Reti RBF*

Corso di MOBD

Roma 30 Ottobre 2017

* Grippo, L., and M. Sciandrone. *”Metodi di ottimizzazione per le reti neurali.* Rapporto Tecnico (2003): 09-03.

Funzioni di base radiali

Reti RBF
● Funzioni di base radiali
● Funzioni di base radiali
● Interpolazione II
● Reti RBF regolarizzate
● RBF regolarizzate
● RBF regolarizzate
● RBF generalizzate
● Esempio

Addestramento RBF

Problema di interpolazione: dati P punti $\{x^p \in \mathbb{R}^n,\, p = 1, \ldots, P\}$ e un corrispondente insieme di numeri reali $\{y^p \in \mathbb{R},\, p = 1, \ldots, P\}$ determinare una funzione $y : \mathbb{R}^n \Rightarrow \mathbb{R}$ in una classe di funzioni assegnate tale che

$$y(x^p) = y^p, \quad p = 1, \ldots, P$$

Funzioni di base radiali

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Problema di interpolazione: dati P punti $\{x^p \in \mathbb{R}^n,\, p = 1, \dots, P\}$ e un corrispondente insieme di numeri reali $\{y^p \in \mathbb{R},\, p = 1, \dots, P\}$ determinare una funzione $y : \mathbb{R}^n \Rightarrow \mathbb{R}$ in una classe di funzioni assegnate tale che

$$y(x^p) = y^p, \quad p = 1, \dots, P$$

Una particolare tecnica di interpolazione consiste nella scelta di y della forma:

$$y(x) = \sum_{p=1}^{P} w_p \phi\left(\|x - x^p\|\right)$$

dove $\phi : \mathbb{R}^+ \Rightarrow \mathbb{R}^+$ è una funzione continua, chiamata **funzione di base radiale**. I coefficienti w_p sono detti **pesi** e i i punti di interpolazione **centri**.

Funzioni di base radiali

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Esempi di funzioni di base radiali sono:

Funzioni di base radiali

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Esempi di funzioni di base radiali sono:

1. $\phi(r) = r$ lineare

Funzioni di base radiali

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Esempi di funzioni di base radiali sono:

1. $\phi(r) = r$ lineare
2. $\phi(r) = e^{-r^2}$ gaussiana

Funzioni di base radiali

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Esempi di funzioni di base radiali sono:

1. $\phi(r) = r$ lineare
2. $\phi(r) = e^{-r^2}$ gaussiana
3. $\phi(r) = \left(r^2 + \sigma^2\right)^{-1/2}$ multiquadrica inversa

Funzioni di base radiali

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Esempi di funzioni di base radiali sono:

1. $\phi(r) = r$ lineare
2. $\phi(r) = e^{-r^2}$ gaussiana
3. $\phi(r) = \left(r^2 + \sigma^2\right)^{-1/2}$ multiquadrica inversa
4. $\phi(r) = \left(r^2 + \sigma^2\right)^{1/2}$ multiquadrica diretta

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Funzioni di base radiali

Esempi di funzioni di base radiali sono:

1. $\phi(r) = r$ lineare
2. $\phi(r) = e^{-r^2}$ gaussiana
3. $\phi(r) = \left(r^2 + \sigma^2\right)^{-1/2}$ multiquadrica inversa
4. $\phi(r) = \left(r^2 + \sigma^2\right)^{1/2}$ multiquadrica diretta

dove $r \geq 0$ e σ è una costante positiva.

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Funzioni di base radiali

Esempi di funzioni di base radiali sono:

1. $\phi(r) = r$ lineare
2. $\phi(r) = e^{-r^2}$ gaussiana
3. $\phi(r) = \left(r^2 + \sigma^2\right)^{-1/2}$ multiquadrica inversa
4. $\phi(r) = \left(r^2 + \sigma^2\right)^{1/2}$ multiquadrica diretta

dove $r \geq 0$ e σ è una costante positiva.

Scelta la funzione ϕ, usando l'espressione di $y(x)$ si ottiene la condizione di interpolazione:

$$\Phi w = y$$

dove $y = (y^1, \ldots, y^P)^T$, Φ (**matrice di interpolazione**) è la matrice $P \times P$ il cui generico elemento è

$$\Phi_{ij} = \phi\left(\|x^i - x^j\|\right)$$

Interpolazione II

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Sotto opportune ipotesi sui punti di interpolazione (che devono essere tutti distinti), si dimostra che la matrice Φ è non singolare.

Interpolazione II

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Sotto opportune ipotesi sui punti di interpolazione (che devono essere tutti distinti), si dimostra che la matrice Φ è non singolare.

In questo caso si dimostra l'invertibilità della matrice Φ per tutte le scelte di ϕ riportate, e per la gaussiana e per la multiquadrica inversa si dimostra che le corrispondenti matrici di interpolazione sono definite positive.

Reti RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Le tecniche di **regolarizzazione** sono studiate nell'ambito dei problemi di approssimazione con insieme di campioni disponibili finito.

Reti RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Le tecniche di **regolarizzazione** sono studiate nell'ambito dei problemi di approssimazione con insieme di campioni disponibili finito.

