Chapter 5

Support Vector Machine with Novel Indefinite Kernel

Contents

Various Euclidean and non Euclidean distance based kernels and their semi definite properties are discussed in this chapter. Some of the kernels are indefinite kernels defined in Kreĭn space which is pseudo Euclidean (pE) space. Classification results are obtained by using various distance based kernels and also by a novel modified Gaussian Kernel. With Support Vector Machine (SVM) as classifier, these kernels are used to diagnose regular skin disorders of Dataset-I and Dataset-II.
Section 5.1, discusses the general introduction of indefinite kernels and various techniques to deal with them. Mathematical formulation of Support Vector Machine with Indefinite Kernels is discussed in section 5.2. In section 5.3, various distance based kernels and modified Gaussian kernel are defined, their semi definite properties are studied using Gram matrix for Dataset-I and Dataset-II. Classification results obtained using F-score are also discussed. Chapter ends with a summary in section 5.4.

5.1 Introduction

In classification problem efforts are made to discover structure in the data. Embedding the data in feature space using kernel functions and with regularization terms, the complexity of the model can be controlled. So, kernels play a vital role in the classification. They measures the similarities among training data for classification and have wide range of applications in many classification problems. Positive (semi) definite kernels are prevailing in machine learning algorithms. These kernels are the traditional requirement of classifier such as Support Vector Machines(SVMs) to achieve global optimum.To get good classification accuracy, sometimes a new kernel required to be determined which may not be positive definite kernel, but it is indefinite kernel. [95].

Distance between two samples can be used as dissimilarity measures. Many researchers have used distance based kernels, some of which are non Mercer's kernels, i.e. they are neither positive definite nor negative definite, though they are giving good classification accuracy. These kernels are known as indefinite kernels. Training a kernel based classifier using indefinite kernels is an active field of research in machine learning. Haasdonk *et.al.* used tangent distance measure in distance based kernels using prior knowledge of transformation invariance into SVM [46]. They have applied their approach to US-Postal-Service digit dataset and obtained comparable better classification accuracy. Cortes *et. al.* used rational kernels, some of which are non Mercer's kernels [23]. Suicheng *et. al.* have used kernel principal component analysis for spectrum modification and then used SVM as classifier [43]. Canu *et. al.* obtained spline solution using indefinite kernels [15]. Luss *et. al.* penalized the distance between proxy kernels and in their approach, learning of original indefinite kernel matrix and support vectors weight learning carried out simultaneously [83].

Techniques to deal with Indefinite Kernel

For indefinite kernels, the optimization problem of SVM becomes non convex and classifier may terminate at local optimum. To make the optimization problem convex, some researchers used spectrum modification techniques viz., CLIP, SHIFT, FLIP, SQUARE etc ([20], [83], [89], [103], [112], [140]). Using these techniques indefinite matrix is converted into definite matrix, which makes the corresponding optimization problem convex. This problem can be solved by regular solver like SMO and global optimum can be obtained.

Following are different approaches for spectrum modification:

- CLIP: In this method, negative eigen values of indefinite kernel matrix are considered as the noise of positive definite matrix and so they are removed. The new kernel matrix thus obtained becomes positive definite and makes the optimization problem convex.
- SHIFT: For small negative eigen values, spectrum can be modified with SHIFT method, in which spectrum is shifted until non negative eigen values are obtained.
- FLIP: In this approach, to remove negative eigen values, absolute value of the eigen values are used.
- SQUARE: In this technique, to remove negative eigenvalues, square of the eigen values are used.

The main drawback of these techniques are, the original kernel change. By removing the negative eigen values, sometimes the relevant information of the Dataset might be lost. Laub *et. al.* discussed that negative eigenvalues also contain some information about the structure in the data [73]. Loosli *et. al.* in have done decomposition of kernel matrix and claimed that the negative part of the kernel can improve the SVM performance [81].

In another approach to deal with indefinite kernels, instead of modifying spectrum, the classifier directly deals with indefinite kernels. For these types of kernels the feature space is not a Hilbert space but it is a pseudo Euclidean space where inner product may not be positive definite. In this approach instead of minimizing the objective function, the emphasis is given to stabilize the optimum value. Some researchers are using the approach in which indefinite kernel matrix is directly considered by the solver without any modification of the spectrum. The resulting optimization problem becomes non convex, which can be solved by solver like LIBSVM.

