# Machine Learning I: Foundations
# Exercise Sheet 3

Prof. Marius Kloft TA: Billy Joe Franks

18.05.2020
Deadline: 19.05.2020

**1) (MANDATORY) 10 Points**
Suppose that $k_1, \ldots, k_n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are kernels. Let $c_1, \ldots, c_n \in \mathbb{R}^+$ and $p \in \mathbb{N}$. Prove that the following functions $k$ are also kernels.

For ease of notation we will further assume that

$$k_i(\mathbf{x}, \mathbf{x}') = \left\langle \Phi_i(\mathbf{x}), \Phi_i(bx') \right\rangle .$$

a) **Scaling**: $k(\mathbf{x}, \mathbf{x}') := c_1 k_1(\mathbf{x}, \mathbf{x}')$

$$k(\mathbf{x}, \mathbf{x}') = c_1 k_1(\mathbf{x}, \mathbf{x}') = c_1 \left\langle \Phi_1(\mathbf{x}), \Phi_1(\mathbf{x}') \right\rangle = \left\langle \sqrt{c_1}\Phi_1(\mathbf{x}), \sqrt{c_1}\Phi_1(\mathbf{x}') \right\rangle$$

Thus by defining $\Phi(\mathbf{x}) := \sqrt{c_1}\Phi_1(\mathbf{x})$, $k$ is a kernel.

b) **Sum**: $k(\mathbf{x}, \mathbf{x}') := k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}')$

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &= k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}') \\ &= \left\langle \Phi_1(\mathbf{x}), \Phi_1(\mathbf{x}') \right\rangle + \left\langle \Phi_2(\mathbf{x}), \Phi_2(\mathbf{x}') \right\rangle \\ &= \left\langle \begin{pmatrix} \Phi_1(\mathbf{x}) \\ \Phi_2(\mathbf{x}) \end{pmatrix}, \begin{pmatrix} \Phi_1(\mathbf{x}') \\ \Phi_2(\mathbf{x}') \end{pmatrix} \right\rangle \end{aligned}$$

where $\begin{pmatrix} \mathbf{x} \\ \mathbf{x}' \end{pmatrix}$ is defined as the vector by concatinating the vectors $\mathbf{x}$ and $\mathbf{x}'$.
Thus by defining $\Phi(\mathbf{x}) := \begin{pmatrix} \Phi_1(\mathbf{x}) \\ \Phi_2(\mathbf{x}) \end{pmatrix}$, $k$ is a kernel.

c) **Linear combination**: $k(\mathbf{x}, \mathbf{x}') := \sum_{i=1}^{n} c_i k_i(\mathbf{x}, \mathbf{x}')$

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{n} c_i k_i(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{n} k_i'(\mathbf{x}, \mathbf{x}')$$

The last equality holds due to part a) of this execise. Now by inductively applying part b) of this exercise over the elements of the sum, $k$ is a kernel.

d) **Product**: $k(\mathbf{x}, \mathbf{x}') := k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}')$

For ease of notation we assume in this exercise $\sum_i$ to sum over the appropriate elements.

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &= k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}') = \big\langle \Phi_1(\mathbf{x}), \Phi_1(\mathbf{x}') \big\rangle \big\langle \Phi_2(\mathbf{x}), \Phi_2(\mathbf{x}') \big\rangle \\ &= \left( \sum_i \Phi_1(\mathbf{x})_i \Phi_2(\mathbf{x}')_i \right) \left( \sum_j \Phi_1(\mathbf{x})_j \Phi_2(\mathbf{x}')_j \right) \\ &= \sum_i \sum_j \Phi_1(\mathbf{x})_i \Phi_1(\mathbf{x})_j \Phi_2(\mathbf{x}')_i \Phi_2(\mathbf{x}')_j \end{aligned}$$

where $\Phi_i(\mathbf{x})_j$ is defined as the $j$-th component of $\Phi_i(\mathbf{x})$. Let $I = \{1, \ldots, i\} \times \{1, \ldots, j\}$, and let $I(i, j)$ be the arbitrary index of $(i, j) \in I$. Thus defining $\forall i, j : \Phi(\mathbf{x})_{I(i,j)} := \Phi_1(\mathbf{x})_i \Phi_2(\mathbf{x})_j$, $k$ is a kernel.

e) **Power**: $k(\mathbf{x}, \mathbf{x}') := k_1(\mathbf{x}, \mathbf{x}')^p$

By inductively applying exercise d) over $p$, $k$ is a kernel.

**2)** In the lecture a few kernels were proposed, and here we will prove them to be kernels. Prove the following statements:

a) **Polynomial kernel**: $k(\mathbf{x}, \mathbf{x}') := (\mathbf{x}^T \mathbf{x}' + c)^d$ is a kernel.

