Optimal Spherical Separability: Towards Optimal Kernel Design

Garimella Rama Murthy
Signal Processing and
Communication Research Center
International Institute of
Information Technology,
Hyderabad, India
Email: rammurthy@iiit.ac.in

Yaparla Ganesh
Signal Processing and
Communication Research Center
International Institute of
Information Technology,
Hyderabad, India
Email: ganesh.yaparla@research.iiit.ac.in

Rhishi Pratap Singh
Signal Processing and
Communication Research Center
International Institute of
Information Technology,
Hyderabad, India
Email: rhishi.pratap@research.iiit.ac.in

***Abstract*—In this research paper, the concept of hyper-spherical/hyper-ellipsoidal separability is introduced. Method of arriving at the optimal hypersphere (maximizing margin) separating two classes is discussed. By projecting the quantized patterns into higher dimensional space (as in encoders of error correcting code), the patterns are made hyper-spherically separable. Single/multiple layers of spherical/ellipsoidal neurons are proposed for multi-class classification. An associative memory based on hyper-ellipsoidal neuron is proposed. The problem of optimal kernel design is discussed.**

I. Introduction

In an effort to model the biological neural network, perceptron provided an important beginning. Rosenblatt proved the convergence theorem associated with the learning law, when the patterns are linearly separable.The notion of linear separability provided the conceptual basis for statistical learning theory based on support vector machines developed by V.Vapnick et. al. Specifically non-linearly separable patterns are mapped to higher dimension space where they become linearly separable by means of suitable kernel. This approach provided a method of arriving at a feature space where the classification is rendered easy.

The authors contemplated the problem of arriving at optimal kernel function which maximizes the concept of margin in higher dimension space where patterns are linearly separable. To progress the investigation, the notion of circular, spherical and hyper-spherical separability concepts are introduced. Using such concept optimal circular/spherical/hyper-spherical separating decision manifolds in 2-class case that maximizes the margin are derived.The ultimate goal is to design optimal kernel function for a given data set that maximizes the margin corresponding to various possible kernel functions between linearly separable classes in higher dimension.

A novel method of multi-state neuron, called Spherical Neuron is proposed. It is reasoned that such a neuron enables classification of certain type of multiple classes(i.e. structured multi-class classification). Efforts are underway to train single/multi-layer networks of such neurons.

II. Circular/Spherical Separable Patterns: Optimal Separating Circle/Sphere

A. Two class classification of circularly/spherically/hyper-spherically separable patterns in 2/3/N dimensional space

Notations used in the section are as following.

- ω_1 and ω_2: classes which are being separated.
- X and Y: data points belonging to classes ω_1 and ω_2 respectively. For N dimension case $X, Y \in \mathcal{R}^N$. Similarly for 2 dimension $X, Y \in \mathcal{R}^2$.
- Point $C(c_1, c_2, \ldots, c_N)$: center of the circle/sphere/hyper-sphere which divides the pattern in two classes. For N dimension case $C \in \mathcal{R}^N$.
- d_{max}, d_{min}: farthest and closest distances from C of points in ω_1 and ω_2 respectively.

1) 2-D separation case: Patterns are circularly separable in 2 dimension if there exists a circle which can separate both the classes.

Let ω_1, ω_2 be circularly separable. There exists a $C \in \mathcal{R}^2$ i.e. $C(c_1, c_2)$. The optimal circle which separates classes is at distance $(d_{max} + d_{min})/2$ from C. When a new data point $Z(p, q)$ is given for classification the decision is taken using following function

$$Z(p,q) = \begin{cases} \omega_1 \text{ if } (p-c_1)^2 + (q-c_2)^2 < (d_{max}+d_{min})/2 \\ \omega_2 \text{ otherwise} \end{cases} \quad (1)$$

2) 3-D separation case: Patterns are spherically separable if there exists a sphere which can separate both the classes.

