MVA "Kernel methods"
Homework 1

Jean-Philippe Vert

Due February 11, 2011

1 Kernel PCA

Let $S_{train} = (x_1, \ldots, x_n)$ and $S_{test} = (x_{n+1}, \ldots, x_{n+p})$ be two sets of points in a space endowed with a positive definite kernel K. Propose an algorithm to project the set S_{test} onto the first principal directions obtained by kernel PCA on the set S_{train}. (*Hint: be careful on how to center the data.*)

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?