ML.exer.ch2

samira moradzade

28 aban 1400

1 Exercise

We have shown that the predictor defined in Equation (2.3) leads to overfitting. While this predictor seems to be very unnatural,the goal of this exercise is to show that it can be described as a thresholded polynomial. That is, show that given a training set $S = \{(x_i, f(x_i))\}_{i=1}^{m} \subseteq (\mathbb{R}^d \times \{0,1\})^m$, there exists a polynomial p_S such that $h_S(x) = 1$ if and only if $p_S(x) \geq 0$, where h_S is as defined in Equation (2.3). It follows that learning the class of all thresholded polynomials using the ERM rule may lead to overfitting.

1.1 solution:

we are searching for the function p_s such that $h_s(x) = 1$ if and only if $p_s \geq 0$. so if $h_s(x) = 0$ the function must take negative values and the function defined as $p_s : X \to \mathbb{R}$ the best way to find the solution of this problem is setting function's value to 0 at points whenever $h_s(x) = 1$ and setting a negative value at the other points. we achieve it by multiplying $(x - x_i)$ s.t $i \in I = \{\ i \in [m] : h_s(x_i) = 1\}$ notice that this $(x - x_i)$ is in a vector space but we need the function to give us the values in R so we have to define the "Norm" function and have the multiplying of $\|x - x_i\|$ s.

on the other hand we know that our function sign should be negative at every points except x_i and takes the maximum value at x_i so we have to consider an even power for each $(x - x_i)$, dont forget to multiply (-1) because $\|x - x_i\|^2 \geq 0$ and also because the values that p_S returns are in R, not a vector type so we use norm function. So that we define the function: $p_s(x) = -\prod_{i \in I} \|x - x_i\|^2$

2 Exercise

Let $\mathcal{H}$ be a class of binary classifiers over a domain $\mathcal{X}$. Let $\mathcal{D}$ be an unknown distribution over $\mathcal{X}$, and let f be the target hypothesis in $\mathcal{H}$. Fix some $h \in \mathcal{H}$. Show that the expected value of $L_S(h)$ over the choice of $S|_x$ equals $L_{(\mathcal{D},f)}(h)$, namely,

$$\mathop{\mathbb{E}}_{S|_x \sim \mathcal{D}^m}[L_S(h)] = L_{(\mathcal{D},f)}(h)$$

2.1 solution:

$$
\begin{aligned}
\mathop{\mathbb{E}}_{S|_x\sim\mathcal{D}^m}[L_S(h)] &= \mathop{\mathbb{E}}_{S|_x\sim\mathcal{D}^m}\left[\frac{1}{m}\sum_{i\in J}[1_{h(x_i)\neq f(x_i)}]\right] \\
&= \frac{1}{m}\sum_{i\in I}\left(\mathop{\mathbb{E}}_{x_i\sim\mathcal{D}}\left[1_{h(x_i)\neq f(x_i)}\right]\right) && \text{linearity} \\
&= \frac{1}{m}\sum_{i\in I}\left(\mathop{\mathbb{P}}_{x_i\sim\mathcal{D}}[h(x_i)\neq f(x_i)]\right) \\
&= \frac{1}{m}\sum_{j\in I}\left(\mathop{\mathbb{P}}_{x\sim\mathcal{D}}[h(x)\neq f(x)]\right) && \text{i.i.d} \\
&= \frac{1}{m}\times m\times L_{\mathcal{D},f}(h) && \text{by the definition} \\
&= L_{\mathcal{D},f}(h)
\end{aligned}
$$