Journal of Insights in Basic and Applied Sciences
مجلة رؤى في العلوم الأساسية والتطبيقية
ISSN: 3106-1737
Volume 2, Issue 4, 2026, Pages: 455-461
Journal homepage: https://jibas.com.ly/index.php/jibas/en

Positive Definite Kernels in Machine Learning: Theoretical Framework (Mercer's Theorem & RKHS) and Empirical Evaluation on Kernel Performance

Aya zidan Mohammed[1], Souad AhmedAbumaryam [2], Tahani Zidan Mohammed[3]
[1,2,3] *Department of Mathematics, Faculty of Sciences, Sirt University,Sirt, Libya.*

النوى الموجبة المحددة في التعلم الآلي: الإطار النظري (نظرية ميرسر و*RKHS*) والتقييم التجريبي لأداء النواة

آية زيدان محمد الزرقي[1*]، سعاد أحمد أبومريم[2]، تهاني زيدان محمد [3]
[1،2،3] قسم الرياضيات، كلية العلوم، جامعة سرت، سرت، ليبيا

*Corresponding author: a.alzwrqi@su.edu.ly

Received: September 18, 2026	Accepted: October 25, 2026	Published: November 20, 2026

Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract:
Positive definite kernels (PDKs) play a critical role in machine learning; this paper provides a combined theoretical and numerical investigation into their behavior. We examine Linear, Polynomial, and RBF kernels, contrasting their characteristics using a benchmark dataset, a comparative analysis was performed, emphasizing the impact of each kernel on enhancing model performance.
The implementation was carried out in Python, emphasizing practical applicability and study demonstrates the positive definite kernels efficacy in handling data complexity structures. The work provides an integrated framework combining theoretical foundations with empirical validation, supporting improved model performance through informed kernel selection tailored to specific applications.

Keywords: Kernel Methods, positive definite kernel, Hilbert Space, Reproducing Kernel Hilbert Space (RKHS), Radial Basis Function (RBF), Machine learning.

الملخص
تؤدي النوى الموجبة المحددة (Positive Definite Kernels - PDKs) دوراً جوهرياً في مجال تعلم الآلة؛ حيث تقدم هذه الورقة بحثاً يجمع بين الجوانب النظرية والعددية لاستكشاف سلوك هذه النوى. ونحن ندرس هنا النوى الخطية (Linear)، ومتعددة الحدود (Polynomial)، ونوى دالة الأساس الشعاعي (RBF)، مع إجراء مقارنة بين خصائصها باستخدام مجموعة بيانات مرجعية، مع التركيز على أثر كل نوع منها في تعزيز أداء النموذج.
وقد تم تنفيذ العمل باستخدام لغة البرمجة "بايثون" (Python) مع التركيز على الجوانب التطبيقية؛ وتُظهر الدراسة مدى فعالية النوى الموجبة المحددة في التعامل مع البيانات ذات البنى المعقدة. كما يقدم هذا العمل إطاراً متكاملاً يجمع بين الأسس النظرية والتحقق التجريبي، مما يدعم تحسين أداء النماذج من خلال اختيار واعٍ للنواة المناسبة التي تتلاءم مع تطبيقات محددة.

الكلمات المفتاحية: النواة المحددة الموجبة، فضاء هيلبرت، إعادة إنتاج فضاءات هيلبرت، تعلم الآلة.

1. Introduction

Significant advancements in machine learning have been made recently, particularly in the area of kernel method. Fundamentally, a kernel function measures how similar two data points are by implicitly calculating their inner product in a higher-dimensional space., allowing this calculation to be performed without explicitly determining the feature vectors themselves.

The choice of kernel function significantly influences model representational strength and performance. Selecting the appropriate kernel is a crucial design decision that can greatly affect the effectiveness of a kernel-based model. This requirement for careful selection simply highlights the important of choosing the inductive bias most suitable to specific situation. When an appropriate kernel is utilized, kernel methods can deliver exceptional performance across various applications. As a result, they are recognized as some of the most effective and widely used techniques in machine learning.

