

Superfici di risposta

- Nel caso di ottimizzazioni complesse (p.e. CFD), il tempo speso per la simulazione di una configurazione può essere notevole.
- Due sono i modi di operare per ridurre i tempi di calcolo :
 - 1) agire sulla velocità di convergenza degli algoritmi di ottimizzazione;
 - 2) estrapolare alcuni valori della funzione da ottimizzare.
- Per agire sul punto 2) e' necessario costruire un modello approssimato della funzione obiettivo.
Superfici di risposta.

Superfici di risposta

- Le superfici di risposta hanno bisogno di un insieme di dati di partenza calcolati “esattamente” (**trainig set**).
- Di conseguenza la creazione di una superficie di risposta avviene:
 - usando i dati di una ottimizzazione precedente;
 - usando i dati di un DOE.

Superfici di risposta

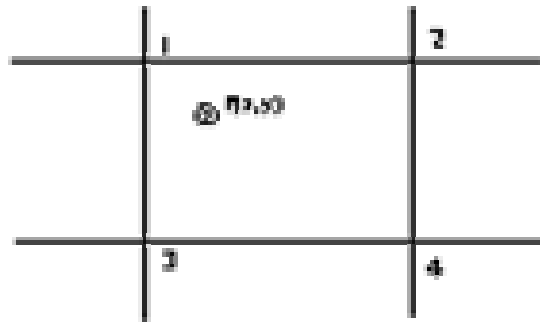
- Ci sono numerosi algoritmi numerici per la creazione di una superficie di risposta:
 - multi lineare;
 - serie di Taylor;
 - k-nearest;
 - reti neurali;
 - kriging;
 - Gaussian processes.

Superfici di risposta

- Bisogna ricordare che ogni caso studiato può essere affrontate da qualunque dei metodi proposti; attenzione comunque da alcuni elementi importanti:
 - difficoltà nel settare i parametri (setting);
 - tempi di calcolo;
 - accuratezza dei risultati.

Interpolazione multi lineare

- Metodo base caratterizzato dalla necessità di avere una griglia regolare cartesiana.



$$z = f(x, y) = g(f_1, f_2, f_3, f_4)$$

$$t = (x - x_1) / (x_2 - x_1)$$

$$u = (y - y_3) / (y_1 - y_3)$$

$$z = f(x, y) = (1-t)(1-u)z_3 + t(1-u)z_4 + tu z_2 + (1-t)u z_1$$

Interpolazione polinomiale

- **Interpolazione bi-cubica** : riprende il concetto di griglia quadra per ottenere un'interpolazione di ordine superiore. Necessita del calcolo delle derivate parziali.

$$y = f(x_1, x_2)$$

$$c[4,4] = g\left(\frac{\partial y}{\partial x_1}\Big|_{1,\dots,4}, \frac{\partial y}{\partial x_2}\Big|_{1,\dots,4}, y\Big|_{1,\dots,4}\right)$$

$$y = f(x_1, x_2) = \sum_{i=1}^4 \sum_{j=1}^4 c_{i,j} t^{i-1} u^{j-1}$$

Serie di Taylor

- La costruzione di una superficie di risposta mediante sviluppo in serie avviene rappresentando la funzione di interesse come una serie di Taylor troncata generalmente al primo termine.
- Si possono definire differenti varianti :

Lineare

$$a(x) = f(x_0) + \sum_{i=1}^m \frac{\partial y}{\partial x_i} \Big|_{x_0} (x^i - x_0^i)$$

Reciproca

$$a(x) = f(x_0) + \sum_{i=1}^m \frac{\partial y}{\partial x_i} \Big|_{x_0} (x^i - x_0^i) \frac{x_0^i}{x^i}$$

