RBF classifier

The method of linear discriminant function will not suffice in some cases. For example, a two class data set may look like the figure below, thus they cannot be separated by linear hyper-plane.

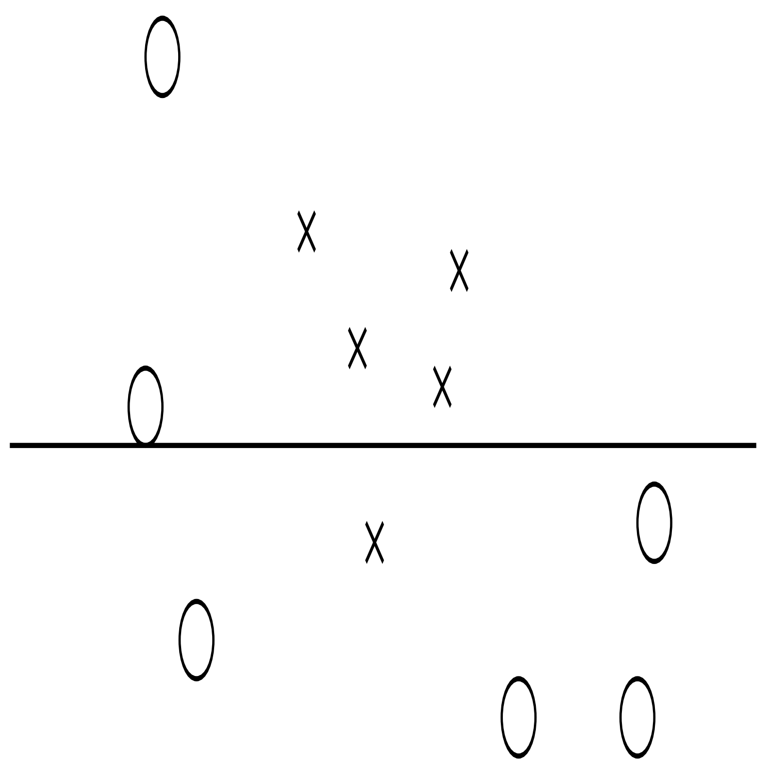

Figure: A two class data set and a linear separating plane.

We'd like to improve the classification performance using a nonlinear classifier as shown in the Figure below.

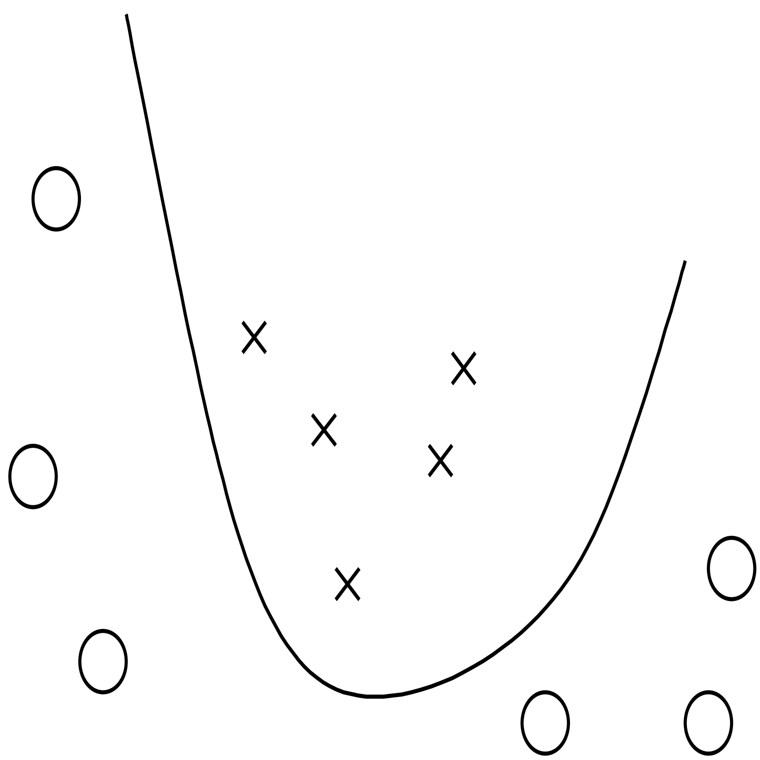

Figure: Nonlinear function is used to separate the two class data set.