(PDKs) serve as a cornerstone for kernel-based algorithms in machine learning.. A PDK is a function which satisfies specific properties of mathematics, making it highly useful tool across a range of applications. These kernels quantify similarity between data points and serve as the backbone of algorithms such as Support Vector Machines (SVMs), kernel ridge regression, Gaussian processes, and Kernel Principal Component Analysis (KPCA). By implicitly map input data across a high dimensional feature space, they enable effective learning and generalization in diverse machine learning tasks underscoring their critical role in the success of kernel methods. From a functional analysis perspective, kernels offer another layer of insight: any kernel K defined on domain X uniquely determines a Hilbert space $\mathcal{H}_k$ of real-valued functions on X. The term "reproducing" highlights the core reproducing property: evaluating any $f \in \mathcal{H}_k$ at $x \in X$ corresponds directly to the inner product $\langle f, K(\cdot, x)\rangle_{\mathcal{H}_k}$. This canonical framework constitutes what is widely recognized as a Reproducing Kernel Hilbert Space (RKHS).

While RKHS theory is well-established, translating these properties into practical algorithms remains challenging due to difficulties in hyperparameter tuning and model selection. This research aims to bridge this gap by offering a comprehensive study of PDKs, focusing on both algorithm generalization and the explicit functional relationship between PDKs and Hilbert spaces.

The kernel trick and Mercer theorem were first introduced by Mercer [1], who established foundational expansion theorems for continuous kernels. This foundational work gave rise to PDKs and linked them to integral equations. Schoenberg [2] then offered crucial advances by identifying which kernels are capable of computing distances in feature spaces, thereby deepening the theoretical understanding of PDKs concepts that later became essential to kernel methods in both function analysis and machine learning. Subsequently, Kolmogorov [3] analyzed kernel representations within linear spaces restricted to countable input domains, Aronszajn [4] later generalized these representation methods to cover arbitrary input domains.

Dunford and Schwartz [5]demonstrated that Mercer's theorem which pertains to PDKs remains valid for general compact spaces. Later, Aizerman, Braverman, and Rozonoer [6] applied Mercer theorem into the realm of machine learning by interpreting kernels as inner products within feature spaces, an insight that became foundational for the development of kernel methods.

In [7], Sabri and his colleagues characterized conditionally PDKs, demonstrating that such kernels are well-suited for use in SVM algorithms. Within the SVM framework, they introduced a novel kernel called the Log kernel, which appears to perform especially well in their image recognition experiments.

Earlier developments by Harald [8] established RKHS formulations over Hermitian positive definite kernels, later generalizing this construction to conditionally positive definite (CPD) kernels. Addressing the numerical stability within CPD-generated spaces, Maryam [9] introduced two distinct methodologies for deriving stable bases. Their framework operates on the native Hilbert space's reproducing kernel, exploiting varied factorizations of the kernel matrix to extract bases possessing specialized algebraic properties.

2. Material and methods

PDKs are fundamental mathematical constructs in machine learning and statistics, providing the foundational framework for a class of algorithms known as kernel methods. These functions are paramount for quantifying the similarity between pairs of data points without the computational burden of explicitly defining their feature mapping. By implicitly computing inner products in a high-dimensional and potentially infinite-dimensional RKHS, they empower algorithms to discover non-linear patterns and relationships. This capability is instrumental in the success of sophisticated learning models including, but not limited to SVM for classification, Kernel Ridge Regression for regression, KPCA for nonlinear dimensionality reduction, and Gaussian Processes for probabilistic modeling. [10]

2.1 Mathematical Definition

Let the function $k: X \times X \to \mathbb{R}$ is said to be a PDK if it is a symmetric function $k(x, z) = k(z, x)$. For any finite $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n} \in \boldsymbol{X}$ and any real coefficients $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha_n} \mathbb{R}$, with $\boldsymbol{n} \in \mathbb{N}$, the following:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{k(x_i, x_j)\alpha_i\alpha_j \geq 0}$$

The definition enables extended to the complex space such that $k: X \times X \to \mathbb{C}$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$ satisfying:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} k(x_i, x_j)\alpha_i\overline{\alpha_j} \geq 0$$

