Complexity Measures

Vladimir Feinberg

May 30, 2017

Complexity measures evaluate the expressiveness of a hypothesis class; they are useful to the extent with which they relate sample and generalization error.

1 Setup

We suppose that our data comes in the form of ordered pairs from $\mathcal{X} \times \mathcal{Y}$. Samples follow a particular distribution $(x, y) \sim D$. A hypothesis class $\mathcal{H}$ is set of functions $\mathcal{X} \to \mathcal{Y}$.

A common approach to supervised learning is empirical risk minimization, where m iid samples from D, S, are used to find the $h \in \mathcal{H}$ minimizing a specified loss $\ell : \mathcal{Y}^2 \to \mathbb{R}$ over this set. Complexity measures then let us quanify exactly how much loss we can expect when sampling from D again.

Thus, it is useful to find bounds on ε, the difference between the generalization loss $\mathbb{E}\left[\ell\left(h(x), y\right)\right]$, where $(x, y) \sim D$, and sample loss, where the loss is the expectation before taken for (x, y) is uniform over S.

Let the gap between the generalization and sample error be ε.

2 Complexity Measures

3 Overview of Results

Proofs can be found in a cogent write-up by Prof. Beckage from the University of Kansas[1], copied into this repository

[1] http://ittc.ku.edu/~beckage/ml800/VC_dim.pdf