



OPEN The impact of the exponential Kernel's bandwidth parameter on learning algorithms

Mahdi A. Almahdawi

Exponential kernels have been used considerably in statistics, machine learning, and artificial intelligence for tasks such as kernel principal component analysis (Kernel PCA), support vector machines(SVM), visualization, clustering, and pattern recognition. Selecting different bandwidth parameters for the exponential kernel can lead to varying insights about the data. Hence, understanding the theoretical impact of the bandwidth parameter is crucial. This paper investigates the influence of the exponential kernel's bandwidth parameter on the approximation of continuous operators by their empirical counterparts, supported by some experimental algorithms.

Keywords Exponential kernel, Bandwidth parameter, Kernel principal component analysis, Reproducing kernel, Support vector machine

Exponential kernels are widely recognized as an effective choice in kernel-based methods. In learning algorithms and artificial intelligence, choosing an appropriate bandwidth parameter for a kernel is as important as selecting the kernel itself^{1,2}. In most cases, statistical learning and data analysis depend on estimating the eigenvalues and eigenvectors of the kernel matrix, which represents the empirical version of a continuous operator. This is because we mostly use kernel functions to map the data to a large space, where the dimension could be infinite, and thus we need to deal with continuous functions. In the last decades, several works studied the connection between the two operators using perturbation theory. The first study that involved defining an integral operator by a kernel appeared in 1998 by Koltchinskii and in 2000 by Koltchinskii and Giné, where they studied the connection between the spectrum of the integral operator and the spectrum of their empirical counterpart. In 2010, another study by Rosasco³, which was focusing on any positive definite kernel has shown how close the eigenvalues of the two operators to each other and the eigenvectors to their continuous counterpart. In a recent study, we were able to apply the same technique to Gaussian kernels⁴. We have shown the impact of the bandwidth parameter of the Gaussian kernels on the closeness of the two versions of the operators.

In this paper, we focus on exponential kernels and their bandwidth parameter. Therefore, it is essential to discuss the Reproducing Kernel Hilbert Space(RKHS) of the exponential kernel and shows its orthonormal basis(ONB) explicitly. With the help of Steinnwart's results, we were able to derive an explicit description of the RKHS induced by the exponential kernel⁵, which is crucial in obtaining some bounds in the following sections. We then study the convergence of certain operators defined by the exponential kernel on the RKHS, and make some bounds that can be extended to the convergence of eigenvalues and eigenfunctions. To support our theoretical results and examine the impact of the exponential kernel's bandwidth parameter, we provide some examples using support vector machine(SVM). These examples illustrate how the bandwidth parameter affects the performance of the nonlinear classifier in support vector machines.

Settings and preliminaries

In this section, we introduce some notations and discuss some properties of the exponential kernel and its corresponding reproducing kernel Hilbert space.

Reproducing Kernel Hilbert spaces induced by exponential kernels

Let \mathcal{X} be a nonempty subset of \mathbb{R}^d , and let $k_\lambda : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be the exponential kernel⁶ given by

$$k_\lambda(x_1, x_2) = e^{-\frac{\langle x_1, x_2 \rangle}{\lambda}}, \quad \lambda > 0, \quad x_1, x_2 \in \mathcal{X}, \quad (1)$$

GRC Department, The Applied College, King Abdulaziz University, Jeddah, Saudi Arabia. email: melmehdwy@kau.edu.sa

where λ is the bandwidth parameter of k_λ . Then, the feature map $\Phi(x_1) : \mathcal{X} \rightarrow \mathcal{H}_{k_\lambda}(\mathcal{X})$ of the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}_{k_\lambda}(\mathcal{X})$ induced by $k_\lambda(x_1, x_2)$ is given by

$$\Phi(x_1) = e^{\frac{\langle x_1, \cdot \rangle}{\lambda}}, \quad \forall x_1 \in \mathcal{X}. \tag{2}$$

$\mathcal{H}_{k_\lambda}(\mathcal{X})$ admits the dot product as follows

$$\langle \Phi(x_1), \Phi(x_2) \rangle_{\mathcal{H}_{k_\lambda}(\mathcal{X})} = \langle e^{\frac{\langle x_1, \cdot \rangle}{\lambda}}, e^{\frac{\langle \cdot, x_2 \rangle}{\lambda}} \rangle_{\mathcal{H}_{k_\lambda}(\mathcal{X})} = e^{\frac{\langle x_1, x_2 \rangle}{\lambda}} = k_\lambda(x_1, x_2). \tag{3}$$

The reproducing property of the kernel k_λ can be defined by

$$f(x_1) = \langle f, e^{\frac{\langle x_1, \cdot \rangle}{\lambda}} \rangle_{\mathcal{H}_{k_\lambda}(\mathcal{X})}, \quad \forall x_1 \in \mathcal{X} \text{ and } f \in \mathcal{H}_{k_\lambda}(\mathcal{X}).$$

Set

$$\kappa_\lambda = \sup \sqrt{k_\lambda(x_1, x_2)} = \sup \sqrt{e^{\frac{\langle x_1, x_2 \rangle}{\lambda}}}, \quad \lambda > 0, \quad x_1, x_2 \in \mathcal{X}. \tag{4}$$

Therefore, all evaluation functions are bounded^{7,8}, that is

$$|f(x_1)| = \sqrt{\langle f, e^{\frac{\langle x_1, \cdot \rangle}{\lambda}} \rangle_{\mathcal{H}_{k_\lambda}(\mathcal{X})}} \leq \sqrt{e^{\frac{\langle x_1, x_1 \rangle}{\lambda}}} \sqrt{\langle f, f \rangle_{\mathcal{H}_{k_\lambda}(\mathcal{X})}} \leq \kappa_\lambda \|f\|_{\mathcal{H}_{k_\lambda}(\mathcal{X})}. \tag{5}$$

Integral operators defined by exponential reproducing kernel

Exponential kernels are commonly used in machine learning applications and learning algorithms. The bandwidth parameters chosen for exponential kernels have a significant impact on the performance of these algorithms. For that reason, our contribution is to provide a theoretical analysis of the extent of this effect. In 2010³, Rosasco considered the case of a positive kernel and established some bounds on eigenvalues and spectral projections; however, he did not investigate the impact of the kernel parameter. In our case, the analysis will be centered on exponential kernels. We derive some bounds that illustrate the effect of these parameters.

First of all, let $\mathcal{X} \subset \mathbb{R}^d$ and $k_\lambda : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be the exponential reproducing kernel defined by

$$k_\lambda(x_1, x_2) = e^{\frac{\langle x_1, x_2 \rangle}{\lambda}}, \quad \lambda > 0, \quad x_1, x_2 \in \mathcal{X},$$

Let ρ be a probability measure on \mathcal{X} and $\mathbb{L}^2(\mathcal{X}, \rho)$ be the space of square-integrable functions with the norm $\|f\|_{\mathbb{L}^2(\mathcal{X}, \rho)} = \langle f, f \rangle_{\mathbb{L}^2(\mathcal{X}, \rho)} = \int_{\mathcal{X}} |f(x_1)|^2 d\rho(x_1)$. Let $L_{k_\lambda} : \mathbb{L}^2(\mathcal{X}, \rho) \rightarrow \mathbb{L}^2(\mathcal{X}, \rho)$ be an integral operator defined by

$$(L_{k_\lambda} f)(x_1) = \int_{\mathcal{X}} e^{\frac{\langle x_1, x_2 \rangle}{\lambda}} f(x_2) d\rho(x_2) \tag{6}$$

for all $x_1, x_2 \in \mathcal{X}$, $\lambda \in \mathbb{R}^+$, and $f \in \mathbb{L}^2(\mathcal{X}, \rho)$. We know that $e^{\frac{\langle x_1, x_2 \rangle}{\lambda}} \leq \kappa_\lambda$ for all $x_1, x_2 \in \mathcal{X}$. Therefore, L_{k_λ} is a bounded operator.

Now assume that we are given a set of points $\{x_1, \dots, x_n\} \subset \mathbb{R}^d$ sampled i.i.d. according to ρ . The entries of the $n \times n$ kernel matrix \mathbb{K}_n are computed by $K_{i,j} = \frac{1}{n} e^{\frac{\langle x_i, x_j \rangle}{\lambda}}$. Let $\mathcal{H}_{k_\lambda}(\mathcal{X})$ be the exponential reproducing kernel Hilbert space and define the operators $\mathcal{T}_{\mathbb{N}}, \mathcal{T}_n : \mathcal{H}_{k_\lambda}(\mathcal{X}) \rightarrow \mathcal{H}_{k_\lambda}(\mathcal{X})$ by

$$\mathcal{T}_{\mathbb{N}} = \int_{\mathcal{X}} \langle \cdot, e^{\frac{\langle x_1, \cdot \rangle}{\lambda}} \rangle_{\mathcal{H}_{k_\lambda}(\mathcal{X})} e^{\frac{\langle x_1, \cdot \rangle}{\lambda}} d\rho(x_1), \tag{7}$$

$$\mathcal{T}_n = \frac{1}{n} \sum_{i=1}^n \langle \cdot, e^{\frac{\langle x_i, \cdot \rangle}{\lambda}} \rangle_{\mathcal{H}_{k_\lambda}(\mathcal{X})} e^{\frac{\langle x_i, \cdot \rangle}{\lambda}}. \tag{8}$$

Donate the inclusion map $\mathcal{H}_{k_\lambda}(\mathcal{X}) \hookrightarrow \mathbb{L}^2(\mathcal{X}, \rho)$ and its adjoint operator by $\mathcal{I}_n, \mathcal{I}_n^*$ respectively.

