THE SEPARABILITY THEORY OF HYPERBOLIC TANGENT KERNELS AND SUPPORT VECTOR MACHINES FOR PATTERN CLASSIFICATION

Mathini Sellathurai and Simon Haykin

McMaster University
Hamilton, Ontario, L8S 4K1, Canada.

ABSTRACT

In this paper, a new theory is developed for the feature spaces of hyperbolic tangent used as an activation kernel for non-linear support vector machines. The theory developed herein is based on the distinct features of hyperbolic geometry, which leads to an interesting geometrical interpretation of the higher-dimensional feature spaces of neural networks using hyperbolic tangent as the activation function. The new theory is used to explain the separability of hyperbolic tangent kernels where we show that the separability is possible only for a certain class of hyperbolic kernels. Simulation results are given supporting the separability theory developed in this paper.

1. INTRODUCTION

The support vector machine (SVM) as a pattern classifier is a recent attraction in the field of neural networks. SVM is capable of finding the optimum decision region with a small set of training points. The SVM uses a linear separating plane to create a pattern classifier. For the patterns that are non-separable in the original space, the machine non-linearly transforms the original input space into a higher-dimensional feature space via a kernel $K(\underline{x}, \underline{x}_i)$. A support vector kernel of particular interest is the hyperbolic tangent kernel, defined by $\tanh(a||x_i||^2 + \beta)$. However, it is known that this class of SVM using the hyperbolic tangent kernel has no feasible solution for many classification problems [5]. The reason for this problem is not known or reported anywhere in the literature.

In this paper we present a new theory for this class of SVM's. In particular, we show that the feature spaces of neural network using hyperbolic activation functions can be studied and explained in indefinite metric hyperbolic spaces but not in Euclidean spaces. The inner product defined in the hyperbolic spaces is the Lorentzian inner product [2]. The Lorentzian inner product leads to many new concepts, in particular, a new concept of length with imaginary lengths being possible. In a Hilbert space, projections always exist in a unique way [1]. In contrast, in a hyperbolic space projections are unique if and only a certain Hessian matrix is non-singular. In a Hilbert space, a quadratic form always has minima or maxima, whereas in a hyperbolic space the quadratic form always has stationary points and further conditions must be met for a stationary point to be a minimum, maximum or a saddle point. The additional conditions are explained by using geometrical explanations that finally produce the range of values which a hyperbolic tangent must satisfy to produce a unique projection.

2. HYPERBOLIC INNER PRODUCT AND HYPERBOLIC SPACES

The inner product in hyperbolic geometry is defined as the Lorentzian inner product. The Lorentzian inner product leads to a new concept of length. In particular, it is possible to have an imaginary length.
Let $\underline{x}$ and $\underline{y}$ be vectors in $\mathcal{R}^n$ with $n > 1$. The *Lorentzian inner product* of $\underline{x}$ and $\underline{y}$ is a real number, denoted by the notation $\circ$ as shown here.

$$\underline{x} \circ \underline{y} = -x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \qquad (1)$$

where x_i and y_i are the i^{th} components of the vectors $\underline{x}$ and $\underline{y}$, respectively. Note that the algebraic sign of the product term is the *Lorentzian norm* of a vector $\underline{x}$ in $\mathcal{R}^n$ is defined as the complex number

$$||\underline{x}|| = \sqrt{(\underline{x} \circ \underline{x})} \qquad (2)$$

where $||\underline{x}||$ is either positive, zero or positive imaginary.

Consider next the three-dimensional case of *Lorentzian inner product*, $< \underline{v}, \underline{v} >= x^2 + y^2 - t^2$. The geometrical form of the Lorentzian inner product is shown in Figure 1, where we may note the following:

(*i*) The cone $x^2 + y^2 - t^2 = 0$ consists of zero-length non-zero vectors referred to as a neural cone.

(*ii*) The inequality $x^2+y^2-t^2<0$ defines the points inside the cone that belong to the negative subspace.

(*iii*) The inequality $x^2+y^2-t^2>0$ defines the points outside the cone that belong to positive subspace.

In a Euclidean space, it may be noted that the inner product is defined as $<\underline{v},\underline{v}>=x^2+y^2+z^2\geq 0$ and the equality occurs iff $x=y=z=0$. Therefore no neural cone exists in the Euclidean space.

