

Lecture 8

RBF center selection methods

The training of the RBF neural networks is often composed of two stages; (i) find the number and the centers c_i and then (ii) find the weights w_i .

So far we assumed that the number and the centers c_i are known, now we will look at possible ways of RBF center selection.

One fact to look at is the localized nature of the hidden nodes, the basis functions. Each basis function is a function of the distance between data points and the centers. This suggests that the centers should be distributed in the range of input data sets.

Another fact is that there is some similarity between RBF neural networks and PNN, in which all data points are centers.

In RBF neural networks, we attempt to use a minimal amount of centers.

Lecture 8

Center selection using a subset of data points

One simple procedure for selecting the basis function centers c_i is to set them equal to a random subset of the input vectors from the training set.

For the width parameter σ , this can be simply set as some constant in the same scale of the data points.

Such ad hoc procedures allow a RBF network to be set up very quickly.

The disadvantage is that it may lead to the use of an unnecessary large number of basis functions in order to achieve adequate performance.

Lecture 8

K-means clustering algorithm

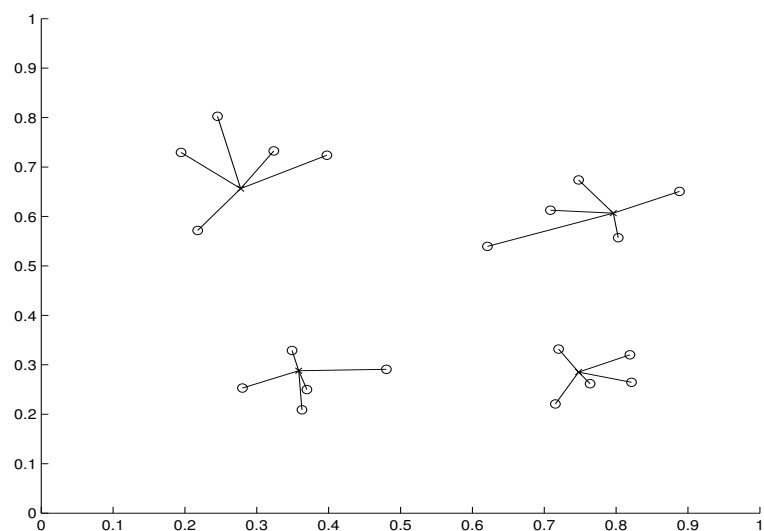
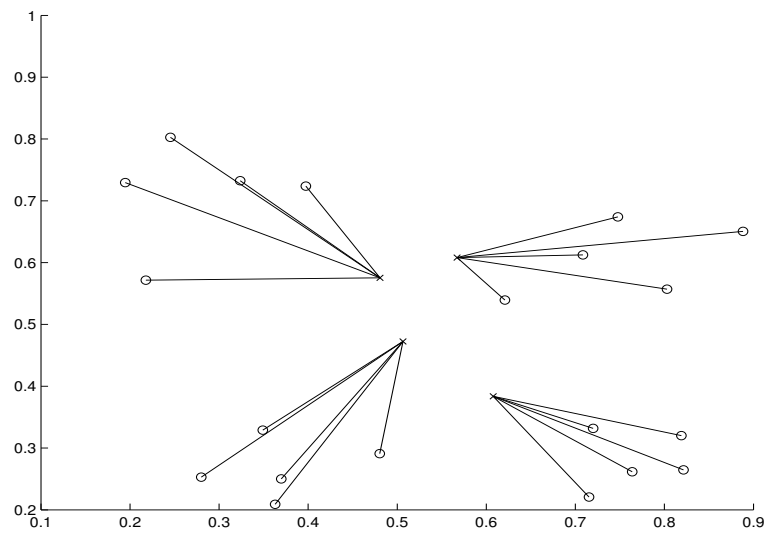
Clustering algorithms can be used to find a set of centers which more accurately reflects the distribution of the data points. The number of centers M is decided in advance, and each center \mathbf{c}_i is supposed to be representative of a group of data points.