Results

In this section, we begin by introducing an explicit orthonormal basis of the exponential reproducing kernel Hilbert space (RKHS), as stated in the following proposition. The result is deduced from Steinwart's results⁵ for the Gaussian kernel.

Proposition 1 Let $n_1, \dots, n_d \in \mathbb{N}$ and $u := (n_1, \dots, n_d)$. Define $e_{n_j} : \mathbb{R} \rightarrow \mathbb{R}$, as follows,

$$e_{n_j}(x_j) = \sqrt{\frac{1}{(n_j!) (\lambda^{n_j})}} x_j^{n_j}, \quad x_j \in \mathbb{R}, \quad n_j \in \{n_1, \dots, n_d\}.$$

Then, the system $e_u(x) = e_{n_1}(x_1) \otimes \cdots \otimes e_{n_d}(x_d)$ serves an orthonormal basis of $\mathcal{H}_{k_\lambda}(\mathbb{R}^d)$, with $k_\lambda(x_1, x_2) = e^{\frac{\langle x_1, x_2 \rangle}{\lambda}}$ being its reproducing kernel.

Proof Let us prove that $k_\lambda(x_1, x_2)$ is the reproducing kernel of $\mathcal{H}_{k_\lambda}(\mathbb{R}^d)$, while the reader can easily verify that the system forms an orthonormal basis using the same technique as in the Gaussian RKHS case⁵. To proceed, let $x_1, x_2 \in \mathbb{R}^d$ with $x_1 = (x_{11}, x_{12} \dots, x_{1d})$ and $x_2 = (x_{21}, x_{22} \dots, x_{2d})$ then the orthonormal basis we obtain

$$\begin{aligned} k_\lambda(x_1, x_2) &= \sum_{n_1, \dots, n_d = 0}^{\infty} e_{n_1} \otimes \cdots \otimes e_{n_d}(x_1) \cdot e_{n_1} \otimes \cdots \otimes e_{n_d}(x_2) \\ &= \sum_{n_1, \dots, n_d = 0}^{\infty} \prod_{j=1}^d \sqrt{\frac{1}{(n_j!)(\lambda^{n_j})}} x_{1j}^{n_j} \sqrt{\frac{1}{(n_j!)(\lambda^{n_j})}} x_{2j}^{n_j} \\ &= \prod_{j=1}^d \sum_{n_1, \dots, n_d = 0}^{\infty} \frac{1}{(n_j!)(\lambda^{n_j})} (x_{1j} \cdot x_{2j})^{n_j} \\ &= \prod_{j=1}^d e^{\frac{x_{1j} \cdot x_{2j}}{\lambda}} \\ &= e^{\frac{\langle x, y \rangle}{\lambda}}. \end{aligned}$$

□

Now, the above proposition leads us to another proposition in which we can compute the exact norm of the Hilbert-Schmidt operator \mathcal{T}_N (7).

Proposition 2 Let $\mathcal{X} = \mathbb{R}$, and $\rho(x_1)$ be a normal distribution with a density $\phi_\sigma(x_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_1^2}{2\sigma^2}}$. Given the assumptions stated above, \mathcal{T}_N is a Hilbert-Schmidt operator, and $\|\mathcal{T}_N\|_{HS}^2 = \frac{1}{\sqrt{1-4\sigma^4/\lambda^2}}$.

Proof First of all, for all $x_1, x_2 \in \mathbb{R}$ we have

$$\begin{aligned} (\mathcal{T}_N f)(x_1) &= \langle k_\lambda(x_1, \cdot), \mathcal{T}_N f \rangle_{\mathcal{H}_{k_\lambda}(\mathbb{R})} \\ &= \langle k_\lambda(x_1, \cdot), \mathcal{I}_n^* \mathcal{I}_n f \rangle_{\mathcal{H}_{k_\lambda}(\mathbb{R})} \\ &= \langle \mathcal{I}_n k_\lambda(x_1, \cdot), \mathcal{I}_n f \rangle_{L^2(\mathbb{R})} \\ &= \int_{\mathbb{R}} k_\lambda(x_1, x_2) f(x_2) d\rho(x_2). \end{aligned}$$

Recall

$$(\mathcal{T}_N f)(x_1) = \int_{\mathbb{R}} e^{\frac{x_1 \cdot x_2}{\lambda}} f(x_2) d\rho(x_2). \tag{9}$$

We know that $\{e_n(x_1) = \sqrt{\frac{1}{\lambda^n \cdot n!}} x_1^n, n = 0, 1, \dots\}$ is an orthonormal basis for the exponential RKHS $\mathcal{H}_{k_\lambda}(\mathbb{R})$. Now we can find the exact value of $\|\mathcal{T}_N\|_{HS}^2$ as follows,

$$\begin{aligned}
 \|\mathcal{T}_{\mathbb{N}}\|_{HS}^2 &= \sum_{n=0}^{\infty} \langle \mathcal{T}_{\mathbb{N}}e_n, \mathcal{T}_{\mathbb{N}}e_n \rangle_{\mathcal{H}_{\kappa_{\lambda}}(\mathbb{R})} = \sum_{n=0}^{\infty} \langle \mathcal{T}_{\mathbb{N}}e_n, \mathcal{I}_n^* \mathcal{I}_n e_n \rangle_{\mathcal{H}_{\kappa_{\lambda}}(\mathbb{R})} \\
 &= \sum_{n=0}^{\infty} \langle \mathcal{I}_n \mathcal{T}_{\mathbb{N}}e_n, \mathcal{I}_n e_n \rangle_{\mathcal{L}^2(\mathbb{R})} = \sum_{n=0}^{\infty} \langle \mathcal{T}_{\mathbb{N}}e_n, e_n \rangle_{\mathcal{L}^2(\mathbb{R})} \\
 &= \sum_{n=0}^{\infty} \int_{\mathbb{R}} \mathcal{T}_{\mathbb{N}}e_n(x_1) \cdot e(x_1) d\rho(x_1) = \sum_{n=0}^{\infty} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} k_{\lambda}(x_1, x_2) e_n(x_2) d\rho(x_2) \right) e_n(x_1) d\rho(x_1) \\
 &= \sum_{n=0}^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}} e_n(x_2) k_{\lambda}(x_1, x_2) e_n(x_1) d\rho(x_2) d\rho(x_1) \\
 &= \frac{1}{2\sigma^2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \sum_{n=0}^{\infty} \frac{1}{\lambda^n \cdot n!} x_2^n e^{-\frac{x_1 \cdot x_2}{\lambda}} x_1^n e^{-\frac{x_2^2}{2\sigma^2}} e^{-\frac{x_1^2}{2\sigma^2}} dx_2 dx_1 \\
 &= \frac{1}{2\sigma^2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2} x_1^2 + \frac{2x_1 x_2}{\lambda} - \frac{1}{2\sigma^2} x_2^2} dx_1 dx_2 = \frac{1}{\sqrt{1 - 4\frac{\sigma^4}{\lambda^2}}}.
 \end{aligned}$$

□

Corollary 1 If $\mathcal{X} = \mathbb{R}^d$, then $\|\mathcal{T}_{\mathbb{N}}\|^2 = \frac{1}{\sqrt{1 - 4\frac{\sigma^4}{\lambda^2}}}$.

The following theorem will provide a new bound that depends on σ and the bandwidth parameter λ for the difference $\mathcal{T}_{\mathbb{N}} - \mathcal{T}_n$.