The basic differences between the properties of a hyperbolic space and a Euclidean space may be summarized as follows: (1) the neural vectors, (2) due to the presence of neural vectors, a special group of vectors called the isotropic vectors which are non-zero vectors lying in a linear space and orthogonal to every element in that linear space and, (3) the unitary twin isometric vectors that are the unit-length vectors forming an orthonormal basis. The unitary isometric spaces are defined as follows:

As imaginary distances are possible in Lorentzian $(n+1)$-space, the hyperbolic n-space can be taken as the sphere of unit imaginary radius:

$$\mathcal{H}^{ni}=\underline{x}\in\mathcal{R}^{n+1}:||\underline{x}||^2=-1 \qquad (3)$$

The set $\mathcal{H}^{ni}$ is a hyperboloid of two unconnected sheets defined by the equation

$$x_1^2-(x_2^2+\ldots+x_{n+1}^2)=1 \qquad (4)$$

The subset of $\mathcal{H}^{ni}$ such that $x_{n+1}>0$ ($x_{n+1}<0$) is called the positive (negative) sheet of $\mathcal{H}^{ni}$. The negative sheet of $\mathcal{H}^{ni}$ is discarded [2].

We now define a twin isometric space with real distances in Lorentzian $(n+1)$-space as:

$$\mathcal{H}^{nr}=\underline{x}\in\mathcal{R}^{n+1}:||\underline{x}||^2=1 \qquad (5)$$

The set $\mathcal{H}^{nr}$ is another hyperboloid of connected sheet (closed hyperbola with open ends) defined by:

$$x_1^2-(x_2^2+\ldots+x_{n+1}^2)=-1 \qquad (6)$$

The real and imaginary isometric spaces consist of isometric vectors in opposite directions with equal magnitude (1 and -1). Figure 1 illustrates the twin hyperbolic isometric space.

3. THE SEPARABILITY OF HYPERBOLIC TANGENT KERNELS

From the inner product defined in hyperbolic geometry, we state the following:

Theorem 1: A kernel $K_H(\underline{x}_j,\underline{x}_k)$, defined in hyperbolic spaces, is separable if and only if the kernel can be expanded in a series given by

$$K_H(\underline{x}_j,\underline{x}_k)=\sum_{i=1}^{\infty}\lambda_i\underline{\phi}_i^T(\underline{x}_j)\underline{\phi}_i(\underline{x}_k) \qquad (7)$$

with $\lambda_1<0$, $|\lambda_1|>|\lambda_i|$, $\forall i$, $i\neq 1$ and $\lambda_i\neq 0$, $\forall i$. Here the functions $\underline{\phi}_i(\underline{x})$ are eigen-functions and numbers λ_i are the corresponding eigenvalues. □

Proof: The matrix of dot product (Hessian or Gram matrix) $\mathbf{H}_{ij}\equiv K(\underline{x}_j,\underline{x}_k)$ [3]. We consider a hyperbolic Hessian $\in\mathcal{R}^n$, $\mathbf{H}=<\mathbf{Y},\mathbf{Y}>$, where $\mathbf{Y}=\{\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_n\}$, and the eigenvalues of the Hessian are $\lambda_i, i=1:n$. The principal eigenvalue must be negative according to the inner product defined in (1). If $\lambda_i=0$ for any i, then, the Hessian is singular. In a Hilbert space, a singular Hessian implies that the solutions are linearly dependent, but there exists a unique projection, $<\underline{y}_i,\underline{y}_i>=0$, iff $\underline{y}_i=0$. In hyperbolic spaces, a singular Hessian means that there exists a linear combination of vectors that are orthogonal to every element in that space, $<\underline{y}_i,\underline{y}_i>=0$ means that the $\underline{y}_i$ is any isotropic vector, therefore, no unique solution.

At this point we say that the hyperbolic tangent $\tanh(a||\underline{x}||^2+\beta)$ has two different classes of geometry according to the sign of β. From this point on, we substitute $\beta=\gamma^2$ to avoid confusion of the two classes of hyperbolic tangent function. Specifically, we state the following:

Theorem 2: The hyperbolic tangent may belong to one of two classes defined by $\tanh(a||\underline{x}||^2-\gamma^2)$ and $\tanh(a||\underline{x}||^2+\gamma^2)$ which have two different geometries. Only the form of $\tanh(a||\underline{x}||^2-\gamma^2)$ can be separated and expressed as in the form of Theorem 1. □

Proof: (*i*) *The geometry of* $\tanh(a||\underline{x}_i||^2-\gamma^2)$:

It can be easily seen that the hyperbolic tangent defined as $\tanh(a||\underline{x}_i||^2-\gamma^2)$, where $\underline{x}_i\in\mathcal{R}^2$, has the same geometry as explained in the previous section.