Where the $\bar{\alpha}$ denotes the complex conjugate ofα_j where $k(x, z) = \overline{k(z, x)}$. [11]

2.2 Common Kernel Functions

Choice of the kernel function is critical, as it implicitly determines feature space in which learning is performed. Below are some of the more widely used PDKs kernels witin machine learning:

- **Linear kernel:** For $x, z \in \mathbb{R}^n$ the linear kernel defined as

$$"k(x, z) = \langle x, z\rangle = x^T z"$$

This kernel induces a feature space identical to the original input space. It is ideal linearly separable problems but lacks the capacity to capture complex, nono-linear relationships.

- **Polynomial kernel:** The derived polynomial kernel is defined as

$$k_d(x, z) = P\big(K(x, z)\big) = (x^T z + c)^d; c \geq 0\,, d \in N$$

$P(.)$ is any polynomial with positive coefficients, c is constant and d denote the degree of the polynomiall. This form defines the non-homogeneous kernel. When $c = 0$, the polynomial kernel is referred as homogenous. The polynomial kernel is capable of distinguishing curved or nonlinear input space by mapping data into a higher dimensional space. As a result, modelling non linearities, with the degree d controlling the complexity of the resulting decision boundary.

- **Gaussian kernel or Radial basis function kernels (RBF):** For $\gamma > 0$

$$k(x, z) = exp(-\gamma\|x - z\|^2)$$

Where $\gamma = \frac{1}{2\sigma^2}$ and σ representing the coefficient of variation. This kernel is widely used due to its ability to capture non-linear relationships. The Gaussian kernel is the most commonly employed type of RBF kernel. RBF kernel projects data into an infinite-dimensional feature space. Its locality-sensitive nature allows it to create highly complex, non-linear decision boundaries, making it exponentially powerful for a wide range of tasks.

3. Reproducing kernel Hilbert Space (RKHS) and Mercer theorem

Kernel methods are fundamentally understood through the functional analysis framework of (RKHS). This approach provides a rigorous mathematical foundation for analyzing feature spaces and their properties in machine learning.

The RKHS framework establishes a critical connection between the selection of a hypothesis space the set of potential functions considered by a learning algorithm and the theory of positive definite kernels. This connection enables optimal performance in both linear and nonlinear learning scenarios through the use of an appropriate positive definite kernel. [12]

Definition 3.1 (Reproducing Kernel Hilbert Space):

Let X be a non-empty domain, and consider a Hilbert space $\mathcal{H}$ composed of real-valued functions $f: X \to \mathbb{R}$. A mapping $k: X \times X \to \mathbb{R}$ serves as a reproducing kernel for $\mathcal{H}$ provided that:

1. $k_x = k(\cdot, x) \in \mathcal{H}$ holds for all $x \in X$.
2. Every $f \in \mathcal{H}$ and $x \in X$ satisfies $f(x) = \langle f, k(\cdot, x)\rangle_{\mathcal{H}}$ (referred to as the reproducing property).

When such a kernel exists, $\mathcal{H}$ is designated as a Reproducing Kernel Hilbert Space (RKHS), frequently denoted by $\mathcal{H}_K(X)$ [13].

3.2 The Theoretical Role of Mercer's Theorem

Mercer's theorem plays a pivotal role in modern learning theory by establishing the mathematical basis for RKHS. These functional spaces are widely utilized throughout machine learning and mathematical analysis, particularly within kernel-centric techniques. On compact domains, this theorem characterizes an RKHS via the range of the square root of an integral operator, establishing a robust foundation for analyzing and implementing kernel methods[14].

Theorem 3.1 (Mercer's Theorem): Consider a compact space X and let $K: X \times X \to \mathbb{R}$ represent a continuous, symmetric, and positive-definite kernel function. The fundamental Mercer conditions are satisfied if and only if K admits the expansion:

$$k(x, z) = \sum_{i=1}^{\infty} \lambda_i \Phi_i(x)\Phi_i(z)$$

Where $\lambda_i \geq 0$ and $\sum_{i=1}^{\infty} \lambda_i < \infty$, and Φ_i here denote the orthonormal system of eigenfunctions associated with the respective integral operator. Furthermore, this series converges both uniformly and absolutely.

