CS7545, Fall 2019: Machine Learning Theory - Homework #5

Jacob Abernethy, Bhuvesh Kumar, and Zihao Hu Due: Tuesday, Dec 3, 2019 at 11:59pm

Homework Policy: Working in groups is fine, but *every student* must submit their own writeup, i.e write the solutions on their own and not submit a shared document. Please write the members of your group on your solutions. There is no strict limit to the size of the group but we may find it a bit suspicious if there are more than 4 to a team. Questions labelled with **(Challenge)** are not strictly required, but you'll get some participation credit if you have something interesting to add, even if it's only a partial answer.

1) **Growth function.** In class we had studied about the growth function for a binary class of functions but the same definition can be used to generalize it to a class of functions that take values in the finite set $\mathcal{Y}$, that is if $H \subseteq \{h | h : \mathcal{X} \to \mathcal{Y}\}$, then

$$\Pi_H(m) = \max_{(x_1, \cdots, x_m) \subseteq \mathcal{X}} |\{(h(x_1), \cdots, h(x_m)) | h \in H\}|$$

Note that $\Pi_H(m) \leq |\mathcal{Y}|^m$ (analogous to 2^m in the binary case.)

(a) Let $H_1 \subseteq \{h|h : \mathcal{X} \to \mathcal{Y}_1\}$ and $H_2 \subseteq \{h|h : \mathcal{X} \to \mathcal{Y}_2\}$ be function classes and let $H_3 \subseteq \{h|h : \mathcal{X} \times \mathcal{X} \to \mathcal{Y}_1 \times \mathcal{Y}_2\}$ such that $H_3 = \{(h_1, h_2)|h_1 \in H_1, h_2 \in H_2\}$. Show that

$$\Pi_{H_3}(m) \leq \Pi_{H_1}(m) \cdot \Pi_{H_2}(m)$$

(b) Let $H_1 \subseteq \{h|h : \mathcal{X} \to \mathcal{Y}_1\}$ and $H_2 \subseteq \{h|h : \mathcal{Y}_1 \to \mathcal{Y}_2\}$ be function classes and let $H_3 \subseteq \{h|h : \mathcal{X} \to \mathcal{Y}_2\}$ such that $H_3 = \{h_2 \circ h_1 | h_1 \in H_1, h_2 \in H_2\}$. Show that

$$\Pi_{H_3}(m) \leq \Pi_{H_1}(m) \cdot \Pi_{H_2}(m)$$

2) **Rademacher Complexity Identities.** For a fixed $m > 0$, prove the following identities for any $\alpha \in \mathbb{R}$ and any two hypothesis sets H and H' of function mappings from $\mathcal{X}$ to $\mathbb{R}$.

(a) $\mathcal{R}_m(\alpha H) = |\alpha| \mathcal{R}_m(H)$ where $\alpha H = \{\alpha h(\cdot) \mid h \in H\}$

(b) $\mathcal{R}_m(H + H') = \mathcal{R}_m(H) + \mathcal{R}_m(H')$ where $H + H' = \{h(\cdot) + h'(\cdot) \mid h \in H, h' \in H'\}$

3) **Rademacher Complexity.** Let a data set $S = (x_1, x_2, \ldots, x_m)$ be a sample of size m and fix $h : \mathcal{X} \to \mathbb{R}$. Denote $\mathbf{y}$ the vector of predictions of h for S: $\mathbf{y} = [h(x_1), h(x_1), \ldots, h(x_m)]^\top$.

Derive an upper bound on the empirical Rademacher complexity $\hat{\mathcal{R}}_S(H)$ of a hypothesis set $H = \{h, -h\}$ in terms of $\|\mathbf{y}\|_2$ and m.

4) **Growth function and Rademacher Complexity.** Let H be the family of threshold functions over the real line: $H = \{x \to 1_{x \leq \gamma}\} \cup \{x \to 1_{x \geq \gamma'}\}$.

(a) Give an upper bound of the growth function of H: $\Pi_H(m) = \max_{x_1, x_2, \ldots, x_m} |\{(h(x_1), \ldots, h(x_m)) : h \in H\}|$. Use that to bound the Rademacher complexity $\mathcal{R}_m(H)$.

(b) Give a high-probability (i.e. true with probability at least $1 - \delta$) upper bound of the true risk that holds for all $h \in H$. The bound should be in terms of the empirical risk, the upper bound of the Rademacher complexity $\mathcal{R}_m(H)$ you get, probability δ, and sample size m.

5) **Massart's Lemma, Take 2.** If you recall when we proved Massart's Lemma, it looks suspiciously similar to the proof behind Hoeffding's Inequality. Indeed, you can reduce it directly, albeit with a slightly worse bound. Prove the following via a reduction to Hoeffding's Inequality.

Let $A \subset [-1, 1]^m$ be a finite set. Let $\sigma_1, \ldots, \sigma_m$ be iid Rademacher random variables (i.e. uniform on $\{-1, 1\}$). Prove that

$$\mathbb{E}_{\sigma_{1:m}} \left[\sup_{\mathbf{a} \in A} \frac{1}{m} \sum_{i=1}^{m} \sigma_i a_i \right] = O\left(\sqrt{\frac{\log m + \log |A|}{m}} \right).$$

(*Hint:* First note that, for any real random variable Z and any real t, we can break an expectation into two pieces $\mathbb{E}[Z] = \mathbb{E}[Z \cdot \mathbf{1}[Z \leq t]] + \mathbb{E}[Z \cdot \mathbf{1}[Z > t]]$. You'll bound the left term by t and the right term via a union bound.)