Lin & Lin have used SMO type decomposition method to solve non convex dual problem for indefinite kernels [78]. Haasdonk has solved the non convex optimization problem of indefinite kernels by minimizing distance between two convex hulls [47]. Loosli *et. al.* have discussed two methods to stabilize the objective function for indefinite kernel. One method is eigen decomposition using SVM (ESVM) in which exact solution is obtained but computational cost of pre-calculation of the whole kernel matrix is high. In the other method instead of eigen value decomposition, emphasis is given on stabilization of optimization problem [81]. In this method, instead of Hilbert space, the feature space is Kreĭn space. This method reduces computational cost of pre-computing kernel matrix but less efficient than ESVM, as in ESVM exact solution is obtained, while in this approach approximate solution is obtained.

5.2 Kreĭn Space: A pseudo-Euclidean(pE) Space

When dataset contains non linearity, it is necessary to use kernels to classify data by linear classifier. If these kernels are indefinite, then instead of Hilbert Space, they are defined in pseudo-Euclidean(pE) space. The space is endowed with a Hilbert topology, called Kreĭn space. Kreĭn space is an indefinite inner product space.

Let $\mathbb{R}^{(p,q)}$, be the pE space with signature (p, q), where $p, q \in \mathbb{N}_0$ with $p + q =$ n, n is the dimension of the pE space. The inner product in pE space is not positive definite but indefinite and defined as follows:

Definition 5.2.1. (Indefinite inner product) [47]
For $\mathbf{x} = (\mathbf{x}_p^T, \mathbf{x}_q^T) \in \mathbb{R}^{(p,q)}$, the inner product is defined as, $\langle \mathbf{x}, \mathbf{y} \rangle_{pE} = \mathbf{x}_p^T \mathbf{y}_p - \mathbf{x}_q^T \mathbf{y}_q$ $= \mathbf{x}^T P \mathbf{y}$, where P = diag$(\mathbf{1}_p, -\mathbf{1}_q)$. So, indefinite inner product is the difference of two standard inner products.

Definition 5.2.2. (pseudo-Euclidean Space)[67]
The pseudo-Euclidean Space is a finite dimensional(n-dimension) real coordinate vector space with an indefinite non-degenerate quadratic form (definition 2.1.7), where inner product is indefinite(definition 5.2.1).

Some basic concepts about Pseudo Euclidean Space.

Following points are discussed in the Encyclopedia of Mathematics [124]:

(a) The norm in pE space is not induced by inner product.

(b) The modulus, $|\mathbf{a}|$ of a vector $\mathbf{a} \in \mathbb{R}^{(p,q)}$ in pE space is a non negative square root $\sqrt{|\langle \mathbf{a}, \mathbf{a} \rangle|}$, where $\langle \mathbf{a}, \mathbf{a} \rangle_{pE}$ defined as (definition 5.2.1), which may be positive or negative.

(c) The number of independent vectors with $|\langle \mathbf{a}, \mathbf{a} \rangle| > 0$ is equal to p and with $|\langle \mathbf{a}, \mathbf{a} \rangle| < 0$ is equal to q. i.e. $p + q =$ n, where n is the dimension of the vector.
The quadratic form in pE space is given by

$$q(x) = (x_1^2 + x_2^2 + ... + x_p^2) - (x_{p+1}^2 + x_{p+2}^2 + ... + x_{p+q}^2)$$

(d) The distance between two points A($\mathbf{x}$) and B($\mathbf{y}$) in pE space is taken to be the modulus of the vector $\overline{AB}$ and is computed as follows:

$$|\overline{AB}|^2 = |\mathbf{y} - \mathbf{x}|^2 = |\langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x} \rangle| = \|\mathbf{x}_p - \mathbf{y}_p\|^2 - \|\mathbf{x_q} - \mathbf{y}_q\|^2.$$

Some Observations:

1. The square distance between two points in pE space is the difference of the square distance in the real and imaginary directions, which can be negative. So, real square root can not be determined.
2. In pE space there are three types of straight lines: Euclidean with $|\langle \mathbf{a}, \mathbf{a} \rangle| > 0$, pseudo-Euclidean $|\langle \mathbf{a}, \mathbf{a} \rangle| < 0$ and isotropic for which $|\langle \mathbf{a}, \mathbf{a} \rangle| = 0$.
3. There are non zero points such that, $\langle x, x \rangle = 0$. i.e. there are non-zero points which are orthogonal to themselves. These are called isotropic points and they form isotropic cone. Isotropic cone separates two regions: positive squared norm and negative squared norm. (figure 5.1).

