ECE 299: STATISTICAL LEARNING THEORY

MAXIM RAGINSKY

Homework 2

Assigned February 23, 2011; due March 17, 2011

---

(1) **Uniform deviations and Rademacher averages.** Let $\mathcal{F}$ be a class of functions $f : \mathsf{Z} \to [0,1]$. Given a distribution $P \in \mathcal{P}(\mathsf{Z})$ and an i.i.d. sample Z^n from P, consider the uniform deviation

$$\Delta_n(Z^n) = \|P_n - P\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(Z_i) - P(f) \right|$$

and the Rademacher average

$$R_n(\mathcal{F}(Z^n)) = \mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i f(Z_i) \right| \right],$$

where σ^n are n i.i.d. Rademacher random variables independent of Z^n. Prove that for any $t > 0$

$$\mathbb{P}\Big(\Delta_n(Z^n) - 2R_n(\mathcal{F}(Z^n)) \ge \mathbb{E}\left[\Delta_n(Z^n) - 2R_n(\mathcal{F}(Z^n))\right] + t \Big) \le e^{-2nt^2/25}$$

and

$$\mathbb{P}\Big(\Delta_n(Z^n) - 2R_n(\mathcal{F}(Z^n)) \ge t \Big) \le e^{-2nt^2/25}.$$

(2) **A simple penalized ERM algorithm.** When choosing a suitable hypothesis space, we often face the following dilemma: If the hypothesis space is not "rich" enough, even the best hypothesis from it may have unacceptably high expected risk. On the other hand, if it is *too* rich, then there may be no guarantee of good behavior of the uniform deviations of empirical means from true means. A great deal of effort in statistical learning theory is devoted to finding ways of coping with this dilemma. One such way is to use *multiple* hypothesis spaces, run ERM on each of them, and then choose the ERM solution that achieves a good trade-off between the empirical risk and some measure of complexity of the hypothesis space at hand. In this problem, you will investigate a very simple penalized ERM algorithm that chooses between finitely many hypothesis classes, where the complexity of each class is measured in a data-driven way by means of Rademacher averages.

Let $\mathcal{F}_1, \ldots, \mathcal{F}_M$ be a finite collection of hypothesis spaces, where each $\mathcal{F}_m$ consists of functions from some space Z into $[0,1]$. Let Z^n be an i.i.d. sample from an unknown distribution $P \in \mathcal{P}(\mathsf{Z})$. For each $m = 1, \ldots, M$ let

$$\widehat{f}_n^{(m)} \triangleq \operatorname*{arg\,min}_{f \in \mathcal{F}_m} P_n(f)$$

be an empirical risk minimizer over $\mathcal{F}_m$, and let $\widehat{f}_n = \widehat{f}_n^{(\widehat{m})}$, where $\widehat{m} \in \{1, \dots, M\}$ is set to

$$\widehat{m} = \underset{1 \le m \le M}{\arg\min} \left[P_n(\widehat{f}_n^{(m)}) + 2R_n(\mathcal{F}_m(Z^n)) \right].$$

Prove that

$$P(\widehat{f}_n) \le \min_{1 \le m \le M} \left[\inf_{f \in \mathcal{F}_m} P_n(f) + 2R_n(\mathcal{F}_m(Z^n)) \right] + \sqrt{\frac{25 \log\left(\frac{M}{\delta}\right)}{2n}}$$

with probability at least $1 - \delta$.

Hint. You may need to use the bounds from the previous problem separately for each m and then combine them.