Il problema lineare dei minimi quadrati

APPLICAZIONE:

Il polinomio di migliore approssimazione nel senso dei minimi quadrati

Felice Iavernaro

Dipartimento di Matematica
Università di Bari

15 Gennaio 2009

Premessa

Sia V sottospazio vettoriale di $\mathbb{R}^n$ e $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ una sua base.
Abbiamo visto che:

$$\begin{aligned} &\mathbb{R}^n = V \oplus V^\perp \\ &\dim(V) + \dim(V^\perp) = n. \end{aligned}$$

Inoltre, detta A la matrice le cui colonne sono i vettori $\mathbf{v}_i$, cioè:
$A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$, sappiamo che:

$$\begin{aligned} V &= \mathrm{Im}(A); \\ V^\perp &= \ker(A^T). \end{aligned}$$

Possiamo allora concludere con la seguente fondamentale proprietà dell'algebra delle matrici:

$$\begin{aligned} &\mathbb{R}^n = \mathrm{Im}(A) \oplus \ker(A^T); \\ &\dim\left(\mathrm{Im}(A)\right) + \dim\left(\ker(A^T)\right) = n \qquad \text{num. di } \underline{\text{righe}} \text{ di } A. \end{aligned}$$

Vettore residuo

Consideriamo il sistema lineare di m equazioni in n incognite:

$$A\mathbf{x} = \mathbf{b}$$

Considerato per ogni $\mathbf{x} \in \mathbb{R}^n$ il vettore residuo

$$\mathbf{r}(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$$

osserviamo che

$$\mathbf{x} \text{ è soluzione del sistema } \iff \mathbf{r}(\mathbf{x}) = \mathbf{0} \iff ||\mathbf{r}(\mathbf{x})|| = 0,$$

dove $||\cdot||$ è una qualsiasi norma vettoriale.
Se $m > n$ il sistema si dice sovradimensionato poiché contiene più equazioni che incognite (nelle applicazioni è spesso $m \gg n$).
In tal caso, in generale, il sistema non sarà compatibile poiché, salvo casi eccezionali, risulterà $\mathrm{rank}(A) \neq \mathrm{rank}([A, \mathbf{b}])$.
Non potendo allora essere $\mathbf{r}(\mathbf{x}) = \mathbf{0}$, ci si chiede allora se esiste un vettore $\mathbf{x}^*$ che minimizza la norma del vettore residuo.

Il problema lineare dei minimi quadrati

Se come norma scegliamo la norma 2, il problema prima esposto prende il nome di *problema lineare dei minimi quadrati*.

Problema

Siano dati: $A \in \mathbb{R}^{m\times n}$ *e* $\mathbf{b} \in \mathbb{R}^m$*, con* $m \geq n$*. Determinare* $\mathbf{x}^*$ *tale che:*

$$||\mathbf{b} - A\mathbf{x}^*||_2 = \min_{\mathbf{x}\in\mathbb{R}^n} ||\mathbf{b} - A\mathbf{x}||_2.$$

Teorema

Se il rango di A *è massimo (cioè* $\mathrm{rank}(A) = n$*), il problema dei minimi quadrati ammette un'unica soluzione* $\mathbf{x}^*$*, che può essere ottenuta come soluzione del seguente sistema di* n *equazioni in* n *incognite:*

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

Tale sistema prende il nome di sistema normale.

Osservazione

Cominciamo ad osservare che, se $\mathrm{rank}(A) = n$, il sistema normale

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

ammette un'unica soluzione, cioè $\det(A^T A) \neq 0$. Per mostrare ciò, faremo vedere che:

$$A^T A\mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{x} = \mathbf{0},$$

cioè che l'unica soluzione del sistema omogeneo $A^T A\mathbf{x} = \mathbf{0}$ è quella nulla. Infatti, se per assurdo esistesse una soluzione non nulla $\mathbf{x} \neq \mathbf{0}$, premoltiplicando il sistema per $\mathbf{x}^T$ otterremmo:

$$\mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \cdot \mathbf{0} \Leftrightarrow (A\mathbf{x})^T(A\mathbf{x}) = 0 \Leftrightarrow ||A\mathbf{x}||_2^2 = 0 \Leftrightarrow A\mathbf{x} = \mathbf{0}.$$

Ricordando che $A\mathbf{x}$ è una comb. lineare delle colonne di A, dall'ultima uguaglianza segue che le colonne di A sono linearmente dipendenti e questo è assurdo poiché implicherebbe $\mathrm{rank}(\mathrm{A}) < n$.

