



# DD2437 – Artificial Neural Networks and Deep Architectures (annda)

## Lecture 5: **Radial basis function NN and introduction to competitive learning**

Pawel Herman

Computational Science and Technology (CST)  
KTH Royal Institute of Technology

January 2018

- Interpolation problem and RBFs
- RBF networks – hybrid learning
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## Lecture overview

- Interpolation problem and radial-basis functions (RBFs)
- RBF networks – hybrid learning
- Weight interpretation
- Competitive mechanisms for unsupervised learning

- **Interpolation problem and RBFs**
- RBF networks – hybrid learning
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The separability of patterns

## *Cover's theorem*

*“A complex pattern classification problem, projected nonlinearly to a high-dimensional space, is more likely to be linearly separable than in low dimensional space, especially if it is not populated too densely.”*

Cover, 1965

- Interpolation problem and RBFs
- RBF networks – hybrid learning
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The separability of patterns

## *Cover's theorem*

*“A complex pattern classification problem, projected nonlinearly to a high-dimensional space, is more likely to be linearly separable than in low dimensional space, especially if it is not populated too densely.”*

Cover, 1965

So, we need to have:

- 1) Many nonlinear mappings  $\varphi_i(\mathbf{x} \in \mathbb{R}^M): \mathbb{R}^M \rightarrow \mathbb{R}^1$ , where  $i = 1, \dots, N$  (large)
- 2) Linear function for separability in high  $N$ -dimensional space

- **Interpolation problem and RBFs**
- RBF networks – hybrid learning
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The radial-basis-function (RBF) technique

Nonlinear mapping with the use of *radial-basis-functions* (RBFs):

$$\varphi_i \left( \left\| \mathbf{x} - \mathbf{x}_i \right\| \right)$$

$\mathbf{x}_i$  – RBF centre

$\|\cdot\|$  – vector norm, often Euclidean

$\varphi_i(r)$  – kernel function, often Gaussian

- Interpolation problem and RBFs
- RBF networks – hybrid learning
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The radial-basis-function (RBF) technique

Nonlinear mapping with the use of *radial-basis-functions* (RBFs):

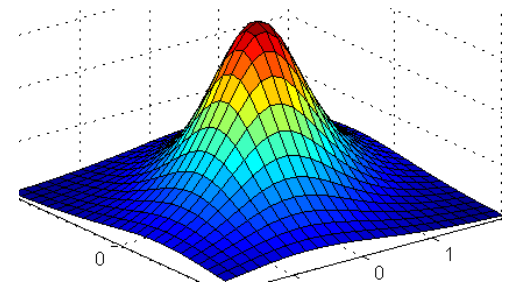
$$\varphi_i \left( \left\| \mathbf{x} - \mathbf{x}_i \right\| \right)$$

$\mathbf{x}_i$  – RBF centre

$\|\cdot\|$  – vector norm, often Euclidean

$\varphi_i(r)$  – kernel function, often Gaussian

$$\varphi_i \left( r = \left\| \mathbf{x} - \mathbf{x}_i \right\| \right) = \exp \left( -\frac{r^2}{2\sigma^2} \right)$$



- **Interpolation problem and RBFs**
- RBF networks – hybrid learning
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The radial-basis-function (RBF) technique

Nonlinear mapping with the use of *radial-basis-functions* (RBFs):

