Mehryar Mohri
Foundations of Machine Learning 2014
Courant Institute of Mathematical Sciences
Homework assignment 1
September 17, 2014
Due: September 30, 2014

A. PAC learning of n-dimensional rectangles

Give a PAC-learning algorithm for C, the set of axis-aligned n-dimensional rectangles in $\mathbb{R}^n$, that is $C = \{[a_1, b_1] \times \cdots \times [a_n, b_n] \colon a_i, b_i \in \mathbb{R}\}$. You should give a careful proof similar to what was given in class for axis-aligned rectangles (case $n = 2$). How does the sample complexity vary as a function of n?

B. Rademacher complexity of regularized neural networks

Let the input space be $X = \mathbb{R}^{n_1}$. In this problem, we consider the family of regularized neural networks defined by the following set of functions mapping X to $\mathbb{R}$:

$$\mathcal{H} = \left\{ \mathbf{x} \mapsto \sum_{j=1}^{n_2} w_j \sigma(\mathbf{u}_j \cdot \mathbf{x}) \colon \|\mathbf{w}\|_1 \leq \Lambda', \|\mathbf{u}_j\|_2 \leq \Lambda, \forall j \in [1, n_2] \right\},$$

where σ is an L-Lipschitz function. As an example, σ could be the sigmoid function which is 1-Lipschitz.

1. Show that $\widehat{\mathfrak{R}}_S(\mathcal{H}) = \frac{\Lambda'}{m} \operatorname{E}_{\boldsymbol{\sigma}} \left[\sup_{\|\mathbf{u}\|_2 \leq \Lambda} \left| \sum_{i=1}^m \sigma_i \sigma(\mathbf{u} \cdot \mathbf{x}_i) \right| \right]$.

2. Use the following form of Talagrand's lemma valid for all hypothesis sets H and L-Lipschitz function Φ:

$$\frac{1}{m} \operatorname*{E}_{\boldsymbol{\sigma}} \left[\sup_{h \in H} \left| \sum_{i=1}^m \sigma_i (\Phi \circ h)(x_i) \right| \right] \leq \frac{L}{m} \operatorname*{E}_{\boldsymbol{\sigma}} \left[\sup_{h \in H} \left| \sum_{i=1}^m \sigma_i h(x_i) \right| \right],$$

to upper bound $\widehat{\mathfrak{R}}_S(\mathcal{H})$ in terms of the empirical Rademacher complexity of $\mathcal{H}'$, where $\mathcal{H}'$ is defined by

$$\mathcal{H}' = \{ \mathbf{x} \mapsto s(\mathbf{u} \cdot \mathbf{x}) \colon \|\mathbf{u}\|_2 \leq \Lambda, s \in \{-1, +1\} \}.$$

3. Use the Cauchy-Schwarz inequality to show that

$$\widehat{\mathfrak{R}}_S(\mathcal{H}') = \frac{\Lambda}{m} \mathop{\mathrm{E}}_{\boldsymbol{\sigma}} \left[\left\| \sum_{i=1}^{m} \sigma_i \mathbf{x}_i \right\|_2 \right].$$

4. Use the inequality $\mathrm{E}[\|\mathbf{X}\|_2] \leq \sqrt{\mathrm{E}[\|\mathbf{X}\|_2^2]}$, which holds by Jensen's inequality to upper bound $\widehat{\mathfrak{R}}_S(\mathcal{H}')$.

5. Assume that for all $\mathbf{x} \in S$, $\|\mathbf{x}\|_2 \leq r$ for some $r > 0$. Use the previous questions to derive an upper bound on the Rademacher complexity of $\mathcal{H}$ in terms of r.