Kernel Functions and the Kernel Trick

1 Introduction to Kernel Functions

Kernel functions are fundamental tools in machine learning and spatial statistics that allow computations in high-dimensional feature spaces without explicitly mapping data points to those spaces. The primary motivation for using kernel functions is to enable linear methods, such as Support Vector Machines (SVMs) or Kriging, to operate in non-linear spaces while maintaining computational efficiency.

2 Inner Product

Given two vectors in two-dimensional space,

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad x' = \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix}$$

their inner product is calculated as:

$$x^\top x' = x_1 x'_1 + x_2 x'_2.$$

More generally, for vectors in n-dimensional space,

$$x^\top x' = x_1 x'_1 + x_2 x'_2 + \cdots + x_n x'_n.$$

This operation gives a single number (a scalar) that tells us how much the two vectors point in the same direction. A larger inner product means the vectors are more similar, while an inner product of zero means they are perpendicular (completely unrelated).

Suppose we have two vectors:

$$x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad x' = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

The inner product is computed as:

$$x^\top x' = (2 \times 4) + (3 \times 1) = 8 + 3 = 11.$$

3 Definition of a Kernel Function

A kernel function $k(x, x')$ computes the similarity between two points x and x' in the input space. Formally, a kernel function satisfies:

$$k(x, x') = \phi(x)^\top \phi(x')$$

where $\phi(x)$ is an implicit mapping to a higher-dimensional feature space.

3.1 Common Kernel Functions

Some widely used kernel functions include:

- **Linear Kernel**: $k(x, x') = x^\top x'$
- **Polynomial Kernel**: $k(x, x') = (x^\top x' + c)^d$
- **Radial Basis Function (RBF) Kernel**: $k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$
- **Gaussian Kernel**: Similar to the RBF kernel, emphasizing smoothness in function approximation.

4 The Kernel Trick

The kernel trick allows us to compute inner products in high-dimensional feature spaces without explicitly performing the transformation $\phi(x)$. Instead of mapping points to a high-dimensional space and computing dot products explicitly, we use a kernel function to obtain the same result directly in the original space.

4.1 Example: Polynomial Kernel Calculation

Consider two data points in one-dimensional space: $x = 2$ and $x' = 3$. If we apply a polynomial kernel of degree $d = 2$:

$$k(x, x') = (xx' + c)^d$$

For example, let $c = 1$:

$$k(2, 3) = (2 \cdot 3 + 1)^2 = (6 + 1)^2 = 49.$$

This result is equivalent to computing the inner product in a transformed quadratic feature space without explicitly mapping x and x' to their higher-order terms.

5 Applications of the Kernel Trick

- **Support Vector Machines (SVMs)**: Enables classification in high-dimensional feature spaces.
- **Gaussian Process Regression**: Uses kernel functions to define covariance structures for probabilistic modeling.
- **Kriging**: Kernel functions describe spatial dependence in geostatistical modeling.
- **PCA and Dimensionality Reduction**: Kernel PCA extends traditional PCA to non-linear feature extraction.

6 Recap

Kernel methods provide transformations implicitly, enabling non-linear models while maintaining computational efficiency. The kernel trick allows high-dimensional computations without explicitly performing the transformations.

7 Kernel Functions and Semivariograms

Kernel functions and semivariograms serve similar roles in quantifying dependencies between data points.

7.1 Quantifying Similarity or Dependence

- a kernel function $k(x, x')$ measures the similarity between two points x and x'. It defines how much influence one observation has on another in function space.
- A semivariogram $\gamma(h)$ quantifies the spatial dependence of a variable over a distance h, describing how variance changes with increasing separation.

7.2 Functional Form

Kernel Function: Typically, kernels are positive definite and often take forms such as:

$$k(x, x') = \exp\left(-\frac{|x - x'|^2}{2\ell^2}\right), \tag{1}$$

which ensures that points closer together have higher similarity (like the squared exponential or Gaussian kernel).

Semivariogram Model: The semivariogram is often modeled as:

$$\gamma(h) = C_0 + C\left(1 - e^{-\frac{h^2}{2\ell^2}}\right), \tag{2}$$

where C_0 is the nugget effect, C is the sill, and ℓ is the range parameter. This structure is similar to the kernel function in GPs.

7.3 Interpretation in Spatial and Functional Spaces

- Kernel Functions define a *covariance structure* in function space. When using a kernel, we assume that function values at nearby points are correlated.
- Semivariogram defines a *variance structure* in spatial fields. When applying kriging, we assume that observations closer together have more similar values.

7.4 Inversion Between the Two

The covariance function in kriging is the direct analog to the kernel function. The semivariogram is related to the covariance function through:

$$\gamma(h) = C(0) - C(h), \tag{3}$$

meaning the semivariogram represents how dissimilarity grows with distance, while covariance measures similarity.

7.5 Key Takeaways

Feature	**Kernel Function (GPs)**	**Semivariogram (Kriging)**
Measures	Similarity	Spatial dependence
Range Parameter	Controls smoothness	Defines spatial correlation
Role in Model	Defines function prior	Defines spatial interpolation
Common Forms	Gaussian (RBF), Matern	Exponential, Spherical

Table 1: Comparison of Kernel Functions and Semivariograms

A **kernel function in Gaussian processes serves the same purpose as a covariance function in kriging**, while the **semivariogram models the inverse of covariance**—measuring how variability increases with distance instead of similarity.