$$\varphi_i \left( \left\| \mathbf{x} - \mathbf{x}_i \right\| \right)$$

$\mathbf{x}_i$  – RBF centre

$\|\cdot\|$  – vector norm, often Euclidean

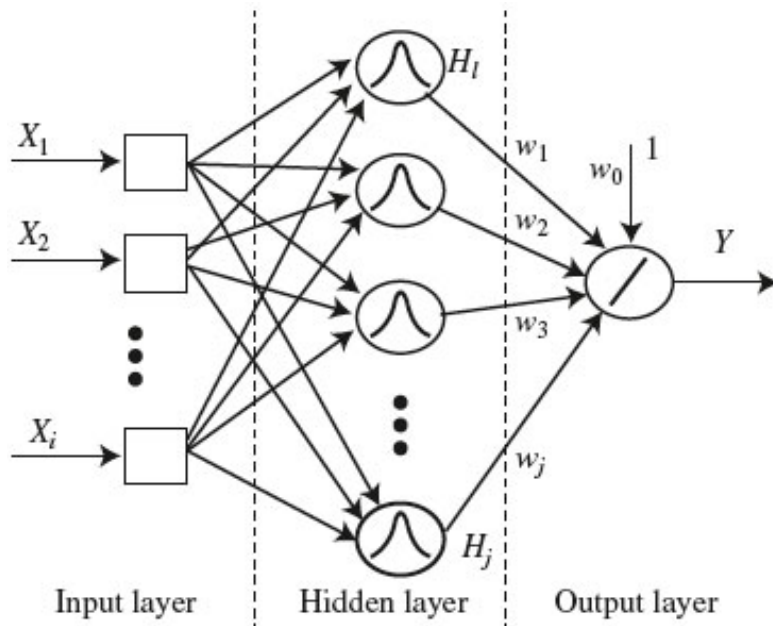
$\varphi_i(r)$  – kernel function, often Gaussian

Linear operation in  $N$ -dimensional space:

$$F(\mathbf{x}) = \sum_{i=1}^N w_i \varphi_i \left( \left\| \mathbf{x} - \mathbf{x}_i \right\| \right)$$

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The RBF neural network concept

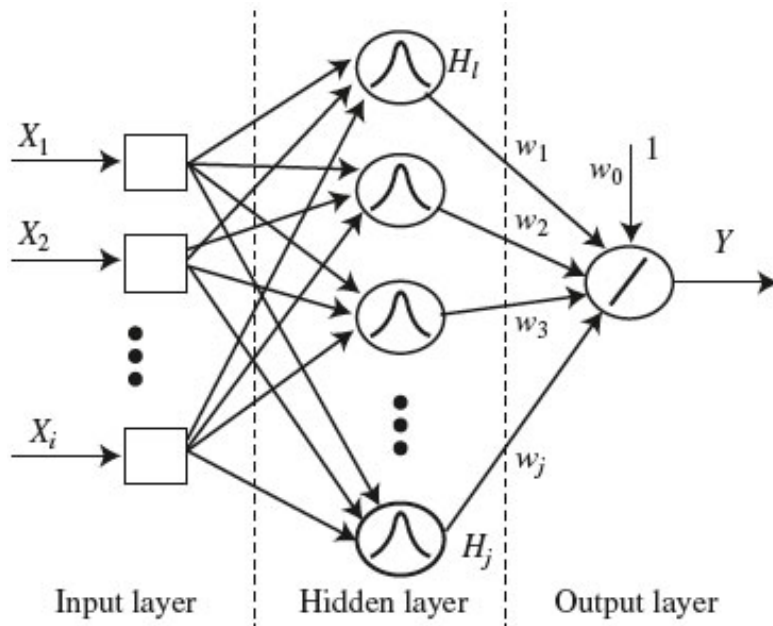


In the *exact interpolation*, the size of the hidden layer,  $N$ , is equal to the number of samples  $n$



- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The RBF neural network concept

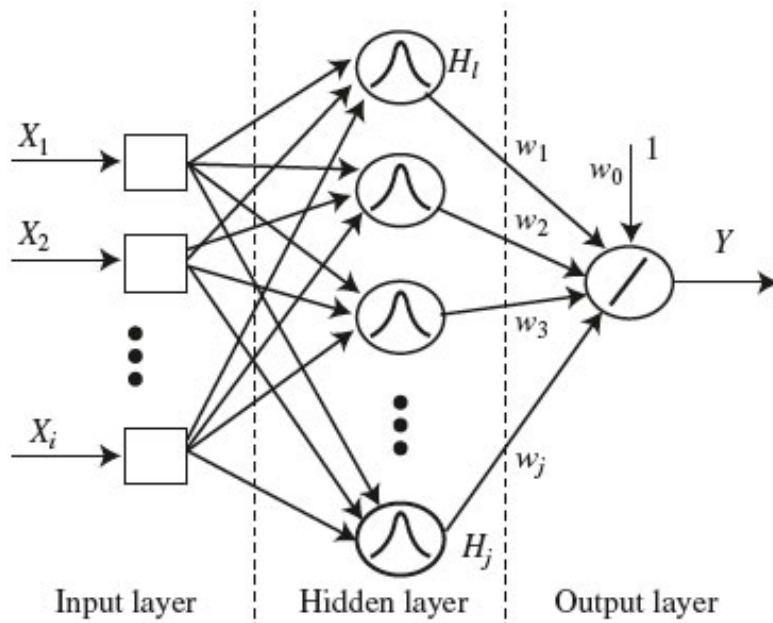


In the *exact interpolation*, the size of the hidden layer,  $N$ , is equal to the number of samples  $n$

