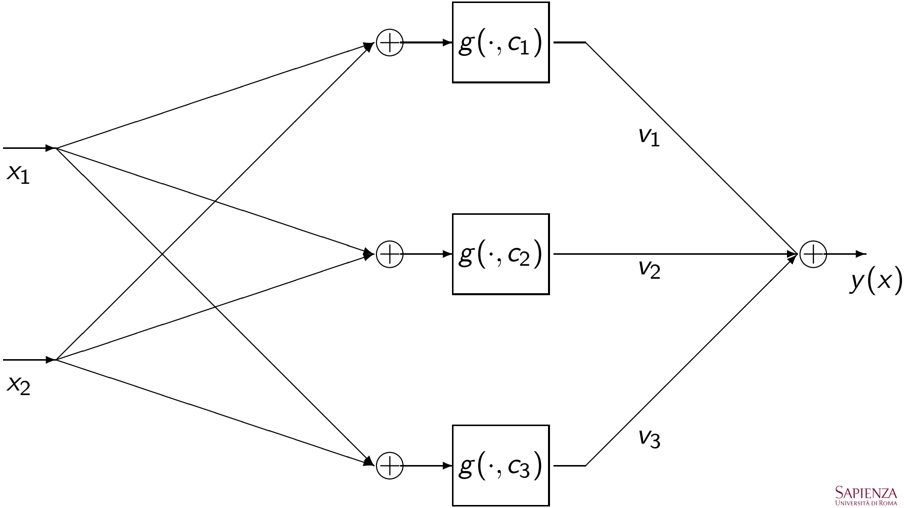
Optimization Methods for Machine Learning
Decomposition methods for RBF networks

Laura Palagi
http://www.dis.uniroma1.it/∼palagi

Dipartimento di Ingegneria informatica automatica e gestionale A. Ruberti
Sapienza Università di Roma

Via Ariosto 25

RBF network



Gradient RBF

Supervised selection of the centers

Step 1. (Minimization with respect to weights v)

Compute v^{k+1} solving LLSQ

Gradient RBF

$\boxed{\text{Supervised selection of the centers}}$

Step 1. (Minimization with respect to weights v)

Compute v^{k+1} solving LLSQ

Step 2. (Minimization with respect to centers c)

Set

$$\hat{c}^{k+1} = \begin{cases} c^k & \text{if } \|\nabla_c E(c^k, v^{k+1})\| \le \xi_2^k \\ & \end{cases}$$

ξ_2^k decreasing tolerance $\xi_2^{k+1} \le \theta \xi_2^k$

Gradient RBF

Supervised selection of the centers

Step 1. (Minimization with respect to weights v)

Compute v^{k+1} solving LLSQ

Step 2. (Minimization with respect to centers c)

Set

$$\widehat{c}^{k+1} = \begin{cases} c^k & \text{if } \|\nabla_c E(c^k, v^{k+1})\| \le \xi_2^k \\ c^k - \eta^k \nabla_c E(c^k, v^{k+1}) & \text{otherwise} \end{cases}$$

with η^k by an Armijo lineseach

$$E(v^{k+1}, c^k + \eta^k d^k) \le E(v^{k+1}, c^k) - \gamma \eta^k \|\nabla_c E(v^{k+1}, c^k)\|^2$$

Gradient RBF

Supervised selection of the centers

Step 1. (Minimization with respect to weights v)

Compute v^{k+1} solving LLSQ

Step 2. (Minimization with respect to centers c)

Set

$$\widehat{c}^{k+1} = \begin{cases} c^k & \text{if } \|\nabla_c E(c^k, v^{k+1})\| \leq \xi_2^k \\ c^k - \eta^k \nabla_c E(c^k, v^{k+1}) & \text{otherwise} \end{cases}$$

with η^k by an Armijo lineseach $E_{ref} = E(v^{k+1}, c^k + \eta^k d^k)$

$$E(v^{k+1}, c^k + \eta^k d^k) \leq E(v^{k+1}, c^k) - \gamma \eta^k \|\nabla_c E(v^{k+1}, c^k)\|^2$$

Gradient RBF

Supervised selection of the centers

Step 1. (Minimization with respect to weights v)

Compute v^{k+1} solving LLSQ

Step 2. (Minimization with respect to centers c)