Theorem 1 $\mathcal{T}_{\mathbb{N}}$ and \mathcal{T}_n are Hilbert-Schmidt operators. Given the assumptions stated above with probability $1 - 2e^{-\tau}$

$$\|\mathcal{T}_{\mathbb{N}} - \mathcal{T}_n\|_{HS} \leq \sqrt{\frac{2\tau}{n}} \left[\kappa_{\lambda} + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right].$$

Proof Assume that we have a sequence of random variables $(\eta_i)_{i=1}^n$ in the Hilbert space of Hilbert-Schmidt operators given by

$$\eta_i = \langle \cdot, e^{\frac{\langle x_i, \cdot \rangle}{\lambda}} \rangle_{\mathcal{H}_{\kappa_{\lambda}}(\mathbb{R})} e^{\frac{\langle x_i, \cdot \rangle}{\lambda}} - \mathcal{T}_{\mathbb{N}}.$$

It is obvious that the mean of the random variable η_i is zero, i.e. $E(\eta_i) = 0$. We can compute that

$$\left\| \langle \cdot, e^{\frac{\langle x_i, \cdot \rangle}{\lambda}} \rangle_{\mathcal{H}_{\kappa_{\lambda}}(\mathbb{R})} e^{\frac{\langle x_i, \cdot \rangle}{\lambda}} \right\|_{HS}^2 = \left\| e^{\frac{\langle x_i, \cdot \rangle}{\lambda}} \right\|_{\mathcal{H}_{\kappa_{\lambda}}(\mathbb{R})}^4 \leq \kappa_{\lambda}^2.$$

From Proposition 2, we have $\|\mathcal{T}_{\mathbb{N}}\|_{HS} = \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}}$ and follows

$$\|\eta_i\|_{HS} \leq \kappa_{\lambda} + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}}.$$

Now, we apply Pinelis's⁹ results on concentration inequalities in Hilbert spaces with confidence $1 - 2e^{-\tau}$

$$\left\| \frac{1}{n} \sum_{i=1}^n \eta_i \right\|_{HS} = \|\mathcal{T}_{\mathbb{N}} - \mathcal{T}_n\|_{HS} \leq \sqrt{\frac{2\tau}{n}} \left[\kappa_{\lambda} + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right].$$

□

As can be seen in Theorem 1, the Hilbert-Schmidt operator of the difference between the operators $\mathcal{T}_{\mathbb{N}}$ and \mathcal{T}_n has been bounded. As a result, we introduce the next proposition, which bounds the ℓ_2 -distance between the spectrum of \mathbb{K}_n and the spectrum of $L_{\kappa_{\lambda}}$.

Proposition 3 There are extended enumerations $\{\beta_j\}_{j \geq 1}$, $\{\hat{\beta}_j\}_{j \geq 1}$ of discrete eigenvalues for the operators $L_{\kappa_{\lambda}}$, \mathbb{K}_n respectively such that

$$\sum_{j \geq 1} (\beta_j - \hat{\beta}_j)^2 \leq \frac{2\tau}{n} \left[\kappa_\lambda + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right]^2,$$

with a probability greater than $1 - 2e^{-\tau}$.

Proof First, the extended enumerations $\{\beta_j\}_{j \geq 1}, \{\hat{\beta}_j\}_{j \geq 1}$ of discrete eigenvalues for the operators $L_{k_\lambda}, \mathbb{K}_n$ respectively, are also extended enumerations of discrete eigenvalues for $\mathcal{T}_\mathbb{N}$, and \mathcal{T}_n respectively³. Thus, the following inequality holds

$$\sum_{j \geq 1} (\beta_j - \hat{\beta}_j)^2 \leq \|\mathcal{T}_\mathbb{N} - \mathcal{T}_n\|_{HS}^2.$$

Now, from Theorem 1 we obtain the following

$$\sum_{j \geq 1} (\beta_j - \hat{\beta}_j)^2 \leq \frac{2\tau}{n} \left[\kappa_\lambda + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right]^2. \quad \square$$

It is evident that Proposition 3 provides us with a bound that contains influential factors such as the number of examples n and the bandwidth parameter λ . Minimizing this bound means that we minimize the difference between the eigenvalues of the operators of $\mathbb{K}n$ and L_{k_λ} , which means that they become closer to each other. For example, when n reaches infinity, the eigenvalues will be almost the same. We can also minimize the targeted bound by maximizing the bandwidth parameter λ . We should remember that approximating the integral operator L_{k_λ} by its empirical version, which is the kernel matrix \mathbb{K}_n is necessary to have a theoretical understanding of what happens in learning algorithms. Since, Proposition 3 shows how we can estimate the eigenvalues of L_{k_λ} by the eigenvalues of \mathbb{K}_n , the following theorem will show how we estimate the eigenfunctions of L_{k_λ} by the eigenvectors of \mathbb{K}_n .

Theorem 2 Assume that \mathcal{M} is the number of the first distinct eigenvalues of the operator L_{k_λ} with a total number of multiplicities r that satisfy $\beta_1 \geq \dots \geq \beta_r > \beta_{r+1}$, and the rank of the kernel matrix \mathbb{K}_n is q . Let $\hat{u}_1, \dots, \hat{u}_q$ be the eigenvectors of \mathbb{K}_n , which correspond to the nonzero eigenvalues in a nonincreasing order, and $\hat{v}_1, \dots, \hat{v}_q$ be the eigenfunctions of \mathcal{T}_n obtained by the Nyström extension.

Denote by $\mathcal{A}_\mathcal{M}$ the orthogonal projection from the Hilbert space of square integrable functions $\mathbb{L}^2(\mathbb{R}^d, \rho)$ onto the span of the eigenfunctions corresponding to the eigenvalues β_1, \dots, β_q . Assume that we are given $\tau > 0$, and n number of examples such that

$$n > \frac{32\tau}{(\beta_r - \beta_{r+1})^2} \left[\kappa_\lambda + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right]^2,$$

then the following inequality holds true

$$\sum_{j=1}^r \|(I - \mathcal{A}_\mathcal{M})\hat{v}_j\|_{\mathbb{L}^2(\mathbb{R}^d, \rho)}^2 + \sum_{j=r+1}^q \|\mathcal{A}_\mathcal{M}\hat{v}_j\|_{\mathbb{L}^2(\mathbb{R}^d, \rho)}^2 \leq \frac{16\tau}{n(\beta_r - \beta_{r+1})^2} \left[\kappa_\lambda + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right]^2$$

with confidence at least $1 - 2e^{-\tau}$.

Proof Assume that the elements of the set $\{u_1, u_2, u_3, \dots\}$ are orthonormal eigenfunctions of the operator L_{k_λ} with corresponding eigenvalues $\{\beta_1, \beta_2, \beta_3, \dots\}$ that are strictly positive, and the elements of the set $\{v_1, v_2, v_3, \dots\}$ are the corresponding orthonormal eigenfunctions of the operator $\mathcal{T}_\mathcal{N}$. Let $\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_q\}$ be another set of orthonormal eigenfunctions of $\mathcal{T}_\mathcal{N}$. Assume that the number of examples we are given is n , and

$$n > \frac{32\tau}{(\beta_r - \beta_{r+1})^2} \left[\kappa_\lambda + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right]^2. \tag{10}$$

Using Theorem 1, Kato's theorem^{10,11}, and Zwald's results¹², with high confidence we derive

$$\begin{aligned} \|\mathcal{T}_n - \mathcal{T}_\mathcal{H}\|_{HS}^2 &\leq \frac{2\tau}{n} \left[\kappa_\lambda + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right]^2 \\ &\leq \frac{(\beta_r - \beta_{r+1})^2}{16} \end{aligned}$$

Define the orthogonal projections $\mathcal{A}^{\mathcal{T}_\mathcal{H}}$, and $\mathcal{A}^{\mathcal{T}_\mathcal{N}}$ as follows

$$\mathcal{A}^{\mathcal{T}\mathcal{H}} = \sum_{j=1}^r \langle \cdot, v_j \rangle_{\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)} v_j, \quad \mathcal{A}^{\mathcal{T}\mathcal{N}} = \sum_{j=1}^r \langle \cdot, \hat{v}_j \rangle_{\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)} \hat{v}_j.$$

Then, we yield

$$\begin{aligned} \|\mathcal{A}^{\mathcal{T}\mathcal{N}} - \mathcal{A}^{\mathcal{T}\mathcal{H}}\|_{HS}^2 &\leq \frac{4}{(\beta_r - \beta_{r+1})^2} \|\mathcal{T}\mathcal{N} - \mathcal{T}\mathcal{H}\|_{HS}^2 \\ &\leq \frac{8\tau}{n(\beta_r - \beta_{r+1})^2} \left[\kappa_\lambda + \left(\frac{1}{1 - 4\frac{\sigma^4}{\lambda^2}} \right)^{\frac{1}{4}} \right]^2, \end{aligned}$$

Now, each one of $\{v_1, v_2, v_3, \dots\}$, and $\{\hat{v}_1, \hat{v}_2, \hat{v}_3, \dots\}$ is an orthonormal basis for $\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)$. Therefore, we obtain

$$\begin{aligned} \|\mathcal{A}^{\mathcal{T}\mathcal{N}} - \mathcal{A}^{\mathcal{T}\mathcal{H}}\|_{HS}^2 &= \sum_{i,j \geq 1} |\langle \mathcal{A}^{\mathcal{T}\mathcal{N}} v_i - \mathcal{A}^{\mathcal{T}\mathcal{H}} v_i, \hat{v}_j \rangle_{\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)}|^2 \\ &= \sum_{j=1}^r \sum_{i \geq r+1} |\langle v_i, \hat{v}_j \rangle_{\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)}|^2 + \sum_{j \geq r+1} \sum_{i=1}^r |\langle v_i, \hat{v}_j \rangle_{\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)}|^2 \\ &\geq \sum_{j=1}^r \sum_{\substack{i \geq r+1 \\ \mathcal{T}\mathcal{H}v_i \neq 0}} |\langle v_i, \hat{v}_j \rangle_{\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)}|^2 + \sum_{j \geq r+1} \sum_{i=1}^r |\langle v_i, \hat{v}_j \rangle_{\mathcal{H}_{\kappa_\lambda}(\mathbb{R}^d)}|^2 \end{aligned}$$