Supponiamo di voler approssimare una funzione $f : \mathbb{R}^n \Rightarrow \mathbb{R}$, e sia dato il training set

$$\mathcal{T} = \{(x^p, y^p), p = 1, \dots, P, y^p = f(x^p)\}$$

Esistono infinite funzioni che lo risolvono e ci possono essere errori di misura nel training set.

Reti RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Le tecniche di **regolarizzazione** sono studiate nell'ambito dei problemi di approssimazione con insieme di campioni disponibili finito.

Supponiamo di voler approssimare una funzione $f : \mathbb{R}^n \Rightarrow \mathbb{R}$, e sia dato il training set

$$\mathcal{T} = \{(x^p, y^p), p = 1, \dots, P, y^p = f(x^p)\}$$

Esistono infinite funzioni che lo risolvono e ci possono essere errori di misura nel training set.

Le tecniche di regolarizzazione determinano la funzione approssimante minimizzando un funzionale costituito da due termini: il primo misura la distanza da $\mathcal{T}$, il secondo penalizza la violazione delle condizioni di regolarità sull'approssimante (sfrutta le informazioni note):

$$\mathcal{E}(y) = \mathcal{E}_1(y) + \mathcal{E}_2(y) = \frac{1}{2}\sum_{p=1}^{P}[y^p - y(x^p)]^2 + \frac{1}{2}\lambda\|\mathcal{P}y\|^2$$

dove λ è il parametro di regolarizzazione e $\mathcal{P}$ è un operatore differenziale che contiene le informazioni che si hanno sulla funzione da approssimare.

RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Sotto opportune ipotesi su $\mathcal{P}$, si dimostra che la $y(x)$ che minimizza $\mathcal{E}(y)$ è

$$y(x) = \sum_{p=1}^{P} w_p \phi(\|x - x^p\|)$$

con ϕ funzione di base radiale e w è soluzione del sistema lineare:

$$(\Phi + \lambda I)w = y$$

RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

Sotto opportune ipotesi su $\mathcal{P}$, si dimostra che la $y(x)$ che minimizza $\mathcal{E}(y)$ è

$$y(x) = \sum_{p=1}^{P} w_p \phi(\|x - x^p\|)$$

con ϕ funzione di base radiale e w è soluzione del sistema lineare:

$$(\Phi + \lambda I)w = y$$

Questa funzione y può essere interpretata come l'uscita di una rete feedforward in cui la funzione di attivazione dei neuroni dello strato nascosto è la funzione ϕ.

RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

1. hanno un solo strato nascosto

RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

1. hanno un solo strato nascosto
2. i neuroni dello strato nascosto hanno come funzione di attivazione una funzione di base radiale e l'argomento della funzione è data dalla distanza tra il vettore di ingresso e il centro dell'unità

RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

1. hanno un solo strato nascosto
2. i neuroni dello strato nascosto hanno come funzione di attivazione una funzione di base radiale e l'argomento della funzione è data dalla distanza tra il vettore di ingresso e il centro dell'unità
3. il numero di neuroni dello strato nascosto è pari al numero di elementi del training set P

RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

1. hanno un solo strato nascosto
2. i neuroni dello strato nascosto hanno come funzione di attivazione una funzione di base radiale e l'argomento della funzione è data dalla distanza tra il vettore di ingresso e il centro dell'unità
3. il numero di neuroni dello strato nascosto è pari al numero di elementi del training set P
4. lo strato di uscita ha un unico neurone che effettua una combinazione lineare delle uscite dei neuroni dello strato nascosto.

RBF regolarizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

1. hanno un solo strato nascosto
2. i neuroni dello strato nascosto hanno come funzione di attivazione una funzione di base radiale e l'argomento della funzione è data dalla distanza tra il vettore di ingresso e il centro dell'unità
3. il numero di neuroni dello strato nascosto è pari al numero di elementi del training set P
4. lo strato di uscita ha un unico neurone che effettua una combinazione lineare delle uscite dei neuroni dello strato nascosto.

Per ogni funzione continua f definita su un sottoinsieme compatto S di $\mathbb{R}^n$ esiste una rete RBF regolarizzata ovvero una funzione $y(x) = \sum_{p=1}^{P} w_p \phi(\|x - x^p\|)$ tale che per ogni $x \in S$ e per ogni $\epsilon > 0$ risulta

$$|f(x) - y(x)| < \epsilon$$

RBF generalizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

La corrispondenza uno a uno tra elementi del training set e i termini della funzione $y(x)$ è troppo onerosa perchè determinare w richiede la soluzione di un sistema $P \times P$ che potrebbe essere proibitivo.

RBF generalizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

La corrispondenza uno a uno tra elementi del training set e i termini della funzione $y(x)$ è troppo onerosa perchè determinare w richiede la soluzione di un sistema $P \times P$ che potrebbe essere proibitivo.

Per questo motivo sono state introdotte le reti RBF generalizzate in cui il numero N dei neuroni dello strato nascosto è di molto minore di P e i centri c_i non coincidono necessariamente con i vettori x^p del training set. La funzione approssimante diventa quindi:

$$y(x) = \sum_{i=1}^{N} w_i \phi(\|x - c_i\|)$$

RBF generalizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

La corrispondenza uno a uno tra elementi del training set e i termini della funzione $y(x)$ è troppo onerosa perchè determinare w richiede la soluzione di un sistema $P \times P$ che potrebbe essere proibitivo.