Let ω_1, ω_2 be spherically separable. There exists a $C \in \mathcal{R}^3$ i.e. $C(c_1, c_2, c_3)$. The optimal sphere which separates classes is at distance $(d_{max} + d_{min})/2$ from C. When a new data point $Z(p, q, r)$ is given for classification the decision is taken using following function

$$Z(p,q,r) = \begin{cases} \omega_1 \text{ if } (p-c_1)^2 + (q-c_2)^2 + (r-c_2)^2 < \\ \qquad (d_{max}+d_{min})/2 \\ \omega_2 \text{ otherwise} \end{cases} \quad (2)$$

3) N-D separation case: Patterns are hyper-spherically separable if there exists a hyper-sphere which can separate both the classes.

Let ω_1, ω_2 be hyper-spherically separable. There exists a $C \in \mathcal{R}^N$ i.e. $C(c_1, c_2, \ldots, c_N)$. The optimal hyper-sphere which separates classes is at distance $(d_{max} + d_{min})/2$ from C. When a new data point $Z(z_1, z_2, \ldots, z_N)$ is given for classification the decision is taken using following function

$$Z(z_1, z_2, \ldots, z_N) = \begin{cases} \omega_1 \text{ if } (z_1 - c_1)^2 + (z_2 - c_2)^2 + \ldots + \\ \qquad (z_N - c_N)^2 < (d_{max} + d_{min})/2 \\ \omega_2 \text{ otherwise} \end{cases} \tag{3}$$

Note 1. It is clear that if patterns belonging to two classes are linearly separable, they are hyper-spherically separable. But hyper-spherical separability does not imply linear separability(by a hyperplane). For instance, if the patterns belonging to two classes are spherically symmetric about the origin, they are clearly not linearly separable.

Note 2. It is well known that the determination of optimal hyperplane (which maximizes the margin in the case of linearly separable patterns) can be formulated as a quadratic programming problem. But the determination of optimal hypersphere separating two spherically separable classes only requires computation of distances to patterns from the center. Thus, we expect reduction in computational complexity in this case.

B. Multi class classification of circularly/spherically/hyper-spherically separable patterns in 2/3/N dimensional space

Notations for the section are as following.

- $\omega_1, \omega_2, \ldots, \omega_M$:$M$ classes which are being separated. Also $\omega = \omega_1 \cup \omega_2 \cup \ldots \cup \omega_M$
- X_i : data points belonging to classes $i \in (1, M)$. For N dimension case $X \in \mathcal{R}^N$.
- Point $C_i(c_{i1}, c_{i2}, \ldots, c_{iN})$: center of the circle/sphere/hyper-sphere which divides the pattern in $i \in (1, M)$ classes. For N dimension case $C \in \mathcal{R}^N$.
- A class t_i where $i \in (1, M)$ is introduced which contains all the points which lie inside circle/sphere/hyper-sphere by which ω_i is enclosed. Also $t = t_1 \cup t_2 \cup \ldots \cup t_M$
- d_{i1}, d_{i2}: farthest and closest distances from C_i of points in t_i and $t - t_i$ respectively.

1) 2-D case: Let $\omega_1, \omega_2, \ldots, \omega_M$ be M classes which are circularly separable. We use one vs rest approach to classify the data points. The optimal circle which separates classes is at distance $(d_{i1} + d_{i2})/2$ from the center. When a new data point $Z(a, b)$ is given for classification, then

$$Z(a, b) \in t_i \text{ if } (a - c_{ix})^2 + (b - c_{iy})^2 < (d_{i1} + d_{i2})/2 \quad \forall i \in (1, n) \tag{4}$$

Let g be the set which contains all the indexes x in t_x in which $Z(a, b)$ is present

$$Z(a, b) \in \omega_i \text{ if } dp_1 \leq dq_1 \forall \quad p, q \in g \tag{5}$$

Similar approach can be followed for 3-D and N-D cases.

III. Two class classification of elliptical/ellipsoidal/hyper elliptically separable patterns in 2/3/N dimensions respectively

The method we propose to optimally solve the problem of two class classification works only if one of the classes is bounded. Bounded in the sense that there should be a 2/3/N dimensional surface which encloses all the data points of a class.

