Radamacher complexity and Uniform Laws of Large Numbers

Advisor: Ambrus Tamás

Béla Szlovicsák

Glivenko-Cantelli

Theorem (Glivenko-Cantelli)

For any distribution, the empirical cumulative distribution function (CDF) $\hat{F}_n(t)$ is a strongly consistent estimator of the population CDF in the uniform norm, that is

$$||\hat{F}_n - F||_\infty \xrightarrow{a.s.} 0.$$

Generalization of Glivenko-Cantelli

Definition

Let $\mathcal{F}$ be a class of integrable real-valued functions over the domain $\mathcal{X}$, and let $X_1 \dots X_n$ be a collection of i.i.d. random variables taking values in $\mathcal{X}$, drawn according to the same distribution $\mathbb{P}$. We then use the notation:

$$||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} := \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_i)] \Big|.$$

We say that $\mathcal{F}$ is a **Glivenko-Cantelli** class for $\mathbb{P}$ if $||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}}$ converges to zero in probability as $n \longrightarrow \infty$.

- Using the function class $\mathcal{F} = \{\mathbb{1}_{(-\infty,t]} \mid t \in \mathbb{R}\}$ gives back the Glivenko-Cantelli theorem.
- In estimation problems one can check whether the class of loss functions is a Glivenko-Cantelli class for any given probability distribution.

Radamacher complexity

Definition

Given $\mathcal{F}$, a class of real-valued functions, a collection of real numbers $\mathbf{x} := (x_1, ..., x_n)$ and a collection of i.i.d. Radamacher random variables $(\epsilon_1 ... \epsilon_n)$, that is $\mathbb{P}[\epsilon_i = -1] = \mathbb{P}[\epsilon_i = 1] = \frac{1}{2}$. Then the **empirical Radamacher complexity** is given by applying $\mathcal{F}$ to $\mathbf{x}$:

$$\mathcal{R}(\mathcal{F}(\mathbf{x})/n) := \mathbb{E}_\epsilon \left[\sup_{f \in \mathcal{F}} \left|\frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(x_i)\right|\right]$$

Definition

Given a class of real-valued functions $\mathcal{F}$, as before and $\mathbf{X} := (X_1, ..., X_n)$ a collection of i.i.d. samples, the empirical Radamacher complexity $\mathcal{R}(\mathcal{F}(\mathbf{X})/n)$ is a random variable. The expectation of this random variable is the **Radamacher complexity** of the function class $\mathcal{F}$, denoted by $\mathcal{R}_n(\mathcal{F})$.

Theorem

For any b-uniformly bounded class of functions $\mathcal{F}$*, any positive integer* $n \geq 1$ *and any real* $\delta \geq 0$*, we have*

$$||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} \leq 2\mathcal{R}_n(\mathcal{F}) + \delta,$$

with probability at least $1 - \exp(-\frac{n\delta^2}{2b^2})$*. Consequently, if* $\mathcal{R}_n(\mathcal{F}) = o(1)$*, we have* $||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} \xrightarrow{a.s.} 0$.

Conclusions

- The last theorem serves as a Uniform Law of Large numbers for classes of uniformly bounded functions.
- From this Theorem the classical Glivenko-Cantelli theorem follows as a corollary.
- This also makes it clear that being able to upper bound Radamacher complexity is crucial to be able to establish meaningful results for different function classes.