Set

$$\widehat{c}^{k+1} = \begin{cases} c^k & \text{if } \|\nabla_c E(c^k, v^{k+1})\| \le \xi_2^k \\ c^k - \eta^k \nabla_c E(c^k, v^{k+1}) & \text{otherwise} \end{cases}$$

with η^k by an Armijo lineseach $E_{ref} = E(v^{k+1}, c^k + \eta^k d^k)$ allow more freedom

Gradient RBF

Supervised selection of the centers

Step 1. (Minimization with respect to weights v)

Compute v^{k+1} solving LLSQ

Step 2. (Minimization with respect to centers c)

Set

$$\widehat{c}^{k+1} = \begin{cases} c^k & \text{if } \|\nabla_c E(c^k, v^{k+1})\| \leq \xi_2^k \\ c^k - \eta^k \nabla_c E(c^k, v^{k+1}) & \text{otherwise} \end{cases}$$

with η^k by an Armijo lineseach $E_{ref} = E(v^{k+1}, c^k + \eta^k d^k)$
allow more freedom

Set c^{k+1}: $E(c^{k+1}, v^{k+1}) \leq E_{ref}$

Convergence of Two-block decompositions in RBF

Theorem (Buzzi, Grippo, Sciandrone 2000, [1])

Let $\{(v^k, c^k)\}$ be an infinite sequence generated the Two-blocks Algorithm. Then:

i) $\{(v^k, c^k)\}$ *has limit points;*

ii) *the sequence* $\{E(v^k, c^k)\}$ *converges to a limit;*

iii) *every limit point of* $\{(v^k, c^k)\}$ *is a stationary point of* E.

SAPIENZA
UNIVERSITÀ DI ROMA

More than two-blocks decomposition in RBF

Even the approximate minimization with respect to the centers can be very expensive when N is large.

More than two-blocks decomposition in RBF

Even the approximate minimization with respect to the centers can be very expensive when N is large. The variables (v, c) are partitioned into the $N+1$ blocks, corresponding to the weights v and the N center positions $c_1, c_2 \ldots, c_N$.

Step 2. (Minimization with respect to centers)

For $j = 1, \ldots, N$.

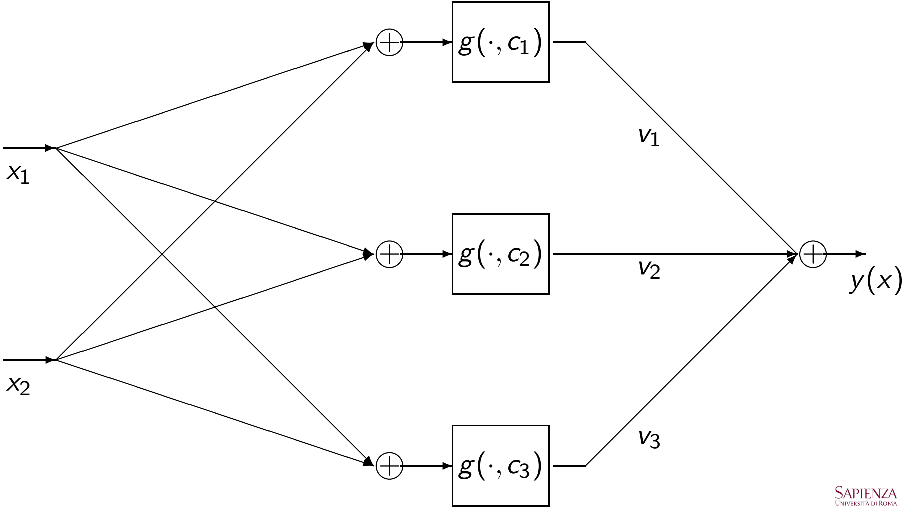
$$c_j^{k+1} = c_j^k - \eta_j^k \nabla_{c_j} E(v^{k+1}, c_1^{k+1}, \ldots, c_{j-1}^{k+1}, c_j, c_{j+1}^k, \ldots, c_N^k)$$