Per questo motivo sono state introdotte le reti RBF generalizzate in cui il numero N dei neuroni dello strato nascosto è di molto minore di P e i centri c_i non coincidono necessariamente con i vettori x^p del training set. La funzione approssimante diventa quindi:

$$y(x) = \sum_{i=1}^{N} w_i \phi(\|x - c_i\|)$$

Devono essere determinati sia w che i centri, quindi il legame ingresso-uscita dipende in modo nonlineare dai parametri come nelle reti multistrato

RBF generalizzate

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF

La corrispondenza uno a uno tra elementi del training set e i termini della funzione $y(x)$ è troppo onerosa perchè determinare w richiede la soluzione di un sistema $P \times P$ che potrebbe essere proibitivo.

Per questo motivo sono state introdotte le reti RBF generalizzate in cui il numero N dei neuroni dello strato nascosto è di molto minore di P e i centri c_i non coincidono necessariamente con i vettori x^p del training set. La funzione approssimante diventa quindi:

$$y(x) = \sum_{i=1}^{N} w_i \phi(\|x - c_i\|)$$

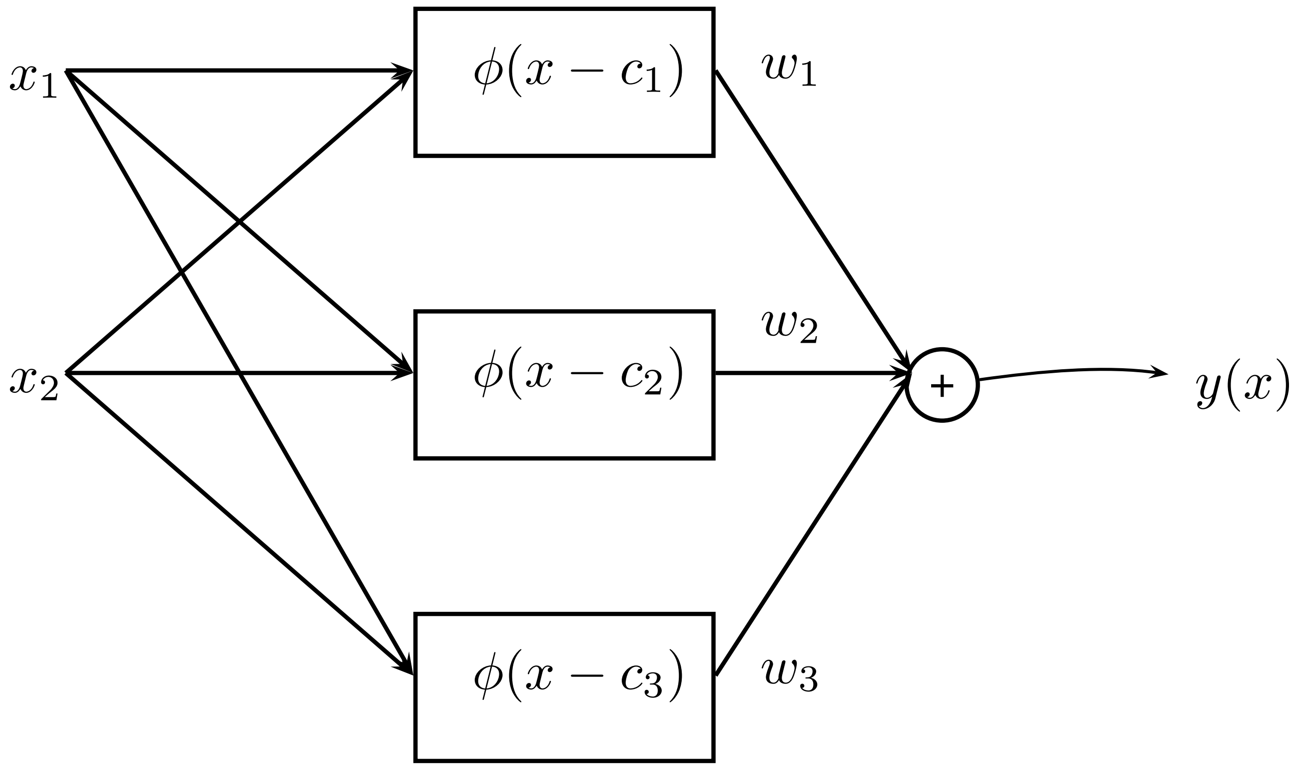
Devono essere determinati sia w che i centri, quindi il legame ingresso-uscita dipende in modo nonlineare dai parametri come nelle reti multistrato

Le RBF generalizzate includono le RBF regolarizzate, quindi sono approssimatori universali

Esempio

Reti RBF
- Funzioni di base radiali
- Funzioni di base radiali
- Interpolazione II
- Reti RBF regolarizzate
- RBF regolarizzate
- RBF regolarizzate
- RBF generalizzate
- Esempio

Addestramento RBF



Addestramento

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Addestrare una rete RBF generalizzata vuol dire individuare il vettore $w \in \mathbb{R}^N$ e il vettore $c \in \mathbb{R}^{nN}$ sulla base del training set

