MLT Homework 6

Ana Borovac
Jonas Haslbeck
Bas Haver

October 29, 2018

Raw 0/1-loss with real-valued predictions:

$$l'(y, \hat{y}) = \frac{1 - \text{sign}(y \cdot \hat{y})}{2}, \quad \text{sign}(u) = \begin{cases} 1 & \text{if } u \geq 0, \\ -1 & \text{if } u < 0 \end{cases}$$

Question 1

Suppose that we have a loss function l based on a raw l' as above that is bounded, i.e. $\sup_{\hat{y} \in \hat{Y}, y \in Y} |l'(y, \hat{y})| < \infty$.

Subquestion 1.1

Suppose that $\mathcal{H}$ consists of just one hypothesis, $\mathcal{H} = \{h\}$. Show that for all samples S, we have that $R(l \circ \mathcal{H} \circ S) = 0$.

Solution

Let's recall the definition of Rademacher complexity:

$$R(l \circ \mathcal{H} \circ S) = \frac{1}{m} \mathbb{E}_{(y_1, \ldots, y_m) \in \{-1, 1\}^m} \left[\sup_{h \in \mathcal{H}} \sum_{y_i} y_i \, l(h, x_i) \right]$$

In our case we have:

$$R(l \circ \mathcal{H} \circ S) = \frac{1}{m} \mathbb{E}_{(y_1, \ldots, y_m) \in \{-1, 1\}^m} \left[\sum_{y_i} y_i \, l(h, x_i) \right]$$

With linearity of expected value:

$$R(l \circ \mathcal{H} \circ S) = \frac{1}{m} \sum_{y_i} l(h, x_i) \, \mathbb{E}_{(y_1, \ldots, y_m) \in \{-1, 1\}^m} [y_i]$$

Let calculate the expected value of y_i:

$$\mathbb{E}[y_i] = -1 \cdot P(y_i = -1) + 1 \cdot P(y_i = 1) = -p + p = 0$$

where $P(y_i = 1) = P(y_i = -1) = p$. Now we can conclude:

$$R(l \circ \mathcal{H} \circ S) = \frac{1}{m} \sum_{y_i} l(h, x_i) \cdot 0 = 0$$

Subquestion 1.2

Consider a hypothesis class $\mathcal{H}$ and some hypothesis h'. Show that

$$R(l \circ \mathcal{H} \cup \{h'\} \circ S) \geq R(l \circ \mathcal{H} \circ S),$$

and explain why it follows that Rademacher complexity is always nonnegative.

Solution

First we are going to write what is $R(l \circ \mathcal{H} \cup \{h'\} \circ S)$:

$$R(l \circ \mathcal{H} \circ S) = \frac{1}{m} \mathbb{E}_{(y_1,\ldots,y_m) \in \{-1,1\}^m} \left[\sup_{h \in \mathcal{H} \cup \{h'\}} \sum_{y_i} y_i \, l(h, x_i) \right]$$

In the previous subquestion we proved that Rademacher complexity of hypothesis class with one hypothesis equals 0, now we proved that if we add a hypothesis to a hypothesis class we get at least as much as we did before. So, Rademacher complexity is nonnegative.

Subquestion 1.3

Rademacher complexity is often used as a tool to show that some given hypothesis class $\mathcal{H}$ is PAC-learnable. Show that Theorem 26.5 in the book implies agnostic PAC-learnability for $\mathcal{H}$ whenever we can prove that expected Rademacher complexity satisfies

$$\lim_{m \to \infty} \sup_D \mathbb{E}_{S \sim D^m} [R(l \circ \mathcal{H} \circ S)] = 0$$

where the supremum is over all probability distributions D that can be defined on $Z = X \times Y$.

Solution

Question 2

(Lipschitz Losses) Consider a raw loss function l' as above with prediction set $\hat{Y}$; here $\hat{Y}$ can either be $\mathbb{R}$ or a closed and bounded interval in $\mathbb{R}$. We say that

l' is $\rho - Lipschitz$ if for all $a, b \in \hat{Y}$ we have

$$\sup_{y \in Y} |l'(y, a) - l'(y, b)| \leq \rho |a - b|$$

We say that l' is Lipschitz if it is ρ-Lipschitz for some $\rho < \infty$. Intuitively, if a loss function is Lipschitz then small changes in the predictions imply small changes in the loss. It turns out that analysis with Rademacher complexity can be done for (raw) loss functions that are either bounded or Lipschitz. For loss functions that are neither bounded nor Lipschitz, it is much more problematic. We will now consider Lipschitzness for some of the most commonly appearing loss functions.

Subquestion 2.1

Show that the 0/1-loss for real-valued predictions is not Lipschitz.

Solution

0/1-loss for real-valued predictions is defined as:

$$l'(y, \hat{y}) = \frac{1 - \mathrm{sign}(y \cdot \hat{y})}{2}, \quad \mathrm{sign}(u) = \begin{cases} 1 & \text{if } u \geq 0, \\ -1 & \text{if } u < 0 \end{cases}$$

Subquestion 2.2

Is the absolute loss on a bounded domain $Y = \hat{Y} = [-1, 1]$ Lipschitz?