(*i*) For $c>a||\underline{x}_i||^2-\gamma^2>0$, $1>\tanh(a||\underline{x}_i||^2-\gamma^2)>0$, where the data subspace spans a feature subspace in the space that belongs to $||\underline{x}||^2>0$, which is a positive subspace in the Lorentz space.

(*ii*) For $-c<a||\underline{x}_i||^2-\gamma^2<0$, $-1<\tanh(a||\underline{x}_i||^2-\gamma^2)<0$, where the data subspace spans a feature subspace in the space that belongs to $||\underline{x}||^2<0$, which is a negative subspace in the Lorentz space.

(*iii*) When $a||\underline{x}_i||^2-\gamma^2=0$, $\tanh(a||\underline{x}_i||^2-\gamma^2)=0$, where the feature space spans on the neural cone $||\underline{x}||^2=0$, which is a null-space having zero-length, non-zero vectors.

(iv) When $a||\underline{x}_i||^2 - \gamma^2 > c, \tanh(a||\underline{x}_i||^2 - \gamma^2) = 1$, the features span a hyperbolic real space.

(v) When $a||\underline{x}_i||^2 - \gamma^2 < -c, \tanh(a||\underline{x}_i||^2 - \gamma^2) = -1$, the features span a hyperbolic imaginary space.

where $c = \tanh^{-1}(1)$. Figure 1 illustrates the geometry of $\tanh(a||\underline{x}_i||^2 - \gamma^2)$. Here, we can see the separability of the hyperbolic tangent function is uniquely defined by the neural cone.

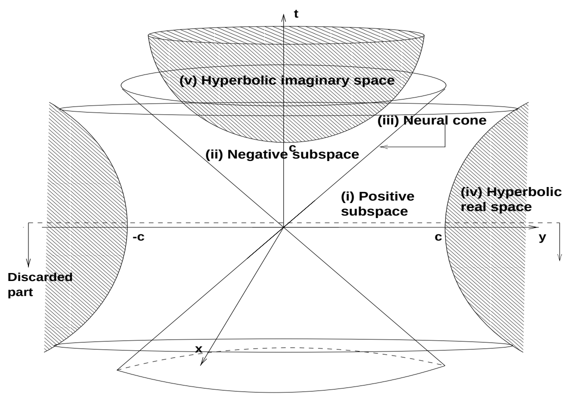

Figure 1: Hyperbolic Geometry/ Space

(ii) *The geometry of* $\tanh(a||\underline{x}_i||^2 + \gamma^2)$:
The geometry of the hyperbolic space $\tanh(a||\underline{x}_i||^2 + \gamma^2)$, $\underline{x}_i \in \mathcal{R}^2$ can be explained in a similar way. In this class of hyperbolic tangents, we can mention two differences from the other class of hyperbolic tangent:

(i) No neural vectors exist, because $\tanh(a||\underline{x}_i||^2 + \gamma^2) = 0$ only if $a||\underline{x}_i||^2 = 0$ and $\gamma^2 = 0$.

(ii) No imaginary isometric vectors exist, because $\tanh(a||\underline{x}_i||^2 + \gamma^2) \geq 0$

The separation sphere of non-isometric - isometric space in this geometry be defined as

$$\mathcal{H}^n = \underline{x} \in \mathcal{R}^{n+1} : ||\underline{x}||^2 = c \tag{8}$$

$$x_1^2 + x_2^2 + \gamma^2 = c \tag{9}$$

where $c = \tanh^{-1}(1)$. The sphere defined by $||\underline{x}||^2 = c$ is a closed sphere. The space outside the sphere is a unitary (real) isometric space, i.e., $\tanh(a||\underline{x}||^2 + \gamma^2) = 1$. The space inside the sphere is $0 < \tanh(a||\underline{x}||^2 + \gamma^2) < 1$. Figure 2 illustrates the geometry of $\tanh(a||\underline{x}||^2 + \gamma^2)$. In this geometry we have only real isometric space and no neural cone, hence no separability. Therefore, the hyperbolic tangent defined by $\tanh(a||\underline{x}||^2 + \gamma^2)$ cannot be used for a SVM. The simulation results presented in [4] also support our theory.