4. Results and discussion

4.1. Synthetic Dataset: Concentric Circles

To empirically evaluate the performance of various positive definite kernels on a complex, non linearly separable problem, a synthetic dataset of two concentric circles was generated. This dataset is specifically designed to be challenging for linear classifiers. The data generation process is defined as follows:

1- Inner Circle (Class 0): Data points are generated at a unit radius, $r_0 = 1$.
2- Outer Circle (Class 1): Points are distributed at a radius defined by $r_1 = 1 +$ separation factor.
3- Angular Distribution: For both classes, the angle θ is uniformly sampled from $[0,2\pi]$ to ensure a symmetric distribution of points.
4- Gaussian Noise Addition: Independent zero-mean Gaussian noise was incorporated into both spatial coordinates (x, y). While the standard deviation σ generally spans uniformly across $[0.05,0.3]$ to simulate real-world measurement variations, it was held constant at $\sigma = 0.05$ for this particular experiment.
5- The final Cartesian coordinates (x, y) for each data point are given by:

$$x = r\,cos(\theta) + \varepsilon \qquad y = r\,sin(\theta) + \varepsilon$$

Where $r \in \{r_0, r_1\}$.

4.2 Experimental Design

- Algorithms: A SVM classifier was utilized to evaluate how different kernels performs.
- Kernels: The following PDKs were compared linear kernel, polynomial kernel and the Gaussian kernel (RBF).
- Evaluation Metrics: Model performance was evaluated by Accuracy and the Macro F1-Score to provide a balanced view of predictive power across both classes.

5. Experimental Results

For the evaluation of the performance of different kernels, the data set which was tested is of two concentric circles which is synthetic and non-linearly separable. Sets of the two concentric circles have been set for testing the strength of the Linear, Polynomial, and RBF kernels considering different conditions of additive Gaussian noise (σ from 0.05 to 0.30) and for different dataset sizes (N=200, 1000, 5000). Performance data which is iterative for different dataset sizes is listed in the table 1 and rest is depicted in the figures 1 to 12. These results show the performance to be in order which is consistent and hierarchical due to the properties of the kernels.

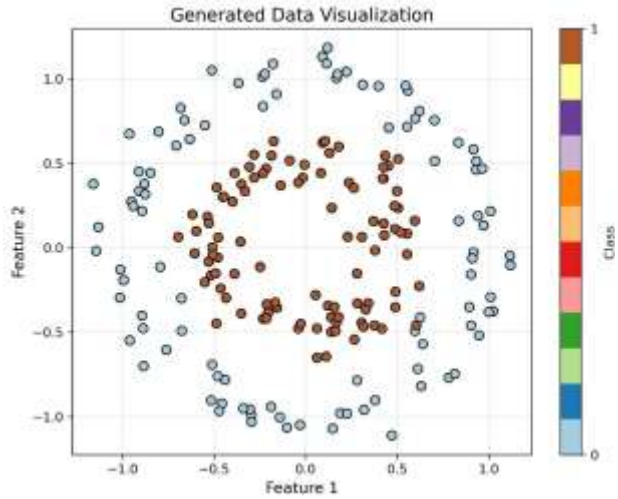


Figure 1 Original Data Visualization (200 datapoints).

