Homework 2

Jean-Philippe Vert

Due February 2, 2009

1 Rademacher complexity

A Rademacher variable is a random variables σ that can take two possible values, -1 and $+1$, with equal probability $1/2$.

1. Let $(u_1, u_2, \ldots, u_N)$ be N vectors in a Hilbert space endowed with an inner product $< ., . >$, and let $\sigma_1, \sigma_2, \ldots, \sigma_N$ be N independent Rademacher variables. Show that:

$$\mathbb{E}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_i\sigma_j < u_i, u_j >\right) = \sum_{i=1}^{N} \parallel u_i \parallel^2 .$$

2. Let K be a positive definite kernel on a space $\mathcal{X}$, $\mathcal{H}_K$ denote the associated reproducing kernel Hilbert space, and $B_R = \{f \in \mathcal{H}_K, \parallel f \parallel_{\mathcal{H}_K} \leq R\}$. Let a set of points $\mathcal{S} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ with $\mathbf{x}_i \in \mathcal{X}$ $(i = 1, \ldots, N)$, and let $\sigma_1, \sigma_2, \ldots, \sigma_N$ be N independent Rademacher variables. Show that:

$$\mathbb{E} \sup_{f \in B_R} \left| \sum_{i=1}^{N} \sigma_i f(\mathbf{x}_i) \right| \leq R \sqrt{\sum_{i=1}^{N} K(\mathbf{x}_i, \mathbf{x}_i)} .$$

2 Conditionally positive definite kernels

Let $\mathcal{X}$ be a set. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called *conditionally positive definite* (c.p.d.) if and only if it is symmetric and satisfies:

$$\sum_{i,j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

for any $n \in \mathbb{N}$, $x_1, x_2, \ldots, x_n \in \mathcal{X}^n$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}^n$ with $\sum_{i=1}^n a_i = 0$
.

1. Show that a positive definite (p.d.) function is c.p.d.

2. Is a constant function p.d.? Is it c.p.d.?

3. If $\mathcal{X}$ is a Hilbert space, then is $k(x, y) = -||x - y||^2$ p.d.? Is it c.p.d.?

4. Let $\mathcal{X}$ be a nonempty set, and $x_0 \in \mathcal{X}$ a point. For any function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, let $\tilde{k} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be the function defined by:

$$\tilde{k}(x, y) = k(x, y) - k(x_0, x) - k(x_0, y) + k(x_0, x_0).$$

Show that k is c.p.d. if and only if $\tilde{k}$ is p.d.

5. Let k be a c.p.d. kernel on $\mathcal{X}$ such that $k(x, x) = 0$ for any $x \in \mathcal{X}$. Show that there exists a Hilbert space $\mathcal{H}$ and a mapping $\Phi : \mathcal{X} \to \mathcal{H}$ such that, for any $x, y \in \mathcal{X}$,

$$k(x, y) = -||\Phi(x) - \Phi(y)||^2.$$

6. Show that if k is c.p.d., then the function $\exp(tk(x, y))$ is p.d. for all $t \geq 0$

7. Conversely, show that if the function $\exp(tk(x, y))$ is p.d. for any $t \geq 0$, then k is c.p.d.

8. **(BONUS)** Show that the opposite of the shortest-path distance on a tree is c.p.d over the set of vertices (a tree is an undirected graph without loops. The shortest-path distance between two vertices is the number of edges of the unique path that connects them). Is it also c.p.d. over general graphs?