A NEW TRIGONOMETRIC KERNEL FUNCTION FOR SVM

A PREPRINT

Sajad Fathi Hafshejani
Department of Math and Computer Science,
University of Lethbridge,
Lethbridge, AB, Canada
sajad.fathihafejani@uleth.ca

Zahra Moberfard
Department of Applied Mathematics
Shiraz University of Technology
Shiraz, Iran
zahra.moaberfard@gmail.com

October 18, 2022

ABSTRACT

In recent years, several machine learning algorithms have been proposed. Among of them, kernel approaches have been considered as a powerful tool for classification. Using an appropriate kernel function can significantly improve the accuracy of the classification. The main goal of this paper is to introduce a new trigonometric kernel function containing one parameter for the machine learning algorithms. Using simple mathematical tools, several useful properties of the proposed kernel function are presented. We also conduct an empirical evaluation on the kernel-SVM and kernel-SVR methods and demonstrate its strong performance compared to other kernel functions.

Keywords Support vector machine · Kernel-method · Trigonometric kernel function

1 Introduction

The Support Vector Machine (SVM) is a supervised learning algorithm mostly used for classification, but it can be used also for regression. SVM was proposed by Vapnik [Cortes and Vapnik, 1995], and it has been utilized in a wide range of real problems such as bioinformatics [Byvatov and Schneider, 2003], biometrics [Vatsa et al., 2005], power systems [Moulin et al., 2004], and chemoinformatics [Doucet et al., 2007]. In SVM, the training data are used for training and building the classification model. This model is then used to classify unknown samples.

Kernel methods are sets of different types of algorithms that are being used for pattern analysis. They map the data into a vector space and employ linear algebra and geometry to discover structure in the data. There are a couple of reasons to map the data into a feature space. By mapping the original data into a higher-dimensional, it is possible to transform nonlinear relations within the data into linear ones [Tharwat, 2019].

It has been proven that the kernel’s theory is based on the structural risk minimization by using the maximum margin idea [Zhou et al., 2022]. Various machine learning algorithms are working based on kernel function that is a key factor to determine the performance of a the machine learning’s algorithms. Kernels allow mapping the data into high dimensional feature space in order to increase the computational power of linear machines. Thus, it is a way of extending linear hypotheses to nonlinear ones, and this step can be performed implicitly. SVM can be categorized into linear and non-linear approaches [Liang et al., 2020]. Trying to learn a nonlinear separating boundary in the input space increases the computational requirements during the optimization phase because the separating surface will be of at least the second order. Instead, SVM maps the data, using predefined kernel functions, into a new but higher-dimensional space, where a linear separator would be able to discriminate between the different classes. The SVM’s optimization phase will thus entail learning only a linear discriminant surface the mapped space. Of course, selection and settings of the kernel function are crucial for SVM optimality. Indeed, kernel-methods transfer the original data into another space that so-called feature space and revile the hidden features of data [Apsemidis et al., 2020].

Various kernel functions for machine learning algorithms have been introduced and some significant properties of them explored. The Gaussian kernel function is a popular kernel function used in most machine learning algorithms [Hoang and Kang, 2019]. Table 1 demonstrates four common kernel functions.
We use the following notational conventions throughout the paper: $\| \cdot \|$ denotes the Euclidean norm of a vector, the non-negative and positive orthants are denoted by $\mathbb{R}_+^n$ and $\mathbb{R}_+^{n+}$ respectively.

### Table 1: Four common kernel functions

| $i$ | $K(x_i, x_j)$ |
|-----|----------------|
| 1   | $K_1(x_i, x_j) = (1 + x_i x_j)^p$ |
| 2   | $K_2(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$ |
| 3   | $K_3(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|)$ |
| 4   | $K_4(x_i, x_j) = \tanh(\alpha + \beta x_i x_j)$ |

The number of parameters used in a kernel function is one of the basic issues, and it can be considered into two important categories. The first category consists of those kernel functions having one free parameter, and the best value for this parameter can be obtained. The second class consists of those kernel functions that have at least two parameters. Finding the best values for these parameters is very challenging problem. In point of fact, the algorithm needs more time to find the best values for these parameters in the training step. During these years, researchers have focused on introducing the new kernel functions having one parameter. Between the kernel functions presented in Table 1, the function $K_3$ has two free parameters, so the algorithm spends more time to discover optimal values for parameters $\alpha$ and $\beta$. Besides, $K_1$ has one parameter, but the results of performing SVM based on this kernel are not fantastic for some data sets. The Gaussian kernel function has been more widely used. This contains one parameter and has shown satisfactory result in practice and works reasonable for most of the data-sets with different dimensional.

