

## Lecture 4c: Kernel Functions and Kernelized Regression

*Lecturer: Jeffrey Varner***Disclaimer:** *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

## 1 Introduction

In this lecture, we will discuss kernel functions and kernel regression. Kernel functions are mathematical tools that enable algorithms to operate in high-dimensional spaces without explicitly computing the coordinates in those spaces. Kernel functions are used in a variety of machine learning algorithms, including Support Vector Machines (SVMs), kernelized regression, and kernelized clustering. Today, we will focus on kernel regression, which is a non-parametric regression technique that uses kernel functions to estimate the relationship between the input and output variables. Let's start by discussing the basic concepts of kernel functions and then consider kernel regression.

## 2 Kernel Functions

Kernel functions in machine learning are mathematical tools that enable algorithms to operate in high-dimensional spaces without explicitly computing the coordinates in those spaces. Kernel functions are used in a variety of machine learning algorithms, including Support Vector Machines (SVMs), kernelized regression, and kernelized clustering. Kernel functions have a few different interpretations. For example, kernel functions can be thought of as similarity measures, i.e., they quantify the similarity between pairs of data points in a high-dimensional space. They are also implicit mappings of data into a high-dimensional space, where the data becomes linearly separable (which is useful for classification algorithms). Thus, kernel functions are powerful tools that we are going to use for many applications in machine learning.

A kernel function  $k : \mathbb{R}^* \times \mathbb{R}^* \rightarrow \mathbb{R}$  is a function that takes a pair of vectors  $\mathbf{z}_i \in \mathbb{R}^*$  and  $\mathbf{z}_j \in \mathbb{R}^*$  as arguments, e.g., a pair of feature vectors, a feature vector and a parameter vector, or any two vectors of compatible size and computes a scalar value that represents the similarity (in some sense) between the two vector arguments. For example, the linear kernel function computes the dot product between two vectors:

$$k(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{z}_i^\top \mathbf{z}_j \quad (1)$$

On the other hand, a polynomial kernel is defined as:

$$k_d(\mathbf{z}_i, \mathbf{z}_j) = (1 + \mathbf{z}_i^\top \mathbf{z}_j)^d \quad (2)$$

where  $d$  is the degree of the polynomial. The radial basis function (RBF) kernel is defined as:

$$k_\gamma(\mathbf{z}_i, \mathbf{z}_j) = \exp(-\gamma \|\mathbf{z}_i - \mathbf{z}_j\|_2^2) \quad (3)$$

where  $\gamma$  is a scaling factor, and  $\|\cdot\|_2^2$  is the squared Euclidean norm; if we define  $\gamma$  as  $1/2\sigma^2$ , the RBF kernel looks like a Gaussian function, without the normalization constant. Of course, not all functions are kernel functions (Defn. 1).

**Definition 1.** (Valid Kernel Function) A function  $k : \mathbb{R}^* \times \mathbb{R}^* \rightarrow \mathbb{R}$  is a valid kernel function if and only if the Gram matrix  $\mathbf{K}$  is positive semidefinite for all possible choices of the data points  $\mathbf{z}_i$ , where  $K_{ij} = k(\mathbf{z}_i, \mathbf{z}_j)$ . This is equivalent to saying that all eigenvalues of the Gram matrix  $\mathbf{K}$  are non-negative. Further, for all real value vectors  $\mathbf{x}$ , the Gram matrix  $\mathbf{K}$  must satisfy  $\mathbf{x}^\top \mathbf{K} \mathbf{x} \geq 0$ .

Kernel functions can also be combined to create more complex kernel functions using the concept of kernel composition. For example, the sum of two valid kernel functions is also a valid kernel function. The product of two valid kernel functions is also a valid kernel function. Multiplying a kernel function by a scalar is also a valid kernel function, etc. See the CS 4780 Lecture Notes (Fall 2018) for more details on kernel composition.

### 3 Kernel Regression

Kernel regression is a non-parametric regression technique that uses kernel functions to estimate the relationship between the input and output variables. Kernel regression is a powerful tool for modeling complex relationships in data and is widely used in machine learning applications. Suppose we have a dataset  $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid i = 1, 2, \dots, n\}$ , where the features  $\mathbf{x}_i \in \mathbb{R}^m$  are  $m$ -dimensional vectors and the target variables are continuous values  $y_i \in \mathbb{R}$ , e.g., the price of a house, the temperature, etc. The basic idea behind kernel regression is to estimate the output variable  $y$  as a weighted average of the output variables of the training data points, where the weights are determined by the kernel functions. The kernel regression function is defined as:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \mathbf{x}) \quad (4)$$

where  $\alpha_i$  are the weights, and  $k(\mathbf{x}_i, \mathbf{x})$  is the kernel function that measures the similarity between the input features  $\mathbf{x}_i$  and  $\mathbf{x}$ . The weights  $\alpha_i$  are determined by minimizing the mean squared error between the predicted output  $f(\mathbf{x})$  and the true output  $y$ :

$$\min_{\alpha} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2 \quad (5)$$

The solution to this optimization problem can be written in matrix form as:

$$\alpha = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y} \quad (6)$$

where  $\mathbf{K}$  is the Gram matrix of the kernel function,  $\mathbf{y}$  is the vector of target variables, and  $\lambda$  is a regularization parameter. The regularization parameter  $\lambda$  controls the smoothness of the regression function and prevents overfitting.

### 4 Summary and Conclusions

In this lecture, we discussed kernel functions and kernel regression. Kernel functions are mathematical tools that enable algorithms to operate in high-dimensional spaces without explicitly computing the coordinates in those spaces. Kernel functions are used in a variety of machine learning algorithms, including Support Vector Machines (SVMs), kernelized regression, and kernelized clustering. They have several interpretations, including similarity measures and implicit mappings of data into high-dimensional spaces. We introduced the linear kernel, polynomial kernel, and radial basis function (RBF) kernel as examples of kernel functions, and discussed the properties of valid kernel functions. Finally, we discussed kernel regression, a non-parametric

regression technique that uses kernel functions to estimate the relationship between the input and output variables. Kernel regression is a powerful tool for modeling complex relationships in data and is widely used in machine learning applications.

## **References**