If two classes are separable in 2/3/N dimensions by ellipse/ellipsoid/hyper ellipse respectively then these classes are said to be elliptically/ellipsoidal/hyper elliptically separable. Following are the steps to algorithm for solving the two class classification problem optimally The ultimate goal is to learn a mapping from input x to output y. $D = (x_i, y_i)_{i=1}^n$. Where D is the training set. We are dealing with two classes, so y_i either belongs to class ω_1 or ω_2.

1) Find the convex hull of that class which is bounded. Let's say it is class ω_1.
2) Try to enclose the convex hull with ellipse/ellipsoid/hyper ellipse, which minimizes the number of points of class ω_2 falling in it and all the points of class ω_1 must satisfy the following

$$= \begin{cases} \frac{x^2}{a^2} + \frac{y^2}{b^2} < 1 \text{ if ellipse} \\ \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \ldots + \frac{x_n^2}{a_n^2} < 1 \text{ if N-D ellipse} \end{cases} \tag{6}$$

   Here a and b represent major and minor axis of the ellipse
3) We need to find whether or not we can solve the problem in current dimension. We know the dimensions $(a_{i1}, a_{i2}, \ldots, a_{iN})$ and center $c(c_{i1}, c_{i2}, \ldots, c_{iN})$, where i is the iteration and N is the dimension. Now we find the nearest point of class ω_2.
   In case of ellipse let $t(p, q)$ be the nearest point of class ω_2.

$$\frac{p^2}{a^2} + \frac{q^2}{b^2} \begin{cases} \leq 1 \text{ proceed to higher dimension} \\ > 1 \text{ patterns are separable in current dimension} \end{cases} \tag{7}$$

   In case of hyper ellipse $t(p_1, p_2, \ldots, p_N)$ be the nearest point of class ω_2,

$$\frac{(p_1 - c_{i1})^2}{a_{i1}^2} + \frac{(p_2 - c_{i2})^2}{a_{i2}^2} + \ldots + \frac{(p_n - c_{in})^2}{a_{in}^2} \begin{cases} \leq 1 \text{ go to higher dimension} \\ > 1 \text{ separable in current dimension} \end{cases} \tag{8}$$

4) If there is a need to solve it in higher dimensions we do projection as given below, else we skip to step 6.
If we are in 2d and want to project it to 3d then, it is done by projecting points normally. We project the points with in the ellipse on upper surface of ellipsoid. We need to choose the third dimension of ellipsoid that is c as a linear function of (a, b) where a and b are major and minor axis. The point at (x, y) with in ellipse maps to (x, y, z) on to ellipsoid where

$$z = \sqrt{c^2(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2})} \tag{9}$$

We do the same for projection of points from N-dim (N+1)dimension. We project hyper ellipse of N-dimension on upper portion of hyper ellipse of (N+1)-dimension. We need to choose the (N+1)-dimension of ellipsoid that is c_{N+1} which is linear function of $c_1, c_2, c_3, ...c_N$.
A point at $(x_1, x_2..x_N)$ with in hyper ellipse of N dimensions to $(x_1, x_2...x_N, x_{N+1})$ on to hyper ellipse of $N + 1$ dimensions where

$$x_{N+1} = \sqrt{(c_n^2(1 - \frac{x_1^2}{c_1^2} - \frac{x_2^2}{c_2^2} - \ldots - \frac{x_n^2}{c_n^2})}. \tag{10}$$

We move to step1 after we increase the dimension.

5) In case of ellipse, $t(p, q)$ is the nearest point of class ω_2. We define a variable $D = \sqrt{(p^2 + q^2)}$. In order to find the optimal margin ellipse we have to find the concentric ellipse which has the data point $t(p, q)$ in it. In order to do that we need to find the dimensions (a, b) i.e. length of major axis and length of minor axis. Let variables (a_1, b_1) be the dimensions of ellipse. As we know that concentric ellipse will have their dimensions multiples of one of them. The factor be k through which it gets multiplied so $a_1 = a * k$, $b_1 = b * k$ where k belongs to real numbers greater than 1.