with η_j^k satisfying

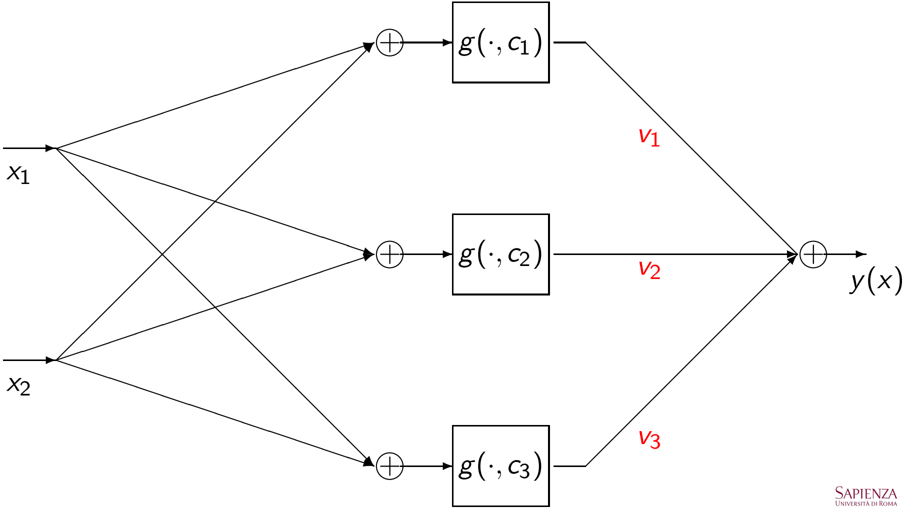
$$E(v^{k+1}, c^{k+1}) \leq E(v^{k+1}, c^k) - \gamma \eta_j^k \|\nabla_{c_j} E(v^{k+1}, c^k)\|^2$$