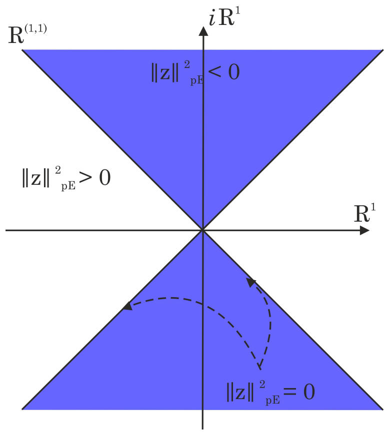

Figure 5.1: Divisions of pE spaces by isotropic cones

(e) pE space is not a metric space as the triangular inequality $|a + b| \leq |a| + |b|$ is not satisfied.

(f)
1. pE is generalised Euclidean space. In particular, pE space with signature (n,0) is Euclidean space.
2. Minkowski space is an example of pE space. It is a space $\mathbb{R}^{(1,3)}$ with $q(x_0, x_1, x_2, x_3) = x_0^2 - x_1^2 - x_2^2 - x_3^2$, where in physics x_0 is ct- coordinate, where c is the speed of light and t is time.
3. The plane $z = x + yi$, consisting of complex numbers equipped with the quadratic form $zz^* = x^2 - y^2$.

4. Kreĭn Space which is indefinite inner product space is also pE space.

Definition 5.2.3. (Kreĭn space)[95]
Kreĭn space $\mathcal{K}$ is an indefinite inner product space spanned by two Hilbert spaces H_+ and H_- such that

1. $\forall f \in \mathcal{K}, f = f_+ - f_-$, where $f_+ \in H_+$ and $f_- \in H_-$
2. $\forall f, g \in \mathcal{K}, \langle f, g \rangle = \langle f_+, g_+ \rangle - \langle f_-, g_- \rangle$

Definition 5.2.4. (Indefinite Kernel in Kreĭn Space)[47]
Let χ be a non empty subset of $\mathbb{R}^n$ and $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\} \in \chi$ be a finite set of m data points. Function $k : \chi \times \chi \rightarrow \mathcal{R}$ is called a Indefinite kernel function, if for some $\mathbf{v} \in \chi$, $\mathbf{v}^T K \mathbf{v} > 0$ and for some $\mathbf{v} \in \chi$, $\mathbf{v}^T K \mathbf{v} < 0$, where $K = k(x_i, x_j)$, $i, j = 1, 2, ..., m$.

5.3 Classification with Indefinite Kernel

In this section the mathematical formulation of classification problem with Indefinite kernel in pE space as well as the Support Vector Machine (SVM) with indefinite kernels is discussed.

When indefinite kernel is used in SVM, the corresponding optimization problem of SVM becomes non convex. Haasdonk has interpreted SVM as optimal hyper plane classifier by minimizing the distance between two convex hulls in pseudo Euclidean Space instead of by maximizing margin [47]. He suggested that for appropriate choice of parameters (low value of regularization term) global optimum can be achieved. In their formulation, they have included only distances and avoided the kernel functions, which induces the distance. Instead of finding global minimum, the emphasis is given on stabilisation. Finding the minimum distance between two reduced convex hulls is equivalent to find maximum margin in SVM [13]. Schölkopf *et. al.* in 2001 have also discussed that maximum margin hyperplane is equivalent to bisecting the shortest line orthogonally connecting to two convex hulls [118]. So, the optimization problem is to minimize the distance between two soft(reduced) convex hulls in such a way that few outliers do not dominate the solution.