**BUT** this is not robust especially if a lot of samples are corrupted with noise!

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# The RBF neural network concept



In the *exact interpolation*, the size of the hidden layer,  $N$ , is equal to the number of samples  $n$

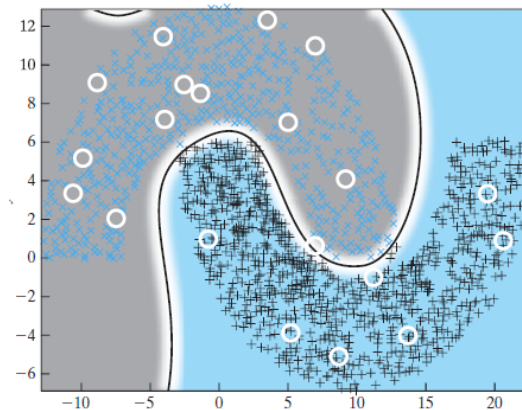
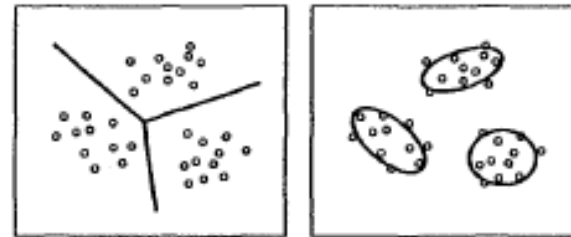
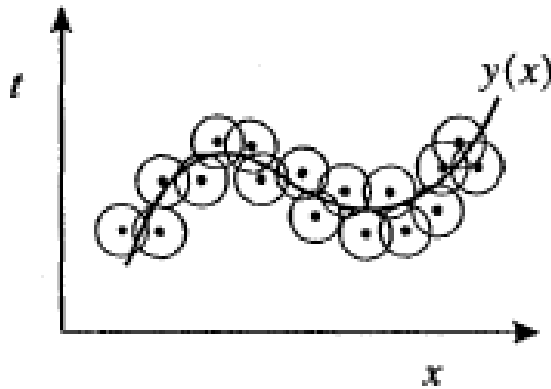
**BUT** this is not robust especially if a lot of samples are corrupted with noise!

We need modifications:

- 1)  $N < n$
- 2) centres  $\mathbf{x}_i$  different from samples
- 3) widths,  $\sigma$ , also differ across RBF nodes
- 4) it is possible to include biases

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

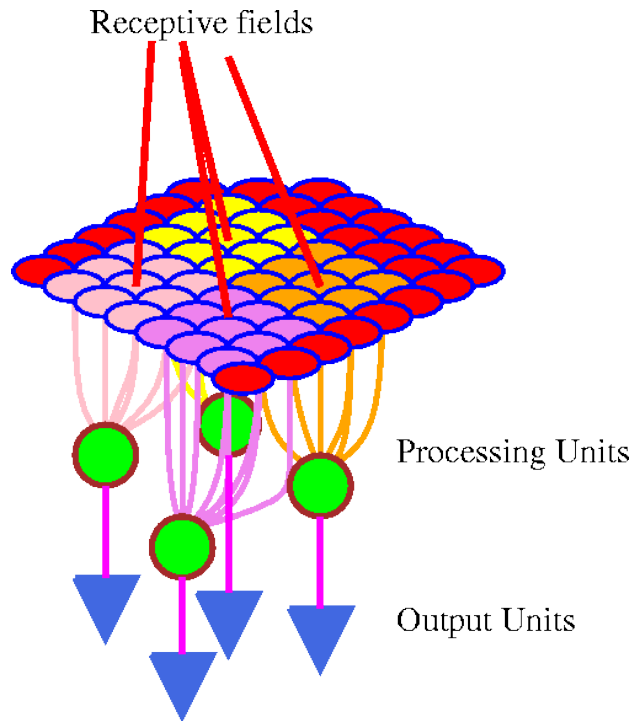
# Examples



See *Bishop (1995)* for Bayesian interpretation of classification with RBF networks

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# Analogy with receptive fields

