

Seminario di matematica applicata avanzata

Problemi ai minimi quadrati

Sanna Mirko

Matricola: 44464

Ingegneria elettrica ed elettronica

Indice

- **Introduzione**
 - Il problema dei minimi quadrati
 - I vari tipi di sistemi lineari e le varie tipologie di soluzioni
- **I metodi**
 - Metodo delle equazioni normali
 - Metodo della fattorizzazione QR
 - Metodo di Householder
- **I problemi ai minimi quadrati in Matlab**
 - Le funzioni Matlab utili per risolvere problemi ai minimi quadrati

Introduzione

I problemi ai minimi quadrati nascono nelle varie applicazioni pratiche dove si riscontrano sistemi lineari (nella forma $Ax=b$, dove A è la matrice dei coefficienti, x il vettore delle incognite e b il vettore dei termini noti) in cui la matrice dei coefficienti A non è quadrata bensì rettangolare.

$$\begin{cases} x + y + z = 2 \\ 2x - 5y = 3 \\ 3y + 7z = 1 \end{cases} \longrightarrow A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -5 & 0 \\ 0 & 3 & 7 \end{pmatrix} \quad \begin{matrix} x \\ y \\ z \end{matrix} \quad b = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \quad \text{Sistema lineare determinato}$$

$$\begin{cases} x + y + z = 3 \\ 3x - 7y = 9 \end{cases} \longrightarrow A = \begin{pmatrix} 1 & 1 & 1 \\ 3 & -7 & 0 \end{pmatrix} \quad \begin{matrix} x \\ y \\ w \end{matrix} \quad b = \begin{pmatrix} 3 \\ 9 \end{pmatrix} \quad \text{Sistema lineare con } A \text{ rettangolare}$$

Mentre nel primo caso si ha una, ed una sola, soluzione, se il determinante di $A \neq 0$, come nell'esempio, nel secondo caso le soluzioni dipendono dalle dimensioni di A . Se A è rettangolare non si può esprimere il sistema come $Ax=b$, perciò si userà $\|Ax = b\|$ e si minimizzerà la norma, il problema vale per qualunque norma, tipicamente si usa la norma 2. Quindi il problema ai minimi quadrati nasce da:

$$\min_{x \in \mathbb{R}^n} \|Ax = b\|_2$$

Ipotizzando A di rango pieno, quindi con determinante della matrice di dimensione più grande $\neq 0$, si dividono i casi in 2, una matrice A del tipo $m \times n$, con m numero di righe ed n di colonne, in cui $m > n$ ed il caso $m < n$.

Quando $m > n$ avrò più equazioni che incognite, quindi avrò un sistema che è detto sovradeterminato. In queste condizioni potrei non avere la soluzione del sistema, ad esempio:

$$\begin{cases} x - y = 2 \\ x + y = 3 \\ x - 3y = 2 \end{cases}$$

Questo caso però è tipico nelle applicazioni in cui si fanno molte misurazioni, poiché essendo i dati affetti da errore è conveniente fare più misurazioni per vedere in che intervallo è più probabile che cada il valore effettivo della misurazione. Un esempio può essere la misura di una tensione ai capi di una resistenza.

Quando $m < n$ avrò più incognite che equazioni, nella geometria classica si hanno ∞^{n-m} soluzioni, ed il sistema è detto sottodeterminato. Con $m < n$ il sistema avrà sempre soluzione.

$$\begin{cases} x - 3y + 2z = 5 \\ 8x + y + 3z = 4 \end{cases}$$

Questo caso è tipico nelle applicazioni dove non è possibile fare troppe misurazioni, per vari fattori: costo, sicurezza, tempo, ecc.

La differenza sostanziale tra le due tipologie è il modo di affrontarle, entrambe danno vita a dei problemi ai minimi quadrati, ma mentre nel caso $m > n$ si ha un problema ai minimi quadrati standard, dove devo solo minimizzare la norma $\|Ax = b\|$, nel caso di $m < n$ invece mi serviranno informazioni aggiuntive, ad esempio Matlab usa l'approccio con il vettore x con più 0 possibili; un'altra soluzione è invece quella di minimizzare anche la norma di x , ottenendo così un problema ai minimi quadrati alla minima norma.