Addestramento

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Addestrare una rete RBF generalizzata vuol dire individuare il vettore $w \in \mathbb{R}^N$ e il vettore $c \in \mathbb{R}^{nN}$ sulla base del training set

Sono possibili due diverse strategie di addestramento:

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Addestramento

Addestrare una rete RBF generalizzata vuol dire individuare il vettore $w \in \mathbb{R}^N$ e il vettore $c \in \mathbb{R}^{nN}$ sulla base del training set

Sono possibili due diverse strategie di addestramento:

1. scelta **non supervisionata** dei centri e **supervisionata** dei pesi

Addestramento

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Addestrare una rete RBF generalizzata vuol dire individuare il vettore $w \in \mathbb{R}^N$ e il vettore $c \in \mathbb{R}^{nN}$ sulla base del training set

Sono possibili due diverse strategie di addestramento:

1. scelta **non supervisionata** dei centri e **supervisionata** dei pesi
2. scelta **supervisionata** sia dei centri che dei pesi

Scelta non supervisionata dei centri

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Si fissano prima i centri o scegliendoli **casualmente** tra i vettori di ingresso del training set oppure con tecniche di **clustering**

Scelta non supervisionata dei centri

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Si fissano prima i centri o scegliendoli **casualmente** tra i vettori di ingresso del training set oppure con tecniche di **clustering**

Si scelgono poi i pesi w minimizzando la funzione di errore

$$E(w) = \frac{1}{2} \sum_{p=1}^{P} \left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p \right)^2 + \frac{\rho}{2} \|w\|^2, \quad \rho > 0$$

Scelta non supervisionata dei centri

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Si fissano prima i centri o scegliendoli **casualmente** tra i vettori di ingresso del training set oppure con tecniche di **clustering**

Si scelgono poi i pesi w minimizzando la funzione di errore

$$E(w) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho}{2}\|w\|^2, \quad \rho > 0$$

Ponendo $y = (y^1, \dots, y^P)^T$ e definendo la matrice $\Phi_{pi} = \phi(\|x^p - c_i\|)$, la minimizzazione di $E(w)$ si riduce al problema ai minimi quadrati lineari:

$$\min_w E(w) = \left\| \begin{pmatrix} \Phi(c) \\ \sqrt{\rho} I \end{pmatrix} w - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|^2.$$

Scelta non supervisionata dei centri

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Si fissano prima i centri o scegliendoli **casualmente** tra i vettori di ingresso del training set oppure con tecniche di **clustering**

Si scelgono poi i pesi w minimizzando la funzione di errore

$$E(w) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i\phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho}{2}\|w\|^2, \quad \rho > 0$$

Ponendo $y = (y^1, \ldots, y^P)^T$ e definendo la matrice $\Phi_{pi} = \phi(\|x^p - c_i\|)$, la minimizzazione di $E(w)$ si riduce al problema ai minimi quadrati lineari:

$$\min_w E(w) = \left\| \begin{pmatrix} \Phi(c) \\ \sqrt{\rho} I \end{pmatrix} w - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|^2.$$

Si possono usare metodi diretti ($x = (AA^T)^{-1}A^T b$) se N non è troppo elevato se no metodi iterativi (gradiente coniugato).

Scelta non supervisionata dei centri

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Si fissano prima i centri o scegliendoli **casualmente** tra i vettori di ingresso del training set oppure con tecniche di **clustering**

Si scelgono poi i pesi w minimizzando la funzione di errore

$$E(w) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho}{2}\|w\|^2, \quad \rho > 0$$

Ponendo $y = (y^1, \dots, y^P)^T$ e definendo la matrice $\Phi_{pi} = \phi(\|x^p - c_i\|)$, la minimizzazione di $E(w)$ si riduce al problema ai minimi quadrati lineari:

$$\min_w E(w) = \left\| \begin{pmatrix} \Phi(c) \\ \sqrt{\rho} I \end{pmatrix} w - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|^2 .$$

Si possono usare metodi diretti ($x = (AA^T)^{-1}A^T b$) se N non è troppo elevato se no metodi iterativi (gradiente coniugato).

Svantaggio: non si tiene conto delle uscite y^p nella scelta dei centri.

Scelta supervisionata

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Si risolve direttamente il problema non vincolato:

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

con $\rho_1, \rho_2 > 0$.

Scelta supervisionata

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Si risolve direttamente il problema non vincolato:

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i\phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

con $\rho_1, \rho_2 > 0$.

Con questa seconda tecnica si ottengono capacità di generalizzazione migliori, rispetto alla scelta non supervisionata dei centri.

Scelta supervisionata

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Si risolve direttamente il problema non vincolato:

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

con $\rho_1, \rho_2 > 0$.

Con questa seconda tecnica si ottengono capacità di generalizzazione migliori, rispetto alla scelta non supervisionata dei centri.

Le dimensioni del problema sono notevolmente più elevate: si utilizzano tecniche di decomposizione.