In the past decades, significant effort has been devoted to developing kernel functions for SVM. Combination of kernel functions is one of the interesting concepts in kernel-based methods. It has proven combination of two or more kernel functions is also a kernel function again and new combined kernel function can significantly improve the performance of the algorithm [Feng et al., 2018].

Motivated by the above works, the goal of this paper is to present a new trigonometric kernel function containing one parameter. In this regard, we first focus on examining some concepts and properties of the new kernel function. Then we show that the proposed kernel function is a positive definite function. Finally, we present some numerical results of performing SVM and SVR based on the proposed kernel function and compare these results with other kernel functions.

The paper is organized as: In Section 2, we talk about kernel function and then introduce the new kernel function and investigate some properties of these functions. Some numerical results of performing kernel-SVM kernel SVR on some data-sets with different features are presented in Section 3. We finally end up the paper by providing some concluding remarks.

We use the following notational conventions throughout the paper: $\| \cdot \|$ denotes the Euclidean norm of a vector, the non-negative and positive orthants are denoted by $\mathbb{R}_+^n$ and $\mathbb{R}_+^{n+}$ respectively.

## 2 A new kernel function

This section is devoted to present a new trigonometric kernel function for SVM and investigate various important properties of the new proposed kernel function. We start this section by examining an example. Let’s investigate exactly how a kernel function makes data classification easier. Let $X$ be a random data-set consisting of two circles and 400 samples. It consists of two classes. As mentioned in Fig. 2(left), it is obvious that a linear SVM can not classify the mentioned data-set into two classes. Now, we utilize a kernel function and map data in new space, Fig. 2(right). After mapping, it is obviously the data-set can be classified by using a linear SVM. Our next goal is to prove that the function defined by (3) is indeed a kernel function for SVM. To this end, we first recall the definition of the kernel function demonstrated in [Van Den Berg et al., 2012].

**Definition 2.1.** Let $X \subseteq \mathbb{R}^d$ be a non-empty set. A function $K : X \times X \rightarrow \mathbb{R}$ on $X$ is a kernel function if there exists a $\mathbb{K}$-Hilbert space $H$ and a map $\Phi : X \rightarrow H$ such that:

$$K(x, x') = \langle \Phi(x), \Phi(x') \rangle, \quad \forall \ x, x' \in X.$$  

(1)

where $\Phi$ is a nonlinear (or sometimes linear) map from the input space $X$ to the feature space $F$, and $\langle ., . \rangle$ is an inner product. Besides, the function $\Phi$ is so-called a feature map and $H$ called a feature space of $K$.

**Definition 2.2.** A kernel function is shift-invariant if it has the following form:

$$K(x, x') = K(x - x').$$  

(2)

Now we are in a position where we can introduce a new kernel function for SVM. In this regard, consider the following function:

$$\psi(x) = \sin(h(x)), \quad h(x) = \frac{\pi}{2 + \alpha x^2},$$  

(3)

...
A new trigonometric kernel function for SVM

where $\sigma$ is a positive real number. The behaviour of the function $\psi(x)$ with different values of parameter $\sigma$, $\psi'(x)$, and function $h(x)$ are plotted in Fig 2.

Next lemma presents some properties of the function given by (3).

**Lemma 2.1.** Let $\psi$ be a function defined by (3). Then, we have:

- $\psi(0) = 1$.
- $\psi(x) > 0$ for all $x \in \mathbb{R}$.
- $\psi'(x) \leq 0$, for all $x \in \mathbb{R}_+$.
- $\psi$ is bounded, i.e., for all $x \in \mathbb{R}$, we have $|\psi(x)| \leq \psi(0)$.