From the lecture we know $k'(\mathbf{x}, \mathbf{x}') := \mathbf{x}^T \mathbf{x}'$ is a kernel. $k''(\mathbf{x}, \mathbf{x}') := c$ is a kernel by just choosing $\Phi''(\mathbf{x}) := \sqrt{c}$. Thus we need only apply 1b) (**sum**) and 1e) (**power**), then $k$ is a kernel.

b) **Limits**: If $k_i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $i \in \mathbb{N}$, are kernels and $k(\mathbf{x}, \mathbf{x}') := \lim_{n\to\infty} k_n(\mathbf{x}, \mathbf{x}')$ exists for all $\mathbf{x}, \mathbf{x}'$, then $k(\mathbf{x}, \mathbf{x}')$ is a kernel. Use the definition of positive semi-definiteness.

Let $K_i$ be the kernel matrix when applying $k_i$ to some data points. Then $K$ the kernel matrix when applying $k$ to these same datapoints can be defined as

$$K = \lim_{n\to\infty} K_n$$

The limit here is applied element-wise withing the matrix. Now it is readily apparent that

$$\mathbf{v}^T K \mathbf{v} = \lim_{n\to\infty} \mathbf{v}^T K_n \mathbf{v}$$

Now since $\forall \mathbf{v} \in \mathbb{R}^d : \mathbf{v}^T K_i \mathbf{v} \geq 0$, the same also holds for $K$, thus $k$ is a kernel.

c) **Exponents**: If $\tilde{k}$ is a kernel, then $k(\mathbf{x}, \mathbf{x}') := \exp(\tilde{k}(\mathbf{x}, \mathbf{x}'))$ is a kernel.

Consider the power series of $\exp(\tilde{k}(\mathbf{x}, \mathbf{x}'))$

$$\exp(\tilde{k}(\mathbf{x}, \mathbf{x}')) := \lim_{n\to\infty} \sum_{i=0}^{n} \frac{(\tilde{k}(\mathbf{x}, \mathbf{x}'))^i}{i!}.$$

Now by applying 1e) (**power**), 1c) (**linear combination**) and lastly 2b) (**limits**), $k$ is a kernel.

d) **Functions**: If $\tilde{k}$ is a kernel and $f : \mathbb{R}^d \to \mathbb{R}$ then $k(\mathbf{x}, \mathbf{x}') := f(\mathbf{x})\tilde{k}(\mathbf{x}, \mathbf{x}')f(\mathbf{x}')$ is a kernel.

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &= f(\mathbf{x})\tilde{k}(\mathbf{x}, \mathbf{x}')f(\mathbf{x}') \\ &= f(\mathbf{x}) \left\langle \tilde{\Phi}(\mathbf{x}), \tilde{\Phi}(\mathbf{x}') \right\rangle f(\mathbf{x}') \\ &= \left\langle f(\mathbf{x})\tilde{\Phi}(\mathbf{x}), f(\mathbf{x}')\tilde{\Phi}(\mathbf{x}') \right\rangle \end{aligned}$$

Thus by defining $\Phi(\mathbf{x}) := f(\mathbf{x})\tilde{\Phi}(\mathbf{x})$, $k$ is a kernel

e) **Gaussian RBF kernel**: $k(\mathbf{x}, \mathbf{x}') := \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2}\right)$ is a kernel.

Note that

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &= \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2}\right) \\ &= \exp\left(-\frac{\|\mathbf{x}\|^2 - 2\mathbf{x}^T\mathbf{x}' + \|\mathbf{x}'\|^2}{2}\right) \\ &= \exp\left(\frac{-\|\mathbf{x}\|^2}{2}\right) \exp\left(\mathbf{x}^T\mathbf{x}'\right) \exp\left(\frac{-\|\mathbf{x}'\|^2}{2}\right) \end{aligned}$$

Now $k'(\mathbf{x}, \mathbf{x}') := \mathbf{x}^T\mathbf{x}'$ is a kernel. By applying 2c) (**exponents**) and 2d) (**functions**), $k$ is a kernel.

**Hint:** Use the results from Exercise 1 above.

**3)** Prove the following lemma:

**Lemma 1** *Let $V$ be a vector space and $I$ a set. Let $f_i : V \to \mathbb{R}$ be a collection of functions indexed by $i \in I$. If $f_i$ is convex for all $i$, then the function*

$$f(x) = \max_{i \in I} f_i(x)$$

*is also convex.*

Just using the definition of convexity we see that

$$\begin{aligned} f(xt + (1-t)y) &= \max_{i \in I} f_i(xt + (1-t)y) \\ &\leq \max_{i \in I} tf_i(x) + (1-t)f_i(y) \\ &\leq \max_{i \in I} tf_i(x) + \max_{j \in I}(1-t)f_j(y) \\ &= tf(x) + (1-t)f(y). \end{aligned}$$

So $f$ is convex by direct application of the definition of convexity.

**4)** Solve programming task 3.