Let the two convex hulls for both classes of training data are

$$\mathbf{x}_+ = \sum_{i,y_i=+1} \alpha_i \phi(x_i), \quad \sum_{y_i=1} \alpha_i = 1, \quad \alpha_i \geq 0 \tag{5.3.1}$$

$$\mathbf{x}_- = \sum_{j,y_j=-1} \beta_j \phi(x_j), \quad \sum_{y_j=-1} \beta_j = 1, \quad \beta_j \geq 0 \tag{5.3.2}$$

5.3.1 Optimal Separation for separable data in pE space

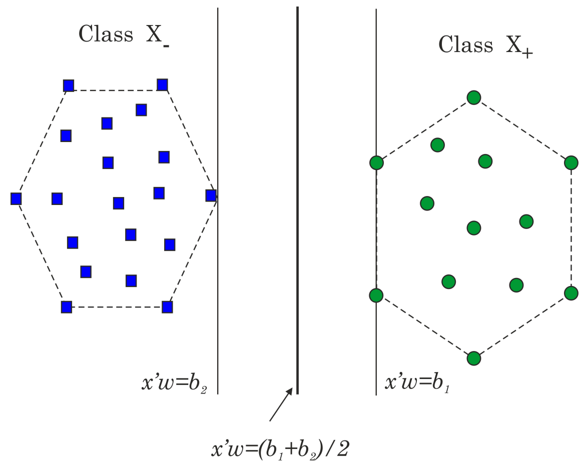

Figure 5.2: Convex Hulls for Separable Data

Let $\mathbf{x}_+$ and $\mathbf{x}_-$ are two convex hulls of two separable datasets. Let $\mathbf{x'w} = \mathbf{b_1}$ and $\mathbf{x'w} = \mathbf{b_2}$ are the two separating planes as shown in the figure (5.2). The distance between two convex hulls (parallel supporting hyperplanes) is given by $\mathbf{x'w} = \frac{\mathbf{b_1 - b_2}}{\|\mathbf{w}\|}$ which can be maximized by minimizing the normal vector(weight vector) $\|\mathbf{w}\|_{\mathbf{pE}}$ and maximizing $(b_1 - b_2)$ [13].
Setting $b_1 - b_2 = 2$, the optimization problem becomes,

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|_{pE}^2 = \mathbf{w}^T P \mathbf{w} \tag{5.3.3}$$

where,

$$\mathbf{w} := \mathbf{x}^+ - \mathbf{x}^- \quad \text{and} \quad P := \text{diag}(\mathbf{1}_p, -\mathbf{1}_q),$$

($p, q \in \mathbb{N}_0$ with $p + q =$ n , n is the dimension of the pE space.)

subject to the constraints,

$$\sum_{i,y_i=1} \alpha_i = 1, \quad \alpha_i \geq 0$$
$$\sum_{j,y_j=-1} \beta_j = 1, \quad \beta_j \geq 0$$

The dual of the optimization problem (5.3.3) is [47]:

$$\max_{\alpha_i,\beta_j} \frac{1}{2} \sum_{i,j} \alpha_i \, \beta_j \, y_i \, y_j \, \langle x_i, x_j \rangle \tag{5.3.4}$$

subject to the constraints,

$$\sum_{i,y_i=+1} \alpha_i y_i = 0, \quad \sum_{j,y_j=-1} \beta_j y_j = 0$$

$$\sum_{i=1}^{m} (\alpha_i + \beta_i) = 2$$
$$0 \leq \alpha_i \leq 1, \quad 0 \leq \beta_j \leq 1.$$

5.3.2 Optimal Separation for Non-Separable data in pE space

For non-separable data, the convex hulls of two different classes will intersect (see figure 5.3). So, using linear classifier it is not possible to separate the two data points.

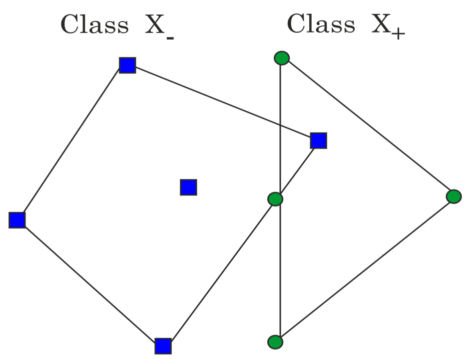

Figure 5.3: Convex Hulls for Non-Separable Data

The effect of outliers can be reduced using reduced convex hulls (refer definition 2.2, figure 5.4) in which an upper bound is kept on the multipliers, in the convex combination of each point of the convex hulls [13].