Dimostrazione del teorema

Poiché $\mathbb{R}^m = \mathrm{Im}(A) \oplus \ker(A^T)$ possiamo effettuare la seguente decomposizione del vettore $\mathbf{b}$:

$$\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2, \qquad \text{con} \;\; \mathbf{b}_1 \in \mathrm{Im}(A) \;\; \text{e} \;\; \mathbf{b}_2 \in \ker(A^T)$$

$$\begin{array}{lll} \mathbf{b}_1 \in \mathrm{Im}(A) & \Longrightarrow & \exists \mathbf{x}^* \in \mathbb{R}^n \text{ tale che } \mathbf{b}_1 = A\mathbf{x}^* \text{ cioè } \mathbf{b}_1 - A\mathbf{x}^* = \mathbf{0} \\ \mathbf{b}_2 \in \ker(A^T) & \Longrightarrow & A^T\mathbf{b}_2 = \mathbf{0} \quad (\text{ricordiamo che } \;\ker(A^T) = (\mathrm{Im}(A))^\perp) \end{array}$$

Consideriamo il vettore residuo: $\mathbf{r}(\mathbf{x}) = \mathbf{b} - A\mathbf{x} = \mathbf{b}_1 - A\mathbf{x} + \mathbf{b}_2 = \mathbf{y} + \mathbf{b}_2$, dove abbiamo posto $\mathbf{y} = \mathbf{b}_1 - A\mathbf{x}$. Osserivamo che il vettore $\mathbf{y} \in \mathrm{Im}(A)$. Calcoliamo la norma 2 di $\mathbf{r}(\mathbf{x})$:

$$\begin{aligned} ||\mathbf{r}(\mathbf{x})||_2^2 = \;& \mathbf{r}^T \cdot \mathbf{r} = (\mathbf{y} + \mathbf{b}_2)^T(\mathbf{y} + \mathbf{b}_2) = (\mathbf{y}^T + \mathbf{b}_2^T) \cdot (\mathbf{y} + \mathbf{b}_2) \\ & = \mathbf{y}^T\mathbf{y} + \mathbf{y}^T\mathbf{b}_2 + \mathbf{b}_2^T\mathbf{y} + \mathbf{b}_2^T\mathbf{b}_2 = ||\mathbf{y}||_2^2 + ||\mathbf{b}_2||_2^2 \end{aligned}$$

Il minimo valore di $||\mathbf{r}(\mathbf{x})||_2$ lo si ottiene in corrispondenza della scelta $\mathbf{x} = \mathbf{x}^*$; infatti si ha: $\mathbf{y} = \mathbf{b}_1 - A\mathbf{x}^* = \mathbf{0} \Rightarrow ||\mathbf{y}||_2 = 0$. Infine:
$A\mathbf{x}^* = \mathbf{b}_1 \Rightarrow A^T A\mathbf{x}^* = A^T\mathbf{b}_1 \Rightarrow A^T A\mathbf{x}^* = A^T\mathbf{b}_1 + A^T\mathbf{b}_2 \Rightarrow$
$A^T A\mathbf{x}^* = A^T(\mathbf{b}_1 + \mathbf{b}_2) \Rightarrow A^T A\mathbf{x}^* = A^T\mathbf{b}$

Limiti dell'interpolazione polinomiale

Siano dati $n+1$ punti del piano (x_i, y_i), $i = 0, \ldots, n$ con i nodi x_i distinti. Sappiamo che esiste ed è unico il polinomio $p_n(x)$ di grado al più n che interpola i dati:

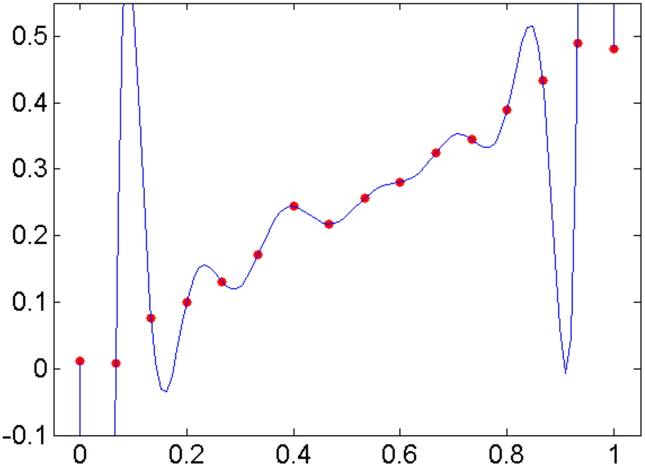
$$p_n(x_i) = y_i, \quad \forall i = 0, \ldots, n \qquad \text{condizioni di interpolazione}$$

Il polinomio interpolante non è tuttavia appropriato quando:

- **il numero di dati è elevato**. In tal caso il polinomio potrà presentare forti oscillazioni tra un nodo e l'altro soprattutto verso gli estremi dell'intervallo, rappresentando male l'andamento dei dati (fenomeno di Runge). Inoltre un polinomio di grado elevato è spesso mal condizionato (es. polinomio di Wilkinson).
- **le ordinate y_i sono affette da errore**. Ciò avviene ad esempio quando le quantità y_i sono ricavate sperimentalmente. In tal caso non avrebbe molto senso far sì che il polinomio passi esattamente per i punti (x_i, y_i).

