STAT 881
Spring 2008

HOMEWORK 4

1. Let $\mathcal{X}$ be a domain and let $K(s,t)$ be a non-negative definite function on $\mathcal{X} \times \mathcal{X}$ with $\int_{\mathcal{X}} \int_{\mathcal{X}} K^2(s,t)dsdt < \infty$. Then, by the Mercer-Hilbert-Schmidt Theorem, there exists an orthonormal set $\{\phi_\nu\}_{\nu=1}^{\infty}$ on $\mathcal{X}$, $\int_{\mathcal{X}} \phi_\mu(s)\phi_\nu(s)ds = I(\mu = \nu)$ and non-negative eigenvalues λ_ν with $\sum_{\nu=1}^{\infty} \lambda_\nu^2 < \infty$ such that $K(s,t) = \sum_{\nu=1}^{\infty} \lambda_\nu \phi_\nu(s)\phi_\nu(t)$. Verify that K is the reproducing kernel of $\mathcal{H}_K$ with inner product

$$\langle f, g \rangle_{\mathcal{H}_K} = \sum_{\nu=1}^{\infty} \frac{(f, \phi_\nu)(g, \phi_\nu)}{\lambda_\nu},$$

where $(u, v) = \int_{\mathcal{X}} u(s)v(s)ds$. Identify the 'feature mapping' induced by K.

2. For a probability distribution of (X, Y) and a sequence of classification rules $\{f_n\}$ based on training data of size n, we say $\{f_n\}$ is consistent if $ER(f_n) \to R^*$ as $n \to \infty$. Show that the consistency of $\{f_n\}$ is equivalent to convergence of $R(f_n)$ to R^* in probability.

3. As an example of scenarios where complete separation of cases into two categories is feasible by simple decision boundaries, consider the following problem.
$P(Y = 0) = P(Y = 1) = 1/2$ and the conditional distribution of X given class 0 is uniform on the unit disk D_0 centered at (-3,0) and that given class 1 is uniform on the unit disk D_1 centered at (3,0). Suppose that we have n iid pairs of (X_i, Y_i) from the specified distribution, and let f_n^k be the k-nearest neighbor rule.

   (a) For the nearest neighbor rule, first show that $P_{\mathcal{D}_n}(f_n^1(X) \neq Y | (X, Y)) = (1/2)^n$ to conclude that $P_{\mathcal{D}_n,(X,Y)}(f_n^1(X) \neq Y) = (1/2)^n$.

   (b) Similarly, when $k = 2k_0 + 1$ with a positive integer k_0, verify that $P_{\mathcal{D}_n,(X,Y)}(f_n^k(X) \neq Y) = (1/2)^n \sum_{j=0}^{k_0} \binom{n}{j}$.
   In this example, the nearest neighbor rule has a strictly lower probability of error than any k-NN rule for $k \neq 1$.

   (c) For a fixed k, comment on how the probability of error for the k-NN rule varies as $n \to \infty$.

4. For universal consistency of kernel classification rules, the regularity conditions of a kernel are assumed. Without reference to any general results on regularity of kernels, show directly that the Gaussian kernel, $K(x) = \exp(-\|x\|^2)$ defined on $\mathbb{R}^d$ is regular.