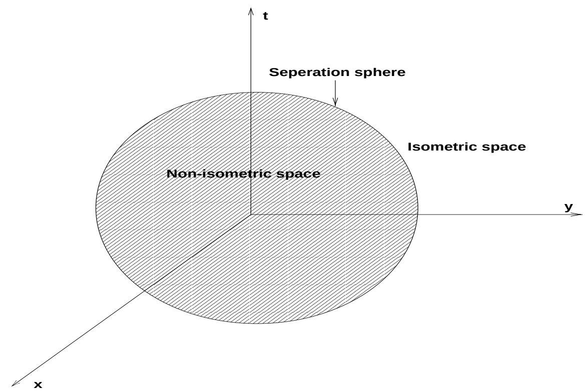

Figure 2: Plot of $\tanh(a||\underline{x}||^2 + \gamma^2)$

3.1. Values of a and γ^2 for $\tanh(a||x_i||^2 - \gamma^2)$

As explained previously, unlike Euclidean spaces, finding the minimum of a degenerate type Hessian is a hard problem. Therefore, we have more restrictions on a solution in the hyperbolic space. For a unique solution, the features must be initiated as non-isotropic and non-isometric vectors.

Condition I: The neural cone consists of isotropic vectors, therefore, initiating the input vectors either in positive or negative subspace will be a solution. Initiating all the features in the positive subspace is impossible because $\tanh(a||\underline{x}_i||^2 - \gamma^2) < 0$ for $a||\underline{x}_i||^2 < \gamma^2$. For continuous data $\in \{0, \infty\}$, more naturally $a||\underline{x}_i||^2_{\min} = 0$ and $a||\underline{x}_i||^2_{\max} = k$ where k is a finite value. Therefore, initiation must be done in the negative subspace.

$$\begin{aligned} \tanh(a||\underline{x}_i||^2{}_{\max} - \gamma^2) &< 0 \\ a||\underline{x}_i||^2{}_{\max} - \gamma^2 &< 0 \\ \gamma^2 &> a||\underline{x}_i||^2{}_{\max} \end{aligned} \tag{10}$$

□

Condition II: In the negative subspace, an imaginary isometric hyperbolic space exists. If the features are initiated in the imaginary hyperbolic space, then there may be no projection and, therefore, the initiation must be done in the time-like Lorentzian vector space.

$$\begin{aligned} \tanh(a||\underline{x}_i{}^2||_{\min} - \gamma^2) &> -1 \\ \gamma^2 &< a||\underline{x}_i||^2{}_{\min} - \tanh^{-1}(-1) \\ \gamma^2 &< \tanh^{-1}(1) \\ a &< \tanh^{-1}(1)/||\underline{x}_i||^2{}_{\max} \end{aligned} \tag{11}$$

□

Finding the optimum solution from the above range is problem-dependent and requires the use of trial and error.

4. REPRESENTATIVE EXAMPLES

In this section, we verify the solutions for a pattern classification task by using SVM that uses the classes of hyperbolic tangent $\tanh(a\|\underline{x}_i\|^2 - \gamma^2)$ and $\tanh(a\|\underline{x}_i\|^2 + \gamma^2)$ as activation functions. The classification task is depicted in Figure 3. All the results presented in this section were obtained with $a = 0.1$, number of points for training =456 and testing=1544.

First, we analyze the eigenvalues (EV) of the Hessian matrices ($\mathbf{H}_{ij} = y_i y_j \tanh(a\underline{x}^T \underline{x}_i \pm \gamma^2)$) formed by both classes of hyperbolic tangents. Typically, for this two-class pattern classification task, a Hessian matrix formed by hyperbolic tangent $\tanh(a\|\underline{x}_i\|^2 - \gamma^2)$, satisfying conditions (10) and (11) has (*i*) a large negative principal eigenvalue, (*ii*) small number of small positive eigenvalues, and, (*iii*) large number of negligible eigenvalues; and the kernels not satisfying these conditions was found to have singular Hessians. The Hessian matrices formed by hyperbolic tangent $\tanh(a\|\underline{x}_i\|^2 + \gamma^2)$ have (*i*) a large positive principal eigenvalue and, (*ii*) small and negligible values.