Non lineare

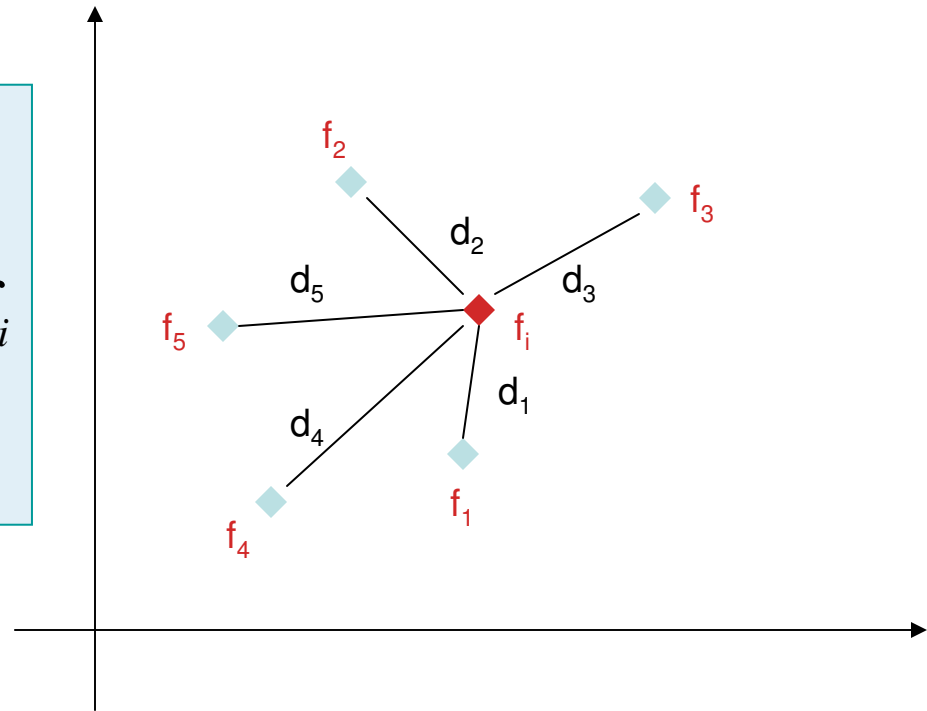
$$a(x) = f(x_0) + \frac{1}{r} \sum_{i=1}^m x_{i,0}^{1-r} \frac{\partial y}{\partial x_i} \Big|_{x_0} (x_i^r - x_{i,0}^r)$$

K-nearest

- Sfrutta il concetto di distanza tra punti nello spazio di definizione delle variabili.

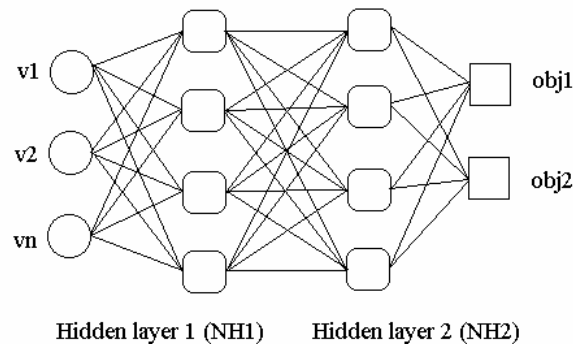
$$f_{app} = \sum_{i=1}^{k_near} \left(\frac{d_i}{\sum_{j=1}^{k_near} d_j} \right)^\alpha f_i$$

Si possono prendere un numero diverso di design (k-nearest)



Reti neurali

- Sfrutta il principio dell'apprendimento per creare un modello matematico di una funzione $y : \mathcal{X}^n \Rightarrow \mathcal{X}^m$
- Non necessita del calcolo delle derivate, ma semplicemente di un set d'addestramento $\left[\left\{ \vec{x}_1, \vec{y}_1 \right\}, \dots, \left\{ \vec{x}_q, \vec{y}_q \right\} \right]$ che puo' essere aggiornato iterativamente.



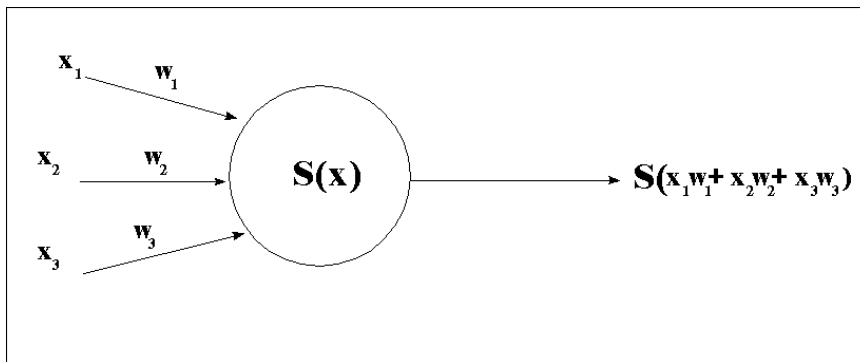
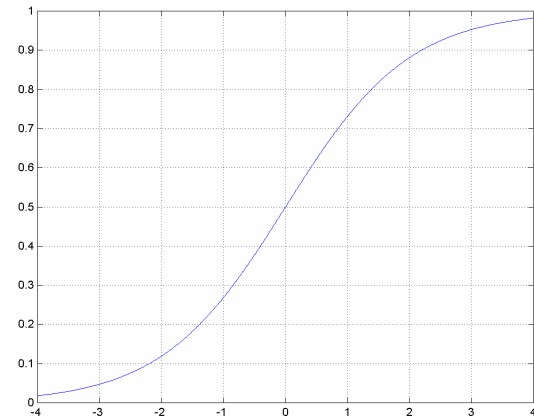
Rete neurale

Fondamentale è la scelta della **funzione attivazione**.

