Selected Solutions for Assignment 6

Machine Learning, Summer term 2014, Ulrike von Luxburg

Exercise 4.

In order to show that a function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is indeed a kernel function we have to show two things:

1. K is symmetric, i.e. $K(x, y) = K(y, x)$ for all $x, y \in \mathcal{X}$.
2. K is positive definite, i.e. for all $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$, $x_1, \ldots x_n \in \mathcal{X}$ we have

$$\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \geq 0.$$

(or in matrix notation: for kernel matrix $\mathbf{K} := [K(x_i, x_j)]_{ij}$ and any vector $\mathbf{c} \in \mathbb{R}^n$: $\mathbf{c}^T \mathbf{K} \mathbf{c} \geq 0$)

- $K = \alpha K_1$ for $\alpha > 0$
  1. $K(x, y) = \alpha K_1(x, y) = \alpha K_1(y, x) = K(y, x)$
  2. $$\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) = \sum_{i,j=1}^{n} c_i c_j \alpha K_1(x_i, x_j) = \underbrace{\alpha}_{>0} \underbrace{\sum_{i,j=1}^{n} c_i c_j K_1(x_i, x_j)}_{\geq 0} \geq 0.$$

  (or in matrix notation: $\mathbf{c}^T \mathbf{K} \mathbf{c} = \mathbf{c}^T (\alpha \mathbf{K}_1) \mathbf{c} = \underbrace{\alpha}_{>0} \underbrace{\mathbf{c}^T \mathbf{K}_1 \mathbf{c}}_{\geq 0} \geq 0$)

- $K = K_1 + K_2$
  1. $K(x, y) = K_1(x, y) + K_2(x, y) = K_1(y, x) + K_2(y, x) = K(y, x)$
  2. $$\begin{aligned}\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) &= \sum_{i,j=1}^{n} c_i c_j \left[K_1(x_i, x_j) + K_2(x_i, x_j)\right] \\ &= \underbrace{\sum_{i,j=1}^{n} c_i c_j K_1(x_i, x_j)}_{\geq 0} + \underbrace{\sum_{i,j=1}^{n} c_i c_j K_2(x_i, x_j)}_{\geq 0} \geq 0.\end{aligned}$$

  (or in matrix notation: $\mathbf{c}^T \mathbf{K} \mathbf{c} = \mathbf{c}^T (\mathbf{K}_1 + \mathbf{K}_2) \mathbf{c} = \underbrace{\mathbf{c}^T \mathbf{K}_1 \mathbf{c}}_{\geq 0} + \underbrace{\mathbf{c}^T \mathbf{K}_2 \mathbf{c}}_{\geq 0} \geq 0$)

- $K = K_1 - K_2$
  K is not necessarily a valid kernel function: If $\mathcal{X} = \mathbb{R}$, $K_1 = 0$ (i.e. $K(x, y) = 0$ for all $x, y \in \mathbb{R}$) and $K_2(x, y) = xy$ (think about that K_2 is indeed a valid kernel!), we have

  $$K(x, x) = -K_2(x, x) = -x^2 < 0 \qquad \forall x \neq 0,$$

  which shows that K is not necessarily positive definite.

  (another example: set $K_2 := 2K_1$ and get that $K = -K_1$ is no kernel for $K_1 \neq 0$)

Web page: `http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/teaching/2014-ss-vorlesung-ml/`
Login: “machine”, Password: “learning”

- $K(x,y) = K_1(x,y)K_2(x,y)$
  1. $K(x,y) = K_1(x,y)K_2(x,y) = K_1(y,x)K_2(y,x) = K(y,x)$
  2. Showing that K is indeed positive definite is a bit more subtle than in the previous cases.