Now, we apply Mercer’s theorem to obtain the following

$$\sum_{i=1}^r |\langle u_i, \hat{v}_j \rangle_{\mathbb{L}^2(\mathbb{R}^d, \rho)}|^2 = \|\mathcal{A}\mathcal{M}\hat{v}_j\|_{\mathbb{L}^2(\mathbb{R}^d, \rho)}^2$$

and

$$\begin{aligned} \sum_{\substack{i \geq r+1 \\ \mathcal{T}\mathcal{H}v_i \neq 0}} |\langle u_i, \hat{v}_j \rangle_{\mathbb{L}^2(\mathbb{R}^d, \rho)}|^2 &= \sum_{\substack{i \geq r+1 \\ L_{\kappa_\lambda} u_i \neq 0}} |\langle u_i, \hat{v}_j \rangle_{\mathbb{L}^2(\mathbb{R}^d, \rho)}|^2 \\ &= \|(I - \mathcal{A}\mathcal{M})\hat{v}_j\|_{\mathbb{L}^2(\mathbb{R}^d, \rho)}^2. \end{aligned} \quad \square$$

As can be seen in Theorem 2, the relationship between the eigenfunctions of the operator L_{κ_λ} and the eigenfunctions of the operator \mathbb{K}_n is similar to the relationship between their eigenvalues. The impact of the number of examples n and the bandwidth parameter λ on the closeness between the eigenfunctions is almost the same.

Learning algorithms using the exponential kernel

In this section, we provide some examples to test the performance of the exponential kernel at different bandwidth values. In fact, the exponential kernel can be used in many learning algorithms, and sometimes performs better than any other kernels. This would depend heavily on how we select the value of the bandwidth parameter. In addition, it is essential that our experiments include testing performance across different learning algorithms. Therefore, we have chosen support vector machine (SVM), and kernel principal component analysis (kernel PCA) as examples for our study. The experiments are conducted on both synthetic and real data.

Support vector machine experiment using the exponential kernel

In machine learning, support vector classifiers are a learning algorithm used to classify data when it has two separable classes. The main idea of support vector classifiers is to construct the optimal hyperplane that best separates the two classes. The hyperplane can take two forms: linear, when the data are linearly separable, and nonlinear, when the data are not linearly separable. In the non-linear case, we need to use a kernel function that embeds the data into a feature space that admits the “dot product” as a similarity measure. Through the kernel functions, the data are mapped to a reproducing kernel Hilbert space, where the dot product between two elements can be easily computed. Therefore, in order to understand the efficiency of any kernel, it is important to study its mathematical properties. In other words, we need to place the arguments about kernel performance in a mathematical framework. In this paper, our goal is to provide a mathematical explanation for the performance of a particular kernel “the exponential kernel(1)” when using different values for its bandwidth parameter λ . Support vector classifiers are considered the best choice for our experiment. This is due to the fact that the support vector machine is a highly useful tool in machine learning, and also because the bandwidth parameter significantly

affects its efficiency. Now, we are in a suitable position to introduce the support vector classifier algorithm with the exponential kernel, and show some examples with different values for the bandwidth parameter.

Support vector classifier algorithm using the exponential kernel

Suppose that our training data are a set of vectors $\{x_1, \dots, x_n\}$, $x_i, x_j \in \mathbb{R}^d$, $i, j \in \{1, \dots, n\}$ with two classes $y_j \in \{-1, +1\}$. Since we are dealing with exponential kernels, we will assume that the input data are not linearly separable. Our goal is to construct a hyperplane that separates the two classes, but in practice it might not exist due to the overlapping points. As a consequence, soft-margin hyperplanes are introduced to allow the violating points to lie within the margin. This is done by adding what they call “slack variables” to the optimization problem of support vector classifiers^{1,13}, which can be seen below.

In order to compute the support vector classifier using the exponential kernel, the decision function will take the form,

$$f(x) = \operatorname{sgn} \left(\sum_{i=1}^n y_i \alpha_i e^{-\frac{\langle x, x_i \rangle}{\lambda}} + b \right) \quad (11)$$

where $\alpha_i \in \mathbb{R}^d$, and b is a real number called the threshold. To compute the optimal support vector classifier that accepts some violating points, we need to solve the following optimization problem for some $C > 0$,

$$\min_{w \in \mathcal{H}_{k_\lambda}(\mathbb{R}^d), \xi \in \mathbb{R}^n} \tau(w, \xi) = \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i, \quad (12)$$

$$\text{subject to } \xi_i \geq 0, \text{ where } i = 1, \dots, n, \quad (13)$$

$$\text{and } y_i (\langle x_i, w \rangle + b) \geq 1 - \xi_i, \quad i = 1, \dots, n. \quad (14)$$

The elements of the sequence ξ_1, \dots, ξ_n are called slack variables. A point (x_i, y_i) satisfying the constraint 14 at $\xi_i = 0$ will not be considered a margin error, while the non-zero ξ_i corresponds to a margin error. The vector w can be calculated as follows:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (15)$$

where the coefficient α_i takes zero if the corresponding point (x_i, y_i) does not meet the constraint 14. In order to compute the coefficients α_i , we need to solve the following quadratic problem:

$$\max_{\alpha \in \mathbb{R}^n} W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j e^{-\frac{\langle x_i, x_j \rangle}{\lambda}}, \quad (16)$$

$$\text{subject to } 0 \leq \alpha_i \leq \frac{C}{n} \text{ for all } i = 1, \dots, n, \quad (17)$$

$$\text{and } \sum_{i=1}^n \alpha_i y_i. \quad (18)$$

The threshold b can be calculated using the following equation:

$$b = y_j - \sum_{i=1}^n y_i \alpha_i e^{-\frac{\langle x_j, x_i \rangle}{\lambda}}, \text{ for all } 0 > \alpha_j > C. \quad (19)$$

Support vector classifier experiment

In this experiment, our goal is to evaluate the performance of the exponential kernel 1 when using different values of its bandwidth parameter λ with the support vector classifier. Therefore, we randomly generate sets of points *’s and +’s within the unit circle, labeling those in the first and third quadrants as +1, and those in the second and fourth quadrants as -1. First, we create 100 points in each quadrant and train the support vector classifier at three different values of λ : 0.2, 0.6, and 1. Second, we create 150 points in each quadrant and train the support vector classifier using the same values of λ . The dataset used in this paper is based on synthetic data created with Simulink (Mathworks, Natick, MA)¹⁴, and modified by the author to fit our findings. These modifications include control signals such as the number of examples in which the hyperplane classifiers are constructed using the exponential kernel with different bandwidth parameters.

As we can see in Figs. 1, 2, and 3, the number of examples is 100 in each quadrant, and the support vector classifiers are the blue lines that separate the green stars from the red pluses. The values we have chosen for the bandwidth parameter λ are 0.20, 0.60 and 1.00, and the out-of-sample misclassification rates are 4.50%, 4.75% and 5.70% respectively. These rates indicate that the exponential kernel performs well for the three different values, but also indicate that decreasing the value of λ leads to overfitting, while increasing its value leads to

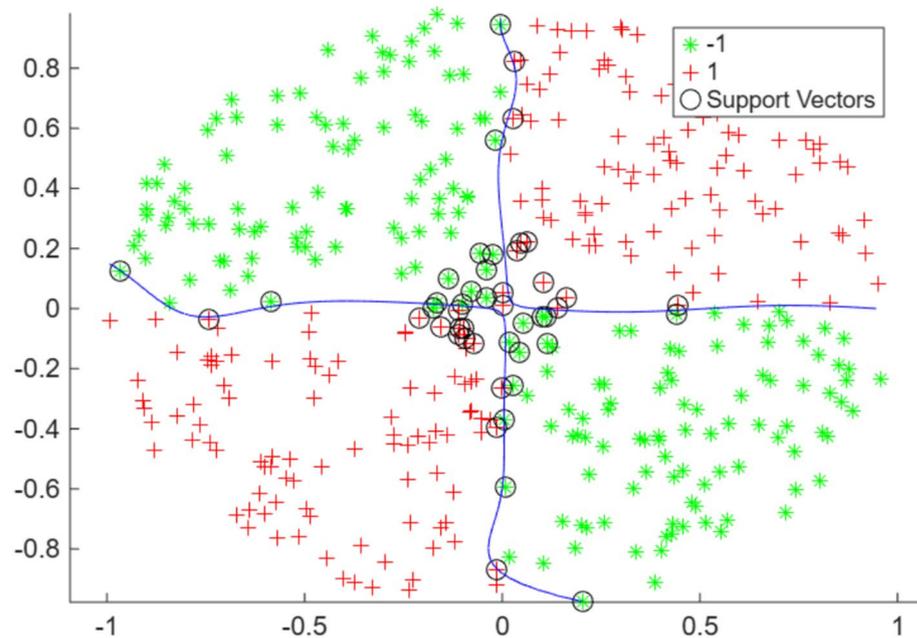


Figure 1. Support vector classifiers (blue lines), when $\lambda = 0.2$ and $n = 100$.