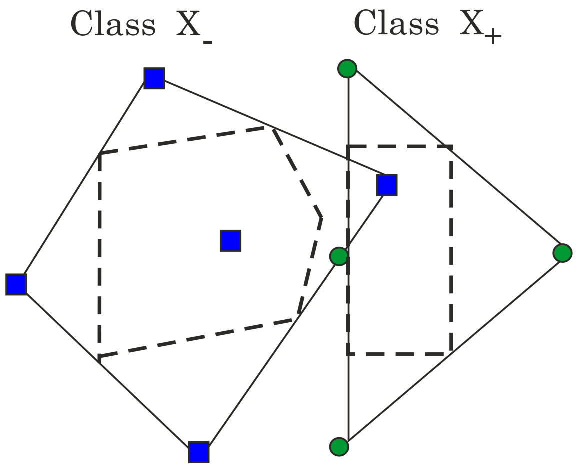

Figure 5.4: Reduced Convex Hulls for Non-Separable Data

The dual of the optimization problem (5.3.3) is:

$$\max_{\alpha_i,\beta_j} \frac{1}{2}\sum_{i,j} \alpha_i\, \beta_j\, y_i\, y_j\, k(x_i, x_j) \tag{5.3.5}$$

subject to the constraints,

$$\begin{aligned} \sum_{i,y_i=1} \alpha_i &= 1, \quad 0 \le \alpha_i \le \mu_1 \\ \sum_{j,y_j=-1} \beta_j &= 1, \quad 0 \le \beta_j \le \mu_2 \end{aligned}$$

where, $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$ and $\phi : \{x_i\}_{i=1}^{n} \to \mathbb{R}^{(p,q)}$ be some nonlinear function.

Training of dual convex hull optimization problem (5.3.3) is performed for different values of μ_1 and μ_2. If classes are not skewed then $\mu_1 = \mu_2 = \mu$ can be considered. The optimization problem (5.3.3) is quadratic but it may not be convex.

The optimal classifier is given by, $\text{sign}(g(\mathbf{x}))$ where,

$$g(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_-\|^2_{pE} - \|\mathbf{x} - \mathbf{x}_+\|^2_{pE}$$

Optimal hyper plane is the minimum distance classifier with respect to closet point of convex hulls.

5.3.3 SVM with Indefinite Kernel

For noisy data, adding soft margin error term the SVM primal problem in pE for non separable sample points is given by:

$$\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}\mathbf{w}^T P\mathbf{w} + C\sum_i \xi_i \tag{5.3.6}$$

subject to the Constraints,

$$y_i \left(\mathbf{w}^T P\phi(x_i) + b\right) \geqslant 1 - \xi_i, \quad \xi_i \geq 0$$

where, $C > 0$ is the regularization parameter.

Optimization problem given in (5.3.6) is not convex. By solving it, stationary point $\mathbf{w} \in \mathbb{R}^{(p,q)}$ and $b \in \mathbb{R}$ and $\xi \in \mathbb{R}^n_+$ can be obtained, which are given by [47],

$$\mathbf{w} = \sum_i \alpha_i y_i \phi(x_i)$$

$$b := -\frac{1}{2}\mathbf{w}^T P\mathbf{w} + \phi(x_i)$$

and ξ_i is given by,

$$\xi_i = \begin{cases} 1 - y_i \left(\mathbf{w}^T P\phi(x_i) + b\right), & \text{if } \alpha_i = C \\ 0, & \text{otherwise.} \end{cases}$$

Some Remarks:

Following points were discussed in [47].

(a) If $\alpha_i < C$ and $\alpha_i = C$, $\xi_i < 1$, then data are correctly classified by SVM. But, if $\alpha_i = C$ and $\xi_i > 1$, then classification is wrong.