I metodi

Il metodo delle equazioni normali

Esistono vari metodi per la risoluzione dei problemi ai minimi quadrati, tra i più utilizzati c'è il metodo che opera risolvendo il sistema normale associato al problema.

Volendo minimizzare una quantità non negativa, conviene considerare la norma quadra:

$$\|Ax = b\|_2^2 = (Ax - b)^T (Ax - b) = x^T A^T Ax - x^T A^T b - b^T Ax + b^T b$$

Sommando i termini simili ($x^T A^T b = b^T Ax$) si ottiene infine:

$$\|Ax = b\|_2^2 = x^T A^T Ax - 2x^T A^T b + b^T b$$

Per minimizzare la norma devo imporre che la derivata prima sia nulla, in questo caso non ho solo un'incognita, devo dunque usare il gradiente, che è definito come: $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right)^T$.

Delle proprietà utili del gradiente sono:

- $\nabla(b^T x) = b$, dove $f(x) = (b^T x)$, con $x, b \in R^n$
- $\nabla(x^T Ax) = Ax + A^T x$, dove A è una matrice quadrata $n \times n$
- $\nabla(x^T Ax) = 2Ax$, se A oltre ad essere quadrata è anche simmetrica

Dunque tornando alla minimizzazione posso scrivere:

$$\frac{1}{2} \nabla(\|Ax = b\|_2^2) = A^T Ax - A^T b = 0$$

Arrivo così al sistema normale: $A^T Ax = A^T b$.

Se A è a rango pieno, la matrice data dalla sua trasposta per se stessa risulta invertibile, ed il sistema ha una soluzione unica data da:

$$x = (A^T A)^{-1} A^T b.$$

Se A invece non ha rango pieno, $A^T A$ risulta singolare, dunque non è possibile invertirla, il sistema però rimane consistente, poiché il vettore $A^T b$ è appartenente all'immagine di A^T , che coincide con l'immagine di $A^T A$. Tra le infinite soluzioni, in questo caso, si adotta la soluzione alla minima norma euclidea.

La matrice $A^+ = (A^T A)^{-1} A^T$ è detta pseudo-inversa, in quanto è un'inversa sinistra di A , ma non destra.

Se la matrice A è a rango pieno allora la sua pseudo-inversa risulta definita positiva ed è possibile risolvere il sistema normale con la fattorizzazione di Cholesky, che ha un basso costo computazionale.

Una volta applicato Cholesky posso calcolare la soluzione del sistema come la soluzione di due sistemi triangolari (quindi di soluzione semplice):

$$\begin{cases} R^T y = A^T b \\ R x = y \end{cases}$$

Un lato negativo di questo approccio è che la matrice $A^T A$ ha un numero di condizionamento pari al quadrato della matrice A , cosa che potrebbe minare alla stabilità.

Il metodo della fattorizzazione QR

La fattorizzazione QR consiste nello scrivere la matrice A come il prodotto di due matrici: una matrice ortogonale Q , ed una matrice triangolare superiore R . La particolarità della matrice ortogonale è che la sua inversa coincide con la trasposta, inoltre Q è trasparente alla norma euclidea, infatti se viene moltiplicata per un vettore non ne altera la norma.

Esistono due algoritmi per calcolare questa fattorizzazione, l'algoritmo di Givens e l'algoritmo di Householder; il primo viene preferito quando la matrice A ha molti 0 ed è in forma quasi triangolare. Si basa sulla matrice

di Givens, detta G. Essa fa compiere una rotazione piana che lascia invariato il sottospazio ortogonale al piano definito dai vettori e_i, e_j . Il secondo invece è il più utilizzato nella maggior parte dei casi, pertanto si vedrà nel dettaglio.

METODO DI HOUSEHOLDER

Il metodo di Householder si fonda sull'usare delle matrici di Householder per la costruzione della matrice ortogonale Q, una matrice di Householder è formata da:

$H=I-2ww^T$, con $w \in R^n$, $\|w\|=1$ (norma euclidea).

$2ww^T$ rappresenta una matrice nxn di rango 1, la matrice H è simmetrica ed ortogonale.

Inserendo un vettore x, determiniamo w in modo che

$Hx=ke_1$, dove e_i sta ad indicare l'i-esimo vettore della base canonica R^n e $k \in R$.

