Introduction to Mercer's theorem and the kernel trick

Abdelwahid Benslimane

wahid.benslimane@gmail.com

Mercer's theorem

Let $\mathcal{X}$ be a compact in $\mathbb{R}^d$ (compact = closed and bounded) and $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a symmetric function also called kernel function. Let v be any Borel measure and $L_2^v(\mathcal{X})$ the space of integrable square functions defined on $\mathcal{X}$. It is also assumed that K is a definite positive kernel, which means that $\forall f \in L_2^v(\mathcal{X})$, the following condition is satisfied:

$\int_{\mathcal{X}} \int_{\mathcal{X}} K(x,y) f(x) f(y) v(dx) v(dy) \geq 0.$

Then there exists a Hilbert space $\mathcal{H}$ and $\phi : \mathcal{X} \to \mathcal{H}$ such that $\forall x, y \in \mathcal{X}$:

$K(x,y) = \displaystyle\sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(y) = \langle \phi(x), \phi(y) \rangle$ (scalar product) with the coefficients $\lambda_k > 0.$

The Hilbert space $\mathcal{H}$ is a data representation space known as feature space and intuitively, the ϕ_k represent its basis.

For the recall, a Hilbert space is a complete normed vector space equipped with a scalar product.

Kernel trick

A valid kernel guarantees the existence of $\mathcal{H}$ and can then be expressed as a scalar product in $\mathcal{H}$. Kernels guarantee the correct application of the kernel trick, which is a well-known approach that implicitly transform a linear method into a non-linear one by replacing the scalar products with a kernel function. Any algorithm on finite-dimensional vectors based on scalar products can be used by replacing the scalar product by any positive definite kernel.The kernel trick is to ignore $\mathcal{H}$ and ϕ if we know they exist!

A kernel is (most of the time) close to a similarity measure. More precisely, if x and y are two vectors, then a kernel defined for these vectors must return a high value if $x \approx y$ and a low value if x and y are very different.

A function $s : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a similarity measure if the following conditions are satisfied:

- $x, y \in \mathcal{X}$ $s(x,y) \geq 0$ (positivity)
- $x, y \in \mathcal{X}$ $s(x,y) = s(y,x)$ (simmetry)
- $\forall y \in \mathcal{X}, y \neq x$ $s(x,y) > s(x,x)$ (uniformity)

- $s(x, y) = s(x, x) \Leftrightarrow x = y$ (identity).

There are several approaches to obtaining kernel functions:

- direct construction (using a ϕ projection)
- transformation of existing kernels,
- combination of existing kernels.

Direct construction:

Let $\mathcal{X}$ be a compact in $\mathbb{R}$. Let $\phi : \mathcal{X} \rightarrow \mathbb{R}$, then $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ given by : $K(x, y) = \phi(x) \cdot \phi(y)$ is a non-negative kernel (called "conformal kernel").

Please note that conformal kernels cannot be interpreted as similarities.

<u>Examples</u>:

$\phi : \mathbb{R} \rightarrow \mathbb{R}, \phi(x) = x \Rightarrow K(x, y) = x \cdot y$ (known as the "linear kernel"),

$\phi : \mathbb{R} \rightarrow \mathbb{R}, \phi(x) = e^x \Rightarrow K(x, y) = e^{x+y}$

Tranformation of kernels:

If $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is definite positive, then $exp(K)$ is also definite positive.

If $K : \mathcal{X} \times \mathcal{X} \rightarrow [-1, 1]$ is definite positive, then $cos(K)$ is also definite positive.

Combination of existing kernels:

If $K_1, K_2 : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ are definie positive and $\alpha_1, \alpha_2 > 0$, then the following kernels are also definite positive:

- linear combination: $K(x, y) = \alpha_1 K_1(x, y) + \alpha_2 K_2(x, y)$
- simple product: $K(x, y) = \alpha_1 K_1(x, y) \cdot \alpha_2 K_2(x, y)$

Of course, K is defined on $\mathcal{X} \times \mathcal{X}$ with values in $\mathbb{R}$.

If $K_1 : \mathcal{X}_1 \times \mathcal{X}_1 \rightarrow \mathbb{R}$ and $K_2 : \mathcal{X}_2 \times \mathcal{X}_2 \rightarrow \mathbb{R}$ are definite positive, then:

- the direct sum: $K_1 \oplus K_2 = K_1 + K_2$

- the tensor product: $K_1 \otimes K_2 = K_1 \cdot K_2$

are also definite positive.