Normalmente viene usata la **sigmoide**:

$$s_c(x) = \frac{1}{1 + e^{-x}}$$

$$\frac{d}{dx} s(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = s(x)(1 - s(x))$$



Neurone con sigmoide e pesi

Rete neurale

- Per ottenere l'apprendimento della rete neurale sul set d'addestramento viene usato l'algoritmo backpropagation.

- Esso trova il minimo della funzione errore rispetto i pesi w .

$$E = \frac{1}{2} \sum_{i=1}^q \|\vec{o}_i - \vec{y}_i\|^2$$

- Si ottiene quindi

$$\nabla E = \left(\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \dots, \frac{\partial E}{\partial w_l} \right)$$

- A questo punto si possono calcolare i nuovi

pesi
$$\Delta w_i = -\gamma \frac{\partial E}{\partial w_i}$$

Kriging

- Metodo che ha origine dallo studio statistico sulla minimizzazione della varianza sul valore della funzione da estrapolare.
- La base della metodologia sta nel creare una funzione dipendente dalla distanza del punto di cui si vuole estrapolare il valore, con gli n punti più vicini (k-nearest).

Kriging

$$f_{app} = \sum_{i=1}^{kn} w_i \cdot f_i$$

$$\begin{cases} w_1 \gamma(d_{1,1}) + w_2 \gamma(d_{1,2}) + \dots + w_{kn} \gamma(d_{1,kn}) + \lambda = \gamma(d_{1,0}) \\ w_1 \gamma(d_{2,1}) + w_2 \gamma(d_{2,2}) + \dots + w_{kn} \gamma(d_{2,kn}) + \lambda = \gamma(d_{2,0}) \\ \vdots \\ w_1 \gamma(d_{kn,1}) + w_2 \gamma(d_{kn,2}) + \dots + w_{kn} \gamma(d_{kn,kn}) + \lambda = \gamma(d_{kn,0}) \end{cases}$$

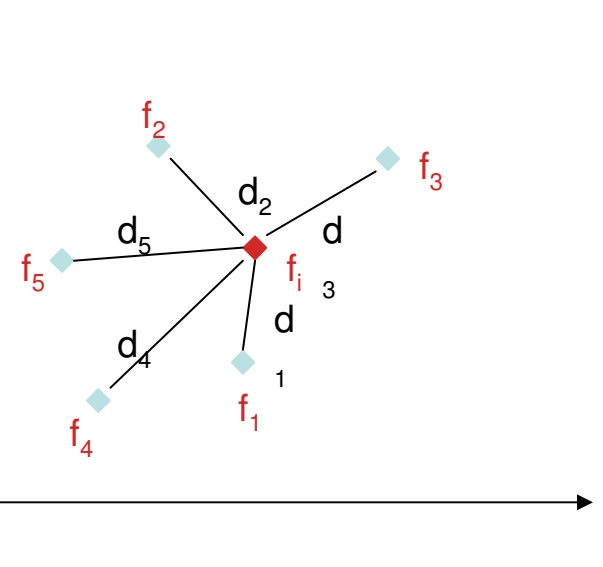
$$\sum_{i=1}^{kn} w_i + \lambda = 1$$

γ viene chiamata funzione di covarianza

esistono molte funzioni γ , le più usate sono le:

lineare $\gamma(d) = d$

esponenziale $\gamma(d) = c_0 + c_1(1 - \exp(-d))$



Kriging attivo

- L'utilità maggiore del metodo di Kriging sta nel fatto che si può riuscire a determinare (in modo statistico) l'errore nell'estrapolazione:

$$f_{reale} = f_{app} \pm 2\sigma \quad \text{con probabilità al 95\%}$$

$$f_{app} = \sum_{i=1}^{kn} w_i f_i$$

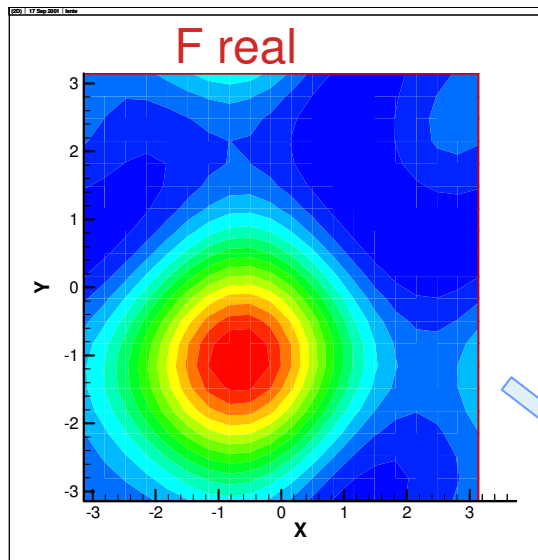
$$\sigma = \sqrt{\sum_{i=1}^{kn} w_i \gamma(d_{i,0})}$$