$$\frac{x^2}{a_1^2} + \frac{y^2}{b_1^2} = 1$$

Replacing $a_1 = ak, b_1 = bk$

$$\frac{x^2}{a^2k^2} + \frac{y^2}{b^2k^2} = 1.$$

$$k = \sqrt{(\frac{x^2}{a^2} + \frac{y^2}{b^2})} \tag{11}$$

the optimal margin hyper ellipse which separates two classes is concentric to the ellipse we had and has dimensions $((a + ak)/2, (b + bk)/2$.
In case of hyper ellipse, $A(t_1, t_2, \ldots, t_N)$ is the nearest point of class ω_2. We define a variable $D = \sqrt{(p_1^2 + p_2^2 + p_3^2 + \ldots + p_N^2)}$. In order to find the optimal margin hyper ellipsoid we have to find the concentric hyper ellipsoid which has the data point $A(t_1, t_2, \ldots, t_N)$ on it. In order to do that we need to find the dimensions$(c_1, c_2, \ldots, c_N)$ of hyper ellipsoid. Let variables $(q_1, q_2, q_3, \ldots, q_N)$ be the dimensions of the concentric hyper ellipsoid on which $A(t_1, t_2, \ldots, t_N)$ lies. As we know that concentric hyper ellipsoids will have their dimensions multiples of one of them. The factor be k through which it gets multiplied so $q_1 = c_1k, q_2 = c_2k, q_N = c_Nk$ where k belongs to real numbers greater than 1.

$$\frac{t_1^2}{q_1^2} + \frac{t_2^2}{q_2^2} + \ldots + \frac{t_N^2}{q_N^2} = 1$$

$$\frac{t_1^2}{c_1^2 * k^2} + \frac{t_2^2}{c_2^2 * k^2} + \ldots + \frac{t_N^2}{c_N^2 * k^2} = 1$$

$$k = \sqrt{(\frac{t_1^2}{c_1^2} + \frac{t_2^2}{c_2^2} + ... + \frac{t_N^2}{c_N^2})} \tag{12}$$

The optimal margin hyper ellipsoid which separates two classes is concentric to the hyper ellipsoid we had and has dimensions $(c_1 + c_1k)/2, (c_2 + c_2k)/2, \ldots, (c_N + c_Nk)/2)$.

Given a new data point to classify we do the following: Let the data point be $X \in \mathcal{R}^K$ or $X = (x_1, x_2, \ldots, x_K)$. Lets say the problem we had, has been solved by an hyper ellipsoid of N-dimension. We know the center and dimensions of all the ellipse/ellipsoid/hyper elliptical structures through which we have separated the data set at each and every step using above mentioned steps. If we start in K'th dimension and our problem gets solved in N'th dimension, where $N > K$, we do the following:
if

$$(\frac{(x_1 - a_{11})^2}{c_{11}^2} + \frac{(x_2 - a_{12})^2}{c_{12}^2} + \ldots + \frac{(x_k - a_{1k})^2}{c_{1k}^2} > 1$$

or

$$\frac{(x_1 - a_{21})^2}{c_{21}^2} + \frac{(x_2 - a_{22})^2}{c_{22}^2} + \ldots + \frac{(x_{k+1} - a_{2(k+1)})^2}{c_{2(k+1)}^2} > 1$$

$\vdots$ or

$$\frac{(x_1 - a_{(N-k)1})^2}{c_{(N-k)1}^2} + \frac{(x_2 - a_{(N-k)2})^2}{c_{(N-k)2}^2} + \ldots + \frac{(x_{(N)} - a_{(N-k)N})^2}{c_{(N-k)N}^2} > 1 \tag{13}$$

The point belongs to class ω_2 else point belongs to class ω_1 .

Note 3. In the theory of error correcting codes, an information word is mapped to the associated codeword using an encoder. Also the coding spheres at hamming distances less than or equal to the minimum distance of the code are disjoint. Using this idea, we project quantized patterns from lower dimension space (in the spirit of SVMs), where they become spherically separable. In the following section we summarize the known results from earlier literature.