**Proof.** Using this fact $\sin(\pi) = 1$, we can conclude that the first item is true.

To prove the second item, we know that $||\cdot||$ is non-negative, so the function $h_1(x) = \frac{\pi}{2 \pi \sigma \|x\|} \in (0, \pi]$ for all $x$, which implies that $\sin(h_1(x)) \in (0, 1]$. To prove the third item, we have:

$$\psi'(x) = h'(x) \cos(h_1(t))$$
It is clear that \( h(x) \) is a decreasing function, so we have \( h'(x) < 0 \). On the other hand, \( \cos(h(x)) \) is a positive function for all \( x \in (0, \frac{\pi}{2}] \), i.e., the first derivative of the function given by (3) is negative.

The fourth item of lemma is obtained by using the fact that \( \psi(0) = 1, \psi(x) \leq 0 \) and \( \psi(x) \geq 0 \) for all \( x \geq 0 \). \( \Box \)

Now, we are in the position to define the new trigonometric kernel function, which has not been introduced so far.

**Definition 2.3.** Suppose \( x, x' \in X \), and the feature map function \( \psi(x) \) given by (3), then the new proposed trigonometric kernel function is defined by:

\[
K(x, x') = \sin\left(\frac{\pi}{2 + \sigma^2\|x - x'\|^2}\right)
\]

where \( \sigma \) is a real positive constant.

Using the inner product symmetric property, the first property of the new proposed kernel function can be expressed as follows:

\[
K(x, x') = \sin\left(\frac{\pi}{2 + \sigma^2\|x - x'\|^2}\right) = \sin\left(\frac{\pi}{2 + \sigma^2\|x' - x\|^2}\right) = K(x', x),
\]

that is, the proposed kernel function has symmetric property. So, we first define the following function:

The next definition recalls another important property of the kernel function called positive definite.

**Definition 2.4.** (Positive definite kernel) [Van Den Berg et al. 2012] Let \( X \) be a nonempty set. A function \( K : X \times X \to \mathbb{R} \) is called a positive kernel function if and only if, for any \( c \in \mathbb{R}^m \), the following inequality is true:

\[
\sum_{i,j=1}^{m} c_i c_j K_{i,j} \geq 0,
\]

in which \( K_{i,j} = K(x_i, x_j) \).

**Remark 2.1.** Suppose that \( K : X \times X \to \mathbb{R} \) is a kernel function. Then \( K \) is so-called strictly positive definite if and only if for any \( c \in \mathbb{R}^m \), we have:

\[
\sum_{i,j=1}^{m} c_i c_j K_{i,j} > 0.
\]

Chasing the mentioned conditions is very challenging for some kernel functions. A common way to address these challenges is to utilize the matrix of the kernel function. [Van Den Berg et al. 2012] proved that a kernel function is positive definite if and only if the symmetric matrix \( K \) is positive definite. Note that, components of matrix \( K_{i,j} \) are \( K(x_i, x_j) \), as well as all \( K_{i,i} = K(x_i, x_i) = 1 \). It shows that the entries on the main diagonal matrix \( K \) are equal to one. Now, we are in the position to elaborate on the positive definite property for the proposed kernel function that delineates some properties of the kernel function in this paper.

**Lemma 2.2** [Van Den Berg et al. 2012]. Suppose \( K : X \times X \to \mathbb{R} \) is a kernel function. Then \( K \) is positive definite if and only if

\[
\det(K(x_1, x_j)^{i,j \leq n}) \geq 0, \quad \{x_1, x_2, \ldots, x_n\} \subseteq X.
\]

**Lemma 2.3.** The new proposed kernel functions defined by (3) is positive definite.

**Proof.** We begin by induction on \( n \). Considering the structure of matrix \( K \) as:

\[
K = \begin{pmatrix}
1 & l_{12} & \cdots & l_{1n} \\
l_{21} & 1 & \cdots & l_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
l_{n1} & l_{n2} & \cdots & 1
\end{pmatrix},
\]

where \( 0 \leq l_{ij} < 1, \ i \neq j, 1 \leq i, j \leq n \), and \( l_{ij} = K(x_i, x_j) \). The property \( K(x_i, x_j) = K(x_j, x_i) \) implies that that \( K \) is a symmetric matrix. For the matrix \( K \), we have:

\[
\begin{align*}
n &= 1, & \det(K_{1,1}) &= 1 > 0; \\
n &= 2, & \det(K_{2,2}) &= 1 - l_{12}^2 > 0,
\end{align*}
\]