(b) When $\mathbf{w}^T P\mathbf{w} > 0$, then points are correctly classified. But, when $\mathbf{w}^T P\mathbf{w} < 0$, the points have negative squared distance from the points of other convex hull. This violets the nice property of distance that lower distance then high similarity. So, by theoretical aspect and geometrical interpretation, $\mathbf{w}^T P\mathbf{w} > 0$ is the requirement for relevant solution. So, $\mathbf{w}^T P\mathbf{w}$ should be positive. Therefore, when kernels are non-CPD, linear separability is not enough but positive norm is also required, which is obtained by lower value of the regularization parameter C.

(c) We can obtain only stationary point in case of indefinite kernels. The stationary point is global optimum if, there is a positive square distance between two convex hulls. Also the point have positive square distance to all points of their corresponding convex hulls.

(d) When $\mathbf{w}^T P\mathbf{w} < 0$, $\mathbf{w}^T P\mathbf{w}$ can diverges to $-\infty$, so margin maximization is not right interpretation for indefinite kernel, instead optimal separation of convex hulls is better.

(e) For $P = I_n$, optimization problem (5.3.6) is equivalent to standard SVM primal problem.

5.4 Classification using Various Kernels

5.4.1 Distance Based Kernels

Standard Radial Basis kernel function and Polynomial kernel function are defined as follows:

- Radial Basis kernel function (RBF): $exp(-\gamma\|\mathbf{x}-\mathbf{y}\|^2)$,
  where parameter γ is the radius of influence of support vectors.
- Polynomial kernel function: $(\alpha\mathbf{x}^T\mathbf{y}+a_0)^d$,
  where d is the degree of the polynomial.

Euclidean distance is considered to be a powerful tool for similarity measure, it is not the only solution of all types of data or pattern to be compared [16]. In this study the Euclidean distance in RBF kernel is replaced with many other distances where, some distances are not satisfying metric property so they are termed as pseudo metric. Also, the standard inner product in polynomial kernel is replaced by different inner products, some of which are not positive definite. In this study Euclidean distance/standard inner product are replaced by some of the distances/inner products discussed in [16]. Their effects on classification results are also studied. Here, we have proposed a new kernel function which is Modified Gaussian kernel [102]. We have studied the definiteness property of each of these kernels by finding Eigen values of Gram matrix(definition 2.1.9) of each kernel where, Some of these kernels are indefinite(definition 5.2.4). .

We have carried out empirical verification by using two different Datasets: Dataset-I and Dataset-II (Appendix-A). For training-testing purpose the datasets are divided into 70%-30% random data partitions. Parameters of the kernels are set using grid search method (2.2.2) and LIBSVM [76] is used to solve optimization problem. LIBSVM gives convergence not only for convex optimization problem but also converge to a stationary point for indefinite kernels of non convex optimization problem [47].

Table 5.1: Performance of SVM using Various Distance Kernels

No	Distance Type	Definition	Gram matrix	Classification Accuracy Data set I	Classification Accuracy Data set II
		Various Distances in RBF Kernel			
		Distances from Minkowski Family			
1.	City Block L_1	$\sum_{i=1}^{m} \lvert x_i - y_i \rvert$	Positive definite	90.78%	97.26%
2.	Euclidean L_2	$\sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}$	Positive definite	90.07%	97.26%
3.	Minkowski $L_p, p = 3, 4, 5$	$\sqrt[p]{\sum_{i=1}^{m} \lvert x_i - y_i \rvert^p}$	Indefinite	89.36%	95.89%
		Distances from L_1 Family			
4.	Sorensen	$\frac{\sum_{i=1}^{m} \lvert x_i - y_i \rvert}{\sum_{i=1}^{m} (x_i + y_i)}$	Positive definite	89.36%	98.63%
5.	Gower	$\frac{1}{m} \sum_{i=1}^{m} \lvert x_i - y_i \rvert$	Positive definite	90.78%	98.63%
6.	Kulczynski	$\frac{\sum_{i=1}^{m} \lvert x_i - y_i \rvert}{\sum_{i=1}^{m} min(x_i + y_i)}$	Indefinite	90.07%	98.63%
7.	Canberra	$\sum_{i=1}^{m} \frac{\lvert x_i - y_i \rvert}{x_i + y_i + \varepsilon}$	Positive definite	90.78%	98.63%
8.	Lorentzian	$\sum_{i=1}^{m} \frac{\log(1 + \lvert x_i - y_i \rvert)}{1}$	Positive def.	90.78%	97.26%
		Distances from Intersection Family			
9.	Wave Hedges	$\sum_{i=1}^{m} \frac{\lvert x_i - y_i \rvert}{max(x_i, y_i)}$	Positive definite	90.78%	97.26%
10.	Intersection	$\frac{1}{2} \sum_{i=1}^{m} \lvert x_i - y_i \rvert$	Positive definite	90.78%	97.26%

