of independence. We have $(1-\epsilon/4)^m \le \exp\left(-\frac{m\epsilon}{4}\right)$.
Step 4: The probability that our sample meets the requirements of Step 2 is

$$\begin{aligned}
\mathcal{D}^m(\neg\exists i \in [4] : S|_x \cap R_i = \emptyset) &= 1 - \mathcal{D}^m(\exists i \in [4] : S|_x \cap R_i \neq \emptyset) && (14)\\
&= 1 - \mathcal{D}^m\left(\bigcup_{i\in[4]} \{S|_x \cap R_i \neq \emptyset\}\right) && (15)\\
&\ge 1 - \sum_{i\in[4]} \mathcal{D}^m(S|_x \cap R_i \neq \emptyset) && (16)\\
&\ge 1 - \sum_{i\in[4]} \exp\left(-\frac{m\epsilon}{4}\right) && (17)\\
&= 1 - 4\exp\left(-\frac{m\epsilon}{4}\right) && (18)
\end{aligned}$$

Now take $\delta \ge 4\exp\left(-\frac{m\epsilon}{4}\right)$. Solving for m, we get $m \ge \frac{4\log(4/\delta)}{\epsilon}$.

We have thus shown that for a sample size of at least $m \ge \frac{4\log(4/\delta)}{\epsilon}$, we will be able return a hypothesis with error at most epsilon with probability at least $(1-\delta)$.