UC Berkeley
Department of Electrical Engineering and Computer Science
Department of Statistics

EECS 281B / STAT 241B
ADVANCED TOPICS IN STATISTICAL LEARNING THEORY

Problem Set 3

Spring 2009

Issued: Monday, February 23, 2009 **Due:** Monday, March 9, 2009

Problem 3.1
Given a collection of i.i.d. samples $\{(x^{(i)}, y^{(i)})$ for $i = 1, \ldots, n$, consider the kernelized classification method, based on determining a classifier $\widehat{f}$ by solving the optimization problem

$$\min_{f \in \mathcal{H}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \phi\big(y^{(i)} f(x^{(i)})\big) + \frac{\lambda_n}{2} \|f\|_{\mathcal{H}}^2 \right\}, \tag{1}$$

where $\mathcal{H}$ is some reproducing kernel Hilbert space, and $\|\cdot\|_{\mathcal{H}}$ is the norm in this RKHS. For each of the following choices of surrogate loss ϕ:

(i) First check whether or not ϕ is classification-calibrated.

(ii) Compute the H_ϕ and Ψ functions from our theorem on surrogate losses.

(iii) Derive the dual form of the problem (1). (As discussed in class, your dual problem should be a convex program in $\mathbb{R}^n$.)

Possible choices of surrogate loss:

(a) Logistic loss $\phi(t) = \log[1 + \exp(-t)]$.

(b) Exponential loss $\phi(t) = \exp(-t)$.

Problem 3.2
In this exercise, we explore some challenges with high-dimensional data. For $i = 1, \ldots, n$, let $X^{(i)} \in \mathbb{R}^d$ be i.i.d. random vectors drawn from the uniform distribution on $[0, 1]^d$. Given a new sample X, define the expected minimum distance to the nearest data point

$$\rho_\infty(d, n) \;=\; \mathbb{E}[\min_{i=1,\ldots,n} \|X^{(i)} - X\|_\infty]$$

The goal of this exercise is to understand how $\rho_\infty(d, n)$ behaves for large d and n.

(a) Show that for $t > 0$,

$$\mathbb{P}[\min_{i=1,\ldots,n} \|X^{(i)} - X\|_\infty > t] \;\geq\; 1 - n\,(2t)^d.$$

(b) Suppose that we wanted to be sure that the new sample X was within distance $1/4$ of at least one data point with probability at least $1/2$. Give a lower bound on how large n would have to be as a function of dimension d.

(c) Use part (a) to show that $\rho_\infty(d, n) \geq \frac{d}{2\,(d+1)} n^{-1/d}$. (*Hint:* Recall that for a non-negative random variable Z with a first moment, $\mathbb{E}[Z] = \int_0^\infty \mathbb{P}[Z > t]dt$.)

(d) Compute the value of your lower bound for $d \in \{1, 10, 20\}$ and $n \in \{100, 1000, 10000, 100000\}$.

Problem 3.3

(Covering/packing) Consider a set S on which a metric ρ is defined. The ball of radius ϵ centered at x^* is given by

$$\mathbb{B}_\epsilon(x^*) \quad = \quad \{x \mid \rho(x, x^*) \leq \epsilon\}.$$

An ϵ-*packing* is a collection of balls $\{\mathbb{B}_\epsilon(x_i), i = 1, \ldots, M\}$ centered at points $x_i \in S$ that are all disjoint. The packing number $M(\epsilon; S, \rho)$ is the cardinality of the largest ϵ-packing. An ϵ-*covering* is a collection of balls $\{\mathbb{B}_\epsilon(x_i), i = 1, \ldots, N\}$ such that $S \subseteq \cup_{i=1}^N \mathbb{B}_\epsilon(x_i)$. The covering number $N(\epsilon; S, \rho)$ is the cardinality of the smallest ϵ-covering.

(a) Show that $M(\epsilon; S, \rho) \leq N(\epsilon; S, \rho)$.

(b) Show that $N(2\epsilon; S, \rho) \leq M(\epsilon; S, \rho)$.

Problem 3.4

Not to be graded: In this problem, we study how to obtain some bounds on the packing number $M(\epsilon; S, \|\cdot\|_2)$ of the ℓ_2-ball $S = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 1\}$.

(a) By considering the volumes of ℓ_2 balls in $\mathbb{R}^d$, show that $M(\epsilon; S, \|\cdot\|_2) \leq (\frac{4}{\epsilon})^d$. (*Hint:* The volume of the ℓ_2-ball scales as r^d, where r is its radius.)

(b) We now consider how to derive a lower bound on $M(\epsilon; S, \|\cdot\|_2)$ — say $\epsilon = 1/4$ for concreteness. Consider the following random procedure. Given an integer M, let $z_i, i = 1, \ldots, M$ be i.i.d. Gaussian random vectors in $\mathbb{R}^d$, each distributed as $N(0, I_{d\times d})$. We then define $x_i = z_i/\|z_i\|_2$, and note that x_i is an element of S by construction.

   (i) For each pair $i \neq j$, compute an upper bound on the probability $\|x_i - x_j\|_2 \leq 1/4$. (*Hint:* You may find the following tail bound useful: if $Z \sim \chi^2_d$ is chi-squared with d degrees of freedom, then for $\delta \in (0, 1/2)$, $\mathbb{P}[|Z - d| \geq \delta\, d] \leq 2\exp(-d\delta^2/16)$.)

   (ii) Use your result from part (i) to show that $M(1/4; S, \|\cdot\|_2) \geq c_1 \exp(c_2 d)$ for some positive constants c_1 and c_2.