11.	Tanimoto	$\frac{\sum_{i=1}^{m}(\max(x_i,y_i)-\min(x_i,y_i))}{\sum_{i=1}^{m}(max(x_i,y_i))}$	Pos. def.	90.78%	97.26%
Distances from Chi-squares family					
12.	Squared Euclidean	$\sum_{i=1}^{m}(x_i-y_i)^2$	Positive definite	90.78%	97.26%
13.	Squared χ^2	$\sum_{i=1}^{m}\frac{(x_i-y_i)^2}{2*(x_i+y_i)+\epsilon}$	Positive definite	90.78%	97.26%
Various Inner Product in Polynomial kernel					
Inner Product family					
14.	Standard Inner Product	$\sum_{i=1}^{m}x_iy_i$	Positive definite	90.78%	97.26%
15.	Cosine	$\frac{\sum_{i=1}^{m}x_iy_i}{\sqrt{\sum_{i=1}^{m}x_i^2}\sqrt{\sum_{i=1}^{m}y_i^2}}$	Positive definite	90.07%	89.82%
16.	Kumar-Hassebrool (PCE)	$\frac{\sum_{i=1}^{m}x_iy_i}{\sqrt{\sum_{i=1}^{m}x_i^2}+\sqrt{\sum_{i=1}^{m}y_i^2}-\sum_{i=1}^{m}x_iy_i}$	Positive definite	90.78%	97.26%

5.4.2 A Novel Modified Gaussian Kernel

A proposed novel kernel is defined as:

$$K(\mathbf{x},\mathbf{y}) = \left(\sqrt{\|\mathbf{x}-\mathbf{y}\|^2+\epsilon}\right)\left(e^{-\gamma*\|\mathbf{x}-\mathbf{y}\|^2}\right) \tag{5.4.1}$$

This kernel is indefinite and using SVM 91.50% classification accuracy for Dataset-I and 98.63% accuracy for Dataset-II is obtained. For Dataset-I, in which total 329 training samples, 199 positive and 130 negative Eigen values for the Gram matrix of modified kernel function is obtained and for Dataset-II, total 293 training samples, 72 positive Eigen values, 221 negative Eigen values for the Gram matrix of the proposed indefinite kernel function are obtained [102]. Figure (5.5) and Figure (5.6) show the classification accuracy using F-scores for distance substitution kernels and Modified Gaussian Kernel for Dataset-I and Dataset-II respectively.