Set $\xi_j^{k+1} \leq \theta \xi_j^k$

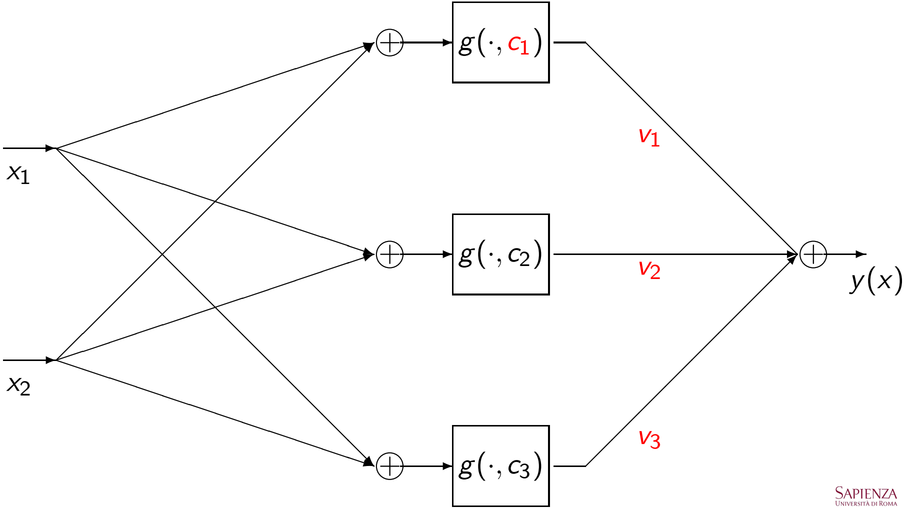
RBF network



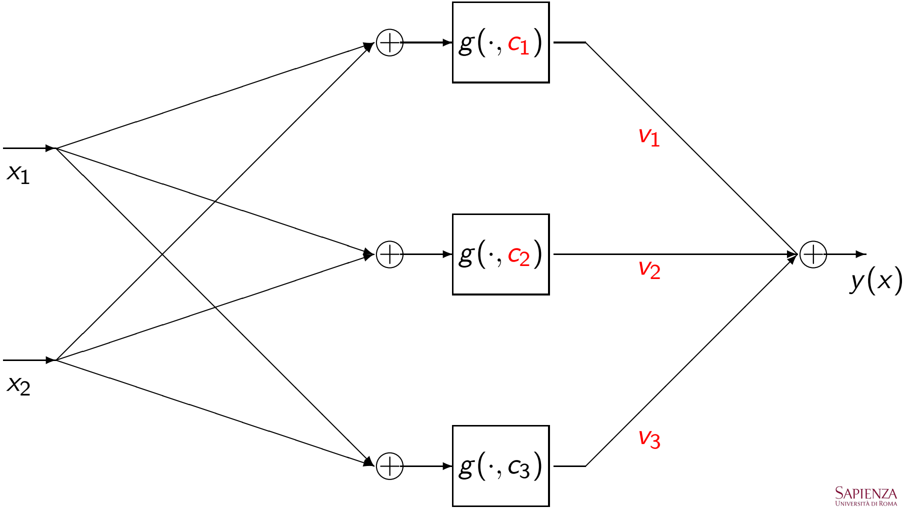
RBF network



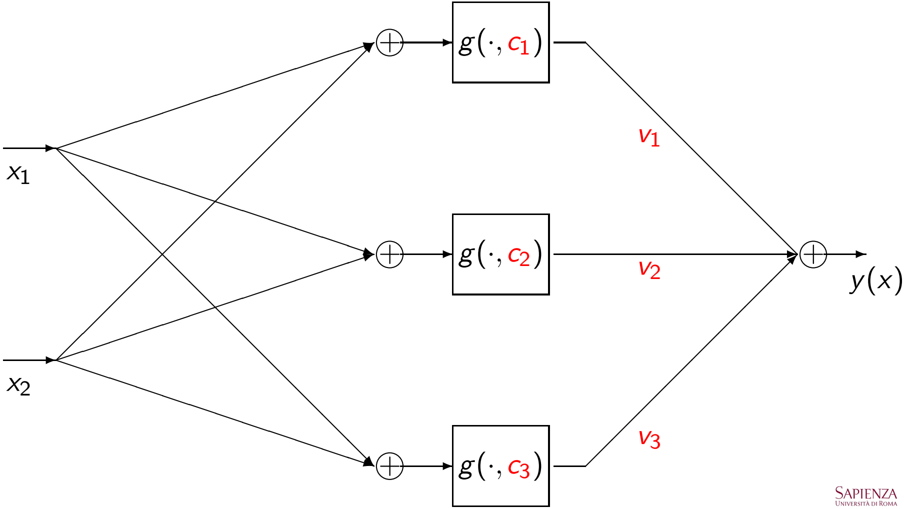
RBF network



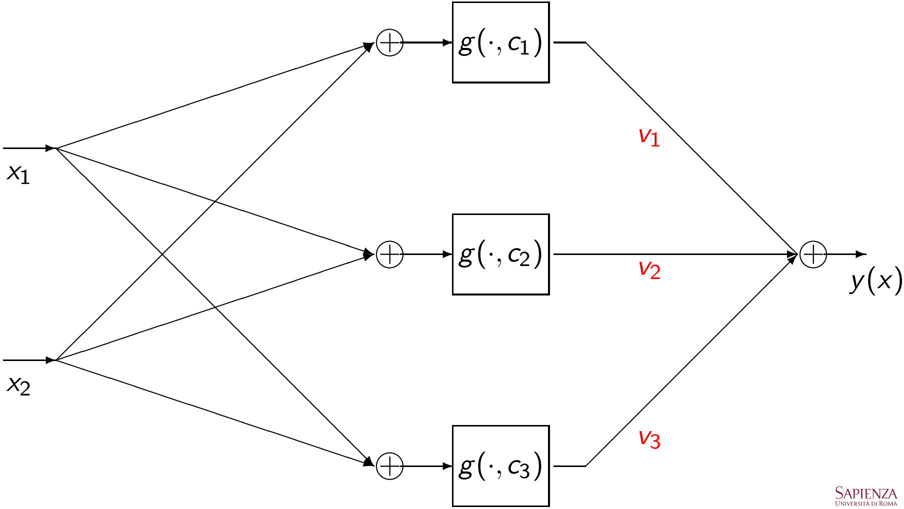
RBF network



RBF network



RBF network



Decomposition methods

PROs

- Problem can be decomposed into much smaller subproblems
- Gradient type iteration
- Convergence holds w/out particular assumption on the functions E_p

SAPIENZA
UNIVERSITÀ DI ROMA

Decomposition methods

PROs

- Problem can be decomposed into much smaller subproblems
- Gradient type iteration
- Convergence holds w/out particular assumption on the functions E_p

$\boxed{\text{Computational price}}$

Decomposition methods

PROs

- Problem can be decomposed into much smaller subproblems
- Gradient type iteration
- Convergence holds w/out particular assumption on the functions E_p

Computational price

Cost of line-search

References

C Buzzi, L Grippo, and Marco Sciandrone.
Convergent decomposition techniques for training rbf neural networks.
Neural Computation, 13(8):1891–1920, 2001.

Luigi Grippo and Marco Sciandrone.
Globally convergent block-coordinate techniques for unconstrained optimization.
Optimization methods and software, 10(4):587–637, 1999.