Kriging attivo

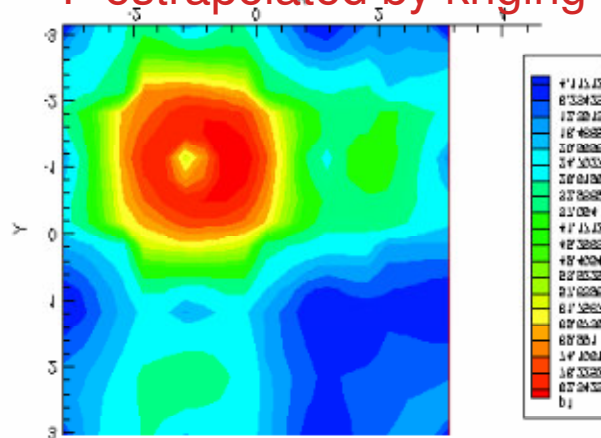
- In questo modo si può riuscire a determinare in maniera più appropriata il DOE iniziale per la creazione della superficie di risposta.
- Infatti i punti del DOE da calcolare saranno quelli dove la superficie di risposta ha l'errore maggiore in estrapolazione.
- Massimizzazione del parametro

$$I = f_{app} \cdot \sigma$$

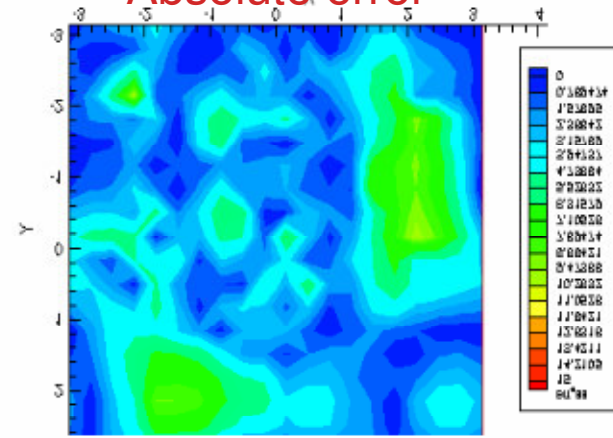
Kriging attivo



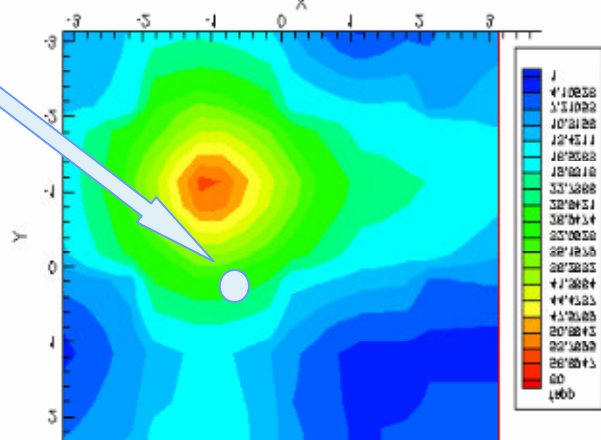
F estrapolated by kriging



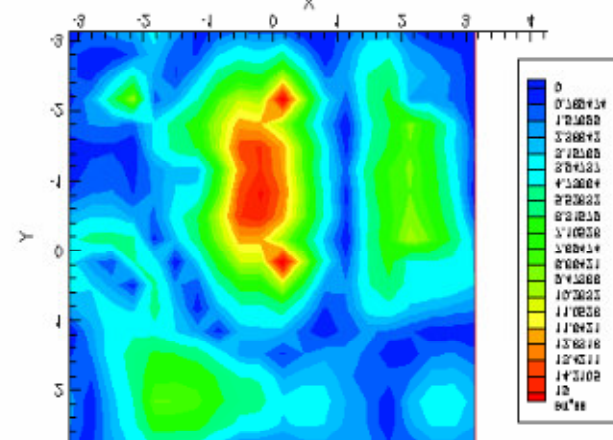
Absolute error



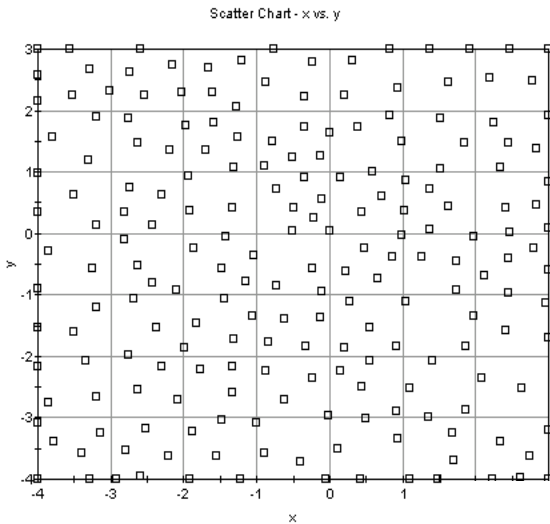
Index of absolute error



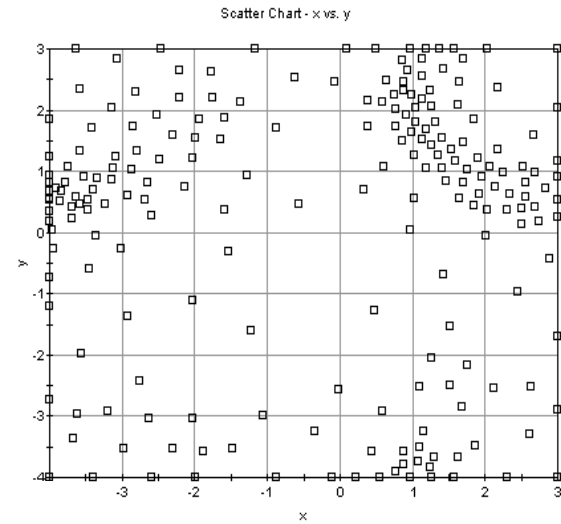
New absolute error



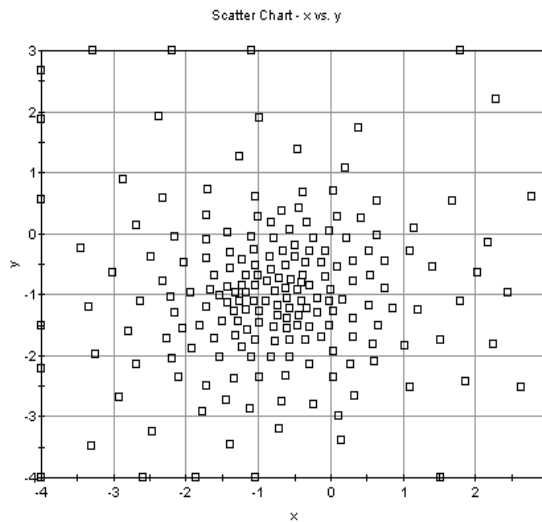
Esempi numerici



Plain Crossvalidation



Relative Error Crossvalidation



Absolute Error Crossvalidation

Gaussian processes