Solution

The absolute loss is defined as:

$$l'(y, \hat{y}) = |y - \hat{y}|$$

It follows:

$$\sup_{y \in Y} |l'(y, a) - l'(y, b)| = \sup_{y \in Y} ||y - a| - |y - b||$$

If $a = b$ we get $\sup_{y \in Y} |l'(y, a) - l'(y, b)| = 0$, so it is bounded with $\rho |a - b|$.

Without loss of generality we can say $a < b$. Now, we have three possible situations:

- $y \leq a < b$: If follows $(y - a) \leq 0$ and $(y - b) < 0$, so:

$$\sup_{y \in Y} ||y - a| - |y - b|| = \sup_{y \in Y} |-y + a + y - b| \leq \underbrace{1}_{\rho} \cdot |a - b|$$

- $a < y \leq b$: If follows $(y - a) > 0$ and $(y - b) \leq 0$, so:

$$\sup_{y \in Y} ||y - a| - |y - b|| = \sup_{y \in Y} |y - a + y - b|$$

$$= \sup_{y \in Y} \left| (a - b) \left(\frac{y - a}{a - b} + \frac{y - b}{a - b} \right) \right|$$

$$= |a - b| \sup_{y \in Y} \left| \left(\frac{y - a}{a - b} + \frac{y - b}{a - b} \right) \right|$$

$$\leq |a - b| \sup_{y \in Y} \left(\left| \frac{y - a}{a - b} \right| + \left| \frac{y - b}{a - b} \right| \right)$$

$$= |a - b| \sup_{y \in Y} \left(\frac{|y - a|}{|a - b|} + \frac{|y - b|}{|a - b|} \right)$$

From $a < y \leq b$ we know:

$$|y - a| < |a - b| \quad \Rightarrow \quad \frac{|y - a|}{|a - b|} < 1$$

$$|y - b| \leq |a - b| \quad \Rightarrow \quad \frac{|y - b|}{|a - b|} \leq 1$$

So:

$$\sup_{y \in Y} ||y - a| - |y - b|| = |a - b| \sup_{y \in Y} \left(\frac{|y - a|}{|a - b|} + \frac{|y - b|}{|a - b|} \right)$$

$$\leq |a - b| \left(\sup_{y \in Y} \frac{|y - a|}{|a - b|} + \sup_{y \in Y} \frac{|y - b|}{|a - b|} \right)$$

$$\leq |a - b|(1 + 1)$$

$$= \underbrace{2}_{\rho} \cdot |a - b|$$

- $a < b < y$: If follows $(y - a) > 0$ and $(y - b) > 0$, so:

$$\sup_{y \in Y} ||y - a| - |y - b|| = \sup_{y \in Y} |y - a - y + b| \leq \underbrace{1}_{\rho} \cdot |a - b|$$

To sum up, we can conclude that:

$$\sup_{y \in Y} |l'(y, a) - l'(y, b)| \leq \underbrace{2}_{\rho} \cdot |a - b|$$

So, yes, the absolute loss is Lipschitz.

Subquestion 2.3

Is the unbounded absolute loss on unbounded domain $Y = \hat{Y} = \mathbb{R}$ Lipschitz?

Solution

Yes, we can repeat the proof from bounded domain.

Subquestion 2.4

Is the squared error loss $l'(y, \hat{y}) = (y - \hat{y})^2$, defined on bounded domain $Y = \hat{Y} = [-1, 1]$ Lipschitz?

Solution

From the below:

$$\sup_{y \in Y} |l'(y, a) - l'(y, b)| = \sup_{y \in Y} |(y - a)^2 - (y - b)^2| =$$
$$= \sup_{y \in Y} |y^2 - 2ay + a^2 - y^2 + 2by - b^2| =$$
$$= \sup_{y \in Y} |a^2 - b^2 - 2y(a - b)| =$$
$$= \sup_{y \in Y} |(a - b)(a + b) - 2y(a - b)| =$$
$$= \sup_{y \in Y} |(a - b)(a + b - 2y)| =$$
$$= \sup_{y \in Y} |a - b||a + b - 2y| =$$
$$= |a - b| \sup_{y \in Y} |a + b - 2y| \leq$$
$$\leq |a - b| \sup_{y \in Y} (|a + b| + |2y|) \leq$$
$$\leq |a - b|(|a + b| + \sup_{y \in Y} |2y|) \leq$$
$$\leq |a - b|(2 + 2) =$$
$$= \underbrace{4}_{\rho} \cdot |a - b|$$

it follows that the squared error loss defined on bounded domain is Lipschitz.

Subquestion 2.5

Is the unbounded squared error loss defined on unbounded domain $Y = \hat{Y} = \mathbb{R}$ Lipschitz?

Solution

No. If it was, we should be able to bound:

$$\sup_{y \in Y} |a + b - 2y|$$

which we can no, since Y is not bounded.