ECE 299: STATISTICAL LEARNING THEORY

MAXIM RAGINSKY

Homework 4

Assigned April 11, 2011; due April 18, 2011

---

(1) **Bernstein's inequality: tighter concentration under variance constraints.** Let $Z_1, \dots, Z_n$ be n independent random variables taking values in some space Z. Consider a bounded function $f : \mathsf{Z} \to \mathbb{R}$ and let

$$b = \|f\|_\infty \equiv \sup_{z \in \mathsf{Z}} |f(z)|$$

and

$$v = \mathbb{E}\left[\sum_{i=1}^n f^2(Z_i)\right].$$

Bernstein's inequality says that, for any $t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^n f(Z_i) \ge \mathbb{E}\left[\sum_{i=1}^n f(Z_i)\right] + t\right) \le \exp\left(-\frac{t^2}{2(v + bt/3)}\right).$$

(a) Prove that, with probability at least $1 - \delta$,

$$\sum_{i=1}^n f(Z_i) \le \mathbb{E}\left[\sum_{i=1}^n f(Z_i)\right] + \sqrt{2v \log(1/\delta)} + \frac{2b \log(1/\delta)}{3}.$$

(b) Bernstein's inequality may give tighter bounds than Hoeffding's inequality because of the additional information on the variance. Use the result of (a) to determine a sufficient condition for this to be the case.

(2) **Talagrand's inequality: concentration of uniform deviations under variance constraints.** As before, let $Z_1, \dots, Z_n$ be n independent Z-valued random variables. Let $\mathcal{F}$ be a class of functions $f : \mathsf{Z} \to \mathbb{R}$, and let

$$b = \sup_{f \in \mathcal{F}} \|f\|_\infty$$

and

$$v = \sup_{f \in \mathcal{F}} \sum_{i=1}^n \mathrm{Var}\, f(X_i).$$

It has been a longstanding open question whether one can obtain a concentration inequality similar to Bernstein's for the uniform deviation $\sup_{f \in \mathcal{F}} |\sum_{i=1}^n (f(Z_i) - \mathbb{E} f(Z_i))|$. The first such result was obtained by Michel Talagrand in 1994.

In its improved form, *Talagrand's inequality* says the following: There exists an absolute constant $C \ge 1$, such that for any $t > 0$

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \left|\sum_{i=1}^n (f(Z_i) - \mathbb{E} f(Z_i))\right| \ge \mathbb{E} \sup_{f \in \mathcal{F}} \left|\sum_{i=1}^n (f(Z_i) - \mathbb{E} f(Z_i))\right| + C(\sqrt{vt} + bt)\right) \le e^{-t}.$$

(a) Use Talagrand's inequality to prove that, for any $t > 0$,

$$\frac{1}{n} \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} (f(Z_i) - \mathbb{E}f(Z_i)) \right| \le 2\mathbb{E}R_n(\mathcal{F}(Z^n)) + \frac{C}{n}(\sqrt{vt} + bt)$$

with probability at least $1 - e^{-t}$.

(b) Suppose that $Z_1, \ldots, Z_n$ are i.i.d. and that the class $\mathcal{F}$ has $b = 1$. Use the result of (a) to derive a bound analogous to Corollary 2 in the lecture notes on ERM. When will this bound be tighter than the bound of Corollary 2?