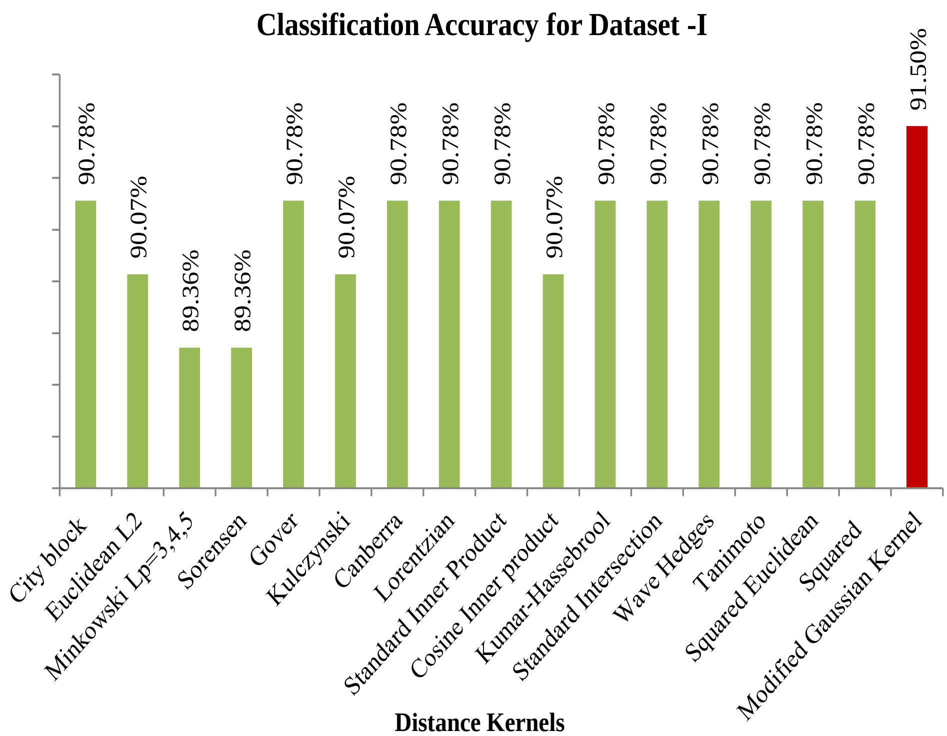

Figure 5.5: Graph of Classification Accuracy for Dataset-I

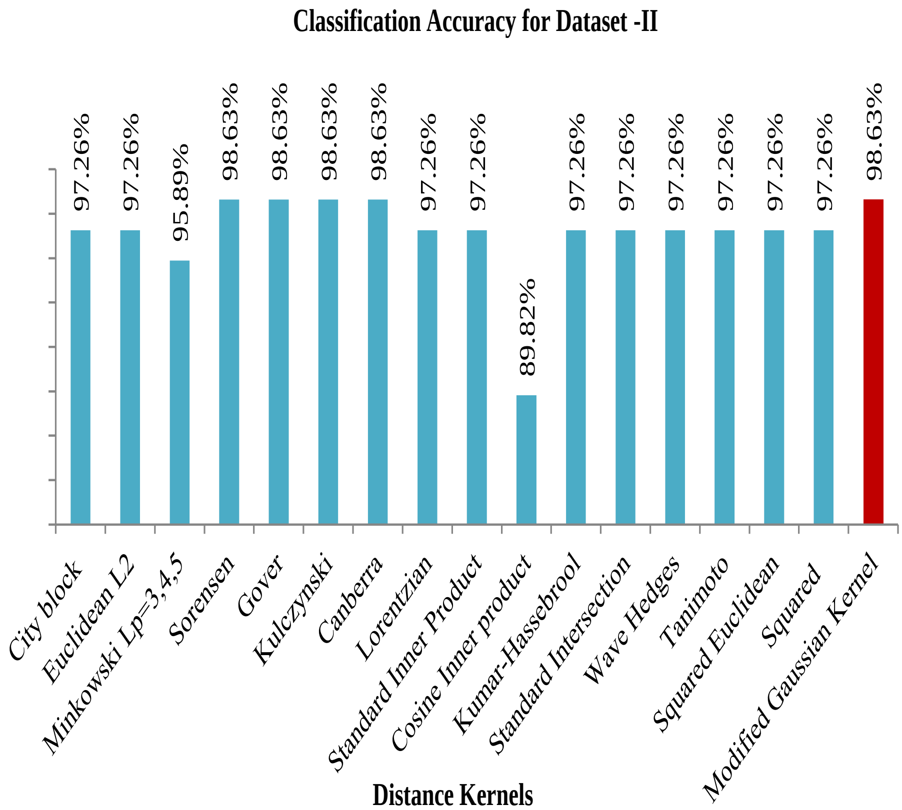

Figure 5.6: Graph of Classification Accuracy for Dataset-II.

5.5 Summary

In this chapter a novel modified indefinite Gaussian Kernel is introduced and classification of various dermatological skin disorders are carried out using Support Vector Machine. Various types of distances and inner products are substituted in lieu of standard Euclidean distance and inner product in Radial basis kernel function and Polynomial kernel function respectively. Classification accuracy is obtained using these distance substitution kernels and for newly developed kernel. Definiteness of these various kernels are computed using Gram matrix. It is observed that the novel Gaussian kernel exhibits the highest classification accuracy. It is an indefinite kernel, but it has robust mathematical theory in Kreĭn space with better F-scores and thus it is acceptable for the Datasets under study.