Tecniche di decomposizione

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Si partizionano le variabili in blocchi e si minimizza rispetto al singolo blocco di variabili: se si sta risolvendo il problema $\min_x f(x)$ si partiziona il vettore x in NB blocchi $x_1, \ldots, x_{NB}$ e si definisce

$$x_i^{k+1} = \arg\min_y f(x_1^{k+1} \ldots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \ldots, x_{NB}^k)$$

Tecniche di decomposizione

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Si partizionano le variabili in blocchi e si minimizza rispetto al singolo blocco di variabili: se si sta risolvendo il problema $\min_x f(x)$ si partiziona il vettore x in NB blocchi $x_1, \ldots, x_{NB}$ e si definisce

$$x_i^{k+1} = \arg\min_y f(x_1^{k+1} \ldots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \ldots, x_{NB}^k)$$

In generale, non converge, esiste un controesempio in 3 variabili per il metodo delle coordinate con ricerca di linea esatta.

Tecniche di decomposizione

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Si partizionano le variabili in blocchi e si minimizza rispetto al singolo blocco di variabili: se si sta risolvendo il problema $\min_x f(x)$ si partiziona il vettore x in NB blocchi $x_1, \dots, x_{NB}$ e si definisce

$$x_i^{k+1} = \arg\min_y f(x_1^{k+1} \dots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \dots, x_{NB}^k)$$

In generale, non converge, esiste un controesempio in 3 variabili per il metodo delle coordinate con ricerca di linea esatta.

Converge in uno dei seguenti casi:

Tecniche di decomposizione

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Si partizionano le variabili in blocchi e si minimizza rispetto al singolo blocco di variabili: se si sta risolvendo il problema $\min_x f(x)$ si partiziona il vettore x in NB blocchi $x_1, \ldots, x_{NB}$ e si definisce

$$x_i^{k+1} = \arg\min_y f(x_1^{k+1} \ldots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \ldots, x_{NB}^k)$$

In generale, non converge, esiste un controesempio in 3 variabili per il metodo delle coordinate con ricerca di linea esatta.

Converge in uno dei seguenti casi:

1. Il vettore delle variabili è decomposto in $NB = 2$ blocchi

Tecniche di decomposizione

Si partizionano le variabili in blocchi e si minimizza rispetto al singolo blocco di variabili: se si sta risolvendo il problema $\min_x f(x)$ si partiziona il vettore x in NB blocchi $x_1, \ldots, x_{NB}$ e si definisce

$$x_i^{k+1} = \arg\min_y f(x_1^{k+1} \ldots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \ldots, x_{NB}^k)$$

In generale, non converge, esiste un controesempio in 3 variabili per il metodo delle coordinate con ricerca di linea esatta.

Converge in uno dei seguenti casi:

1. Il vettore delle variabili è decomposto in $NB = 2$ blocchi
2. la funzione obiettivo f è pseudo-convessa ($\forall x, y \in \mathbb{R}^n, x \neq y$ si ha che $\nabla f(x)^T(y-x) \geq 0$ implica $f(y) \geq f(x)$) e insiemi di livello sono compatti

Tecniche di decomposizione

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Si partizionano le variabili in blocchi e si minimizza rispetto al singolo blocco di variabili: se si sta risolvendo il problema $\min_x f(x)$ si partiziona il vettore x in NB blocchi $x_1, \dots, x_{NB}$ e si definisce

$$x_i^{k+1} = \arg\min_y f(x_1^{k+1} \dots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \dots, x_{NB}^k)$$

In generale, non converge, esiste un controesempio in 3 variabili per il metodo delle coordinate con ricerca di linea esatta.

Converge in uno dei seguenti casi:

1. Il vettore delle variabili è decomposto in $NB = 2$ blocchi
2. la funzione obiettivo f è pseudo-convessa ($\forall x, y \in \mathbb{R}^n, x \neq y$ si ha che $\nabla f(x)^T(y - x) \geq 0$ implica $f(y) \geq f(x)$) e insiemi di livello sono compatti
3. la funzione obiettivo f è strettamente convessa rispetto a $NB - 2$ blocchi.

Tecniche di decomposizione

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Si partizionano le variabili in blocchi e si minimizza rispetto al singolo blocco di variabili: se si sta risolvendo il problema $\min_x f(x)$ si partiziona il vettore x in NB blocchi $x_1, \ldots, x_{NB}$ e si definisce

$$x_i^{k+1} = \arg\min_y f(x_1^{k+1} \ldots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \ldots, x_{NB}^k)$$

In generale, non converge, esiste un controesempio in 3 variabili per il metodo delle coordinate con ricerca di linea esatta.

Converge in uno dei seguenti casi:

1. Il vettore delle variabili è decomposto in $NB = 2$ blocchi
2. la funzione obiettivo f è pseudo-convessa ($\forall x, y \in \mathbb{R}^n, x \neq y$ si ha che $\nabla f(x)^T(y - x) \geq 0$ implica $f(y) \geq f(x)$) e insiemi di livello sono compatti
3. la funzione obiettivo f è strettamente convessa rispetto a $NB - 2$ blocchi.

In alternativa esistono metodi di discesa a blocchi in cui non è richiesta la minimizzazione globale rispetto a ogni blocco di variabili.