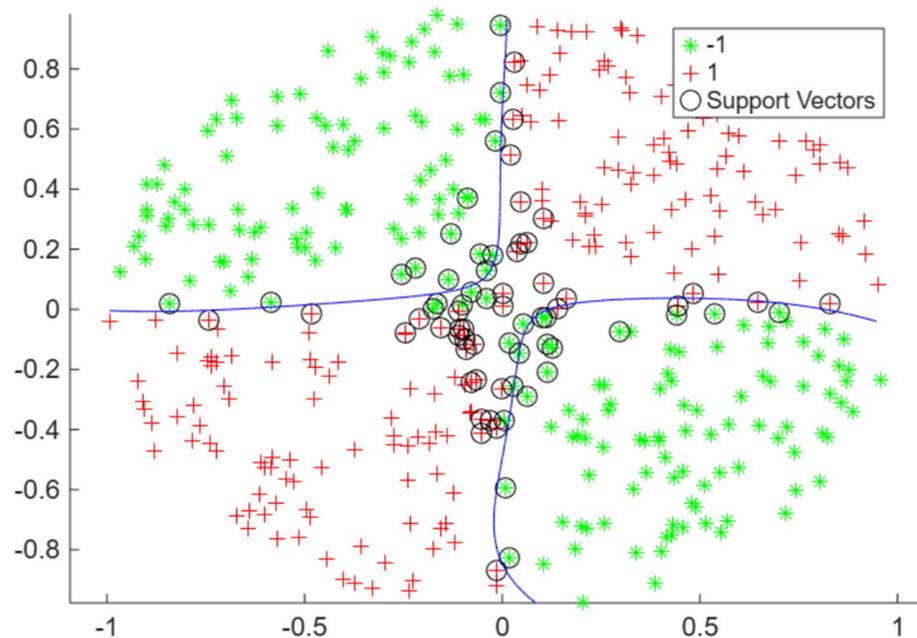


Figure 2. Support vector classifiers (blue lines), when $\lambda = 0.6$ and $n = 100$.

underfitting. In practical applications, it is very important to choose a value for the bandwidth parameter λ that helps constructing the best hyperplanes for generalizations. In our case, 0.60 seems to be the best value we can choose for the bandwidth parameter λ from the three values, and we avoid 0.2, and 1.00 due to overfitting and underfitting respectively.

On the other hand, our theoretical results show how the closeness between the integral operator L_{k_λ} and the operator \mathbb{K}_n can be controlled by the value of λ and the number of examples n . We know that the operator L_{k_λ} works on a large space, which is the Hilbert space of square integrable functions $\mathbb{L}^2(\mathbb{R}^d, \rho)$. Mapping data into that space means that we can use a high dimensional space that helps detecting more information about data. Kernel functions including the exponential kernel are used to embed the data into a high dimensional space. The kernel matrix \mathbb{K}_n is just an empirical version of the operator L_{k_λ} . As previously noted that the closeness between the two operators increases by increasing the number of examples. In Figs. 4, 5, and 6, we have increased

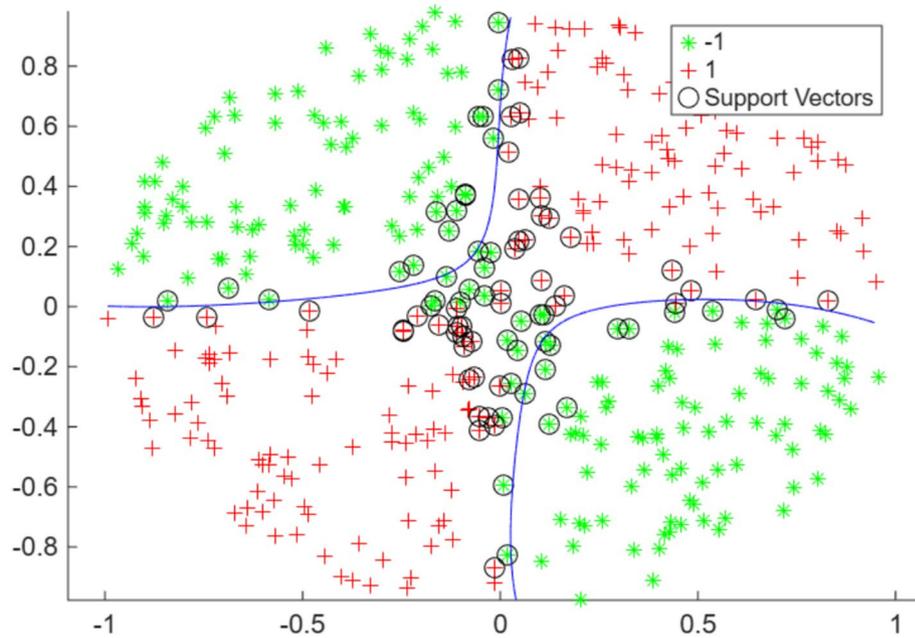


Figure 3. Support vector classifiers (blue lines), when $\lambda = 1$ and $n = 100$.

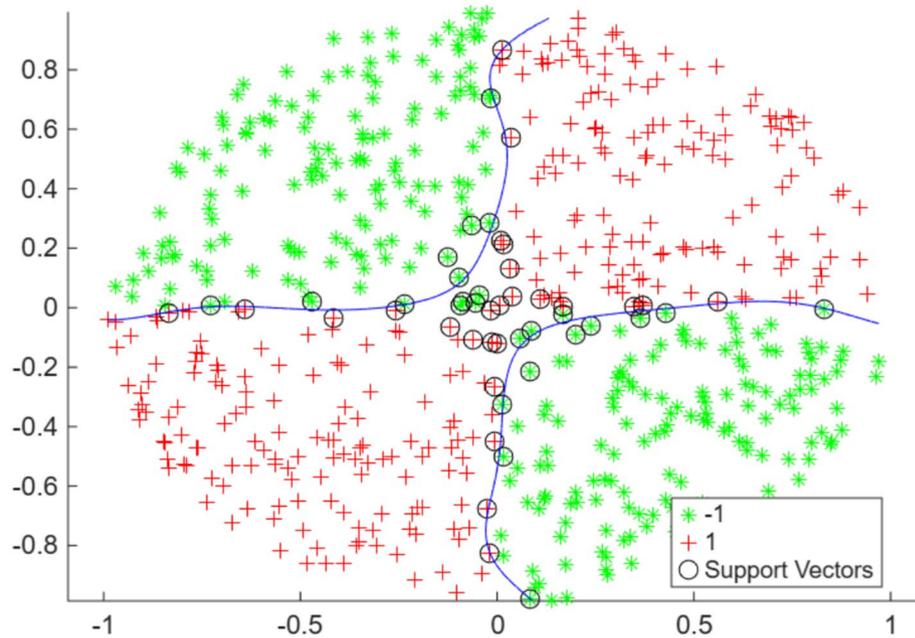


Figure 4. Support vector classifiers (blue lines), when $\lambda = 0.2$ and $n = 150$.

the number of examples from 100 to 150 using the same values of λ , which has decreased the out-of-sample misclassification rates from 4.50%, 4.75%, and 5.70% to 3.67%, 3.67%, and 4.50% respectively.