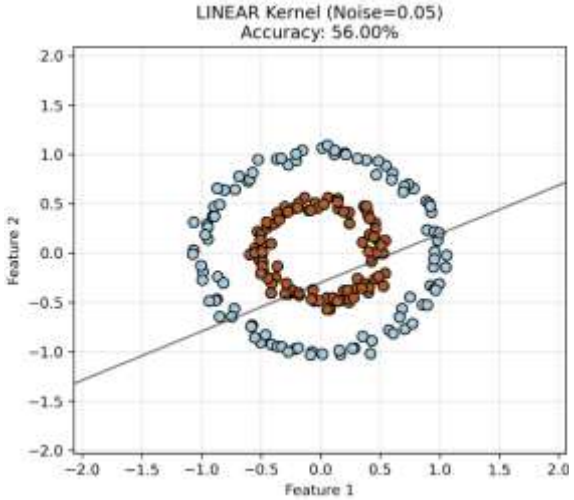


Figure 2 Linear kernel decision boundary (200 data points)

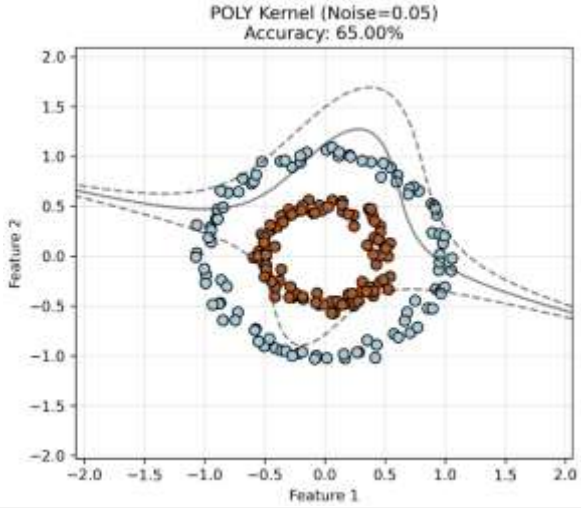


Figure 3 Polynomials kernel decision boundary (200 data points)**.**

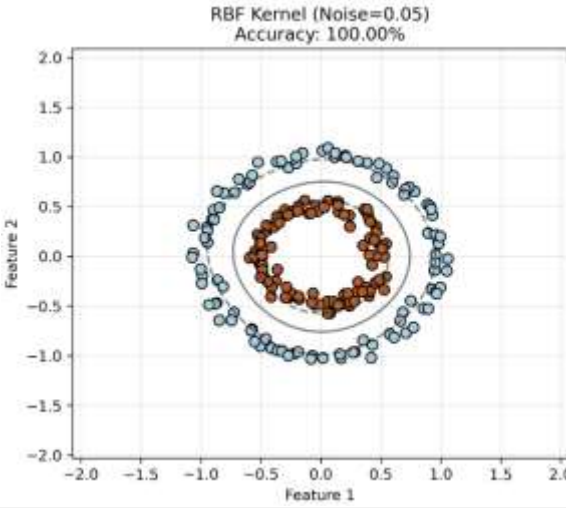


Figure 4 Gaussian kernel decision boundary (200 data points)

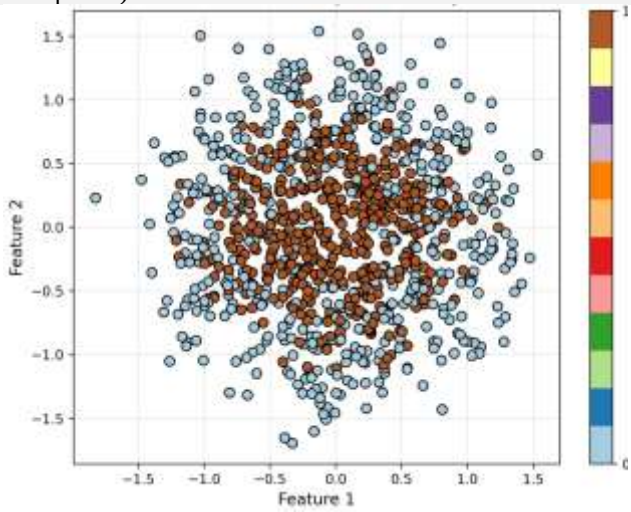


Figure 5 Original Data Visualization (1000 datapoints)