     Let $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$ and $x_1, \ldots x_n \in \mathcal{X}$. Set

     $$\mathbf{K}_1 = [K_1(x_i, x_j)]_{i,j=1}^n \qquad \text{and} \qquad \mathbf{K}_2 = [K_2(x_i, x_j)]_{i,j=1}^n.$$

     Since K_1 and K_2 are kernel functions, the matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are symmetric positive semi-definite and hence - as a consequence of the spectral theorem - can be written as

     $$\mathbf{K}_1 = A^T A \qquad \text{and} \qquad \mathbf{K}_2 = B^T B$$

     for some real $n \times n$-matrices A and B. Hence, we have

     $$\begin{aligned}
     \sum_{i,j=1}^n c_i c_j K(x_i, x_j) &= \sum_{i,j=1}^n c_i c_j K_1(x_i, x_j) K_2(x_i, x_j) \\
     &= \sum_{i,j=1}^n c_i c_j \widetilde{K}_{ij}^1 \widetilde{K}_{ij}^2 \\
     &= \sum_{i,j=1}^n c_i c_j \left(A^T A\right)_{ij} \left(B^T B\right)_{ij} \\
     &= \sum_{i,j=1}^n c_i c_j \left(\sum_{k=1}^n A_{ik}^T A_{kj}\right) \left(\sum_{l=1}^n B_{il}^T B_{lj}\right) \\
     &= \sum_{i,j,k,l=1}^n c_i c_j A_{ki} A_{kj} B_{li} B_{lj} \\
     &= \sum_{k,l=1}^n \sum_{i=1}^n c_i A_{ki} B_{li} \sum_{j=1}^n c_j A_{kj} B_{lj} \\
     &= \sum_{k,l=1}^n \underbrace{\left(\sum_{i=1}^n c_i A_{ki} B_{li}\right)^2}_{\geq 0} \geq 0.
     \end{aligned}$$

- $K(x,y) = f(x)K_1(x,y)f(y)$ for any function $f : \mathcal{X} \to \mathbb{R}$
  1. $K(x,y) = f(x)K_1(x,y)f(y) = f(y)K_1(y,x)f(x) = K(y,x)$
  2. $$\sum_{i,j=1}^n c_i c_j K(x_i, x_j) = \sum_{i,j=1}^n c_i c_j f(x_i) K_1(x_i, x_j) f(x_j) = \sum_{i,j=1}^n \underbrace{c_i f(x_i)}_{\tilde{c}_i} \underbrace{c_j f(x_j)}_{\tilde{c}_j} K_1(x_i, x_j) \geq 0.$$

  Remark: Note that we do *not* have to show that $f(x)K_1(x,y)f(y) \geq 0$ for all $x, y \in \mathcal{X}$, which indeed would not hold. Neither we have to show that $\sum_{i,j=1}^n c_i c_j f(x) K_1(x_i, x_j) f(y) \geq 0$, which would also not hold. Just plugin the definition $K(x,y) := f(x)K_1(x,y)f(y)$ into the definition of positive definiteness to get that we have to show that for all choices of $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$, $x_1, \ldots x_n \in \mathcal{X}$ it holds that:

  $$\sum_{i,j=1}^n c_i c_j K(x_i, x_j) = \sum_{i,j=1}^n c_i c_j f(x_i) K_1(x_i, x_j) f(x_j) \overset{!}{\geq} 0.$$

  Let us for fun consider the following special case: Let $\mathcal{X} = \mathbb{R}$ and $K_1(x,y) = 1$ be a (quite trivial) kernel function. However, proving that it indeed is a kernel is not *that* trivial, since we have to show that $\sum_{i,j=1}^n c_i c_j K_1(x_i, x_j) = \sum_{i,j=1}^n c_i c_j \overset{!}{\geq} 0$ for all choices of $c_1, \ldots, c_n \in \mathbb{R}$. One way to do this is by writing this double sum as

$$\sum_{i,j=1}^{n} c_i c_j = \sum_{i=1}^{n} c_i \underbrace{\sum_{j=1}^{n} c_j}_{=:s} = s \cdot \sum_{i=1}^{n} c_i = s^2 \geq 0$$

Another way would be to see that the sum can be written in matrix notation as $\mathbf{c}^T \mathbf{1}\mathbf{1}^T \mathbf{c} = (\mathbf{1}^T \mathbf{c})^T (\mathbf{1}^T \mathbf{c}) = \|\mathbf{1}^T \mathbf{c}\|^2 \geq 0$. A third way is by noting that the $n \times n$-all-ones-matrix $\mathbf{1}\mathbf{1}^T =: J$ has 0 as its onliest eigenvalue, thus, is positive definite, thus, $c^T J c \geq 0$ for all c. Indeed, $J = \mathbf{K}_1$ is the corresponding kernel matrix for any n points.

It is now a nice exercise to show explicitly for the special case $K(x, y) = f(x)f(y)$ that it is indeed a kernel for any choice of f...