In terms of our theoretical results on the bandwidth parameter λ and how they clarify the variation in the performance of support vector classifiers when using different values of λ , we have stated in “Results” that increasing the value of the bandwidth would minimize the norm of the difference between the operators L_{k_λ} and \mathbb{K}_n , which means we succeeded in embedding the data into high dimensional space, but they are mapped very closely. As a consequence, the ability of the exponential kernel of detecting any information about the data will be nearly lost. For instance, if λ goes to infinity, the exponential kernel will be equal to 1, and will be no longer able to distinguish between any two points. On the other hand, a very small value of λ increases the difference between the operators L_{k_λ} and \mathbb{K}_n , but it leads to very high values for the exponential kernel that are not appropriate to learn about the data. One of the significant advantages of our theoretical results is that we can

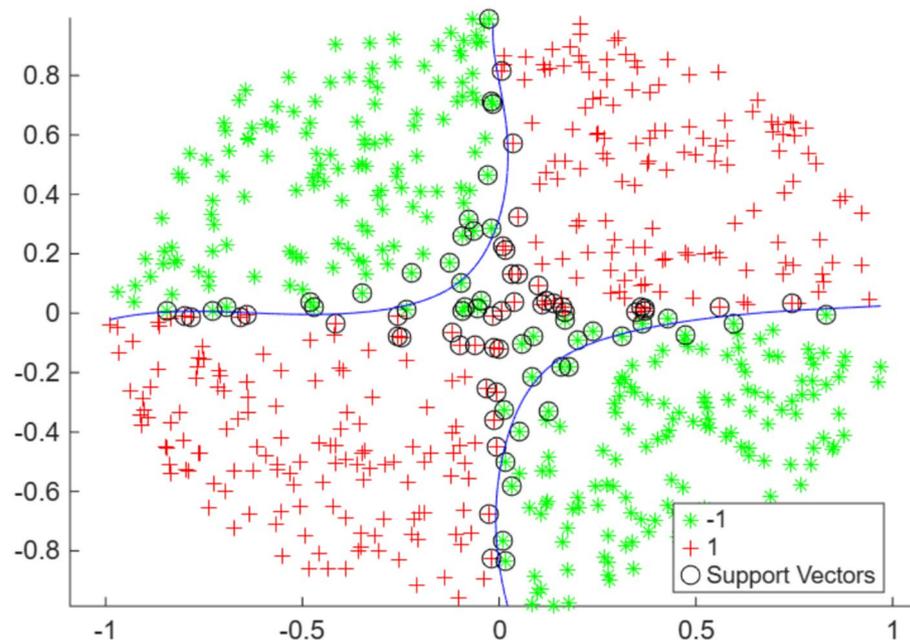


Figure 5. Support vector classifiers (blue lines), when $\lambda = 0.6$ and $n = 150$.

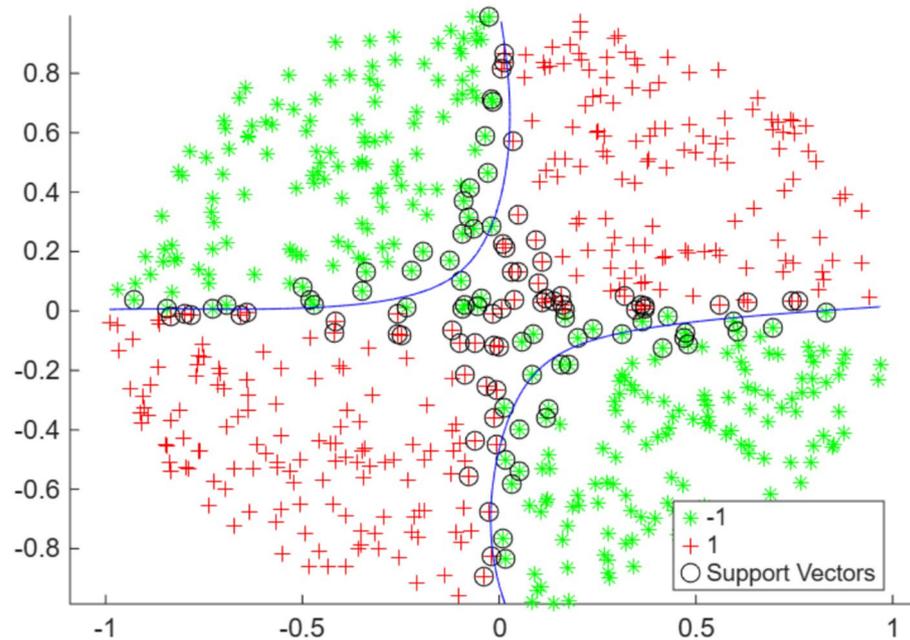


Figure 6. Support vector classifiers (blue lines), when $\lambda = 1$ and $n = 150$.

select the optimal value for the bandwidth parameter λ by avoiding values that can make a huge impact on the difference between the operators L_{k_λ} , and \mathbb{K}_n . Therefore, in our examples we have avoided the values that are very close to zero or larger than one due to the fact that they can make a significant undesirable change in the performance of support vector classifiers.

A multi-class support vector classifier experiment

In this section, we have so far explained only a binary classification, where the class labels can be only $+1$ or -1 . However, in many cases, real data can have more than two classes. Therefore, it is important to test the impact of the value of the bandwidth parameter on the performance of the exponential kernel when dealing with multiple classes. It is known that the multi-class support vector classifiers are just an extension of the binary classification technique.

As can be seen in the Figs. 7 and 8, the Fisher Iris dataset was used to construct a support vector classifier that separates three species of 150 flowers, which are setosa, versicolor, and virginica with the colors red, blue, and green consecutively¹⁵. We applied the SV classifier using the exponential kernel at two different values for the bandwidth parameter λ , which are 0.29 and 2. At $\lambda = 0.29$, the SV classifier achieved an accuracy rate of 78%, which is very reasonable because the noise is significant, as we can see the interference between the blue and green points in Fig. 7, and the three groups are represented in a 2-dimensional space¹⁶. In Fig. 8, we increased λ to 2, which decreased the accuracy rate to 71%. This is because the large values of λ lead to almost a linear boundary decision that separates the data sets with more errors.

Note that an accuracy rate of more than 95% can be easily achieved when representing the data in a 3-dimensional space, but testing the performance of the exponential kernel when dealing a lot of noise is important. In this example, the results show that small values for the bandwidth parameter λ lead to overfitting, while large values lead to underfitting, which is consistent with our example on the binary classification. In both examples, the optimal values for λ were between 0.2 and 0.7. The relationship between the results of this example and the theoretical results is similar to what we have explained in the binary classification case.

Kernel principal component analysis experiment

Our second choice of a kernel-based method for our experiment is kernel principal component analysis (kernel PCA). This is because kernel PCA is a very useful technique, in which we compute principal components in high-dimensional spaces that are related to input space by kernels, which means reducing the dimensionality of nonlinear data. The exponential kernel is one of the best options for kernel PCA method. Furthermore, different values of the bandwidth parameter λ can show different information about the data. In this section, we present some experiments on kernel PCA using the exponential kernel on two datasets which are Fisher Iris dataset and Breast Cancer Wisconsin dataset. The Fisher Iris dataset has 150 samples from three species, which are setosa, versicolor, and virginica with colors blue, orange, and green consecutively¹⁵. The Breast Cancer Wisconsin dataset contains 569 samples and 30 features, and the samples are labeled as malignant or benign with color blue and red consecutively¹⁷.

First, we computed the first eigenvectors of the exponential kernel matrix in which we projected the original data for visualization. In the case of Iris dataset, the first eigenvectors were computed at three different values for the bandwidth parameter λ , which are 3.33, 5, and 11.11 as can be seen in Figs. 9, 10, and 11 consecutively¹⁶. Clearly, this shows that a small value of λ tells us more about clusters, while a large value tells us more about the locations, and improves the separability of the data. In the case of Breast Cancer Wisconsin dataset, a similar result was observed in Figs. 12, and 13, where the data exhibited a high degree of clustering at $\lambda = 100$, and a high degree of separability at $\lambda = 100000$.

The interpretation of our theoretical results for these observations is that a small value of λ leads to a high clustering, because the data are mapped into a Hilbert space very closely. Meanwhile, a large value of λ shows more variability and improves the separability of the data. This is because the data are embedded into a high-dimensional Hilbert space while preserving the differences among the examples. In the theoretical sense, increasing the value of λ leads to a high closeness between the kernel matrix \mathbb{K}_n , and the operator L_{k_λ} , which is obvious from the bound of their difference in the previous section.

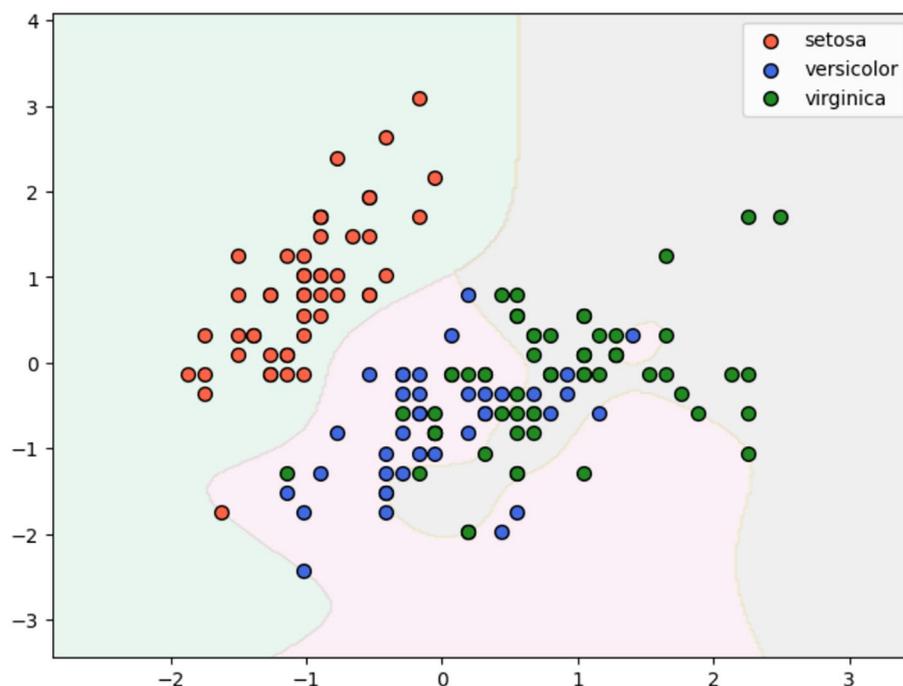


Figure 7. A multi-class support vector classifier (a light line separates the three species) with $\lambda = 0.29$.

