Projection-Based Design Method for
Radial Basis Function Networks

Mehmet Kerem Muezzinoglu

CI Lab. Seminar
February 18, 2004
Radial Basis Function (RBF) Networks

One of the most popular feedforward architectures.

They represent the decomposition of a continuous mapping
\( \varphi(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m \) into \( \ell \) RBFs as

\[
\mathbf{y} = \mathbf{W} \cdot \begin{bmatrix} \phi_1(\mathbf{x}) & \cdots & \phi_\ell(\mathbf{x}) \end{bmatrix}^T \approx \varphi(\mathbf{x}).
\]
Radial Basis Functions

The power of RBF networks is due to selective receptiveness provided by the underlying RBFs.

An admissible RBF $\phi_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$

1. attains its global maximum at its center $\mathbf{c}^i \in \mathbb{R}^n$,
2. decreases along each sequence in $\mathbb{R}^n$ that retreats $\mathbf{c}^i$,
3. is bounded from below, typically by 0.

These properties render an RBF node *sensitive* only to a convex sub-region of $\mathbb{R}^n$ located around its center.

Note that a Gaussian function

$$\phi_i (\mathbf{x}) = \exp \left( -\gamma^i \cdot \| \mathbf{x} - \mathbf{c}^i \|_2^2 \right)$$

constitutes an admissible RBF for all $\gamma^i > 0$. 
RBF Networks are 

*Perfect Interpolators*

An RBF network with \( N \) Gaussian RBF nodes (\( \ell = N \)) can be designed as follows to implement exactly any given combination of \( N \) training pairs \( \{(x^i, y^i)\}_{i=1}^N \).

- Choose \( \gamma^i > 0 \) sufficiently small,
- Assign \( c^i = x^i \),
- Assign \( w_i = y^i, \ i = 1, 2, \ldots, N \),

for each RBF node \( i = 1, 2, \ldots, N \).
RBF Networks are *Universal Approximators*

For each consistent training set $\{(x^i, y^i)\}_{i=1}^{N} \subset \mathbb{R}^n \times \mathbb{R}^m$, and for each $\varepsilon > 0$ there exists an RBF network, possibly with $\ell < N$, such that

$$R := \frac{1}{N} \sum_{i=1}^{N} \| y^i - W \cdot \Phi(x^i) \|^2 < \varepsilon.$$  

RBF network model is one of the two known universal approximators among feed-forward neural networks.
RBF Network Design

Any design method, except of the ones that employ Support Vector techniques, aims an optimum collection of parameters

\[ W_*, c_1^*, \ldots, c_*^*, \gamma_1^*, \ldots, \gamma_*^* \]

which minimizes the empirical risk \( R \).

- Since \( R \) is quadratic and convex in \( W \), calculating the weight matrix is relatively easier, in contrast to other parameters.

- Most design procedures focus on center allocation problem, which could be solved in either supervised or unsupervised way.

- Adjusting widths is usually out of the scope of the methods, or, at most, considered as a post-trimming procedure.

- Network model is always considered as an input to the problem.
Unsupervised Design Methods

Clustering the input data is the most popular center allocation procedure.

Though there are many effective methods that adapt this idea, the clustering problem itself has not been completely solved yet.
Supervised Design Methods

Located initially at random points in the input space, centers can be updated according to a gradient-descent iteration

\[ c^i[k + 1] = c^i[k] - \eta \cdot \nabla_{c^i} R, \]

as done in back-propagation for general feed-forward networks. Note that this idea could be applied also for trimming the widths:

\[ \gamma^i[k + 1] = \gamma^i[k] - \eta \cdot \frac{\partial R}{\partial \gamma^i}. \]

In fact, a gradient-descent weight update scheme

\[ W[k + 1] = W[k] - \eta \cdot \nabla_W R, \]

could be incorporated, and these three iterations could perform sequentially the search of optimum parameters.
Optimum Widths is a Linearly-Constrained Problem

• Considered width update mechanisms do not restrict widths to the positive half-space $\gamma^i \geq 0$, $i = 1, 2, \ldots, \ell$.

• However, the admissible Gaussian RBFs are the ones with positive widths.
Projection Operator

If the inequality constraints of a problem define an affine border on the search space, then any attempt to exceed the border can be projected onto the border via a linear operator $P$.

$$P = I - A^T \cdot (A \cdot A^T)^+ \cdot A$$
Proposed Design Technique

Apply sequentially the three iterations

\[
\begin{align*}
  \mathbf{W}[k + 1] &= \mathbf{W}[k] - \eta \cdot \nabla_{\mathbf{W}} R, \\
  \mathbf{c}^i[k + 1] &= \mathbf{c}^i[k] - \eta \cdot \nabla_{\mathbf{c}^i} R, \quad i = 1, 2, \ldots, \ell, \\
  \gamma[k + 1] &= \gamma[k] - \eta \cdot \mathbf{G} \cdot \nabla_{\gamma} R,
\end{align*}
\]

where

\[
\mathbf{G} = \begin{cases} 
  \mathbf{I}, & \text{if } \gamma^i \geq 0 \ \forall i \\
  \mathbf{P}, & \text{otherwise}
\end{cases}
\]
Some RBF Nodes are Pruned

Note that the projection mechanism keeps the widths in the closed half-space $\gamma^i \geq 0$, $i = 1, 2, \ldots, \ell$.

Therefore, some of the widths may stick to zero along the iterations. Such nodes no longer implement an RBF, but a constant function, namely $\phi = 1$.

We prune them from the hidden-layer, so simplify the architecture.
Thank You!