Seminar 10

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.

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

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

Hint. Use the inequality $R(f) - R(f^*) \geqslant \mathbb{E}\left[f(X) - f^*(X)\right]^2$, which holds for all $f \in \mathcal{F}$ due to the strong convexity of the squared loss.

Definition 1. The process

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

is called offset Rademacher process.

Remark 2. 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 $\Phi(x) = e^{\lambda x}$, $\lambda > 0$).