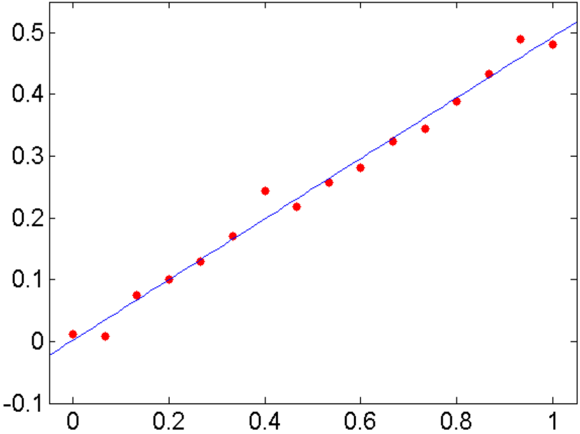
Esempio: n=15 (16 punti)

Il polinomio interpolante $p_n(x)$ di grado 15, pur passando per i punti (x_i, y_i) non approssima bene l'andamento globale degli stessi dati.



Esempio: n=15 (16 punti)

Ad occhio, la funzione che i dati descrivono evidenzia una crescita mediamente lineare. È naturale approssimare i dati mediante una retta che, pur non passando esattamente per i punti (x_i, y_i), se ne *avvicini quanto più possibile*.



Il polinomio di migliore approssimazione

Siano dati $m+1$ punti del piano (x_i, y_i), $i = 0, \ldots, m$ con i nodi x_i distinti. Sia poi $\mathcal{P}_n$ lo spazio vettoriale dei polinomi di grado al più n, e sia $p_n \in \mathcal{P}_n$ rappresentato lungo la base delle potenze:

$$p_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n$$

Supporremo $n \leq m$. Se imponiamo le condizioni di interpolazione $p_n(x_i) = y_i$ per $i = 0, \ldots, m$ otteniamo il sistema di $m+1$ equazioni nelle $n+1$ incognite $a_0, a_1, \ldots, a_n$:

$$\left\{ \begin{array}{rcl} a_0 x_0^n + a_1 x_0^{n-1} + \ldots + a_{n-1} x_0 + a_n & = & y_0 \\ a_0 x_1^n + a_1 x_1^{n-1} + \ldots + a_{n-1} x_1 + a_n & = & y_1 \\ \vdots & & \vdots \\ a_0 x_m^n + a_1 x_m^{n-1} + \ldots + a_{n-1} x_m + a_n & = & y_m \end{array} \right.$$

Se, in particolare $m > n$ il sistema sarà sovradeterminato e quindi, in generale, incompatibile.

Il polinomio di migliore approssimazione

La matrice dei coefficienti $(m+1) \times (n+1)$:

$$A = \begin{pmatrix} x_0^n & x_0^{n-1} & \dots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \dots & x_1 & 1 \\ \vdots & & \vdots & & \vdots \\ x_m^n & x_m^{n-1} & \dots & x_m & 1 \end{pmatrix}$$

è detta matrice di Vandermonde. Si può dimostrare che se i nodi x_i sono distinti, il suo rango è massimo (ovvero $\mathrm{rank}(A) = n+1$).

Ha senso allora risolvere il sistema rettangolare nel senso dei minimi quadrati, cioè ricavare la soluzione che minimizza la norma 2 del residuo $||\mathbf{b} - A\mathbf{x}||_2$.

Il polinomio $p_n^*(x)$ che corrisponde a tale soluzione prende il nome di *polinomio di migliore approssimazione nel senso dei minimi quadrati*.

Il polinomio di migliore approssimazione

Si verifica facilmente (per esercizio) che in questo caso la norma 2 del vettore residuo è:

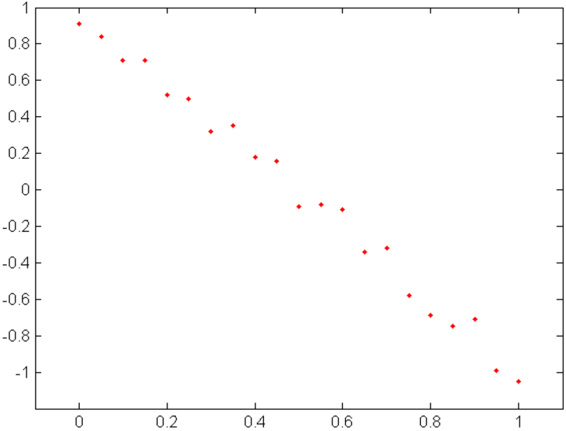
$$||\mathbf{b} - A\mathbf{x}||_2 = \sqrt{\sum_{i=0}^{m} (p_n(x_i) - y_i)^2}$$

Tale quantità prende il nome di *scarto quadratico medio*.

