

**STAT 672**  
**Statistical Learning II**

John Karasev  
Homework 2  
January 31, 2020

## 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  $x_1, \dots, x_n$ , be  $n$  points in  $\mathcal{X}$ . Let  $K_c$  like “K centered” be another kernel over  $\mathcal{X}$  defined by

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

1. Verify that  $K_c$  is a positive definite kernel;  
Let  $H_c$  be the RKHS with reproducing kernel  $K_c$ .

*Proof.*

for any  $n$  in  $\mathbb{N}$ ,  $(a_1, \dots, a_n)$  in  $\mathbb{R}^n$ ,  
 $(x_1, \dots, x_n) \in \mathcal{X}^n$

$$\begin{aligned} & \sum_{i,j=1}^n a_i a_j K_c(x_i, x_j) \\ &= \sum_{i,j=1}^n a_i a_j \langle K(., x_i) - \bar{f}(.), K(., x_j) - \bar{f}(.) \rangle_{\mathcal{H}} \\ &= \left\langle \sum_{i=1}^n a_i [K(., x_i) - \bar{f}(.)], \sum_{j=1}^n a_j [K(., x_j) - \bar{f}(.)] \right\rangle_{\mathcal{H}} \\ &= \left\| \sum_{i=1}^n a_i [K(., x_i) - \bar{f}(.)] \right\|_{\mathcal{H}}^2 \geq 0 \end{aligned}$$

□

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

$$\begin{aligned} & \frac{1}{n} \sum_{l=1}^n f(x_l) = 0 \quad (2) \\ & f(x) = \sum_{m=1}^n \alpha_m K_c(x, y_m) \end{aligned}$$

*Proof.*

$$\begin{aligned}
\frac{1}{n} \sum_{l=1}^n f(x_l) &= 0 \\
\frac{1}{n} \sum_{l=1}^n \sum_{m=1}^n \alpha_m K_c(x_l, y_m) &= 0 \\
\frac{1}{n} \sum_{l=1}^n \sum_{m=1}^n \alpha_m \left\langle K(., x_l) - \frac{1}{n} \sum_{i=1}^n K(., x_i), K(., y_m) - \frac{1}{n} \sum_{i=1}^n K(., x_i) \right\rangle_{\mathcal{H}} &= 0 \\
\frac{1}{n} \left\langle \sum_{l=1}^n K(., x_l) - \sum_{l=1}^n \frac{1}{n} \sum_{i=1}^n K(., x_i), \sum_{m=1}^n \alpha_m K(., y_m) - \sum_{m=1}^n \alpha_m \frac{1}{n} \sum_{i=1}^n K(., x_i) \right\rangle_{\mathcal{H}} &= 0 \\
\left\langle \sum_{l=1}^n K(., x_l) - \frac{1}{n} \sum_{i=1}^n K(., x_i), \sum_{m=1}^n \frac{\alpha_m K(., y_m)}{n} - \sum_{m=1}^n \frac{\alpha_m}{n} \sum_{i=1}^n \frac{K(., x_i)}{n} \right\rangle_{\mathcal{H}} &= 0 \\
\left\langle 0, \sum_{m=1}^n \frac{\alpha_m K(., y_m)}{n} - \sum_{m=1}^n \frac{\alpha_m}{n} \sum_{i=1}^n \frac{K(., x_i)}{n} \right\rangle_{\mathcal{H}} &= 0 \\
0 &= 0
\end{aligned}$$

□

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

$$p(f) = C e^{-\frac{\|f\|^2}{2}} \quad (3)$$

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

$$\begin{aligned}
K_c(x, y) &= \left\langle K(., x) - \frac{1}{n} \sum_{i=1}^n K(., x_i), K(., y) - \frac{1}{n} \sum_{i=1}^n K(., x_i) \right\rangle \\
&= \left\langle K(., x), K(., y) \right\rangle + \left\langle K(., x), \frac{1}{n} \sum_{i=1}^n K(., x_i) \right\rangle + \\
&\quad \left\langle K(., y), \frac{1}{n} \sum_{i=1}^n K(., x_i) \right\rangle + \left\langle \frac{1}{n} \sum_{i=1}^n K(., x_i), \frac{1}{n} \sum_{j=1}^n K(., x_j) \right\rangle \\
&= K(x, y) + \frac{1}{n} \sum_{i=1}^n K(x, x_i) + \frac{1}{n} \sum_{i=1}^n K(y, x_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j)
\end{aligned}$$

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

$$\begin{aligned}
K_c(x, y) &= K(x, y) + \frac{1}{n} \sum_{i=1}^n K(x, x_i) + \frac{1}{n} \sum_{i=1}^n K(y, x_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j) \\
&= K(x, y) - \frac{1}{n} \mathbf{K}_{\mathbf{y}}^T \mathbf{1} - \frac{1}{n} \mathbf{K}_{\mathbf{x}}^T \mathbf{1} + \frac{1}{n} \mathbf{1}^T \mathbf{K} \mathbf{1}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{K} &= [K(x_i, x_j)]_{ij \ (n,n)} \\
\mathbf{K}_{\mathbf{y}} &= [K(y, x_i)]_{i \ (1,n)}
\end{aligned}$$