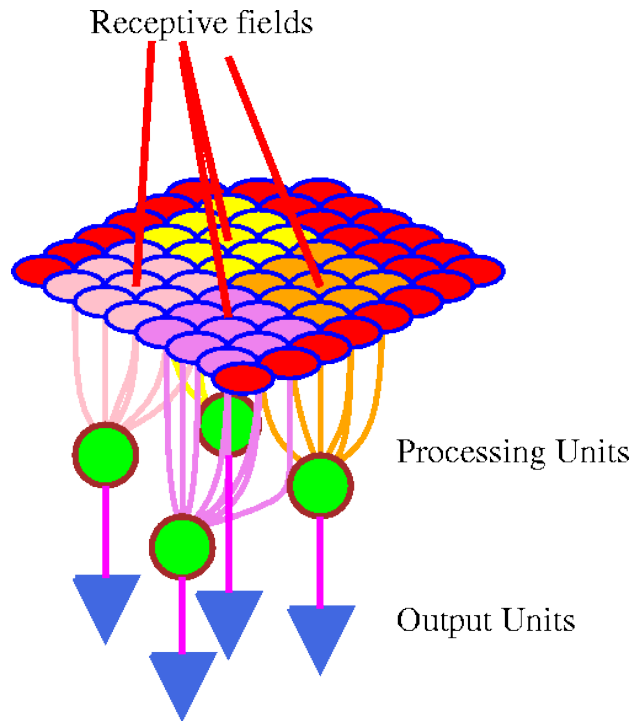


The **receptive field** of a unit is the region of the input space from which a stimulus pattern evokes a response.

Haykin, 2009

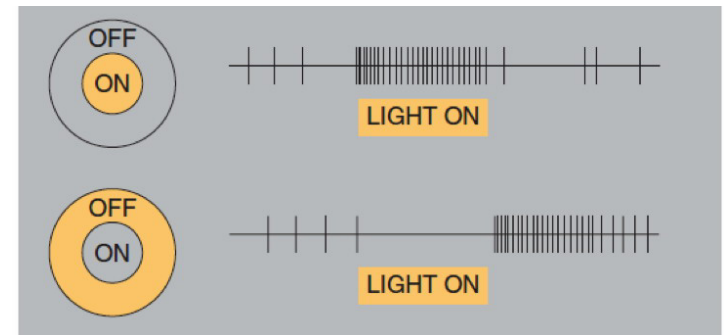
- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# Analogy with receptive fields