in which \( K_{i,i} \) denotes a square matrix with dimension \( i \). Let us suppose that the lemma be true for \( n - 1 \). We prove that \( \det(K_{n,n}) > 0 \). Using the fact that \( K_{11} = 1 > 0 \) and by subtracting \( K_{1,1} \) times the first column from the \( K \)-th column, \( K = 2, \ldots, n \), the new matrix where the first column remained unchanged and other columns are changed by using the
relation \( K'_{jk} = K_{jk} - K_{1k}K_{1j} \) for \( k \geq 2 \) can be obtained. In addition, the new matrix has the same principal minors as the matrix \( K \), so we have:

\[
det(K'_{j,k \leq p}) = det(K_{j,k \leq p}), \quad p = 1, 2, \ldots, n. \tag{10}
\]

So, the matrix \( K' \) can be rewritten as:

\[
K' = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & k'_{22} & \cdots & k'_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
0 & k'_{2n} & \cdots & k'_{nn}
\end{pmatrix}. \tag{11}
\]

Determinant of the matrix \( K \) can be calculated as follows:

\[
det(K) = 1 \ast det \begin{pmatrix}
k'_{2,2} & \cdots & k'_{2,n} \\
\vdots & \ddots & \vdots \\
k'_{2,n} & \cdots & k'_{n,n}
\end{pmatrix} > 0,
\]

where the inequality is obtained from the fact that determinate of matrix \( K_{n-1,n-1} \) is positive. \( \square \)

Lemma 2.3 implies that the proposed kernel functions is positive definite.

### 2.1 Improvement of the proposed kernels function

It has been proven that when two kernel functions are combined, the generated function is indeed a kernel function and inherits all properties of the kernel function [Kung, 2014]. Moreover, the generated kernel function meets better results in terms of accuracy than the original functions. Here, we utilize this idea and suggest a new combination kernel function using Gaussian kernel function as well as the proposed kernel function given by (1). To this end, we first recall a lemma allowing us to combine two or more kernel functions together.

**Lemma 2.4.** Let \( f \) and \( g \) be two kernel functions. Then, a convex combination of \( f \) and \( g \), that is, \( \beta f + (1 - \beta)g \) is also a kernel function for all \( \beta \in [0, 1] \).

Using Lemma 2.4 we introduce an extension of the new proposed kernel function as fellows:

\[
K_c(x, x') := \beta \sin(h_1(x, x')) + (1 - \beta)e^{\frac{-||x-x'||^2}{2\sigma^2}} \tag{12}
\]

where

\[
h_1(x, x') = \frac{\pi}{2 + \sigma^2||x-x'||^2}, \quad \text{and} \quad \beta \in [0, 1]. \tag{13}
\]

### 3 Numerical results

This section is devoted to present some numerical results obtained by performing the kernel-SVM and kernel-SVR on various data sets. We implement the SVM based on the proposed kernel function, Gaussian kernel function, and a convex combination of the proposed kernel function and Gaussian kernel function. As mentioned in the most of the papers, the Gaussian kernel function leads to better algorithm performance in terms of accuracy. In all expreriments, we first divide data into two separate categories of training (80%) and testing (20%). To achieve the best values of parameters \( c \) and \( \sigma \), the values of parameters \( c \) and \( \sigma \) are selected from the set \( \{2^{-5}, 2^{-4}, \ldots, 2^1\} \). After finding the best values of \( c \) and \( \sigma \), we perform the SVM with test data set. Table 3 gives information about 24 datasets used in this section. Moreover, “Name” denotes the name of data-set, “F” presents the number of the features of data set, “N” is applied to denote the number of samples.

The accuracy results of performing SVM based on three mentioned kernel functions are presented in Fig 3. The label of x-axis denotes the data-set number (i.e., \( i \)) in Table 3. “Gua” denotes the results for the Gaussian kernel function, “Sin” and “Sin+Gua” are denoted the results for the new proposed kernel function and kernel function given (12) respectively. In addition, we set \( \beta = 0.5 \) for the kernel function given by (12).
A new trigonometric kernel function for SVM

Moreover, we plot the behaviour of the kernel functions given by (3) and (12) based on different values of parameter $\sigma$ in terms of accuracy in Fig. 4. Fig 4 shows that the proposed kernel function given by (3) achieves the best results when $0 < \sigma \leq 0.5$. On the other hands, the best results for the proposed kernel function defined by (12) occurs when $\sigma = 0.1$. Moreover, when $0 < \sigma \leq 1$, the obtained accuracy is more than 94%.