$$\mathbf{K}_{\mathbf{y}} = [K(x, x_i)]_{i \in (1, n)}$$

$$\mathbb{1} = [1]_{i \in (1, n)}$$

The centered kernel matrix can be written as:

$$\begin{aligned} \mathbf{K}_c &= \mathbf{K} - \mathbf{U}\mathbf{K} - \mathbf{K}\mathbf{U} - \mathbf{U}\mathbf{K}\mathbf{U} \\ &= (\mathbf{I} - \mathbf{U})\mathbf{K}(\mathbf{I} - \mathbf{U}) \end{aligned}$$

where

$$\mathbf{U} = \left[ \frac{1}{n} \right]_{ij \in (n, n)}$$

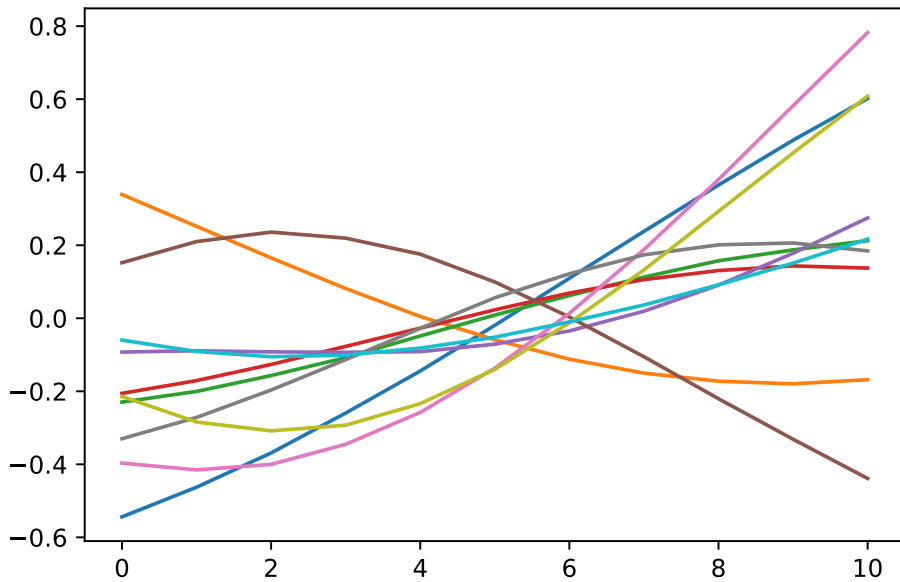
$\mathbf{I}_{(n, n)}$  is the identity matrix

4. Choose the Gaussian kernel with  $\tau = 10$ , and show 10 samples. Check that for each curve  $f$ ,

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

up to numerical errors.

$$\frac{1}{11} \sum_{i=-10}^0 \mathbf{f}(i) = \frac{1}{11} \sum_{i=-10}^0 \begin{bmatrix} f_1(i) \\ f_2(i) \\ f_3(i) \\ \vdots \\ f_9(i) \\ f_{10}(i) \end{bmatrix} = \begin{bmatrix} -0.00050791 \\ -0.02180374 \\ -0.03592173 \\ -0.0435211 \\ -0.04306572 \\ -0.03498534 \\ -0.01940224 \\ 0.00311266 \\ 0.03170139 \\ 0.06432604 \end{bmatrix}$$



## 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  $\mathcal{X}$ ,

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

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

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

1. verify that

$$\langle f_i, f_k \rangle_H = \delta_{ik}$$

where  $\delta_{ik} = 1$  if  $i = k$  and  $\delta_{ik} = 0$  if  $i \neq k$ .

$$\begin{aligned} \langle f_i, f_k \rangle_{\mathcal{H}} &= \lambda_i^{-\frac{1}{2}} \mathbf{u}_i^T \mathbf{K} \mathbf{u}_k \lambda_k^{-\frac{1}{2}} = \lambda_i^{-\frac{1}{2}} \mathbf{u}_i^T \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \mathbf{u}_k \lambda_k^{-\frac{1}{2}} \\ &= \lambda_i^{-\frac{1}{2}} \mathbf{e}_i^T \mathbf{\Lambda} \mathbf{e}_k \lambda_k^{-\frac{1}{2}} = \lambda_i^{-\frac{1}{2}} \lambda_i \mathbf{e}_i^T \mathbf{e}_k \lambda_k^{-\frac{1}{2}} = \frac{\lambda_i^{\frac{1}{2}}}{\lambda_k^{-\frac{1}{2}}} \mathbf{e}_i^T \mathbf{e}_k \\ &= \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{else} \end{cases} = \delta_{ik} \end{aligned}$$

where

$$\mathbf{e}_k = [1 \text{ if } k = j \text{ else } 0]_{(j,1)}$$

2. Show that the orthogonal projection of any  $f \in H$ , the RKHS with kernel  $K$  onto

$$V = \text{span}\{f_1, \dots, f_n\}$$

is

$$\pi_v(f) = \langle f, f_1 \rangle_H f_1 + \dots, \langle f, f_n \rangle_H f_n$$

Now, let us project the feature functions of  $x_1, \dots, x_n$ , that is  $K(., x_1), \dots, K(., x_n)$ . Show that

$$\langle K(., x_k), f_i \rangle = \lambda_i^{\frac{1}{2}} u_{ki}$$

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.

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