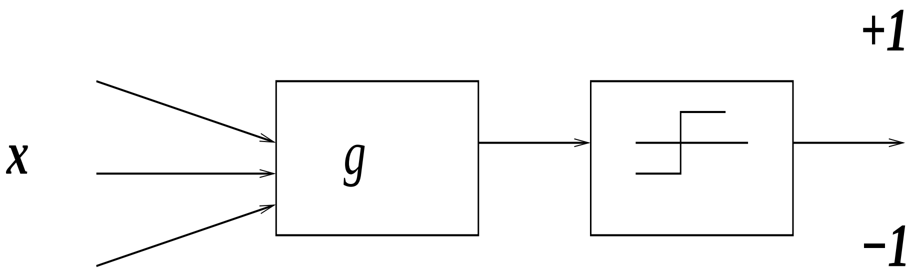

Figure: A classifier using discriminant function.

For the classifier depicted above, we attempt to model the discriminant function $g(\mathbf{x})$ by using the RBF neural networks given by

$$g(\mathbf{x}) = \sum_{i=1}^{M} w_i \exp\left(-\frac{(\|\mathbf{x} - \mathbf{c}_i\|)^2}{2\sigma^2}\right)$$

For a new sample $\mathbf{x}$ and a given discriminant function, we can decide on $\mathbf{x}$ belongs to Class 1 if $g(\mathbf{x}) > 0$, otherwise it's Class 2.

For now we assume that the number and location of the centres $\mathbf{c}_i$ are appropriately chosen (the method of center selection will be discussed next week).

A set of n pairs $(\mathbf{x}_1, t_1)$,..., $(\mathbf{x}_n, t_n)$ is given, where $\mathbf{x}_i$ is the outcome of the set of measurements made upon the ith individual and takes real values. t_i takes values of [-1,1]. We will look at the estimation of weights w_i, for a set of M given centers $\mathbf{c}_i$ through an example, in which σ is predetermined as $\sigma = 1$.

For example, a set of 10 data samples $(\mathbf{x}_1, t_1)$,..., $(\mathbf{x}_{10}, t_{10})$ as given in the following Table.

Note that $\mathbf{x}_i = [x_{1,i}, x_{2,i}]^T$, $i = 1, ..., 10$.

i	1	2	3	4	5
$x_{1,i}$	0.5	0.4	0.6	0.6	0.8
$x_{2,i}$	0.7	0.5	0.6	0.4	0.6
t_i	-1	-1	-1	-1	-1

i	6	7	8	9	10
$x_{1,i}$	0.2	0.1	0.9	0.8	0.3
$x_{2,i}$	0.8	0.7	0.3	0.1	0.1
t_i	1	1	1	1	1