Suppose there are n data points $\{\mathbf{x}_j, j = 1, \dots, n\}$ in total, and we wish to find M representative vectors $\mathbf{c}_i, i = 1, \dots, M$. The algorithm seeks to partition the data points $\{\mathbf{x}_j, j = 1, \dots, n\}$, into M disjoint subsets S_i containing N_i data points, in such a way as to minimize the sum-of-squares clustering function given by

$$J = \sum_{i=1}^M \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \mathbf{c}_i\|^2$$

where \in denotes belongs to.

Lecture 8



Figures above; the sum-of-squares clustering function J is the total distance of the lines. When it is minimized, c_i should be in the center of a group of data points.

Lecture 8

J is minimized when

$$\mathbf{c}_i = \frac{1}{N_i} \sum_{\mathbf{x}_j \in S_i} \mathbf{x}_j.$$

The batch K-means algorithm begins by assigning the points at random to M sets and then computing the mean vectors of the points in each set.

Next each point is reassigned to a new set according to which is the nearest mean vector. The means of the set are then recomputed.

This procedure is repeated until there is no further change in the grouping of the data points.

Lecture 8

On-line K-means clustering algorithm:

The initial centers are randomly chosen from the training data set. As each data point \mathbf{x}_j is presented.

1. Find the nearest \mathbf{c}_i to \mathbf{x}_j ; ($i = 1, \dots, M$). Suppose that this is found to be \mathbf{c}_k .

2. $\mathbf{c}_k^{new} = \mathbf{c}_k^{old} + \eta(\mathbf{x}_j - \mathbf{c}_k^{old})$

3. Set \mathbf{c}_k^{new} as \mathbf{c}_k^{old} ;

where $\eta > 0$ is a small number called learning rate. The process repeated until there is no more changes in all centers.