Our next goal in this section is to investigate efficiency of the proposed kernel functions for the support vector regression (SVR). In this regard, we implement the SVR based on the kernel functions given by (3) and (12) on a set of data points generated by using the following function:

$$f(x) = \sin(x) \exp(-0.2 * x) + \epsilon$$  \hspace{1cm} (14)

where $\epsilon$ is a random number. We generate 200 samples. The Fig. 5 denotes the behaviour of the SVR based on the mentioned kernel functions.
3.1 Results

Based on Table 3 and Fig. 3, 4, and 5, we can conclude that:

- A variety of data-sets with different numbers of samples and features have been used.
- The proposed kernel function improves the accuracy obtained by Gaussian kernel function for the most of data-sets.
- The improved kernel function meets the best results in terms of accuracy for most of the data-sets.
- The values of $\sigma \in (0, 1]$ work well for the SVM method.
- The proposed kernel functions are able to predict the appropriate graph that

4 Concluding remarks

In this paper, we suggested a new trigonometric kernel function for the machine learning. Various properties of the proposed kernel function were investigated, for example, we proved that the new kernel function is a positive kernel function. Then we combined the proposed kernel function with the Gaussian kernel function and introduced a new kernel function. The numerical results confirmed that the new proposed trigonometric kernel function improve the accuracy of the SVM for the most of considered data-set. Moreover, the combined kernel function achieved the best results in terms of accuracy.
Declarations

Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their constructive comments.

Funding

No funding was received for conducting this study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The dataset used in this paper is available from the following link https://paperswithcode.com/dataset/orl

Conflict of Interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Competing interests

The authors have declared that no competing interests exist.

References

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Evgeny Byvatov and Gisbert Schneider. Support vector machine applications in bioinformatics. *Applied bioinformatics*, 2(2):67–77, 2003.

Mayank Vatsa, Richa Singh, and Afzel Noore. Improving biometric recognition accuracy and robustness using dwt and svm watermarking. *IEICE Electronics Express*, 2(12):362–367, 2005.

LS Moulin, AP Alves Da Silva, MA El-Sharkawi, and Robert J Marks. Support vector machines for transient stability analysis of large-scale power systems. *IEEE Transactions on Power Systems*, 19(2):818–825, 2004.

Jean-Pierre Doucet, Florent Barbault, Hairong Xia, Annick Panaye, and Botao Fan. Nonlinear svm approaches to qspr/qsar studies and drug design. *Current Computer-Aided Drug Design*, 3(4):263–289, 2007.

Alaa Tharwat. Parameter investigation of support vector machine classifier with kernel functions. *Knowledge and Information Systems*, 61(3):1269–1302, 2019.

Jingyue Zhou, Ye Tian, Jian Luo, and Qianru Zhai. Novel non-kernel quadratic surface support vector machines based on optimal margin distribution. *Soft Computing*, 26(18):9215–9227, 2022.

Zhenhu Liang, Shuai Shao, Zhe Lv, Duan Li, Jamie W Sleigh, Xiaoli Li, Chongyang Zhang, and Jianghong He. Constructing a consciousness meter based on the combination of non-linear measurements and genetic algorithm-based support vector machine. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(2):399–408, 2020.

Anastasios Apseimidis, Stelios Psarakis, and Javier M Moguerza. A review of machine learning kernel methods in statistical process monitoring. *Computers & Industrial Engineering*, 142:106376, 2020.
Duy Tang Hoang and Hee Jun Kang. A motor current signal-based bearing fault diagnosis using deep learning and information fusion. *IEEE Transactions on Instrumentation and Measurement*, 69(6):3325–3333, 2019.

Xinxin Feng, Xianyao Ling, Haifeng Zheng, Zhonghui Chen, and Yiwen Xu. Adaptive multi-kernel svm with spatial–temporal correlation for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(6):2001–2013, 2018.

C Van Den Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic analysis on semigroups: theory of positive definite and related functions*, volume 100. Springer Science & Business Media, 2012.

Sun Yuan Kung. *Kernel methods and machine learning*. Cambridge University Press, 2014.