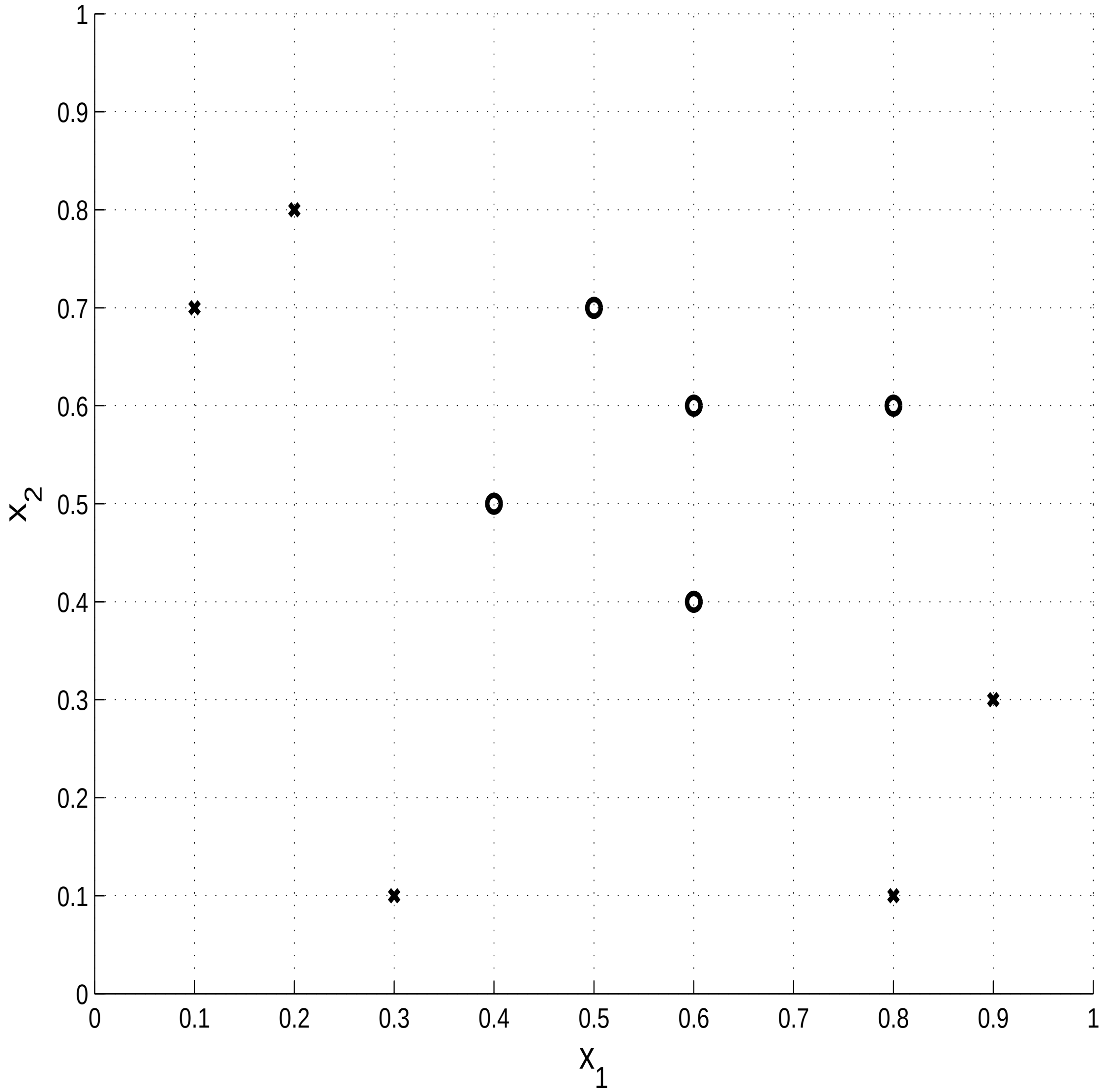

Figure: The input data set.

Again we assume centers $\mathbf{c}_i$ are given (to be discussed next week).

For this example we simply use 4 centers as $\mathbf{c}_1 = [0.5, 0.7]^T$, $\mathbf{c}_2 = [0.6, 0.4]^T$, $\mathbf{c}_3 = [0.2, 0.8]^T$ and $\mathbf{c}_4 = [0.9, 0.3]^T$. Note that these are 4 randomly selected data samples ($i = 1, 4, 6, 8$) Here we also set $\sigma = 1$. This gives us four basis functions.

$$\phi_1(\mathbf{x}) = \exp\left(-\frac{(x_1 - 0.5)^2 + (x_2 - 0.7)^2}{2}\right)$$

$$\phi_2(\mathbf{x}) = \exp\left(-\frac{(x_1 - 0.6)^2 + (x_2 - 0.4)^2}{2}\right)$$

$$\phi_3(\mathbf{x}) = \exp\left(-\frac{(x_1 - 0.2)^2 + (x_2 - 0.8)^2}{2}\right)$$

and

$$\phi_4(\mathbf{x}) = \exp\left(-\frac{(x_1 - 0.9)^2 + (x_2 - 0.3)^2}{2}\right)$$

Over the given ten data samples, form the matrix Φ given by

$$\Phi = \begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \phi_{1,3} & \phi_{1,4} \\ \phi_{2,1} & \phi_{2,2} & \phi_{2,3} & \phi_{3,4} \\ \vdots & \vdots & \ldots & \vdots \\ \phi_{9,1} & \phi_{9,2} & \phi_{9,3} & \phi_{9,4} \\ \phi_{10,1} & \phi_{10,2} & \phi_{10,3} & \phi_{10,4} \end{pmatrix}$$

with

$$\phi_{i,1} = \exp\left(-\frac{(x_{1,i} - 0.2)^2 + (x_{2,i} - 0.8)^2}{2}\right),$$

$$\phi_{i,2} = \exp\left(-\frac{(x_{1,i} - 0.5)^2 + (x_{2,i} - 0.7)^2}{2}\right),$$

$$\phi_{i,3} = \exp\left(-\frac{(x_{1,i} - 0.6)^2 + (x_{2,i} - 0.4)^2}{2}\right),$$

$$\phi_{i,4} = \exp\left(-\frac{(x_{1,i} - 0.2)^2 + (x_{2,i} - 0.8)^2}{2}\right),$$

$$i = 1, 2, 3..., 10$$

We can write ten linear equations

$$\begin{cases} \phi_{1,1}w_1 + \phi_{1,2}w_2 + \phi_{1,3}w_3 + \phi_{1,4}w_4 = t_1 \\ \phi_{2,1}w_1 + \phi_{2,2}w_2 + \phi_{2,3}w_3 + \phi_{2,4}w_4 = t_2 \\ \phi_{3,1}w_1 + \phi_{3,2}w_2 + \phi_{3,3}w_3 + \phi_{3,4}w_4 = t_3 \\ \cdots\cdots\cdots \\ \phi_{10,1}w_1 + \phi_{10,2}w_2 + \phi_{10,3}w_3 + \phi_{10,3}w_4 = t_{10} \end{cases}$$