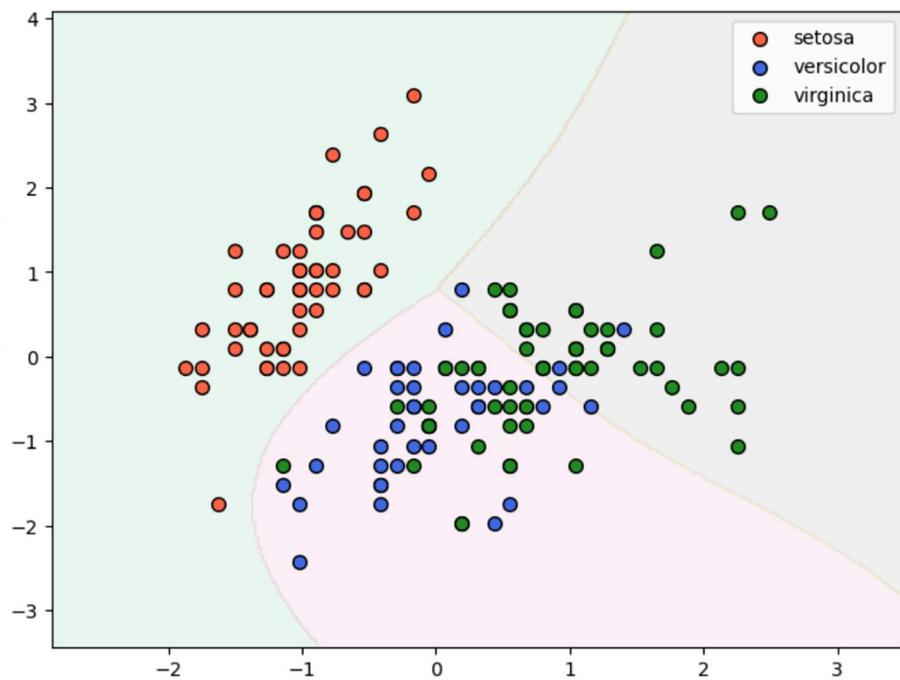


Figure 8. A multi-class support vector classifier (a light line separates the three species) with $\lambda = 2$.

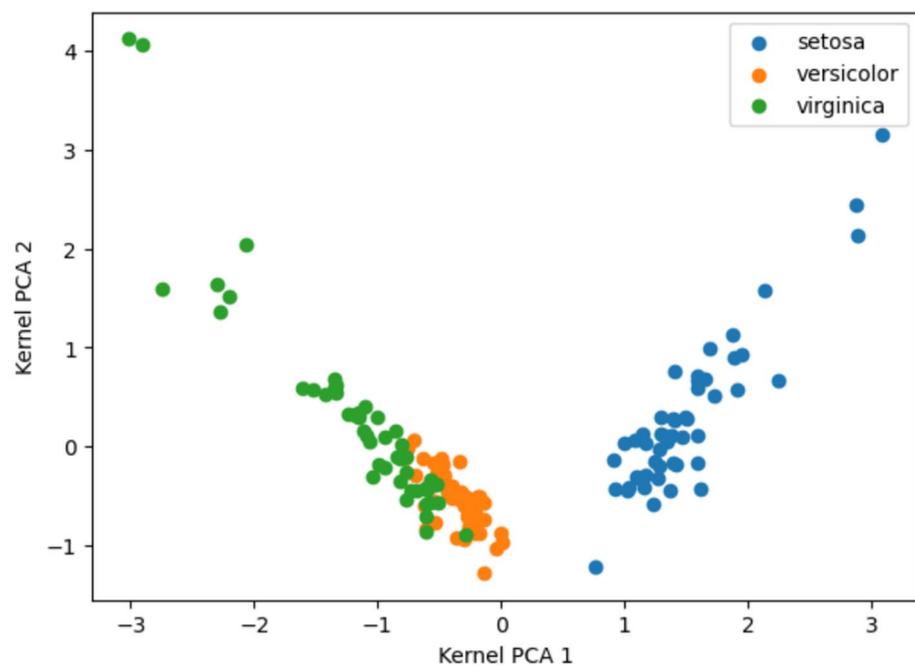


Figure 9. The first two eigenvectors of the exponential kernel matrix with $\lambda = 3.33$.

Conclusion and future scope

In this paper, we investigated the impact of the bandwidth parameter of the exponential kernel on estimating the integral operator L_{k_λ} by its empirical version, which is the kernel matrix \mathbb{K}_n as well as the number of examples from a theoretical perspective. Therefore, we have established a bound that controls the difference between their eigenvalues, and another bound controls the difference between their eigenfunctions. These bounds contain the bandwidth parameter λ , and the number of examples n , which means we can use them to reach the targeted bound for many learning algorithms. In fact, the more examples we use the more closeness between the two operators we obtain, while selecting small values for the bandwidth parameter λ can enlarge the difference between the operators.

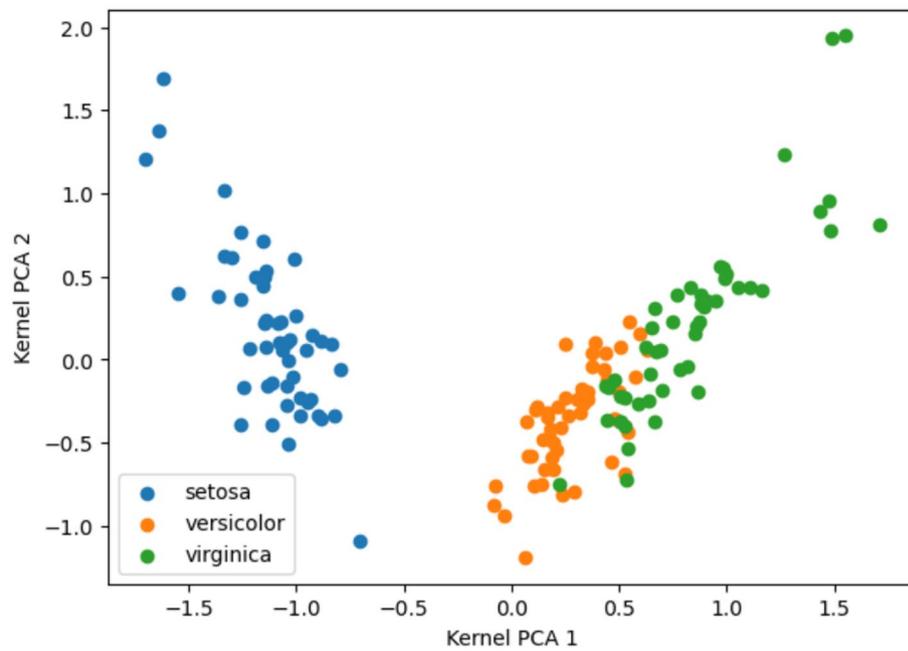


Figure 10. The first two eigenvectors of the exponential kernel matrix with $\lambda = 5$.

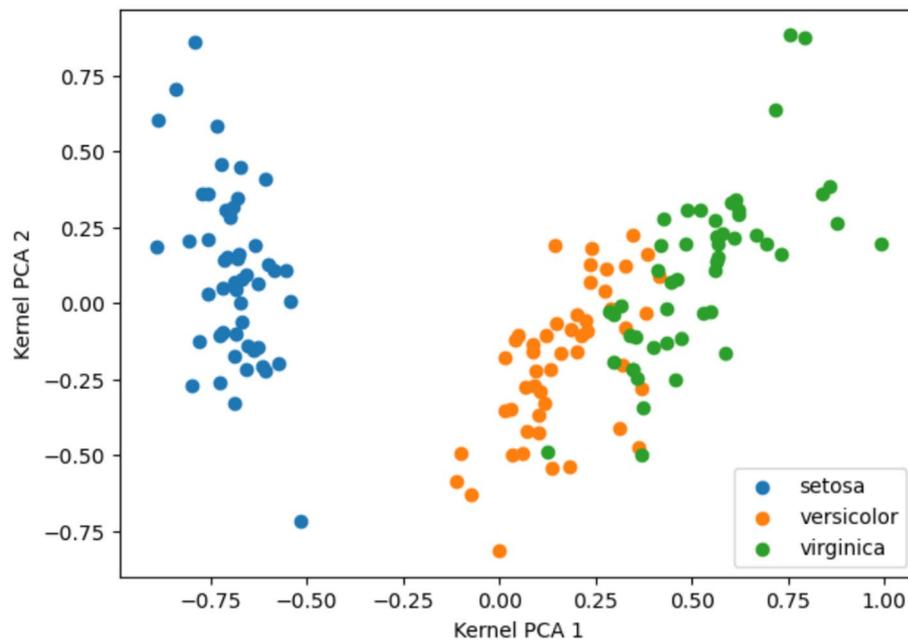


Figure 11. The first two eigenvectors of the exponential kernel matrix with $\lambda = 11.11$.