- Metodo statistico, valuta la distribuzione condizionata per il punto da estrapolare \mathbf{x}_{N+1} dato il set $D = \{\mathbf{x}_n, t_n\}_{n=1, \dots, N}$ di punti noti.

- Si utilizza il concetto di **processo gaussiano**, inteso come collezione di variabili \mathbf{t} avente una distribuzione del tipo:

$$P(\mathbf{t} | \mathbf{C}, \{x_n\}) = \frac{1}{Z} \exp\left(-\frac{1}{2}(\mathbf{t} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{t} - \boldsymbol{\mu})\right)$$

- Partendo da questo presupposto si ottiene la distribuzione predittiva del punto da estrapolare

$$P(t_{N+1} | D, C(\mathbf{x}, \Theta), \mathbf{x}_{N+1}) = \frac{1}{Z} \exp\left(-\frac{(t_{N+1} - \hat{t}_{N+1})^2}{2\sigma_{N+1}^2}\right)$$

Processi gaussiani

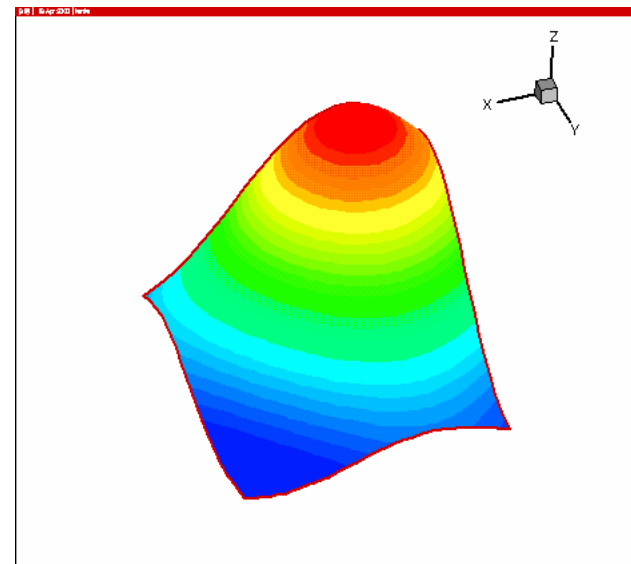
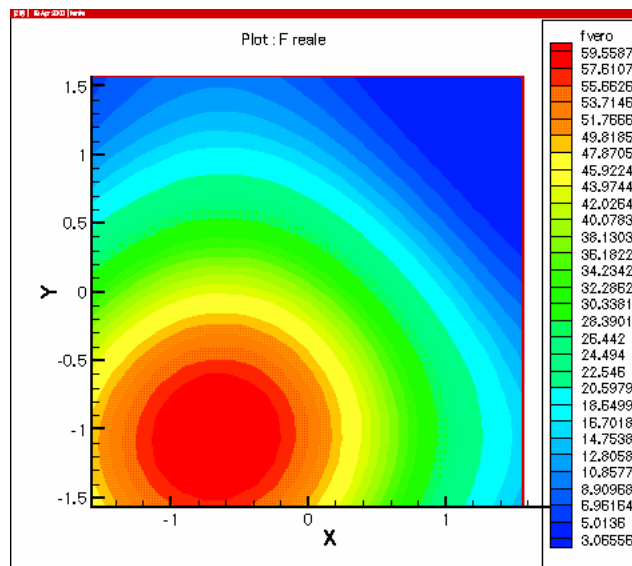
- Gli elementi della matrice C possono essere scelti della forma :

