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 $x_1, \ldots, x_n$, 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 K_c.

   Proof.

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

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

   □

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

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

$$f(x) = \sum_{m=1}^{n} \alpha_m K_c(x, y_m)$$

Proof.

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

□

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

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

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

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

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

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

where

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

$$\boldsymbol{K_y} = [K(x, x_i)]_{i\ (1,n)}$$
$$\mathbb{1} = [1]_{i\ (1,n)}$$

The centered kernel matrix can be written as:

$$\begin{aligned}\boldsymbol{K_c} &= \boldsymbol{K} - \boldsymbol{UK} - \boldsymbol{KU} - \boldsymbol{UKU} \\ &= (\boldsymbol{I} - \boldsymbol{U})\boldsymbol{K}(\boldsymbol{I} - \boldsymbol{U})\end{aligned}$$

where

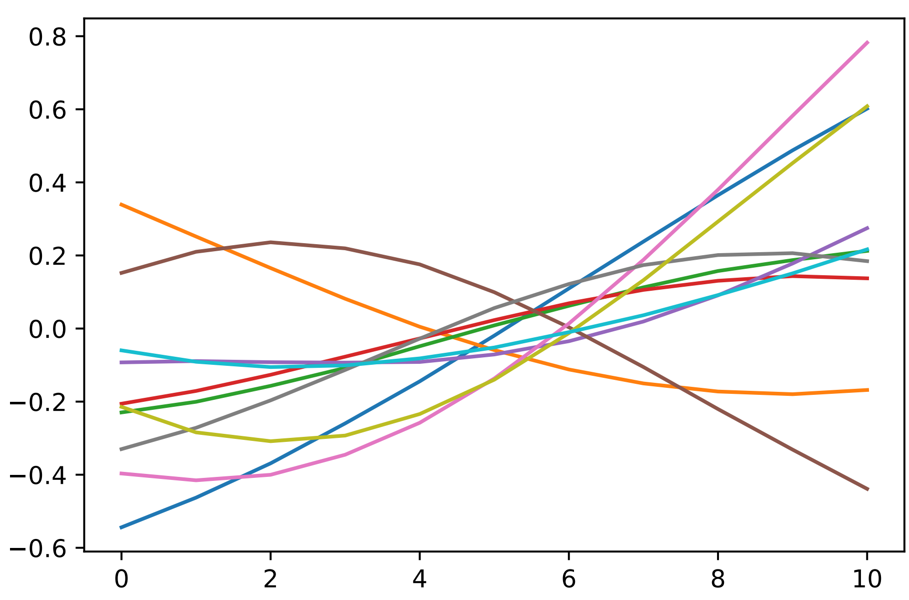
$$\boldsymbol{U} = \left[\frac{1}{n}\right]_{ij\ (n,n)}$$
$$\boldsymbol{I}_{\ (n,n)} \text{ is the indentity matrix}$$

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

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

up to numerical errors.

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



2 Kernel PCA

$\mathcal{X}$ a non empty set, $x_1, \ldots, x_n \in \mathcal{X}$, 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 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ the e-values and $u_1, \ldots, u_n$ the corresponding e-vectors. We have seen in class that the principal directions $f_1, \ldots, f_n$ 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 $i \neq k$.

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

where

$$e_k = [1 \ if \ k = j \ else \ 0]_{(j,1)}$$

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

$$V = span\{f_1, \ldots, f_n\}$$

is

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

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