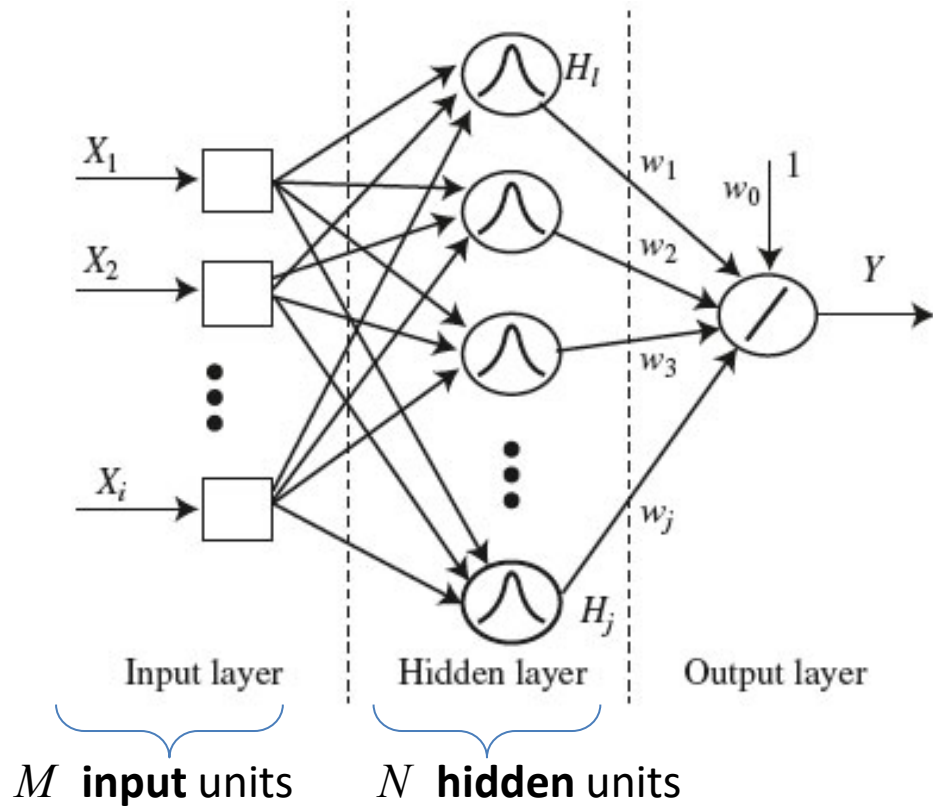
## ***Example:***

Retinal ON-centre and OFF-centre cells



- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

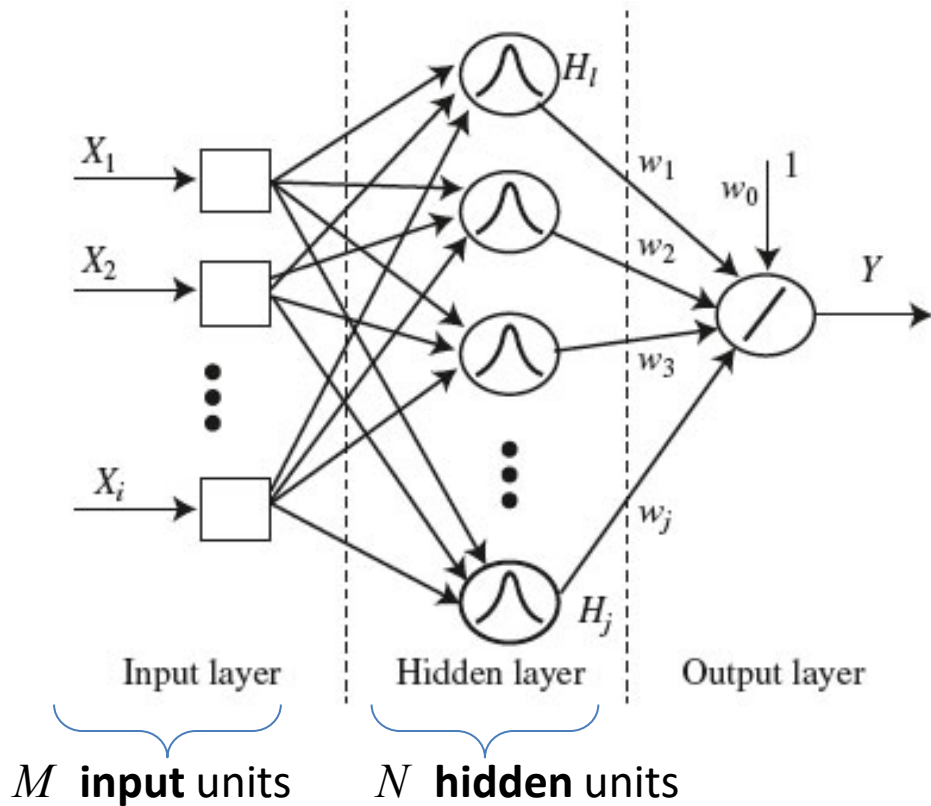
# Hybrid learning algorithm



$n$  **training** samples (each sample is  $N$ -dimensional)