$$C_{n,m} = C(\mathbf{x}_n, \mathbf{x}_m; \Theta) = \theta_0 \exp \left\{ -\frac{1}{2} \sum_{i=1}^d \theta_i (x_n^i - x_m^i)^2 \right\} \\ + \theta_{d+1} + \theta_{d+2} \sum_{i=1}^d x_n^i x_m^i + \theta_{d+3} \delta(n, m)$$

- I parametri Θ ideali sono determinati ottimizzando la distribuzione di probabilita' $P(\Theta|D, C)$ utilizzando per esempio il metodo del simplesso.

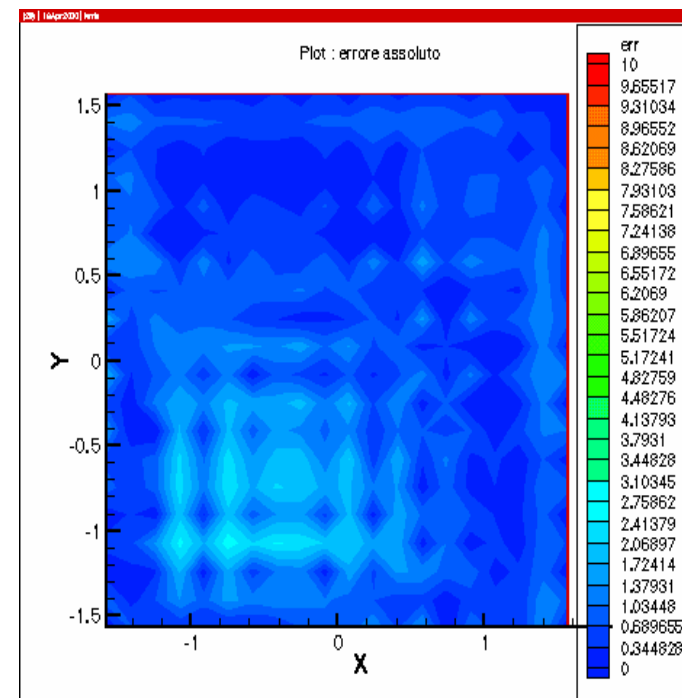
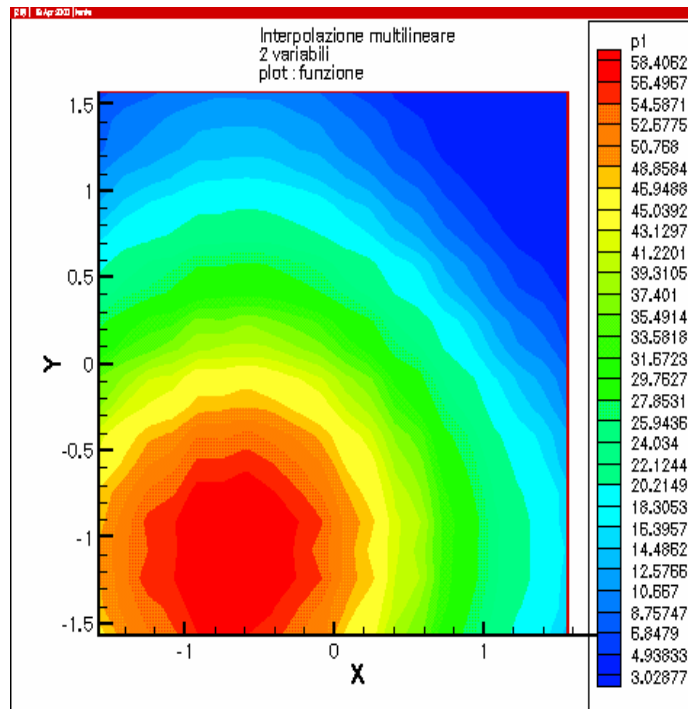
Esempi di superfici di risposta

- La funzione ad essere estrapolata e' la TEST1.
- Vengono mostrati due casi : 2 variabili, 8 variabili.
- In entrambi i casi e' stato usato un set d'addestramento di 100 punti definiti nel dominio.