In our experiments, increasing the number of examples have shown clear improvement in the performance of support vector classifiers. The value of the bandwidth parameter λ would be the value that ensures the bound between the operators L_{k_λ} and \mathbb{K}_n is not too large or too small. In the mathematical sense, the data should be mapped to the perfect locations in a high dimensional space that are suitable for learning. In kernel PCA experiments, we can use small values for the bandwidth parameter λ to learn about clusters, and large values to learn about differences and separability in the data. Minimizing the closeness between the operators \mathbb{K}_n and L_{k_λ} can be obtained by increasing the value of λ , which preserves and highlights the differences, providing a better visualization.

Our future studies will be extended to other kernels such as the rational quadratic kernel and the spherical kernel. These studies will provide a deep understanding of the Hilbert spaces established by such kernels as well

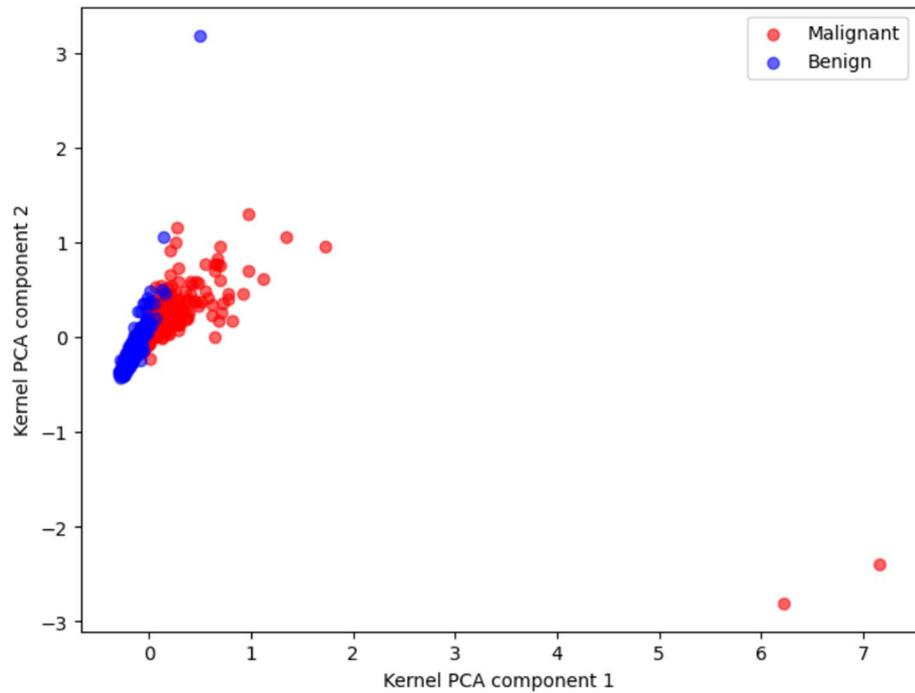


Figure 12. The first two eigenvectors of the exponential kernel matrix with $\lambda = 100$.

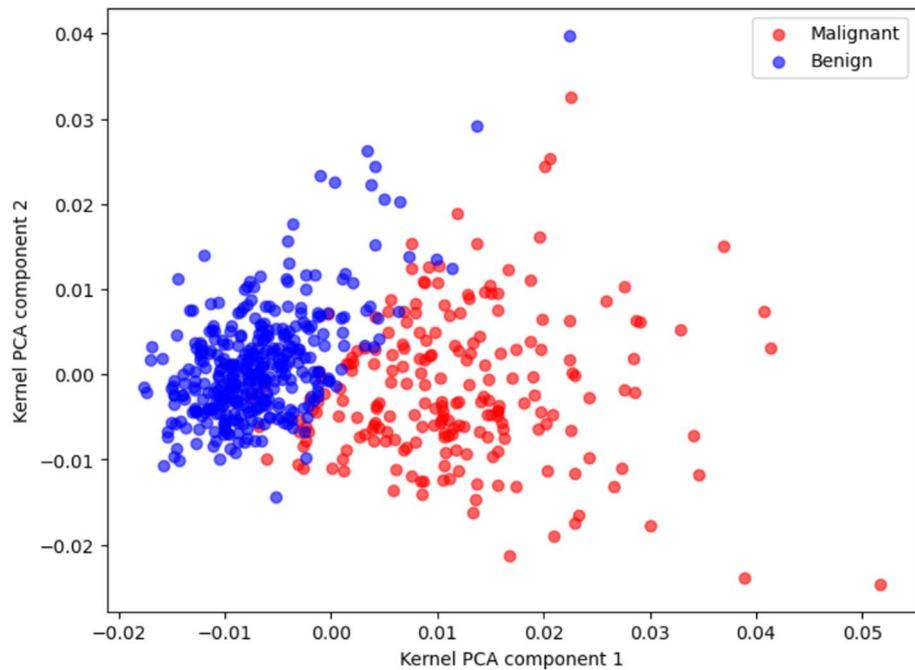


Figure 13. The first two eigenvectors of the exponential kernel matrix with $\lambda = 100000$.

as the impact of their bandwidth parameters on learning algorithms. These techniques are very important to improve the performance of any kernel in any kernel-based method.

Data availability

All the datasets are available online. We have used three datasets one of them is synthetic and two are real as follows : 1. The synthetic dataset generated and/or analysed during the current study are publicly available at <http://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html#buax656-1>. 2. The second dataset is Fisher's Iris dataset. It is available on: https://scikit-learn.org/1.5/auto_examples/datasets/plot_

[iris_dataset.html](#). 3. The third dataset is Breast Cancer Wisconsin dataset. It is available on: https://scikit-learn.org/stable/modules/generated/sklearn.dataset.load_breast_cancer.html.

Received: 15 August 2025; Accepted: 15 October 2025

Published online: 20 November 2025

References

- Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2009).
- Shawe-Taylor, J. & Cristianini, N. *Kernel Methods for Pattern Analysis* (Cambridge University Press, New York, 2004).
- Rosasco, L., Belkin, M. & Vito, E. D. On learning with integral operators. *J. Mach. Res.* **11**, 905–934 (2010).
- Almahdawi, M. A. & Cabrera, O. D. Learning with multiple kernels. *IEEE Access* **12**, 56973–56980 (2024).
- Steinwart, I., Hush, D. & Scovel, C. An explicit description of the reproducing kernel hilbert spaces of gaussian rbf kernels. *IEEE Trans. Inf. Theory* **35**, 4635–4643 (2006).
- Isteinwart, I. & Andreas, C. (eds) *Support Vector Machines* (Springer, New York, 2008).
- Aronszajn, N. Theory of reproducing kernels. *Trans. Am. Math. Soc.* **68**, 337–404 (1950).
- Cucker, F. & Zhou, D. X. *LEARNING THEORY: An Approximation Theory Viewpoint* (Cambridge University Press, New York, 2007).
- Pinelis, I. An approach to inequalities for the distributions of infinite-dimensional martingales. In: Dudley, R.M., Hahn, M.G., Kuelbs, J. (eds.) *Probability in Banach Spaces, 8: Proceedings of the Eighth International Conference*, vol. 30, pp. 128–134. Birkhäuser Boston, Boston, MA (1992).
- Kato, T. Variation of discrete spectra. *Commun. Math. Phys.* **111**, 501–504 (1987).
- Kato, T. *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
- Zwald, L. & Blanchard, G. On the convergence of eigenspaces in kernel principal component analysis. In *Advances in Neural Information Processing Systems* Vol. 18 (eds Weiss, Y. et al.) 1649–1656 (MIT Press, Cambridge, MA, 2006).
- Schölkopf, B. & Smola, J. A. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, 2002).
- The MathWorks, I. Support Vector Machines for Binary Classification, 25.1, (2025). <https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html#buax656-1>.
- Anderson, E. The species problem in irises. *Ann. Missouri Bot.* **23**, 457–509 (1936).
- Pedregosa, F., Varoquaux, G. & Gramfort, A., et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- Wolberg, W., Mangasarian, O., Street, N. & Street, W. Breast Cancer Wisconsin (Diagnostic)[Dataset].UCI Machine Learning Repository, 3.13.1, (1995). <https://doi.org/10.24432/C5HP4Z>.

Author contributions

This work is done by Mahdi A. Almahdawi.

Funding

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (IPP: 995-156-2025). The author, therefore, acknowledges with thanks DSR for technical and financial support.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025