IV. Hybrid Neural Networks: Spherical Separability : Spherical Neuron

Bruck et.al have shown that hopfield neural network is naturally associated with a graph-theoretical code in the sense that code words are associated with stable states[1]. They generalize the result to linear error correcting codes and non-linear error correcting codes. Effective code words are associated with stable states and vice-versa in the sense that they are the local/global optima of an energy function (associated with the encoder of an error correcting code). Thus effectively a one step associate memory(realized by encoder) performs clustering of data points. In [2], one of the authors proposed hybrid neural networks where encoders are cascaded with multi-layer perceptron, a feedforward network. It is clear that the patterns in different coding spheres are spherically separable. Figure 1 depicts the idea.

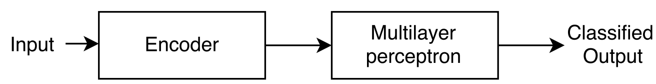


Fig. 1: Hybrid Neural Networks

Spherical Neuron: Multi-class classification

In the above discussion, in two dimension circularly separable patterns(belonging to two classes) are classified. More generally, hyper spherically separable classes are classified into 2-classes in higher dimensions. Motivated by the idea of ceiling neuron [3], we introduce a novel neuron called Spherical neuron, which performs classification of patterns belonging to multiple classes with certain restrictions/structure. The details of such neuron are presented below.

The inputs belong to M classes. They are clustered around the center $(a_1, a_2, \ldots, a_N) = \bar{A}$. It could be the origin of N-dimensional euclidean space. Let an input vector be $(x_1, x_2, \ldots, x_N)$. The output class y is determined in following manner

$$y = \begin{cases} \text{class 1 if } d(\bar{X}, \bar{A}) \leq r_1 \\ \text{class 2 if } d(\bar{X}, \bar{A}) \leq r_2 \\ \vdots \\ \text{class } M \text{ if } d(\bar{X}, \bar{A}) \geq r_M \end{cases} \tag{14}$$

Here $d(\bar{X}, \bar{A})$ is the euclidean distance between the vectors $\bar{X}$ and $\bar{A}$. Note here that spherical neuron activation function has resemblance to the neurons utilized in radial basis function networks. It is multi-valued in this case.

Several such neurons are placed in one or more layers to perform fine classification of input patterns.

Note 4. We can conceive of Ellipsoidal neuron, where the distance $d(\bar{X}, \bar{A})$ is computed in the following manner (with $\bar{A} = \bar{0}$).

$$d(\bar{X}, \bar{0}) = (\frac{x_1}{b_1})^2 + (\frac{x_2}{b_2})^2 + \ldots + (\frac{x_N}{b_N})^2 \tag{15}$$

Single or multiple layers of ellipsoidal neurons are utilized for fine classification of input data.

V. Hyper-Ellipsoidal Neuron: Associative Memory

Hopfield proposed an associative memory [4] based on McCulloch-Pitts neuron (which assumes $+1$ or -1 values). In literature, associative memories are proposed based on multi state neuron [5].In this research paper, we propose an associative memory based on hyper-ellipsoidal neuron. let the state space of such an associative memory be the bounded lattice i.e. each component of the state vector , $\bar{V}(n)$ assumes values in the set $\{0, 1, 2, \ldots, M\}$. Let there be N neurons and let the symmetric synaptic weight matrix be $\bar{W}$. The i'th component of the state vector at time $n+1$ is computed in the following manner

$$V_i(n+1) = f(net) \tag{16}$$

where

$$net = w_{i1}V_1^2(n) + w_{i2}V_2^2(n) + \ldots + w_{iN}V_N^2(n)$$

$$f(net) = \begin{cases} 0 \text{ if } net < T_1, Threshold \\ 1 \text{ if } T_1 \leq net < T_2 \\ \vdots \\ M-1 \text{ if } T_{M-1} \leq net < T_M \\ M \text{ if } net \geq T_M \end{cases}$$

As in case of Hopfield associative memory, the neural network is operated in the serial mode (state of only one neuron is updated at a given time i.e. asynchronously) or fully parallel mode(state of all the neurons is updated at any given time i.e. fully synchronously)

It is clear that the network dynamics is periodic (because of choice of $f(.)$). Convergence to stable state (i.e. cycle of length one) or a cycle of certain length is currently being investigated. Energy function based approach to investigate nature of dynamics is being currently pursued.