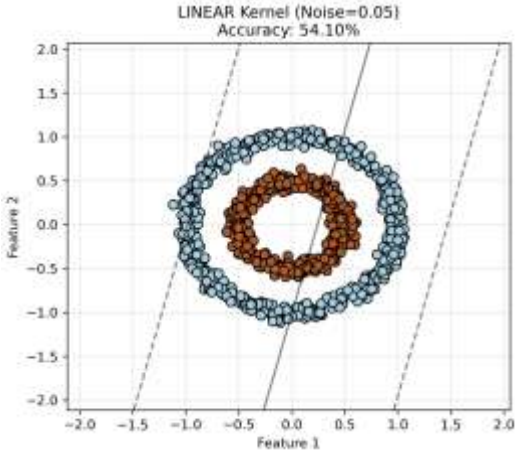


Figure 6 Linear kernel decision boundary (1000 data points)

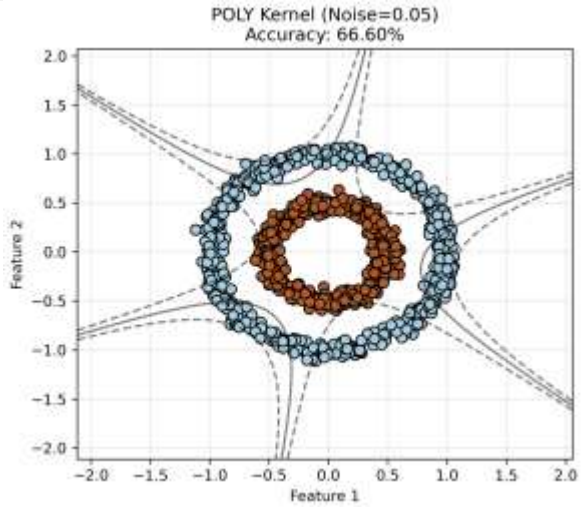


Figure 7 Polynomials kernel decision boundary (1000 data points)

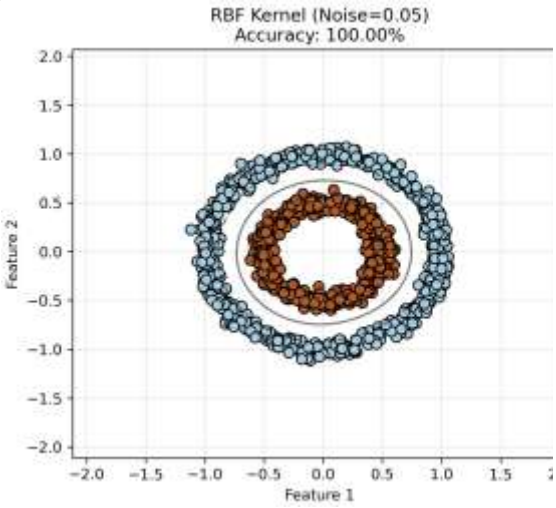


Figure 8 Gaussian kernel decision boundary (1000 data points)

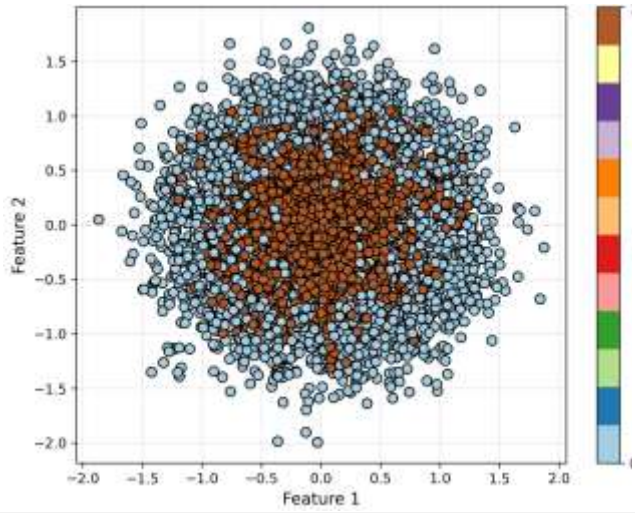


Figure 9 Original Data Visualization (5000 datapoints).

