Teoria dell'Approssimazione e Applicazioni, A.A. 2012/13

Esercitazione di laboratorio del 23 maggio 2013

Prof. Stefano De Marchi

1 Esercitazione proposta

1.1 Approssimazione ai minimi quadrati quando i centri differiscono dai "data sites"

Supponiamo che la funzione f sia nota su data sites $X = \{x_1, ..., x_N\}$ mentre le funzioni di base siano centrate in $\Xi = \{\xi_1, \ldots, \xi_M\}$ con $M \le N$. L'approssimante RBF sarà

$$Q_f(x) = \sum_{i=1}^{M} c_j \Phi(x, \xi_j), \quad x \in \mathbb{R}^s. \tag{1}$$

I coefficienti c_j si potranno determinare, come noto, come soluzione ai minimi quadrati del sistema $A\mathbf{c} = \mathbf{f}$ (con A matrice rettangolare $N \times M$ con componenti $A_{j,k} = \Phi(x_j, \xi_k)$, $\mathbf{c}$, $\mathbf{f}$, vettori di lunghezze M ed N, rispettivamente) minimizzando il funzionale quadratico $\|Q_f - f\|_2^2$.

1. Usare lo script `RBFApproximation2D.m`, che utilizza due insiemi di punti, i "data sites" e i centri, per approssimare la funzione `sinc` in $[0,1]^2$ usando come RBF la funzione gaussiana con $\epsilon = 1$. Aumentare o diminuire ϵ e vedere come cambia l'errore e il rango della matrice A. Fare anche esperimenti con diversi insiemi di punti (sia data sites che centri): equispaziati e di Chebyshev.

2. Usare poi lo script `RBFApproximation2Dlinear.m` (che consente di usare funzioni di base *Thin Plate Splines (TPS)*, definite nella funzione `tps.m`, con l'aggiunta di un termine polinomiale di primo grado, lineare appunto) per approssimare *dati affetti da errore*. La simulazione proposta consiste nel campionare la *funzione di Franke* sull'insieme $X \subset [0,1]^2$ dei data sites, quindi aggiungendo un rumore random uniformemente distribuito di lunghezza variabile (la riga dello script che fa questo è `rhs=rhs+0.03*randn(size(rhs))`, ovvero un'aggiunta del 3% di rumore).

   (a) La prima proposta è di determinare l'approssimazione ai minimi quadrati di tanti punti con poche funzioni di base (come fatto coi minimi quadrati al punto precedente). Questo approccio in statistica si chiama *regressione con splines*. In questo esperimento i centri coincidono con i data sites. Si noterà dai plot delle approssimanti, che questa strategia funziona meglio che non l'interpolazione.

   (b) La seconda strategia per "lisciare" dati affetti da rumore è il cosidetto *ridge regression method*, cioè il metodo dei *minimi quadrati regolarizzati* che fa uso di un parametro ω di smoothing, risolvendo non il sistema sovradeterminato $A\mathbf{c} = \mathbf{f}$ ma

$$\left(A + \frac{1}{2\omega} I\right) \mathbf{c} = \mathbf{f}\,.$$

Per questo, usare lo script `TPS_RidgeRegression2D.m`, che usa le sempre le TPS per approssimare ancora la funzione di Franke, facendo variare ω. Mediante il metodo di *cross-validation* trovare il parametro ω^* che rende lo smoothing migliore.

1.2 Approssimazione di Shepard

L'*approssimante di Shepard* è data come

$$\mathcal{P}_f(\mathbf{x}) = \sum_{i=1}^{N} f(\mathbf{x}_i) \frac{w(\mathbf{x}_i, \mathbf{x})}{\sum_{j=1}^{N} w(\mathbf{x}_i, \mathbf{x})}, \quad \mathbf{x} \in \mathbb{R}^s . \tag{2}$$

In $\mathbb{R}^2$, si considerino le funzioni peso

$$w_1(\mathbf{x}_i, \mathbf{x}) = \mathrm{e}^{-\epsilon^2 \|\mathbf{x}-\mathbf{x}_i\|}, \quad \texttt{gaussiana, globale}$$

$$w_2(\mathbf{x}_i, \mathbf{x}) = (1 - \epsilon\|\mathbf{x} - \mathbf{x}_i\|)_+^4 (4\epsilon\|\mathbf{x} - \mathbf{x}_i\| + 1), \quad \texttt{Wendland, locale}$$

e si voglia approssimare la `funzione di Franke`, campionata su punti equispaziati di $[0, 1]^2$.

1. **Caso non-stazionario**. Usando $\epsilon = 3$, si produca per ognuna delle funzioni peso, una tabella

	stazionario			non-stazionario	
N	RMS-errors	r	N	RMS-error	r
3					
5					
9					
17					
33					
65					

dove N indica il numero di punti equispaziati in *ciascuna direzione spaziale* (quindi complessivamente saranno N^2) e r l'ordine di convergenza determinato usando la formula

$$r_k = \frac{\ln(e_{k-1}/e_k)}{\ln(h_{k-1}/h_k)}, \quad k = 2, 3, \ldots \tag{3}$$

dove e_k indica il RMS-error per l'esperimento k-esimo e h_k la *fill-distance* corrispondente.

Si noti, che per punti uniformemente distribuiti il rapporto delle fill-distances è sempre 2.

Infine, gli errori RMS si possono calcolare su una griglia di riferimento 40×40 di punti equispaziati su $[0, 1]^2$.

2. **Caso stazionario**. Partendo da $\epsilon = 3$, lo si moltiplichi per il fattore 2 ad ogni dimezzamento della fill-distance.

   Si dovrebbe verificare che il metodo di Shepard, nel caso stazionario, ha un ordine $\mathcal{O}(h)$.

• Ripetere i calcoli con la funzione `f=f+0.03*rand(size(f))` ovvero con la funzione di Franke perturbata del 3%. In particolare, nel caso stazionario usare i valori $\epsilon = 48$ e $\epsilon = 12$.

Usare i programmi `Shepard2D.m` e `Shepard_CS.m` per fare i calcoli richiesti.

Ricordo che il link dove trovare i files delle funzioni Matlab è:
`http://www.math.unipd.it/~demarchi/TAA2010`