VI. Linear Transformation of Non-Hyper-spherically separable patterns to spherically separable patterns: Quadratic Neuron

Traditionally, single artificial neuron called perceptron was based on the concept of linear separability. Rosenblatt proposed a learning law which converges(i.e. the synaptic weights converge), when the patterns are linearly separable. The resulting hyperplane is one among various possible hyperplanes that separate the patterns into two classes.

Vapnick, by introducing the concept of margin, showed that the problem of synthesizing optimal hyperplane(i.e. a hyperplane which maximizes the margin) separating two classes can be formulated as a Quadratic optimization problem.

These two approaches remained as the basis for research related to artificial neural networks(e.g. classification problem).

The authors contemplated on the possibility of combining the logical basis of above two approaches for classification. They succeeded in such an effort by introducing hyper-spherical separability concept. The details are summarized below.

In McCulloh-Pitts neuron, the net contribution is computed using the inner product of weight vector and the vector of the inputs. This net contribution is operated on by by signum activation function, to arrive at the neuron output. Such a model of neuron is utilized to classify linearly separable patterns (by a hyperplanes). Generalizing this idea, several researchers proposed a neuron where higher order synaptic operations(e.g. quadratic synaptic operations) are utilized to arrive at the net contribution which is operated on by signum activation function[6].

In such a neuron model, the activation function is retained as signum function. It is thus clear that such models of neuron classify non-linearly separable patterns. Specifically, let W be a symmetric $M\text{X}M$ matrix and $\bar{X}$ be a $M\text{X}1$ vector of inputs. The output of neurons

$$y = signum\{\bar{X}^T W \bar{X} - T_0\} \tag{17}$$

Here T_0 is a threshold value.

Assumption: W be a positive symmetric matrix. Hence by cholesky decomposition we have $W = NN^T$, where N is a

Applying it in equation (17)

$$\bar{X}^T W \bar{X} = \bar{X}^T N N^T \bar{X} = Z^T Z = \sum_{i=1}^{M} {z_i}^2 \tag{18}$$

where $Z = N^T \bar{X}$. Thus output of such a neuron is given by

$$y = signum\{\sum_{i=1}^{M} {z_i}^2 - T_0\} \tag{19}$$

Claim: The patterns arrived at by the above linear transformation are hyper-spherically separable.

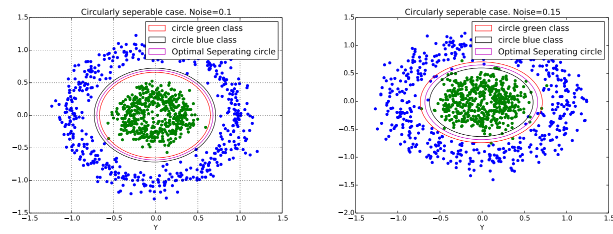
Note 5. Using above idea, first documented in research monograph [7], NP-hard problem of maximum cut computation is reduced to multi-linear objective function optimization over hypercube[8].

Note 6. It is well known that homogeneous multivariate polynomial (of degree higher than two) can be expressed in terms of symmetric tensor. Using cholesky type decomposition of symmetric tensor, the results in this section can be generalized.

Note 7. The approach proposed in this section naturally leads to the idea of transforming the patterns by a non-linear transformation(when they are separable by certain manifold) such that they become spherically separable(without projecting to higher dimensions.)

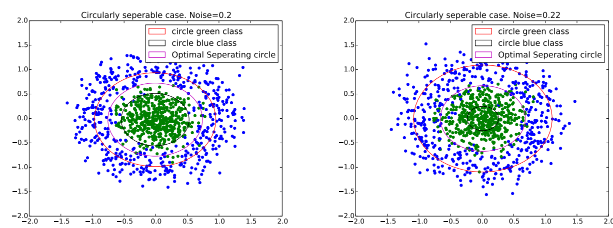
VII. Experiments and Results

Following are the results for optimal circular separation case. Two circularly separable classes, having 1000 data points were generated randomly. We have compared our method against SVM with linear, polynomial and rbf kernels and K nearest neighbors algorithm with k=3. The result is repeated with different noise level(mixing of classes). Figure 2 and 3 show the graphical representation of the results.