Lecture 8

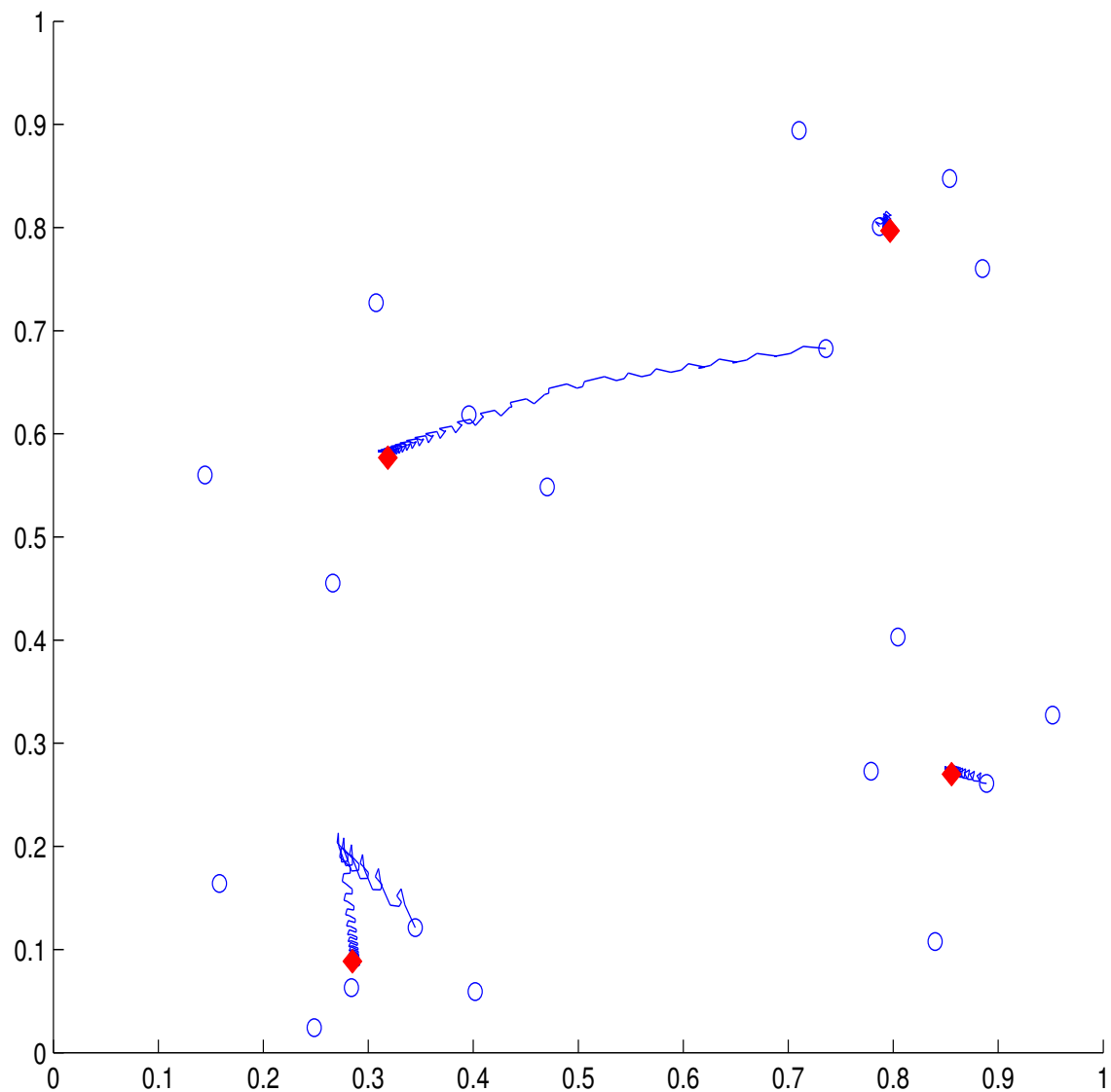


Figure: The trajectory of the centers updating from an on-line K-means clustering algorithm. Code *kms.m* is used.

Lecture 8

A set of n pairs $(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_n, t_n)$ is given, where \mathbf{x}_i is the outcome of the set of measurements made upon the i th individual and takes real values. t_i takes values of $[-1, 1]$.

To illustrate the k-means clustering algorithm for RBF center selection, we will look at finding a set of M centers \mathbf{c}_i using k-means clustering algorithm, followed by the estimation of weights w_i through a previous example.

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

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

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

Lecture 8

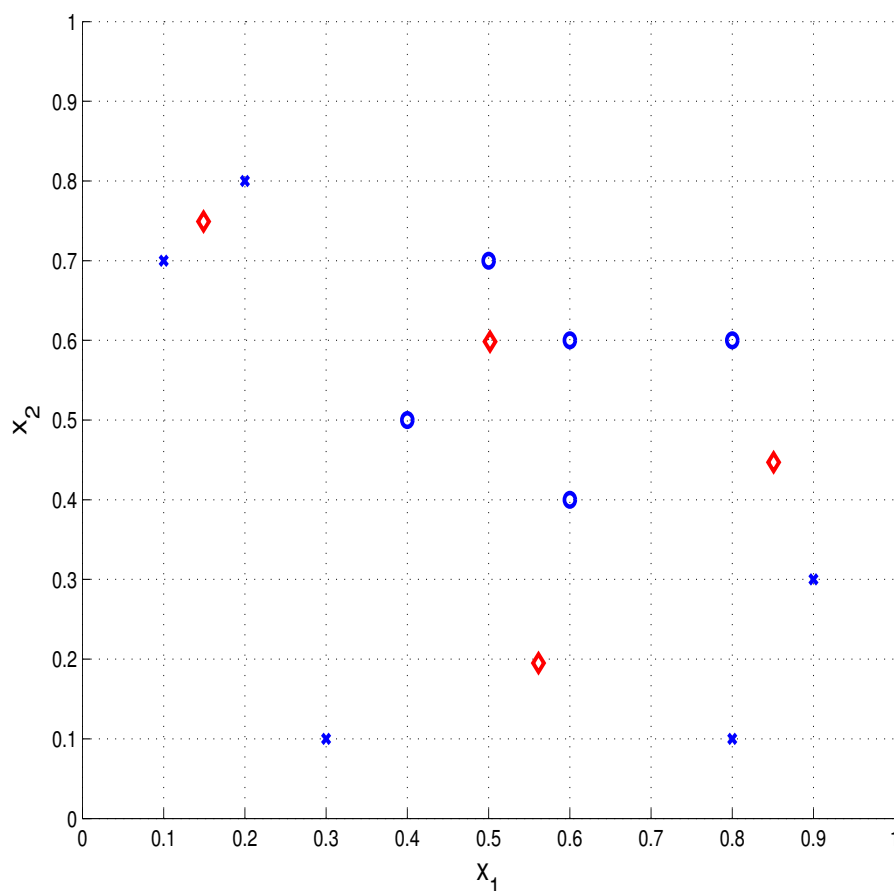


Figure: The input data set and centers(Diamond) found by using k-means clustering algorithms.

