Formalize Kernels

Here are the most common kernels:

Linear	$\mathsf{K}(\mathbf{x},\mathbf{z})=\mathbf{x}^\top\mathbf{z}$
RBF	$\mathsf{K}(\mathbf{x},\mathbf{z})=e^{-\frac{\|\|\mathbf{x}-\mathbf{z}\|\|^2}{\sigma^2}}$
Polynomial	$\mathsf{K}(\mathbf{x},\mathbf{z})=\left(1+\mathbf{x}^\top\mathbf{z}\right)^p$

A kernel is called "well-defined" if it corresponds to an inner-product function of mappings of feature vectors $\phi(\mathbf{x})$ and $\phi(\mathbf{z})$. A kernel is well-defined if and only if the corresponding kernel matrix K is "positive semidefinite". We omit the proof in this course. Positive semidefinite is equivalent to any of the following:

1. All eigenvalues of K are non-negative.
2. there exists a real matrix L s.t. $\mathsf{K}=L^\top L$.
3. $\mathbf{c}^\top\mathsf{K}\mathbf{c}\geq 0$ for all real vectors $\mathbf{c}$.

Exercise: Using any of these definitions, prove that the linear kernel matrix $\mathsf{K}\left(\mathbf{x}_i,\mathbf{x}_j\right)=\mathbf{x}_i^\top\mathbf{x}_j$ is positive semi-definite.

Answer

Arrange $\mathbf{x}_1,\ldots,\mathbf{x}_n$ vectors in a matrix $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]\in\mathbb{R}^{d\times n}$. By inspection, $\mathsf{K}=\mathbf{X}^\top\mathbf{X}$. Thus a real matrix $L=\mathbf{X}$ exists such that $\mathsf{K}=L^\top L$.

Proving a kernel as well-defined is not always simple. A simple work-around is to take one of the known kernel functions and combine them to a new kernel function with any one of the following rules:

1. $\mathsf{K}(\mathbf{x},\mathbf{z})=\mathbf{x}^\top\mathbf{z}$
2. $\mathsf{K}(\mathbf{x},\mathbf{z})=c\mathsf{K}_1(\mathbf{x},\mathbf{z})$
3. $\mathsf{K}(\mathbf{x},\mathbf{z})=\mathsf{K}_1(\mathbf{x},\mathbf{z})+\mathsf{K}_2(\mathbf{x},\mathbf{z})$
4. $\mathsf{K}(\mathbf{x},\mathbf{z})=g(\mathsf{K}(\mathbf{x},\mathbf{z}))$
5. $\mathsf{K}(\mathbf{x},\mathbf{z})=\mathsf{K}_1(\mathbf{x},\mathbf{z})\mathsf{K}_2(\mathbf{x},\mathbf{z})$

6. $\mathsf{K}(\mathbf{x}, \mathbf{z}) = f(\mathbf{x})\mathsf{K}_1(\mathbf{x}, \mathbf{z})f(\mathbf{z})$
7. $\mathsf{K}(\mathbf{x}, \mathbf{z}) = e^{\mathsf{K}_1(\mathbf{x}, \mathbf{z})}$
8. $\mathsf{K}(\mathbf{x}, \mathbf{z}) = \mathbf{x}^\top A\mathbf{z}$

where K_1 and K_2 are well-defined kernels, $c \geq 0$, g is a polynomial function with positive coefficients, f is any function and A is positive semi-definite.

Exercise: Using the rules above, prove that the RBF kernel matrix $\mathsf{K}\left(\mathbf{x}_i, \mathbf{x}_j\right) = e^{-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{\sigma^2}}$ is positive semi-definite.

Answer

$\mathsf{K}_1\left(\mathbf{x}_i, \mathbf{x}_j\right) = \frac{2\mathbf{x}_i^\top \mathbf{x}_j}{\sigma^2}$ is a valid kernel by rule 2.

$\mathsf{K}_2\left(\mathbf{x}_i, \mathbf{x}_j\right) = e^{\mathbf{K}_1(\mathbf{x}_i, \mathbf{x}_j)} = e^{\frac{2\mathbf{x}_i^\top \mathbf{x}_j}{\sigma^2}}$ is a valid kernel by rule 7.

$\mathsf{K}_3\left(\mathbf{x}_i, \mathbf{x}_j\right) = f\left(\mathbf{x}_i\right)\mathsf{K}_2\left(\mathbf{x}_i, \mathbf{x}_j\right)f\left(\mathbf{x}_j\right) = \left(e^{-\frac{\|\mathbf{x}_i\|^2}{\sigma^2}}\right)\left(e^{\frac{2\mathbf{x}_i^\top \mathbf{x}_j}{\sigma^2}}\right)\left(e^{-\frac{\|\mathbf{x}_j\|^2}{\sigma^2}}\right) = e^{\frac{-\|\mathbf{x}_i\|^2 + 2\mathbf{x}_i^\top \mathbf{x}_j - \|\mathbf{x}_j\|^2}{\sigma^2}}$ is a valid kernel by rule 6.

$\mathsf{K}_3\left(\mathbf{x}_i, \mathbf{x}_j\right) = e^{-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{\sigma^2}}$ is the desired RBF kernel.