../../../psulogo_horiz_bw-eps-converted-to.pdf

STAT 672

Statistical Learning II

Fall 2019
Homework 2
Due February 5^{th} at the beginning of class

1 Kernel centering

Let $\mathcal{X}$ be a non empty set, K a kernel over $\mathcal{X}$ and H the RKHS with kernel K. Let $x_1, \ldots, x_n$, be n points in $\mathcal{X}$.

Let K_c like "K centered" be another kernel over $\mathcal{X}$ defined by

$$K_c(x, y) = \langle K(., x) - \bar{f}(.), K(., y) - \bar{f}(.) \rangle_H, \text{ with } \bar{f}(.) = \frac{1}{n} \sum_{i=1}^{n} K(., x_i) \tag{1}$$

1. Verify that K_c is a positive definite kernel;

   Let H_c be the RKHS with reproducing kernel H_c.

2. (**) Verify that for any $f \in H_c$,

$$\frac{1}{n} \sum_{i=1}^{n} f(x_i) = 0 \tag{2}$$

   In homework 1, you have learned to sample functions from a RKHS with kernel K over the set $\mathcal{X} = \{-m, \ldots, m\}$ according to the probability

$$p(f) = Ce^{-\frac{||f||^2}{2}} \tag{3}$$

   Here, you are asked to do the same thing but over the RKHS of a centered kernel K_c. Specifically, choose $m = 10$, $n = 11$, $x_1 = -10, x_2 = -9, \ldots, x_{11} = 0$.

3. (**) Write K_c in term of K using matrix operations.

4. Choose the Gaussian kernel with $\tau = 10$, and show 10 samples. Figure **??** shows examples of what you should find. Check that for each curve f,

$$\frac{1}{11} \sum_{i=-10}^{0} f(i) = 0 \tag{4}$$

   up to numerical errors.

Figure 1: Example of 10 samples of ''functions" of the RKHS over $\{-10,\dots,10\}$ with the Gaussian kernel, $\tau = 10$ under the centering constraint $\frac{1}{11}\sum_{i=-10}^{0} f(i) = 0$

Figure 2: Example of resu;ts with Kernel PCA

2 Kernel PCA

$\mathcal{X}$ a non empty set, $x_1,\dots,x_n \in \mathcal{X}$, K a centered kernel, that is, starting with a kernel G over $\mathcal{X}$,

$$K(x,y) = \langle G(.,x) - \bar{f}, G(.,y) - \bar{f}\rangle_H, \text{ with } \bar{f} = \frac{1}{n}\sum_{i=1}^{n} G(.,x_i) \tag{5}$$

Assume for simplicity that K is full rank. Notate $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$ the e-values and $u_1,\dots,u_n$ the corresponding e-vectors. We have seen in class that the principal directions $f_1,\dots,f_n$ are

$$f_i = \sum_{j=1}^{n} \alpha_{ij} K(.,x_j), \text{ with } \alpha_i = \lambda_i^{-\frac{1}{2}} u_i \tag{6}$$

1. verify that
$$\langle f_i, f_k\rangle_H = \delta_{ik} \tag{7}$$
where $\delta_{ik} = 1$ if $i = k$ and $\delta_{ik} = 0$ if $i \neq k$.

2. Show that the orthogonal projection of any $f \in H$, the RKHS with kernel K onto
$$V = span\{f_1,\dots,f_n\} \tag{8}$$
is
$$\pi_v(f) = \langle f, f_1\rangle_H f_1 + \dots, \langle f, f_n\rangle_H f_n \tag{9}$$
Now, let us project the feature functions of $x_1,\dots,x_n$, that is $K(.,x_1),\dots K(.,x_n)$. Show that
$$\langle K(.,x_k), f_i\rangle = \lambda_i^{\frac{1}{2}} u_{ki} \tag{10}$$

3. Perform kernel PCA on the MNIST dataset. Choose the digits 1 and 7. Start with the linear kernel $G(x,y) = x^T y$. Sample $n = 500$ digits. Show 8 projections onto the first 8 principal directions. Do not forget to center the kernel. your result should look like Figure **??**.

4. Redo the same but this time with a non linear kernel of your choice.