Decomposizione per RBF

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Un primo schema di decomposizione si ottiene decomponendo il problema

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

rispetto a due blocchi: pesi e centri. (nota: gli insiemi di livello sono compatti)

Decomposizione per RBF

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Un primo schema di decomposizione si ottiene decomponendo il problema

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

rispetto a due blocchi: pesi e centri. (nota: gli insiemi di livello sono compatti)

Si iterano due fasi:

Decomposizione per RBF

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Un primo schema di decomposizione si ottiene decomponendo il problema

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

rispetto a due blocchi: pesi e centri. (nota: gli insiemi di livello sono compatti)

Si iterano due fasi:

1. w_{k+1} risolve il problema dei minimi quadrati lineari in corrispondenza di c_k

Decomposizione per RBF

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Un primo schema di decomposizione si ottiene decomponendo il problema

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

rispetto a due blocchi: pesi e centri. (nota: gli insiemi di livello sono compatti)

Si iterano due fasi:

1. w_{k+1} risolve il problema dei minimi quadrati lineari in corrispondenza di c_k
2. c_{k+1} viene individuato tramite algoritmo di discesa con ricerca unidimensionale (Armijo) lungo l'antigradiente

Decomposizione per RBF

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Un primo schema di decomposizione si ottiene decomponendo il problema

$$\min_{w,c} E(w,c) = \frac{1}{2}\sum_{p=1}^{P}\left(\sum_{i=1}^{N} w_i \phi(\|x^p - c_i\|) - y_p\right)^2 + \frac{\rho_1}{2}\|w\|^2 + \frac{\rho_2}{2}\|c\|^2,$$

rispetto a due blocchi: pesi e centri. (nota: gli insiemi di livello sono compatti)

Si iterano due fasi:

1. w_{k+1} risolve il problema dei minimi quadrati lineari in corrispondenza di c_k
2. c_{k+1} viene individuato tramite algoritmo di discesa con ricerca unidimensionale (Armijo) lungo l'antigradiente

SI dimostra la convergenza a punti stazionari e che i pesi di uscita corrispondono a un minimo globale nello spazio di w per la data scelta di c.

Decomposizione in 2 blocchi

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $k = 0$.

Decomposizione in 2 blocchi

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}, w_0 \in \mathbb{R}^N, k = 0.$

While $\nabla E(w_k, c_k) \neq 0$

Decomposizione in 2 blocchi

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}, w_0 \in \mathbb{R}^N, k = 0.$

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Decomposizione in 2 blocchi

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^{N}$, $k = 0$.

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Passo 2 Minimizzazione rispetto ai centri:

(a) calcola $d_k = -\nabla_c E(w_{k+1}, c_k)$

Decomposizione in 2 blocchi

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}, w_0 \in \mathbb{R}^N, k = 0.$

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Passo 2 Minimizzazione rispetto ai centri:

(a) calcola $d_k = -\nabla_c E(w_{k+1}, c_k)$

(b) calcola η_k con una ricerca unidimensionale opportuna (Armijo)

Decomposizione in 2 blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $k = 0$.

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Passo 2 Minimizzazione rispetto ai centri:

(a) calcola $d_k = -\nabla_c E(w_{k+1}, c_k)$

(b) calcola η_k con una ricerca unidimensionale opportuna (Armijo)

(c) scegli c_{k+1} tale che $E(w_{k+1}, c_{k+1}) \leq E(w_{k+1}, c_k + \eta_k d_k)$

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Decomposizione in 2 blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $k = 0$.

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Passo 2 Minimizzazione rispetto ai centri:

(a) calcola $d_k = -\nabla_c E(w_{k+1}, c_k)$

(b) calcola η_k con una ricerca unidimensionale opportuna (Armijo)

(c) scegli c_{k+1} tale che $E(w_{k+1}, c_{k+1}) \leq E(w_{k+1}, c_k + \eta_k d_k)$

Passo 3 $k = k + 1$

Decomposizione in 2 blocchi

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $k = 0$.

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Passo 2 Minimizzazione rispetto ai centri:

(a) calcola $d_k = -\nabla_c E(w_{k+1}, c_k)$

(b) calcola η_k con una ricerca unidimensionale opportuna (Armijo)

(c) scegli c_{k+1} tale che $E(w_{k+1}, c_{k+1}) \leq E(w_{k+1}, c_k + \eta_k d_k)$

Passo 3 $k = k + 1$

End

Decomposizione in 2 blocchi

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $k = 0$.

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Passo 2 Minimizzazione rispetto ai centri:

(a) calcola $d_k = -\nabla_c E(w_{k+1}, c_k)$

(b) calcola η_k con una ricerca unidimensionale opportuna (Armijo)

(c) scegli c_{k+1} tale che $E(w_{k+1}, c_{k+1}) \leq E(w_{k+1}, c_k + \eta_k d_k)$

Passo 3 $k = k + 1$

End

La sequenza $\{(w_k, c_k)\}$ generata ha punti di accumulazione, $\{E(w_k, c_k)\}$ converge e ogni punto di accumulazione di $\{(w_k, c_k)\}$ è un punto stazionario.

Decomposizione in $N + 1$ blocchi

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Per problemi di dimensioni molto elevate, la minimizzazione approssimata rispetto a c può essere troppo costosa.

Decomposizione in $N + 1$ blocchi

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Per problemi di dimensioni molto elevate, la minimizzazione approssimata rispetto a c può essere troppo costosa.