(a) (b)

Fig. 2: Classification with low pattern noise



(a) (b)

Fig. 3: Classification with high pattern noise

Table I and figure 4 show the comparison among various methods. It is clear that our method provides competitive performance with simple computation.

Error	**.1**	**.15**	**.2**	**.22**
Polynomial	0.5350	0.5749	0.4799	0.4774
RBF	1.0	0.9825	0.9575	0.9375
Linear	0.6474	0.4799	0.5324	0.4774
KNN	1.0	0.9825	0.9499	0.9425
Our	1.0	0.9849	0.9599	0.9425

TABLE I: Comparison of classification results among SVM with polynomial, rbf and linear kernels, KNN(n=3) and our procedure.

VIII. Conclusion

In this research paper the concept of hyper-spherical/ hyper-ellipsoidal separability of patterns is proposed. Method of arriving at optimal hyper-sphere/optimal hyper-ellipsoidal separating two classes of patterns is discussed. It is reasoned how quantized patterns can be made hyper-spherically separable by means of a suitable encoder. Spherical/ellipsoidal neuron is proposed for multi-class classification. Briefly, associative memory based on hyper-ellipsoidal neuron is proposed.

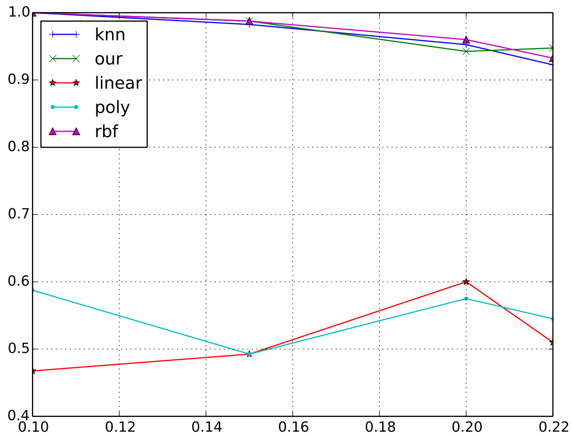


Fig. 4: Comparison various results.

IX. Future Work

Towards optimal kernel design: When patterns are not linearly separable in lower dimension, one of various possible kernels, such as polynomial kernel, radial basis function kernel, spherical kernel, are chosen to project the points to a higher dimension space. In traditional support vector machine design, no serious attention is paid to optimally select the kernel function so that the margin between projected points is made as large as possible(prior to optimally selecting the support vector that maximizes the margin between linearly separable classes in higher dimension space) i.e. project points belonging to 2 classes such that they are separated as far as possible. For example optimal polynomial kernel design, from among various possible polynomial kernels in an interesting problem.

References

[1] Jehoshua Bruck and Mario Blaum. Neural networks, error-correcting codes, and polynomials over the binary n-cube. *IEEE Transactions on information theory*, 35(5):976–987, 1989.

[2] G Rama Murthy. Hybrid neural networks. In *Proceedings of International Conference on Power System Analysis, Control and Optimization (PSACO2008)*.

[3] G. Rama Murthy. Novel ceiling neuronal model: Artificial neural networks. Technical report, 2015.

[4] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[5] Stanislaw Jankowski, Andrzej Lozowski, and Jacek M Zurada. Complex-valued multistate neural associative memory. *IEEE Transactions on Neural Networks*, 7(6):1491–1496, 1996.

[6] Zeng-Guang Hou, Ki-Young Song, Madan M Gupta, and Min Tan. Neural units with higher-order synaptic operations for robotic image processing applications. *Soft Computing*, 11(3):221–228, 2007.

[7] G Rama Murthy. *Multidimensional Neural Networks: Unified Theory*. New Age International, 2008.

[8] G Rama Murthy. Np hard problems: Multi-linear objective optimization problem. Manuscript in prepration.