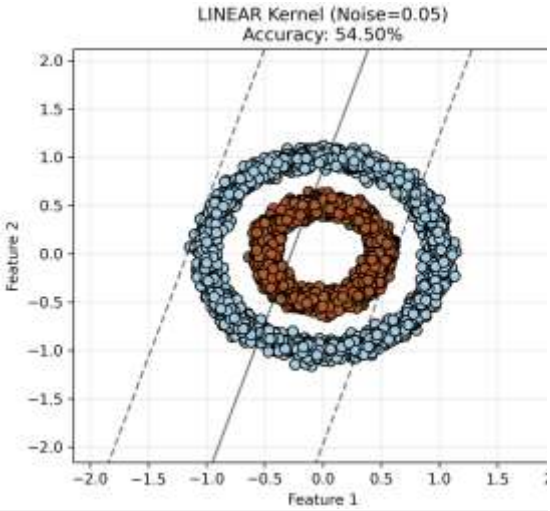


Figure 10 Linear kernel decision boundary (5000 data points)

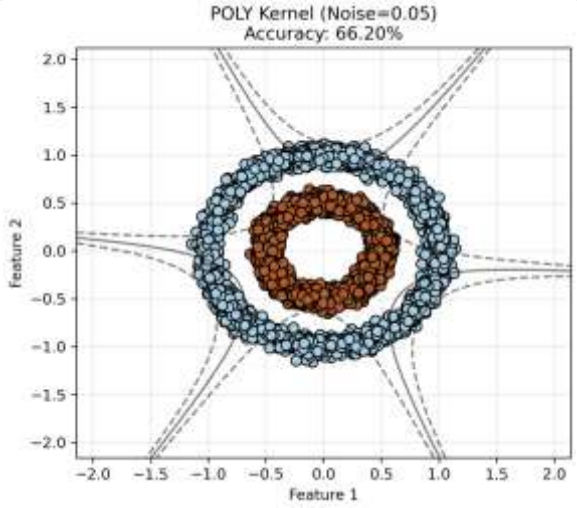


Figure 11 Polynomials kernel decision boundary (5000 data points)

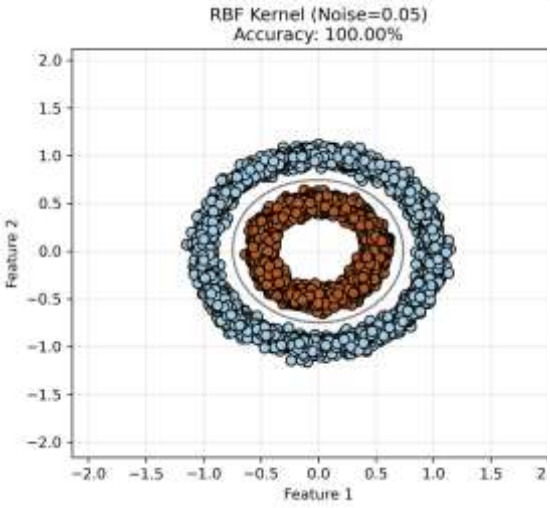


Figure 12 Gaussian kernel decision boundary (5000 data points)

5.1 Accuracy of SVM Kernels at Different Noise Levels

Table 1 provides a detailed comparison of accuracy levels achieved by different kernels under varying noise conditions. The findings indicate that the RBF kernel is more robust to noise, consistently attaining higher classification accuracy than both the linear and polynomial kernels.

Table 1 Comparative accuracy (%) of SVM kernels under varying noise levels and sample sizes.

NOISE LEVEL (σ)	SAMPLE	LINEAR KERNEL	POLYNOMIAL KERNEL	RBF KERNEL
0.05	N=200	56.00	65.00	100.00
	N=1000	54.10	66.60	100.00
	N=5000	54.50	66.20	100.00
0.10	N=200	58.00	63.00	99.60
	N=1000	54.10	64.30	99.60
	N=5000	54.88	62.82	99.40
0.15	N=200	56.00	62.00	94.50
	N=1000	54.30	61.20	94.60
	N=5000	54.96	60.28	95.38
0.20	N=200	57.00	64.00	89.50
	N=1000	53.90	59.00	88.40
	N=5000	55.50	58.92	89.14
0.25	N=200	58.00	63.00	86.00
	N=1000	54.20	59.50	83.30
	N=5000	55.24	58.61	83.68