Si decompone ulteriormente il passo 2 in una successione di minimizzazioni sequenziali rispetto ai singoli centri

Decomposizione in $N + 1$ blocchi

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Per problemi di dimensioni molto elevate, la minimizzazione approssimata rispetto a c può essere troppo costosa.

Si decompone ulteriormente il passo 2 in una successione di minimizzazioni sequenziali rispetto ai singoli centri

Le variabili vengono partizionate in $NB = N + 1$ blocchi costituiti dal vettore dei pesi w e dai vettori dei centri c_i, $i = 1, \ldots, N$.

Decomposizione in $N + 1$ blocchi

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Per problemi di dimensioni molto elevate, la minimizzazione approssimata rispetto a c può essere troppo costosa.

Si decompone ulteriormente il passo 2 in una successione di minimizzazioni sequenziali rispetto ai singoli centri

Le variabili vengono partizionate in $NB = N + 1$ blocchi costituiti dal vettore dei pesi w e dai vettori dei centri c_i, $i = 1, \ldots, N$.

Come nell'algoritmo precedente i pesi vengono aggiornati risolvendo il problema dei minimi quadrati lineare, mentre per i centri si possono usare metodi di discesa per minimizzare rispetto ai centri.

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Decomposizione in $N + 1$ blocchi

Per problemi di dimensioni molto elevate, la minimizzazione approssimata rispetto a c può essere troppo costosa.

Si decompone ulteriormente il passo 2 in una successione di minimizzazioni sequenziali rispetto ai singoli centri

Le variabili vengono partizionate in $NB = N + 1$ blocchi costituiti dal vettore dei pesi w e dai vettori dei centri c_i, $i = 1, \ldots, N$.

Come nell'algoritmo precedente i pesi vengono aggiornati risolvendo il problema dei minimi quadrati lineare, mentre per i centri si possono usare metodi di discesa per minimizzare rispetto ai centri.

Per garantire la convergenza la linesearch deve soddisfare ipotesi più forti, in particolare va bene Armijo combinata con una limitazione superiore del passo.

Decomposizione in $N + 1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \dots, N$.

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Decomposizione in $N+1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \ldots, N$.

While $\nabla E(w_k, c_k) \neq 0$

Decomposizione in $N+1$ blocchi

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \dots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Decomposizione in $N + 1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}, w_0 \in \mathbb{R}^N, \tau_i > 0, k = 0, \xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \ldots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:

Decomposizione in $N + 1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \ldots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \ldots, N$

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

Decomposizione in $N+1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \ldots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \ldots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$

Decomposizione in $N+1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \dots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \dots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$
else

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Decomposizione in $N+1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \ldots, N$.

While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:

calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.

Passo 2 Minimizzazione rispetto ai centri:

For $i = 1, \ldots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$

else

(a) calcola $d_i^k = -\nabla_{c_i} E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}, c_i^k, \ldots, c_N^k)$

Decomposizione in $N + 1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \dots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \dots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$
else
(a) calcola $d_i^k = -\nabla_{c_i} E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}, c_i^k, \dots, c_N^k)$
(b) calcola η_i^k con una ricerca unidimensionale opportuna (Armijo con limitazione superiore sul passo)

Decomposizione in $N + 1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \ldots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \ldots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$
else
(a) calcola $d_i^k = -\nabla_{c_i} E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}, c_i^k, \ldots, c_N^k)$
(b) calcola η_i^k con una ricerca unidimensionale opportuna (Armijo con limitazione superiore sul passo)
(c) scegli $\tilde{c}_i^k$ tale che $E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}), \tilde{c}_i^k, \ldots, c_N^k) \leq$
$E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}, c_i^k + \eta_i^k d_i^k, \ldots, c_N^k)$

Decomposizione in $N+1$ blocchi

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N+1$ blocchi
- Decomposizione in $N+1$ blocchi
- Proprietà

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k=0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \ldots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \ldots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$
else
(a) calcola $d_i^k = -\nabla_{c_i} E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}, c_i^k, \ldots, c_N^k)$
(b) calcola η_i^k con una ricerca unidimensionale opportuna (Armijo con limitazione superiore sul passo)
(c) scegli $\tilde{c}_i^k$ tale che $E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}), \tilde{c}_i^k, \ldots, c_N^k) \leq E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}, c_i^k + \eta_i^k d_i^k, \ldots, c_N^k)$
If $E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}), \tilde{c}_i^k, \ldots, c_N^k) \leq E(w_{k+1}, c_1^{k+1}, \ldots, c_{i-1}^{k+1}, c_i^k, \ldots, c_N^k) - \tau_i \|\tilde{c}_i^k - c_i^k\|^2$ **then** poni $c_i^{k+1} = \tilde{c}_i^k$

Decomposizione in $N + 1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \dots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \dots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$
else
(a) calcola $d_i^k = -\nabla_{c_i} E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}, c_i^k, \dots, c_N^k)$
(b) calcola η_i^k con una ricerca unidimensionale opportuna (Armijo con limitazione superiore sul passo)
(c) scegli $\tilde{c}_i^k$ tale che $E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}), \tilde{c}_i^k, \dots, c_N^k) \leq$
$E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}, c_i^k + \eta_i^k d_i^k, \dots, c_N^k)$
If $E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}), \tilde{c}_i^k, \dots, c_N^k) \leq$
$E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}, c_i^k, \dots, c_N^k) - \tau_i \|\tilde{c}_i^k - c_i^k\|^2$ **then** poni $c_i^{k+1} = \tilde{c}_i^k$
else poni $c_i^{k+1} = c_i^k + \eta_i^k d_i^k$

