Seminar 11

High-dimensional Probability and Statistics

HSE University, spring 2023

Offset Rademacher complexity

Let $\mathcal{F}$ be a convex class of functions taking their values in $\mathbb{R}$. Let $S_n = \{Z_i = (X_i, Y_i) : 1 \leqslant i \leqslant n\}$ be a training sample, consisting of i.i.d. pairs $(X_i, Y_i) \sim P$. Consider an empirical risk minimizer

$$\widehat{f} \in \operatorname*{argmin}_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \big(Y_i - f(X_i)\big)^2.$$

The performance of the estimator is measured with the squared risk:

$$R(f) = \mathbb{E}_{Z \sim P} \big(Y - f(X)\big)^2, \quad f^* \in \operatorname*{argmin}_{f \in \mathcal{F}} R(f),$$

where $Z = (X, Y)$ is generated independently of S_n.

Notation: $\ell(f, Z) = (Y - f(X))^2$.

Lemma 1. *Let $\varepsilon_1, \ldots, \varepsilon_n$ be i.i.d. Rademacher random variables (independent of the training sample). Show that*

$$\mathbb{E}R(\widehat{f}) - R(f^*) \leqslant 4\mathbb{E}\mathbb{E}_\varepsilon \sup_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \left[\ell(f, Z_i) - \ell(f^*, Z_i)\right] - \frac{1}{4n} \sum_{i=1}^{n} \left[f(X_i) - f^*(X_i)\right]^2 \right\}.$$

Definition 2. The complexity measure $\mathcal{R}^\circ(\mathcal{F}) = \mathbb{E}\mathcal{R}_n^\circ(\mathcal{F})$, where

$$\mathcal{R}_n^\circ(\mathcal{F}) = \mathbb{E}\mathbb{E}_\varepsilon \sup_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \left[\ell(f, Z_i) - \ell(f^*, Z_i)\right] - \frac{1}{4n} \sum_{i=1}^{n} \left[f(X_i) - f^*(X_i)\right]^2 \right\},$$

is called offset Rademacher complexity.

Remark 3. Similarly, one can prove that

$$\mathbb{E}\Phi\left(R(\widehat{f}) - R(f^*)\right) \leqslant \mathbb{E}\mathbb{E}_\varepsilon \Phi\left(4 \sup_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \left[\ell(f, Z_i) - \ell(f^*, Z_i)\right] - \frac{1}{4n} \sum_{i=1}^{n} \left[f(X_i) - f^*(X_i)\right]^2 \right\}\right)$$

for any convex increasing function Φ (including $\Phi(x) = e^{\lambda x}$, $\lambda > 0$).

Problem 1. Let $V \subset \mathbb{R}^n$ be a finite set of vectors of cardinality N. Show that

$$\mathbb{E}_\varepsilon \max_{v \in V} \left\{ \frac{1}{n} \sum_{i=1}^n \varepsilon_i v_i - C v_i^2 \right\} \leqslant \frac{\log N}{2Cn}.$$

From now on, we assume that $\mathcal{F}$ is a convex class of functions, satisfying the inequality

$$\int\limits_0^{\operatorname{diam}(\mathcal{F})} \sqrt{\log \mathcal{N}(\mathcal{F}, L_2(P_n), u)}\, \mathrm{d}u < \infty \quad \text{almost surely},$$

where, for any function g.

$$\|g\|_{L_2(P_n)}^2 = \frac{1}{n} \sum_{i=1}^n g(X_i)^2.$$

Fix $r > 0$ and introduce

$$\mathcal{F}_0 = \left\{ f \in \mathcal{F} : \|f - f^*\|_{L_2(P_n)} \leqslant r \right\}, \quad \mathcal{F}_k = \left\{ f \in \mathcal{F} : 2^{k-1} r < \|f - f^*\|_{L_2(P_n)} \leqslant 2^k r \right\}, \quad k \in \mathbb{N}.$$

Problem 2. Assume that, for any $f \in \mathcal{F}$, $|Y - f(X)| \leqslant B$ almost surely. Show that

$$\mathcal{R}_n^\circ(\mathcal{F}_0) \leqslant \frac{48B}{\sqrt{n}} \int\limits_0^r \sqrt{\log \mathcal{N}(\mathcal{F}, L_2(P_n), u)}\, \mathrm{d}u$$

and

$$\mathcal{R}_n^\circ(\mathcal{F}_k) \leqslant \frac{48B}{\sqrt{n}} \int\limits_{2^{k-1}r}^{2^k r} \sqrt{\log \mathcal{N}(\mathcal{F}, L_2(P_n), u)}\, \mathrm{d}u - \frac{4^k r}{16}.$$

Theorem 4. *Let $\mathcal{F}$ be such that $\mathcal{N}(\mathcal{F}, L_2(P_n), u) \leqslant (A/u)^d$ for all $u > 0$. Assume that, for any $f \in \mathcal{F}$, $|Y - f(X)| \leqslant B$ almost surely. Then*

$$\mathbb{E}\mathbb{E}_\varepsilon \sup_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^n \varepsilon_i \left[\ell(f, Z_i) - \ell(f^*, Z_i) \right] - \frac{1}{4n} \sum_{i=1}^n \left[f(X_i) - f^*(X_i) \right]^2 \right\} \lesssim \frac{Bd}{n} \log(n/d).$$