Statistical Foundations of Learning - Sample Problems 4

CIT4230004 (Summer Semester 2024)

Sample Problem 4.1: Agnostic PAC Learnability

Let $H \subseteq \{\pm 1\}^X$ be a finite hypothesis class. We know that H is agnostic PAC learnable with respect to 0-1 loss since $\text{VCdim}(H) \leq \log_2(|H|) < \infty$.

Now consider the following loss function $\ell_{c1,c2} : \{\pm 1\} \times \{\pm 1\} \to \{0, c1, c2\}$ such that

$$\ell_{c1,c2}(h(x), y) = \begin{cases} 0 & \text{if } h(x) = y, \\ c1 & \text{if } h(x) = -1 \text{ and } y = +1, \\ c2 & \text{if } h(x) = +1 \text{ and } y = -1. \end{cases}$$

Show that H is agnostic learnable with respect to $\ell_{c1,c2}$.

Since $c1, c2 < \infty$, we consider Empirical Risk Minimization (ERM) for the above loss and adapt the uniform convergence bound for finite H to this case.

Proof:

1. **Define the empirical risk:**

$$\hat{L}_{c1,c2}(h) = \frac{1}{m} \sum_{i=1}^{m} \ell_{c1,c2}(h(x_i), y_i)$$

2. **Expected risk:**

$$L_{c1,c2}(h) = \mathbb{E}_{(x,y)\sim D}[\ell_{c1,c2}(h(x), y)]$$

3. **Uniform convergence:** Using Hoeffding's inequality for bounded losses $\ell_{c1,c2}$:

$$\mathbb{P}\left(\sup_{h \in H} |L_{c1,c2}(h) - \hat{L}_{c1,c2}(h)| > \epsilon\right) \leq 2|H| \exp\left(-\frac{2m\epsilon^2}{(c2 - c1)^2}\right)$$

For ϵ small enough:

$$\mathbb{P}\left(|L_{c1,c2}(h) - \hat{L}_{c1,c2}(h)| > \epsilon\right) \leq \delta$$

This ensures that the empirical risk minimizer $\hat{h}$ converges to the true risk minimizer as $m \to \infty$.

Thus, H is agnostic PAC learnable with respect to $\ell_{c1,c2}$.

Sample Problem 4.2: On-average-replace-one Stability for ERM

Let H be a hypothesis class with $\mathrm{VCdim}(H) = d$, let ℓ be the 0-1 loss, and let A be the ERM learner for H. Use uniform convergence results to derive an upper bound β_m for

$$\mathbb{E}_{S\sim D^m,(x',y')\sim D,i\sim \mathrm{Unif}(m)}\left[\ell(A_{S^i}(x_i), y_i) - \ell(A_S(x_i), y_i)\right]$$

Show that $\lim_{m\to\infty} \beta_m = 0$.

Proof:

1. **Uniform convergence results:** By uniform convergence, the empirical risk converges to the true risk as $m \to \infty$:

$$\mathbb{P}\left(\sup_{h\in H} |L(h) - \hat{L}(h)| > \epsilon\right) \leq 2|H| \exp\left(-\frac{2m\epsilon^2}{1}\right)$$

2. **On-average-replace-one stability:** Let S^i be the dataset S with the i-th example replaced by (x', y'). The stability bound is:

$$\mathbb{E}_{S,(x',y'),i}\left[\ell(A_{S^i}(x_i), y_i) - \ell(A_S(x_i), y_i)\right] \leq \frac{c}{m}$$

where c is a constant dependent on the complexity of H.

3. **As $m \to \infty$:**

$$\lim_{m\to\infty} \frac{c}{m} = 0$$

Therefore, $\lim_{m\to\infty} \beta_m = 0$.

Sample Problem 4.3: Rademacher Complexity of Sets

The Rademacher complexity of a subset $X \subset \mathbb{R}^m$ is defined as

$$R_m(X) = \frac{1}{m}\mathbb{E}_\sigma\left[\sup_{x\in X}\langle \sigma, x\rangle\right]$$

where the expectation is with respect to m independent Rademacher random variables $\sigma = (\sigma_1, \ldots, \sigma_m) \in \{\pm 1\}^m$. Furthermore, we define the convex hull of a set X as

$$\mathrm{conv}(X) = \left\{\sum_{i=1}^N \lambda_i x_i \mid x_i \in X, \lambda_i \geq 0, \sum_{i=1}^N \lambda_i = 1, N > 0\right\}$$

Show that $R_m(X) = R_m(\mathrm{conv}(X))$.

Proof:

1. **Rademacher complexity of X:**

$$R_m(X) = \frac{1}{m}\mathbb{E}_\sigma \left[\sup_{x \in X} \langle \sigma, x \rangle\right]$$

2. **Rademacher complexity of the convex hull $\text{conv}(X)$:**

$$R_m(\text{conv}(X)) = \frac{1}{m}\mathbb{E}_\sigma \left[\sup_{z \in \text{conv}(X)} \langle \sigma, z \rangle\right]$$

3. **Convex combination:** For $z \in \text{conv}(X)$, we have $z = \sum_{i=1}^N \lambda_i x_i$ where $x_i \in X$ and $\sum_{i=1}^N \lambda_i = 1$.

Thus:

$$\langle \sigma, z \rangle = \left\langle \sigma, \sum_{i=1}^N \lambda_i x_i \right\rangle = \sum_{i=1}^N \lambda_i \langle \sigma, x_i \rangle$$

By the linearity of expectation and the fact that $\sum_{i=1}^N \lambda_i = 1$, we have:

$$\sup_{z \in \text{conv}(X)} \langle \sigma, z \rangle = \sup_{\sum_{i=1}^N \lambda_i = 1} \sum_{i=1}^N \lambda_i \langle \sigma, x_i \rangle = \sup_{x \in X} \langle \sigma, x \rangle$$

Therefore:

$$R_m(X) = R_m(\text{conv}(X))$$