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# Hybrid learning algorithm

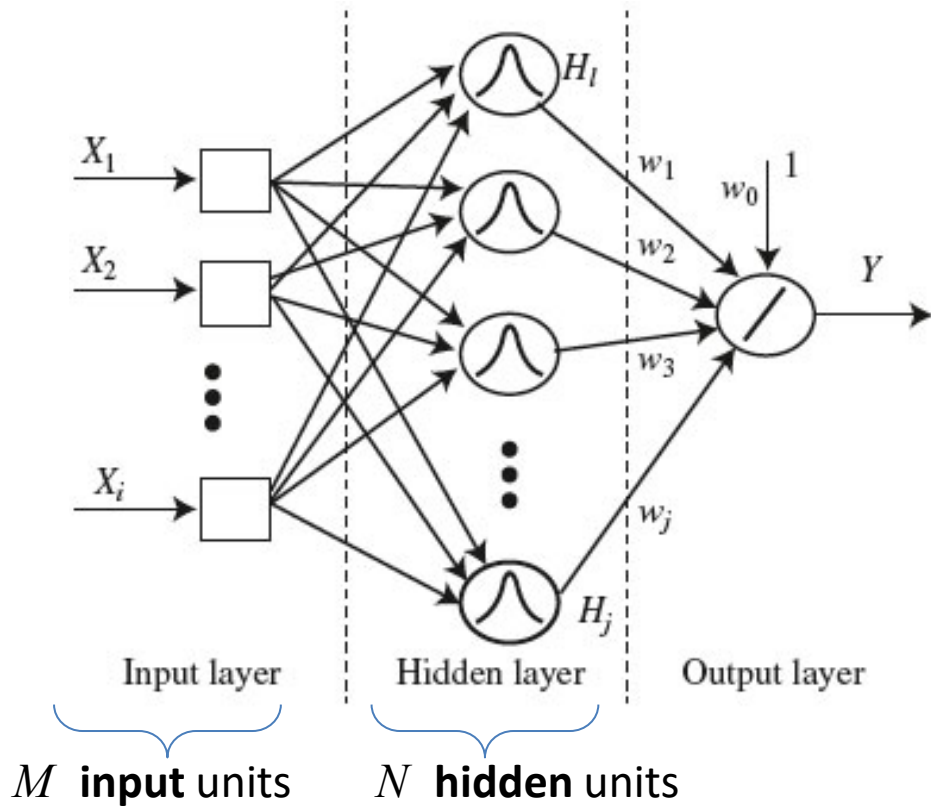


Size of the **input layer** determined by the dimensionality of the input.

$n$  **training** samples (each sample is  $N$ -dimensional)

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# Hybrid learning algorithm



Size of the **input layer** determined by the dimensionality of the input.

## Hidden layer

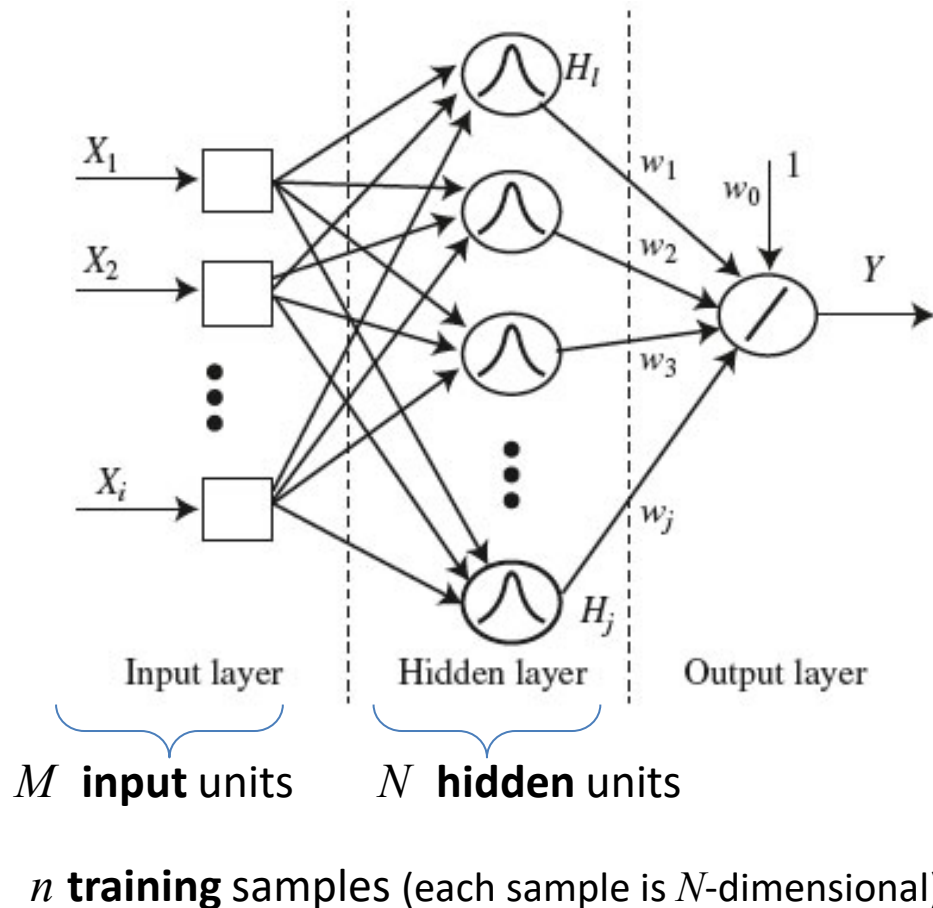
- $N$  has to be decided
- centres and widths of hidden units have to be identified

$n$  **training** samples (each sample is  $N$ -dimensional)



- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

# Hybrid learning algorithm



Size of the **input layer** determined by the dimensionality of the input.