It can be seen that for the cases given in Table 1, with $\tanh(a\|\underline{x}_i\|^2 + \gamma^2)$, the principal eigenvalues are positive; therefore, in this case there are no Lorentzian inner products. For this case no unique solution is obtained for the separating hyper plane. For the cases given in Table 2, with $\tanh(a\|\underline{x}_i\|^2 - \gamma^2)$, the principal eigenvalues are negative and the remaining eigenvalues are positive. For this case, unique separating hyper planes were obtained. We show the separating hyperplanes obtained for $\tanh(a\|\underline{x}_i\|^2 - \gamma^2)$, $a = 0.1$ and $\gamma^2 = 2, 3$ and 4 in Figure 3. According to the range derived in (10) and (11) for a and γ^2 of $\tanh(a\|\underline{x}_i\|^2 - \gamma^2)$ for $a = 0.1$, $4 > \gamma^2 > 1.0$. The simulation results show a better separation (see Figure 3 and Table 2) for $\gamma^2 = 2$ compared to $\gamma^2 = 3$ and $\gamma^2 = 4$. It may be noted, we approximate $\tanh^{-1}(1) \simeq 5.0$.

Table 1: Performance of $\tanh(a\underline{x}^T \underline{x}_i + \gamma^2)$

γ^2	1	2	3	4
principal EV	439.64	453.75	455.69	455.69
second EV	4.93	0.69	0.09	0.09
third EV	2.03	0.28	0.04	0.04
fourth EV	-0.72	-0.10	0.01	-0.01
no. of zeros in EVs	0	0	0	0

Table 2: Performance of $\tanh(a\underline{x}^T \underline{x}_i - \gamma^2)$

γ^2	1	2	3	4
principal EV	-336.04	-437.26	-453.40	-455.64
second EV	32.51	5.84	0.82	0.11
third EV	13.55	2.39	0.34	0.04
fourth EV	3.12	0.75	0.11	0.01
no. of zeros in EVs	357	0	0	0
support vectors	No sol.	56	78	119
raw error (%)	-	3.30	4.15	5.51

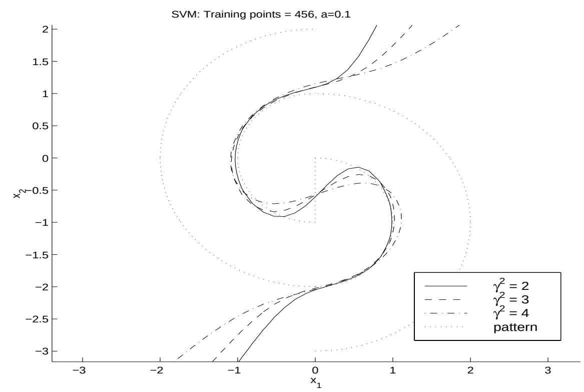

Figure 3: Plot of optimum hyper-planes

5. CONCLUSIONS

In this paper, we have shown that the separability of hyperbolic tangent can be uniquely achieved only for a certain class of hyperbolic tangent given by $\tanh(a\|\underline{x}_i\|^2 - \gamma^2)$ for certain values of a and γ^2 that provide a non-singular Hessian. A unique separation cone known as the neural cone is used for separating pattern which exists only for the above class of hyperbolic tangent. Finally, the necessary conditions to choose the parameters a and γ^2 in $\tanh(a\|\underline{x}_i\|^2 - \gamma^2)$ for a unique solution were derived.

6. ACKNOWLEDGMENT

The authors wish to thank T. Ratnarajah for the useful discussions to initiate the idea of using the hyperbolic space. The research reported herein was supported by the Natural Sciences and Engineering Research council of Canada.

7. REFERENCES

[1] Courant, R and Hilbert, D., Methods of Mathematical Physics, Interscience, 1953.

[2] Fenchler, W., Elementary Geometry in Hyperbolic space, Walter de Gruyter and Co., Berlin, 1989.

[3] Horn, A., Charles, R., Matrix Analysis, Cambridge University Press, 1985.

[4] Schlkopf, B., Burges, C., and Vapnik, V., Extracting support data for a given task, Proc., First Int.Conference on Knowledge Discovery and Data Mining, 1995.

[5] Vapnik,V., The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.