Decomposizione in $N + 1$ blocchi

Dati. $c_0 \in \mathbb{R}^{nN}$, $w_0 \in \mathbb{R}^N$, $\tau_i > 0$,$k = 0$, $\xi_i^k > 0$ tali che $\xi_i^k \to 0$ per $i = 1, \dots, N$.
While $\nabla E(w_k, c_k) \neq 0$

Passo 1 Minimizzazione rispetto ai pesi:
calcola $w_{k+1} = \arg\min_w E(w, c_k)$ risolvendo il problema ai minimi quadrati in w.
Passo 2 Minimizzazione rispetto ai centri:
For $i = 1, \dots, N$

If $\|\nabla_{c_i} E(w_{k+1}, c_k)\| \leq \xi_i^k$ **then** poni $c_i^{k+1} = c_i^k$
else
(a) calcola $d_i^k = -\nabla_{c_i} E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}, c_i^k, \dots, c_N^k)$
(b) calcola η_i^k con una ricerca unidimensionale opportuna (Armijo con limitazione superiore sul passo)
(c) scegli $\tilde{c}_i^k$ tale che $E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}), \tilde{c}_i^k, \dots, c_N^k) \leq$
$E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}, c_i^k + \eta_i^k d_i^k, \dots, c_N^k)$
If $E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}), \tilde{c}_i^k, \dots, c_N^k) \leq$
$E(w_{k+1}, c_1^{k+1}, \dots, c_{i-1}^{k+1}, c_i^k, \dots, c_N^k) - \tau_i \|\tilde{c}_i^k - c_i^k\|^2$ **then** poni $c_i^{k+1} = \tilde{c}_i^k$
else poni $c_i^{k+1} = c_i^k + \eta_i^k d_i^k$

Passo 3 $k = k + 1$

Proprietà

Reti RBF

Addestramento RBF

- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

◁ Non si possono più effettuare le minimizzazioni parziali con qualunque algoritmo di discesa, poichè i punti tentativo $\tilde{c}_i^k$ possono non essere accettati se producono spostamenti troppo grandi dal punto c_i^k.

Proprietà

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

- Non si possono più effettuare le minimizzazioni parziali con qualunque algoritmo di discesa, poichè i punti tentativo $\tilde{c}_i^k$ possono non essere accettati se producono spostamenti troppo grandi dal punto c_i^k.
- Il test al passo 2 fa si che la minimizzazione parziale rispetto al centro c_i è evitata quando il corrispondente gradiente parziale è piccolo, risparmiando in termini di tempo

Proprietà

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

- Non si possono più effettuare le minimizzazioni parziali con qualunque algoritmo di discesa, poichè i punti tentativo $\tilde{c}_i^k$ possono non essere accettati se producono spostamenti troppo grandi dal punto c_i^k.
- Il test al passo 2 fa si che la minimizzazione parziale rispetto al centro c_i è evitata quando il corrispondente gradiente parziale è piccolo, risparmiando in termini di tempo
- Il calcolo della funzione di errore in corrispondenza di un nuovo centro c_i può essere fatto in modo molto efficiente, memorizzando le uscite della rete corrispondenti a c_i^k. Infatti

$$y(x^p; w_{k+1}, c_1^{k+1}, \ldots, c_i, \ldots, c_N^k) = y(x^p; w_{k+1}, c_1^{k+1}, \ldots, c_i^k, \ldots, c_N^k)$$
$$-w_i^k \phi(\|x^p - c_i^k\|) + w_i^k \phi(\|x^p - c_i\|)$$

Proprietà

Reti RBF

Addestramento RBF
- Addestramento
- Scelta non supervisionata dei centri
- Scelta supervisionata
- Tecniche di decomposizione
- Decomposizione per RBF
- Decomposizione in 2 blocchi
- Decomposizione in $N + 1$ blocchi
- Decomposizione in $N + 1$ blocchi
- Proprietà

- Non si possono più effettuare le minimizzazioni parziali con qualunque algoritmo di discesa, poichè i punti tentativo $\tilde{c}_i^k$ possono non essere accettati se producono spostamenti troppo grandi dal punto c_i^k.
- Il test al passo 2 fa si che la minimizzazione parziale rispetto al centro c_i è evitata quando il corrispondente gradiente parziale è piccolo, risparmiando in termini di tempo
- Il calcolo della funzione di errore in corrispondenza di un nuovo centro c_i può essere fatto in modo molto efficiente, memorizzando le uscite della rete corrispondenti a c_i^k. Infatti

$$y(x^p; w_{k+1}, c_1^{k+1}, \ldots, c_i, \ldots, c_N^k) = y(x^p; w_{k+1}, c_1^{k+1}, \ldots, c_i^k, \ldots, c_N^k)$$
$$-w_i^k \phi(\|x^p - c_i^k\|) + w_i^k \phi(\|x^p - c_i\|)$$

- si hanno proprietà di convergenza analoghe a quelle dell'algoritmo a due blocchi