which is

$$\mathbf{\Phi w} = \mathbf{t}$$

Note $\mathbf{t} = [-1, -1, -1, -1, -1, 1, 1, 1, 1, 1]^T$. The least squares estimate is calculated as

$$\mathbf{w} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{t}$$

The RBF classifier is given by

$$g(\mathbf{x}) = \sum_{i=1}^{4} w_i\phi_i(\mathbf{x})$$

$\mathbf{w}$ is found to be

$$\mathbf{w} = [70.5912, 37.4476, -63.3062, -52.7027]^T$$

For any $\mathbf{x}$, the class label is determined by checking $g(\mathbf{x}) > 0$ or $g(\mathbf{x}) < 0$

The classification for these training data samples are

i	1	2	3	4	5
$x_{1,i}$	0.5	0.4	0.6	0.6	0.8
$x_{2,i}$	0.7	0.5	0.6	0.4	0.6
t_i	-1	-1	-1	-1	-1
$sign(g(\mathbf{x}_i))$	-1	-1	-1	-1	-1

i	6	7	8	9	10
$x_{1,i}$	0.2	0.1	0.9	0.8	0.3
$x_{2,i}$	0.8	0.7	0.3	0.1	0.1
t_i	1	1	1	1	1
$sign(g(\mathbf{x}_i))$	1	1	1	1	1

The decision classification boundary is generated and shown in the Figure below.

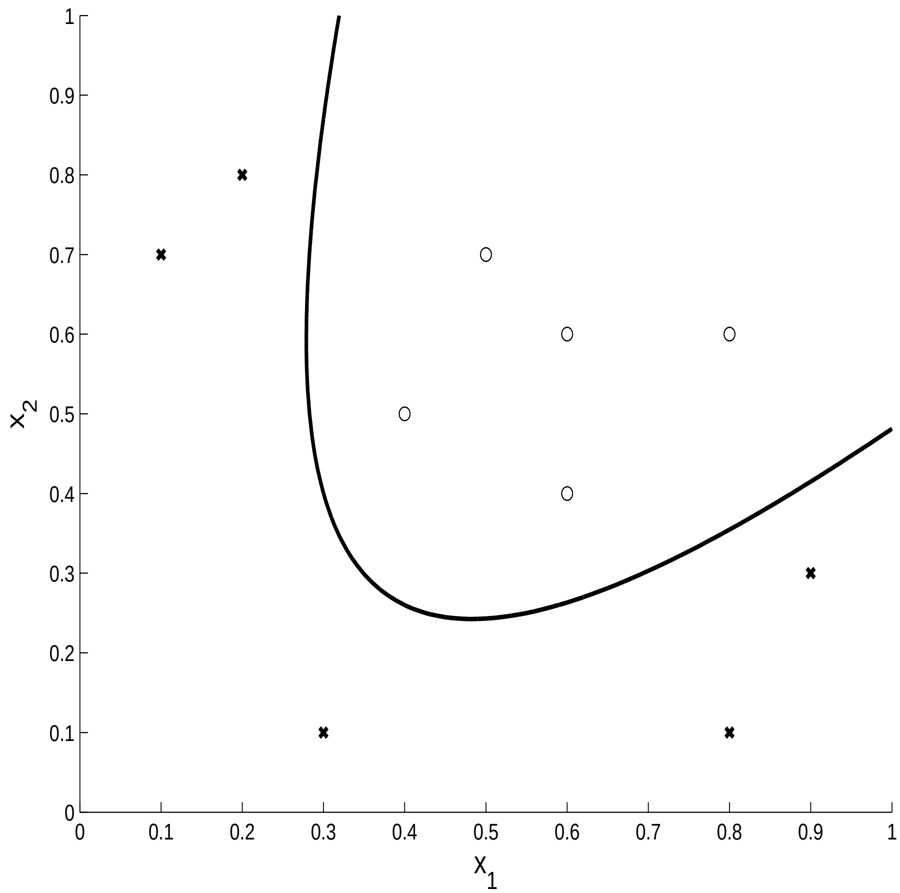

Figure: The input data set and the decision boundary.

The file $RBFCexF.m$ is used for calculation in this example. Both $RBFCex.m$ and $RBFCexF.m$ are downloadable to generate examples.

A summary of the construction of RBF Classifier.

1. Determine the number and the centers $\mathbf{c}_i$;

2. Calculate $\phi_i(\mathbf{x})$ for all training data samples;

3. Form matrix Φ and $\mathbf{t}$;

4. Calculate

$$\mathbf{w} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t}$$

5. Use the resultant classifier $g(\mathbf{x})$ for classification.