## Hidden layer

- $N$  has to be decided
- centres and widths of hidden units have to be identified

**Output layer** performs a linear mapping – e.g., training with least square methods

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \dots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{n1} & \varphi_{n2} & \dots & \varphi_{nN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## Hybrid learning – RBF layer

- Clustering algorithms
  - k-means clustering
  - clustering with Kohonen feature maps (SOMs), vector quantization (VQ)
  - estimate of the cluster width (variance problem)
- Gaussian mixture models: expectation-maximization (EM) alg.

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## Hybrid learning – RBF layer

- Clustering algorithms
  - k-means clustering
  - clustering with Kohonen feature maps (SOMs), vector quantization (VQ)
  - estimate of the cluster width (variance problem)
- Gaussian mixture models: expectation-maximization (EM) alg.
- Supervised tuning of the RBF parameters with gradient descent

$$\frac{\partial E}{\partial \mu_{ji}} = \sum_n \sum_k \{y_k(\mathbf{x}^n) - t_k^n\} w_{kj} \exp\left(-\frac{\|\mathbf{x}^n - \mu_j\|^2}{2\sigma_j^2}\right) \frac{(x_i^n - \mu_{ji})}{\sigma_j^2}$$

$$\frac{\partial E}{\partial \sigma_j} = \sum_n \sum_k \{y_k(\mathbf{x}^n) - t_k^n\} w_{kj} \exp\left(-\frac{\|\mathbf{x}^n - \mu_j\|^2}{2\sigma_j^2}\right) \frac{\|\mathbf{x}^n - \mu_j\|^2}{\sigma_j^3}$$

Bishop, 1995

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## Hybrid learning – RBF layer

- Clustering algorithms
  - k-means clustering
  - clustering with Kohonen feature maps (SOMs), vector quantization (VQ)
  - estimate of the cluster width (variance problem)
- Gaussian mixture models: expectation-maximization (EM) alg.
- Supervised tuning of the RBF parameters with gradient descent

$$\frac{\partial E}{\partial \mu_{ji}} = \sum_n \sum_k \{y_k(\mathbf{x}^n) - t_k^n\} w_{kj} \exp\left(-\frac{\|\mathbf{x}^n - \mu_j\|^2}{2\sigma_j^2}\right) \frac{(x_i^n - \mu_{ji})}{\sigma_j^2}$$

$$\frac{\partial E}{\partial \sigma_j} = \sum_n \sum_k \{y_k(\mathbf{x}^n) - t_k^n\} w_{kj} \exp\left(-\frac{\|\mathbf{x}^n - \mu_j\|^2}{2\sigma_j^2}\right) \frac{\|\mathbf{x}^n - \mu_j\|^2}{\sigma_j^3}$$

Bishop, 1995

- Subset selection with orthogonal least squares

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## Hybrid learning – output layer

- Least-squares fitting – batch approach

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

- Recursive least-squares estimation (see *Haykin*, RLS algorithm)

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## RBF NN vs MLP

- Hidden units in MLP rely on weighted linear summations of inputs (a matter of interpretation)
  - RBFs rely on distance to prototype vectors

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## RBF NN vs MLP

- Hidden units in MLP rely on weighted linear summations of inputs (a matter of interpretation)
  - RBFs rely on distance to prototype vectors
- In MLPs, function approximation is defined as by a nested sum of weighted summations
  - In RBFs, the approximation is defined by a single weighted sum

- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

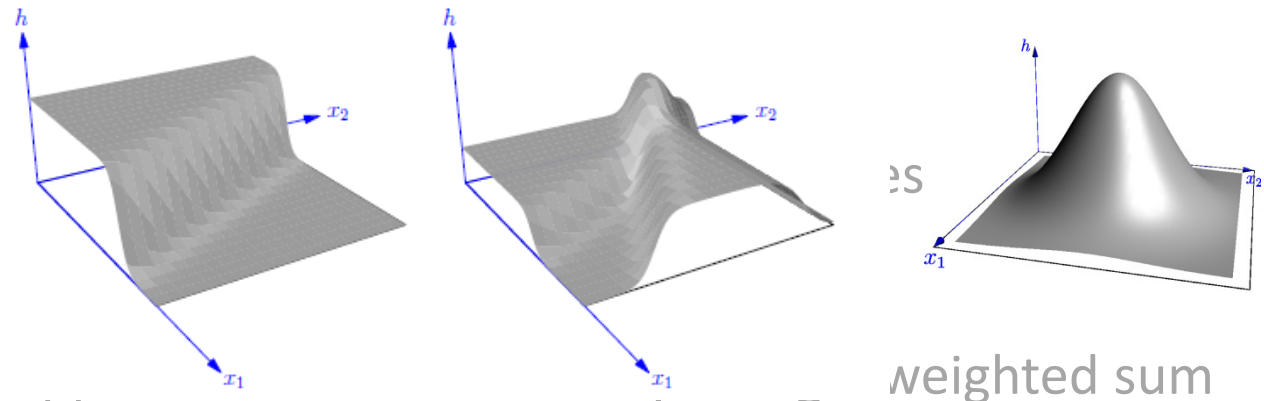
# RBF NN vs MLP

- Hidden units in MLP rely on weighted linear summations of inputs  
(a matter of  $\sum w_i x_i$ )

➤ RBFs rely

- In MLPs, function is a weighted sum

➤ In RBFs, function is a weighted sum



- MLPs form distributed activations (many hidden units contribute to the output for a given input, which partly leads to local minima etc.)
  - In RBFs, very few local basis functions (wrt. input) are activated for a given input



- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## RBF NN vs MLP

- In MLPs, usually all parameters/weights are trained at the same time
  - In RBFs, hybrid two-stage training is used

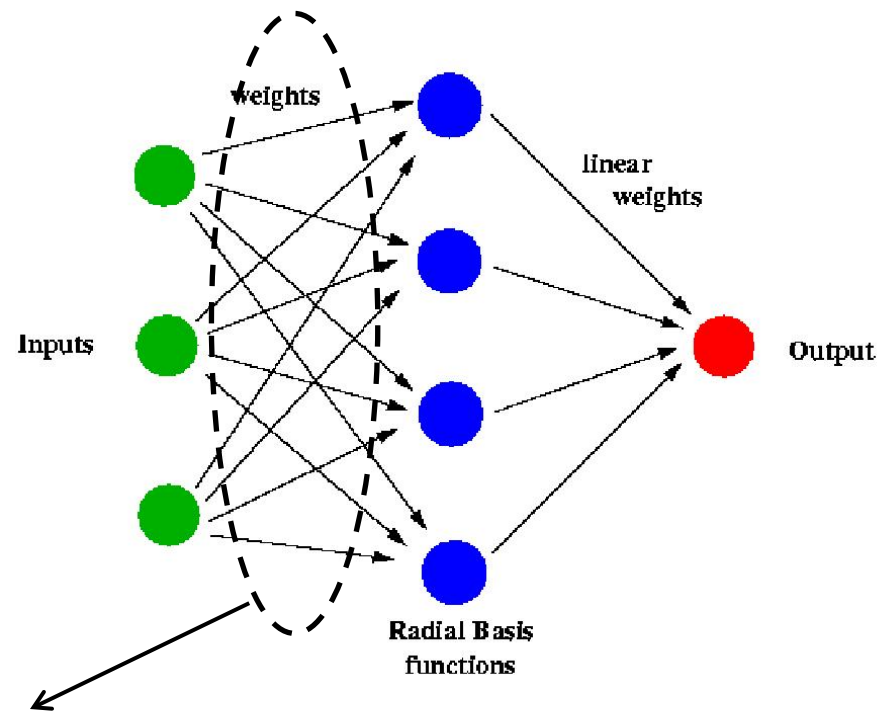
- Interpolation problem and RBFs
- **RBF networks – hybrid learning**
- Weight interpretation in the input space
- Competitive mechanisms for unsupervised learning

## RBF NN vs MLP

- In MLPs, usually all parameters/weights are trained at the same time
  - In RBFs, hybrid two-stage training is used
- MLPs rely on complex multi-layer architecture
  - RBF NNs have simple one-hidden-layer architecture

- Interpolation problem and RBFs
- RBF networks – hybrid learning
- **Weight interpretation in the input space**
- Competitive mechanisms for unsupervised learning

# Interpretation of the weights to hidden layer



Do these connections have weights?