La retta di regressione lineare

Supponiamo che i dati da approssimare evidenzino una dipendenza lineare tra i valori della grandezza y rispetto alla variabile x, come nel seguente esempio (21 punti).

x_i	y_i
0	0.91
0.05	0.84
0.10	0.71
0.15	0.71
0.20	0.52
0.25	0.50
0.30	0.32
0.35	0.35
0.40	0.18
0.45	0.16
0.50	-0.09
0.55	-0.08
0.60	-0.11
0.65	-0.34
0.70	-0.32
0.75	-0.58
0.80	-0.69
0.85	-0.75
0.90	-0.71
0.95	-0.99
1	-1.05



Avrà senso allora approssimare i dati mediante un polinomio lineare (retta)

La retta di regressione lineare

Definizione

La retta che approssima i dati (x_i, y_i), $i = 0, \dots, m$ nel senso dei minimi quadrati, prende il nome di <u>*retta di regressione lineare*</u>.

Deriviamo l'equazione della retta di regressione lineare. Dovremo imporre le condizioni di interpolazione

$$p_1(x_i) = y_i, \qquad i = 0, \dots, m$$

essendo

$$p_1(x) = a_0 x + a_1$$

Otteniamo un sistema di $m + 1$ equazioni nelle 2 incognite a_0 ed a_1.

La retta di regressione lineare

La matrice dei coefficienti $(m+1) \times 2$ del sistema è:

$$A = \begin{pmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}$$

che, come già osservato, avrà rango 2.

Il sistema normale $A^T A\mathbf{x} = A^T\mathbf{b}$ di 2 equazioni e 2 incognite sarà:

$$\begin{cases} \left(\sum_{i=0}^{m} x_i^2\right) a_0 & + & \left(\sum_{i=0}^{m} x_i\right) a_1 & = \sum_{i=0}^{m} x_i y_i \\ \left(\sum_{i=0}^{m} x_i\right) a_0 & + & (m+1)\, a_1 & = \sum_{i=0}^{m} y_i \end{cases}$$

La retta di regressione lineare

Introducendo le quantità:

- $\bar{x} = \dfrac{1}{m+1}\sum_{i=0}^{m} x_i$ (valor medio delle ascisse x_i)
- $\bar{y} = \dfrac{1}{m+1}\sum_{i=0}^{m} y_i$ (valor medio delle ordinate y_i)
- $\mathrm{var}(\mathbf{x}) = \dfrac{1}{m+1}\sum_{i=0}^{m} (x_i - \bar{x})^2$ (varianza di $\mathbf{x}$)
- $\mathrm{cov}(\mathbf{x}) = \dfrac{1}{m+1}\sum_{i=0}^{m} (x_i - \bar{x})(y_i - \bar{y})$ (covarianza di $\mathbf{x}$)

la soluzione del sistema è

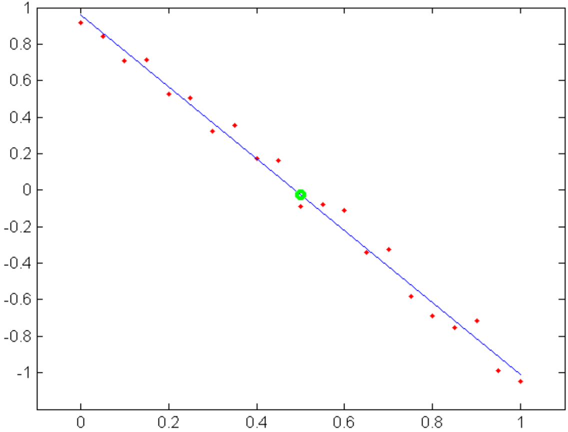
$$a_0 = \frac{\mathrm{cov}(\mathbf{x}, \mathbf{y})}{\mathrm{var}(\mathbf{x})}, \qquad a_1 = \bar{y} - a_0 \bar{x}$$

OSSERVAZIONE: il punto $(\bar{x}, \bar{y})$ appartiene alla retta.

La retta di regressione lineare

Per l'esempio precedente si ha:

$$a_0 = -1.9712 \qquad a_1 = 0.9623$$



La norma 2 del vettore residuo è: $||\mathbf{r}(\mathbf{x})||_2 = 0.2918$.