MVA "Kernel methods"
Homework 2

Jean-Philippe Vert

Due February 4, 2015

Exercice 1. Kernel PCA for data denoising
Let $\mathcal{X}$ be a space endowed with a p.d. kernel K, and $\Phi : \mathcal{X} \to \mathcal{H}$ a mapping to a Hilbert space $\mathcal{H}$ such that for all $x, x' \in \mathcal{X}$,

$$\langle \Phi(x), \Phi(x') \rangle = K(x, x') \,.$$

Let $\mathcal{S} = \{x_1, \ldots, x_n\}$ be a set of points in $\mathcal{X}$, and

$$m = \frac{1}{n} \sum_{i=1}^{n} \Phi(x_i)$$

their barycenter in the feature space.
1. For $x \in \mathcal{X}$, let

$$\Psi(x) = P_d \left(\Phi(x) - m \right) + m$$

where P_d is the projection onto the linear span of the first d kernel principal components of $\mathcal{S}$. Show that $\Psi(x)$ can be expressed as

$$\Psi(x) = \sum_{i=1}^{n} \gamma_i \Phi(x_i) \,,$$

for some γ_i to be explicitly computed.
2. For $y \in \mathcal{X}$, express

$$f(y) = \| \, \Phi(y) - \Psi(x) \, \|^2$$

in terms of kernel evaluations. Explain why minimizing $f(y)$ can be thought of as a method to "denoise" x.

3. Express f and ∇f in the case $\mathcal{X} = \mathbb{R}^p$ and $K(x, x') = \exp\left(-\frac{\| x-x' \|^2}{2\sigma^2}\right)$. Propose an iterative algorithm (for example gradient descent) to find a local minimum of f in that case.

4. Download the USPS ZIP code data from `http://statweb.stanford.edu/~tibs/ElemStatLearn/data.html` Visualize (a subset of) the dataset in two dimensions with kernel PCA, for different kernels. Implement the procedure discussed in question 4, and test it on some data that you have corrupted with noise. Compute how similar the denoised images are from the original (uncorrupted) images as a function of the number of principal components used.