We predetermine the number of centers $M = 4$, 4 centers was located by using the k-means clustering algorithm as $c_1 = [0.1490, 0.7490]^T$, $c_2 = [0.8510, 0.4471]^T$, $c_3 = [0.5615, 0.1950]^T$ and $c_4 = [0.5018, 0.5984]^T$.

Lecture 8

We set $\sigma = 1$ as before, and this gives us four basis functions $\phi_1(\mathbf{x})$ $\phi_2(\mathbf{x})$ $\phi_3(\mathbf{x})$ and $\phi_4(\mathbf{x})$. e.g.

$$\phi_1(\mathbf{x}) = \exp \left(-\frac{(x_{1,i} - 0.1490)^2 + (x_{2,i} - 0.7490)^2}{2} \right)$$

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

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

We can write ten linear equations

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

which is

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

Lecture 8

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

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

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

$$\mathbf{w} = [-74.1191, -65.3503, -8.2930, 138.2853]^T$$

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

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

Lecture 8

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

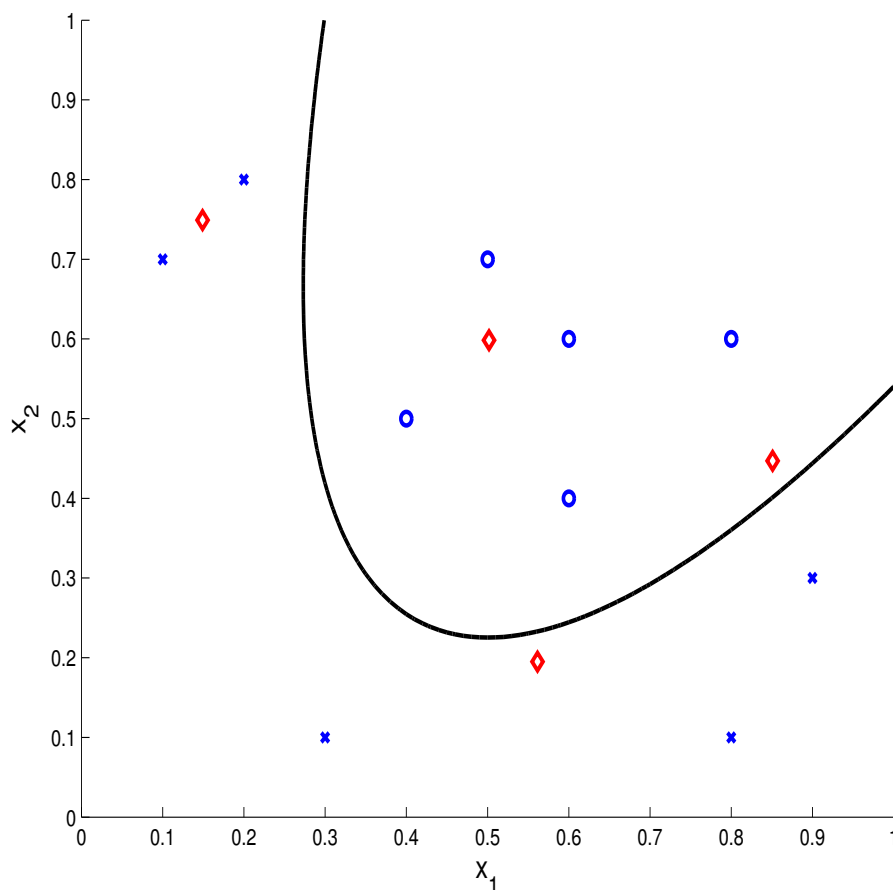


Figure: The input data set, centers and the decision boundary.

The file *kmsRBF.m* is used for calculation in this example.

Lecture 8

A summary of the construction of RBF Classifier with k-means clustering algorithm.

1. Predetermine the number the centers.
2. Use the k-means clustering algorithm to find 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.

Note that k-means clustering algorithm can also be used RBF neural networks center selection for regression.