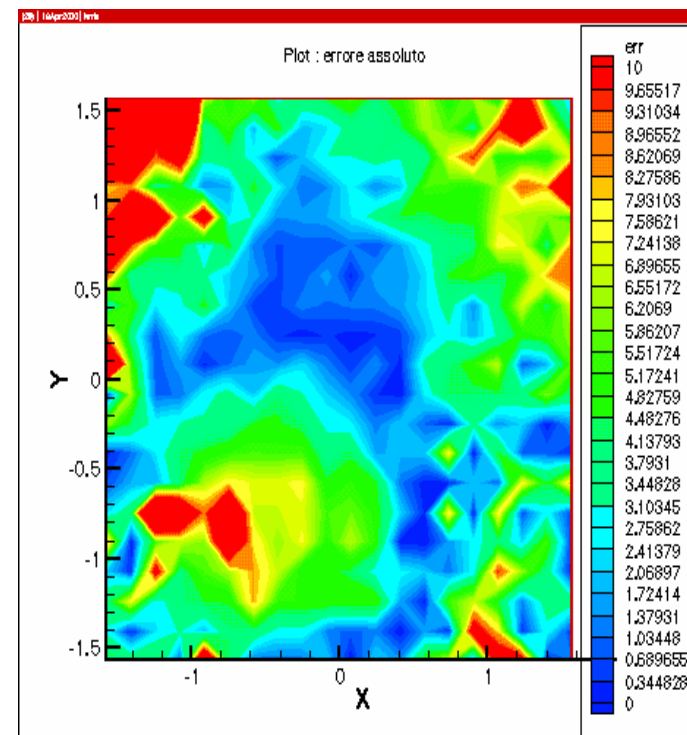
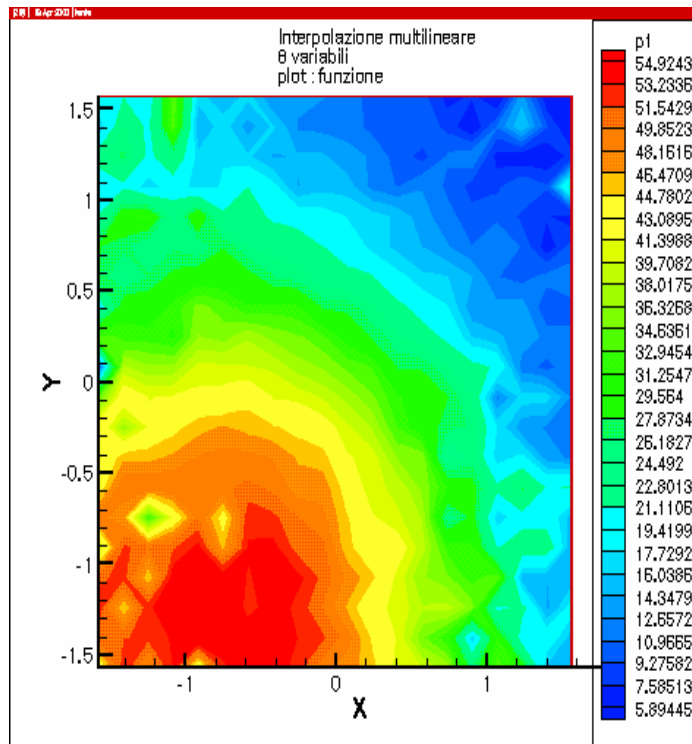
Esempi di superfici di risposta

Interpolazione multi lineare (2 variabili)



