EENG 431 - Homework 1

Franklin She

February 23, 2024

1 Chapter 2

Problem 1. Let's recall the predictor in Equation (2.3),

$$h_S(x) = \begin{cases} y_i & \text{if } \exists i \in [m] \text{ s.t. } x_i = x \\ 0 & \text{otherwise} \end{cases}$$

The goal is to construct $p_S(x)$ in such a way that it outputs a non-negative value if and only if x matches one of the $\mathbf{x}_i$ where $f(\mathbf{x}_i)) = 1$ and to a negative value otherwise.

Let's consider the case of only one positive example in the training set, or $S = \{(\mathbf{x}_1, 1)\}$. Then, we can define $p_S(x)$ as follows:

$$p_S(x) = -\|x - \mathbf{x}_1\|$$

We know that $p_S(\mathbf{x}_1) = 0$ and $p_S(x) < 0$ for all $x \neq \mathbf{x}_1$.

To generalize this to the case of multiple examples, we can define $p_S(x)$ as follows:

$$p_S(x) = -\prod_{i=1:\, y_i=1}^{m} (\|x - \mathbf{x}_i\|)$$

Notice by taking the product of the distances, we ensure that if x matches any of the $\mathbf{x}_i$ where $f(\mathbf{x}_i) = 1$, then $p_S(x) = 0$. Otherwise, $p_S(x) < 0$.

Problem 2.

$$\begin{aligned}
\mathbb{E}_{S|_x \sim D^m}[L_S(h)] &= \mathbb{E}_{S|_x \sim D^m}\left[\frac{1}{m}\sum_{i=1}^{m} \mathbb{1}(h(x_i) \neq y_i)\right] \\
&= \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}_{x \sim D}\left[\mathbb{1}(h(x_i) \neq y_i)\right] && \text{by linearity of expectation} \\
&= \frac{1}{m}\sum_{i=1}^{m} \Pr_{x \sim D}\left[h(x_i) \neq y_i\right] && x_1, \ldots, x_m \text{ are i.i.d} \\
&= \frac{1}{m} \cdot m \cdot L_{D,h}(h) \\
&= L_{D,h}(h)
\end{aligned}$$

Problem 3.

1. To show that A is an ERM, we can show that A returns h^* s.t. $L_{D,f}(h^*) = 0$. By definition, A labels correctly all the positive examples in the training set. Because we also assume realizability and A is the tightest rectangle, all negative examples in the training set are also correctly labeled. Therefore, A labels the whole training set correctly, so A must be an ERM.

2. Let $\mathcal{D}$ be a distribution over $\mathcal{X}$. We will use $R^* = R(a_1^*, a_2^*, a_3^*, a_4^*)$ defined in the hint, and let f be its corresponding hypothesis. Let $R(S)$ be the rectangle returned by the algorithm A given the training set S. We can first notice that $R(S) \subseteq R^*$ because of the way R is defined. Thus we have

$$\begin{aligned}
L_{\mathcal{D},f}(A(S)) &= \mathcal{D}(\{x \in \mathcal{X} : A(S)(x) \neq f(x)\}) \\
&= \mathcal{D}(\{x \in \mathcal{X} : x \notin S \text{ and } f(x) = 1\}) \\
&= \mathcal{D}(R^* \setminus R(S))
\end{aligned}$$

Next, we consider the rectangles R_1, R_2, R_3, and R_4 defined in the hint, which all have a probability mass of $\epsilon/4$. We can deduce $L_{\mathcal{D},f}(A(S)) \leq \epsilon$ if S contains positive examples in all the rectangles R_1, R_2, R_3, and R_4. The probability that S contains no positive examples in any of the rectangles is at most $(1-\epsilon/4)^m \leq e^{(-\epsilon/4)m}$. By the union bound, we have

$$\mathcal{D}(\{S : S \cap R_i = \emptyset \text{ for some } i \in \{1,2,3,4\}\}) \leq 4 \cdot e^{(-\epsilon/4)m}$$

Plugging in the value of m of $\geq \frac{4\log(4/\delta)}{\epsilon}$, we can see that A will return a hypothesis with error of at most ϵ with probability at least $1-\delta$.

2 Chapter 3

Problem 2.

1. If a positive x appears in S, we can return the true h_x. Otherwise, we can return the all-negative hypothesis.

$$h_S(x) = \begin{cases} h_x & \text{if } \exists x \in S \text{ s.t. } f(x) = 1 \\ h^- & \text{otherwise} \end{cases}$$

Because we assume realizability, we have $L_S(h_S) = 0$, so the algorithm that returns the hypothesis h_S is an ERM.

2. Let $\mathcal{D}$ be a distribution over $\mathcal{X}$. First we notice that if the true hypothesis is h^-, our algorithm returns the perfect hypothesis.

Now assume that there exists a positive example x such that $f(x) = 1$. Because of the realizaibility assumption, x is unique. If this x is in our sample, then our algorithm returns the perfect hypothesis again. Also note that if $\mathcal{D}(x) \leq \epsilon$, then $L_{\mathcal{D}}(h) \leq \epsilon$ with probability 1.

To find an upper bound on sample complexity, we are thus interested in the event where x does not appear in our sample, and $\mathcal{D}(x) > \epsilon$. This means that $\mathcal{D}(x') \leq 1-\epsilon$ for all $x' \neq x$. Therefore, sampling m, we have

$$\mathcal{D}^m(\{S : L_{\mathcal{D}}(h_S) > \epsilon\}) \leq (1-\epsilon)^m \leq e^{-\epsilon m}$$

Picking δ such that $e^{-\epsilon m} \leq \delta$, we can solve for m to show that

$$m_{\mathcal{H}}(\epsilon, \delta) \leq \left\lceil \frac{\log(1/\delta)}{\epsilon} \right\rceil$$

Problem 3. Let A, given a training set S produce a hypothesis $h_{\hat{r}}$ corresponding to the smallest circle that encloses all positive examples in S, where $\hat{r}$ denotes the radius of this circle.

Define C^* as the circle corresponding to h^* (realizability assumption hypothesis) with radius r^*, and $C(\hat{r})$ as the circle produced by A with radius $\hat{r}$. We first notice that $C(\hat{r}) \subseteq C^*$.

Let $r_1 \leq r^*$ be such that the circular strip $E = \{x \in \mathbb{R}^2 : r_1 \leq \|x\| \leq r^*\}$ has a probability mass. This implies that the probability that a randomly drawn sample falls within E is exactly ϵ.

If the training set S contains at least one positive example from E, the hypothesis $h_{\hat{r}}$ produced by A will have a generalization error of at most ϵ.

The probability that no sample in S falls within E is at most $(1-\epsilon)^m$, as each sample independently has a probability of ϵ to fall within E.

$$\mathcal{D}^m(\{S : L_{\mathcal{D}}(h_S) > \epsilon\}) \leq (1-\epsilon)^m \leq e^{-\epsilon m}$$

Picking δ such that $e^{-\epsilon m} \leq \delta$, we can solve for m to show that

$$m_{\mathcal{H}}(\epsilon, \delta) \leq \left\lceil \frac{\log(1/\delta)}{\epsilon} \right\rceil$$