0.30	N=200	57.00	58.00	82.00
	N=1000	53.30	57.40	77.90
	N=5000	54.84	57.18	79.26

6. Discussion

The experimental results heavily support the theoretical basis of kernel methods. The linear kernel's consistent failure (54-58% accuracy) across all settings witnesses its incompetence in non-linear separation tasks, particularly this problem of circular decision boundary. The RBF kernel performed better, with perfect separation for low noise levels (100% accuracy at $\sigma = 0.05$) and with robustness (77.9-82% accuracy) even at high noise levels ($\sigma = 0.30$). The reason for this perfection is its ability to create complex, localized decision boundaries through its projection of infinite-dimensional feature space, which is appropriate for the radial organization of our data. The polynomial kernel generated mid-range accuracies (57-67%), showing some non-linearity but poor inductive bias for this specific geometry compared to RBF .These findings emphasize that a decision between kernels must be made based on the intrinsic data geometry and not be reduced to a matter of hyperparameter optimization. The fine agreement between theoretical kernel properties and experimental performance emphasizes the importance of understanding feature space mappings across machine learning tasks.

7. Conclusion

These findings both validate and extend RKHS's theoretical model and Mercer's conditions, validating their practical applicability in real-world applications. The results provide the following direct guidance for practitioners: RBF kernels offer maximum possible performance with challenging non-linear patterns, but simpler kernels are adequate for linearly separable instances.

This work unifies theoretical kernel methodology with practical implementation and suggests promising directions for future research on adaptive kernel designs and parameter learning by machine learning systems. The demonstrated correspondence between theoretical mathematics and empirical performance underscores the necessity for theoretically-informed kernel selection in machine learning systems.

.

8. References

[1] J. Mercer, “Xvi. functions of positive and negative type, and their connection the theory of integral equations,” *Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character*, vol. 209, no. 441–458, pp. 415–446, 1909.

[2] I. J. Schoenberg, “Metric spaces and positive definite functions,” *Trans. Am. Math. Soc.*, vol. 44, no. 3, pp. 522–536, 1938.

[3] A. N. Kolmogorov, *Stationary sequences in Hilbert space*. John Crerar Library] National Translations Center, 1978.

[4] N. Aronszajn, “Theory of reproducing kernels,” *Trans. Am. Math. Soc.*, vol. 68, no. 3, pp. 337–404, 1950.

[5] N. Dunford, “Linear operators II,” *No Title*, 1963.

[6] M. A. Aizerman, E. M. Braverman, and L. I. Rozonoer, “Theoretical foundations of potential function method in pattern recognition,” *Autom. Remote Control*, vol. 25, no. 6, pp. 917–936, 1964.

[7] S. Boughorbel, J.-P. Tarel, and N. Boujemaa, “Conditionally positive definite kernels for svm based image recognition,” presented at the 2005 IEEE International Conference on Multimedia and Expo, IEEE, 2005, pp. 113–116.

[8] *Reproducing Kernel Hilbert Spaces*. University of Houston, 2022.

[9] Maryam,T, “Stable bases for CPD kernel spaces: Factorization and duality,” *SIAM J. Numer. Anal.*, 2024.

[10] A. Z. M. Alzarqiu, “Leveraging Positive Definite Kernels for Improved Performance in Machine Learning Models,” M.S, Sirt, 2024.

[11] H. Yang, *Learning methods in reproducing Kernel Hilbert space based on high-dimensional features*. The University of North Carolina at Chapel Hill, 2016.

[12] F. Dinuzzo and G. D. Nicolao, “Learning functions with kernel methods”.

[13] D. Krebs, “Introduction to kernel methods,” 2007.

[14] A. C. Pipkin, “Hilbert—Schmidt Theory,” in *A course on integral equations*, Springer, 1991, pp. 44–71.

Disclaimer/Publisher’s Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **JIBAS** and/or the editor(s). **JIBAS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.