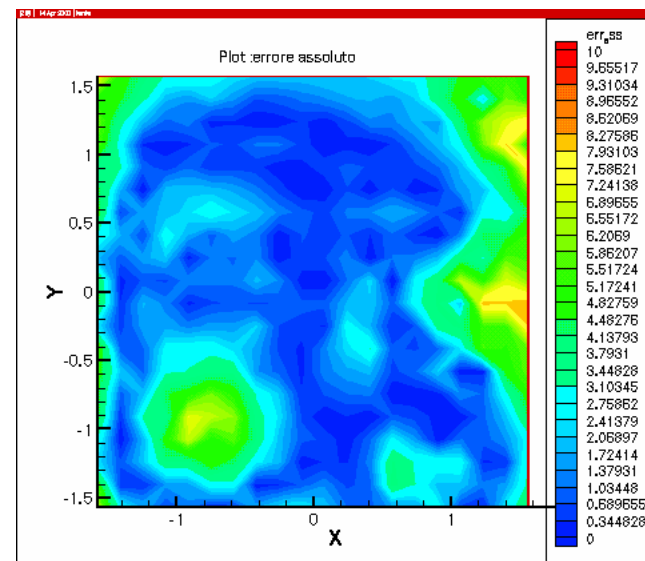
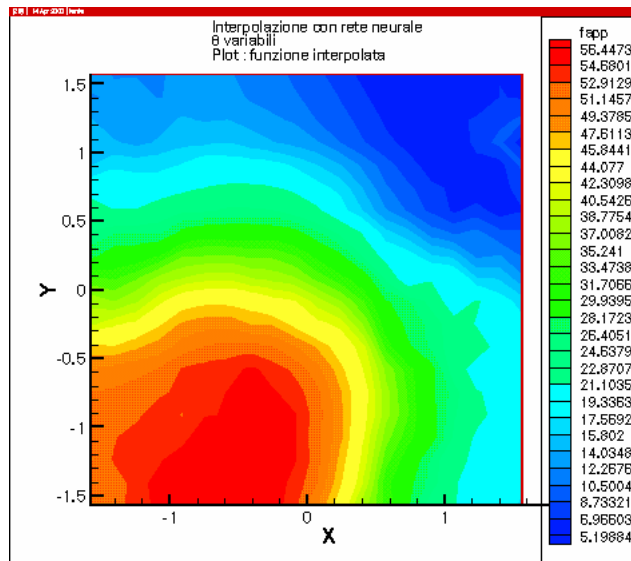
Esempi di superfici di risposta

Interpolazione multi lineare (8 variabili)



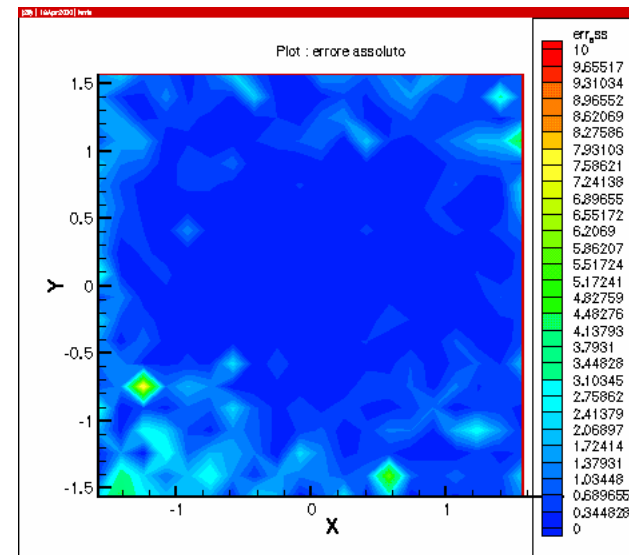
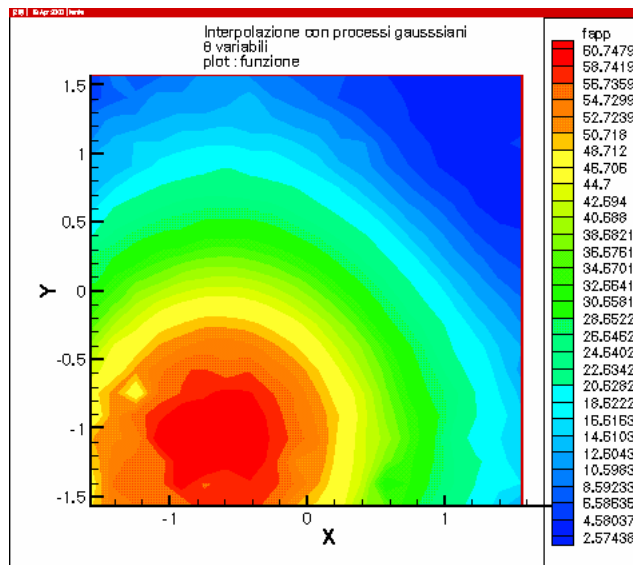
Esempi di superfici di risposta

Interpolazione con rete neurale (8 variabili)



Esempi di superfici di risposta

Interpolazione con processi gaussiani (8 variabili)



Esempi di superfici di risposta

- L'uso delle superfici di risposta e' utile soprattutto per diminuire i tempi di calcolo in caso di ottimizzazioni complesse. Alcune configurazioni verranno calcolate "esattamente" (10-20%) mentre le rimanenti saranno estrapolate utilizzando uno dei metodi visti.
- **Accuratezza** : + processi gaussiani, + sviluppi in serie, - reti neurali.
- In caso di set d'addestramento ridotto, il metodo da privilegiare e' la **rete neurale**.
- Molto interessante l'uso del metodo Kriging attivo, per la possibilità di minimizzare l'errore in estrapolazione.