Rademacher Complexity

Rademacher Complexity

The shattering coefficient is a purely combinatorial object, it does not take into account what the actual probability distribution is. This seems suboptimal.

Definition: Fix a number n of points. Let $\sigma_1, ..., \sigma_n$ be i.i.d. tosses of a fair coin (result is -1 or 1 with probability 0.5 each). The Rademacher complexity of a function class $\mathcal{F}$ with respect to n is defined as

$$\mathrm{Rad}_n(\mathcal{F}) := E \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(X_i)$$

The expectation is both over the draw of the random points X_i and the random labels σ_i.

It measures how well a function class can fit random labels.

Rademacher Complexity

There exist a number of generalization bounds for Rademacher complexities, and they tend to be sharper than the ones by combinatorial concepts like shattering coefficients. They typically look like this:

Theorem 44 (Rademacher generalization bound)

With probability at least $1 - \delta$, for all $f \in \mathcal{F}$,

$$R(f) \leq \mathrm{R_n}(f) + 2\,\mathrm{Rad}_n(\mathcal{F}) + \sqrt{\frac{\log(1/\delta)}{2n}}$$

Generalization Bounds

- Generalization bounds are a tool to answer the question whether a learning algorithm is consistent.
- Consistency refers to the estimation error, not the approximation error.
- Typically, generalization bounds have the following form:
  With probability at least $1 - \delta$, for all $f \in \mathcal{F}$

  $$R(f) \leq R_n(f) + \text{capacity term} + \text{confidence term}$$

  The capacity term measures the size of the function class.
  The confidence term deals with how certain we are about our statement.
- There are many different ways to measure the capacity of function classes, we just scratched the surface.

Generalization Bounds

- Generalizations are worst case bounds: worst case over all possible probability distributions, and worst case over all learning algorithms that pick a function from $\mathcal{F}$.