La costruzione di H è possibile dividerla in 3 fasi:

1. Dato che H è ortogonale :

$$\|Hx\| = \|x\| = |k| \text{ da cui ricaviamo } k=\pm\sigma, \text{ con } \sigma=\|x\|.$$

2. Per la definizione di H, si ottiene:

$$Hx = x-2ww^T x = x-2(w^T x)w = ke_1, \text{ che implica:}$$

$$w = \frac{x-ke_1}{2w^T x} = \frac{x-ke_1}{\| -ke_1 \|} \text{ in quanto } w \text{ ha norma unitaria.}$$

Già con questa fase si è determinato w, la terza fase serve per fare delle considerazioni numeriche.

3. Usando la relazione $x^T Hx = x^T x - 2(w^T x)^2 = kx_1$ si ottiene:

$$(w^T x)^2 = \frac{\sigma^2 - kx_1}{2}$$

ed utilizzando l'espressione per w:

$$\|x - ke_1\| = 2w^T x = \sqrt{2(\sigma^2 - kx_1)}$$

Con questa operazione si riduce il calcolo computazionale sulla normalizzazione di w. Per evitare problemi di cancellazione si sceglie il segno di k, lo si sceglie in modo che $-kx_1$ sia positivo.

Una volta costruite le matrici H è possibile costruire la fattorizzazione QR di A. Ipotizzando che A sia quadrata (non è necessario, la fattorizzazione funziona anche su matrici rettangolari, serve solo a rendere più semplici le spiegazioni), si può generare una successione di matrici $A^{(i)}$, con i che varia da 1 alla dimensione della matrice, in modo che $A^{(n)}$ sia una matrice triangolare superiore.

Scrivendo la matrice A in termini delle sue colonne:

$$A = A^{(1)} = \begin{bmatrix} a_1^{(1)} & a_2^{(1)} & \dots & a_n^{(1)} \end{bmatrix}$$

Seguendo i 3 passi scritti prima, si crea una matrice elementare di Householder tale che:

$$H_1 a_1^{(1)} = k_1 e_1$$

Applicando questa matrice alla sinistra della matrice $A^{(1)}$, si genera la seconda matrice della successione:

$$A^{(2)} = H_1 A^{(1)} = \begin{bmatrix} a_1^{(2)} & a_2^{(2)} & \dots & a_n^{(2)} \end{bmatrix}, \text{ dove } a_1^{(2)} = k_1 e_1, a_j^{(2)} = H_1 a_j^{(1)} \text{ con } j=2, \dots, n.$$

La matrice $A^{(2)}$ appena generata ha la struttura seguente:

$$A^{(2)} = \begin{bmatrix} k_1 & a_{12}^{(2)} & \dots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \dots & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \dots & a_{nn}^{(2)} \end{bmatrix} = \begin{bmatrix} k_1 & v_1^T \\ 0 & \hat{A}^{(2)} \end{bmatrix},$$

con 0 vettore colonna di dimensioni tali che le dimensioni della matrice siano rispettate, con componenti tutte nulle, ed $\hat{A}^{(2)}$ una sottomatrice di dimensione n-1.

Dopo di che si considera la sottomatrice ricavata $\hat{A}^{(2)}$:

$$\hat{A}^{(2)} = \begin{bmatrix} \hat{a}_2^{(2)} & \hat{a}_3^{(2)} & \dots & \hat{a}_n^{(2)} \end{bmatrix}$$

Si costruirà quindi la matrice elementare di Householder di dimensione n-1 in modo che:

$\hat{H}_2 \hat{a}_2^{(2)} = k_2 e_1$, in questo caso e_1 ha le dimensioni adeguate a fare il prodotto tra vettori, quindi ha lunghezza $n-1$.

Dopo questo passaggio si aggiunge una riga ed una colonna alla matrice \hat{H}_2 , la riga e la colonna sono le prime della matrice identità, affinché raggiunga la dimensione n :

$$H_2 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & \hat{H}_2 & \\ \vdots & & & \\ 0 & & & \end{bmatrix} = \begin{bmatrix} 1 & 0^T \\ 0 & \hat{H}_2 \end{bmatrix}$$

E moltiplicando a sinistra per la matrice $A^{(2)}$ si ottiene:

$$A^{(3)} = H_2 A^{(2)} = \begin{bmatrix} 1 & 0^T \\ 0 & \hat{H}_2 \end{bmatrix} \begin{bmatrix} k_1 & v_1^T \\ 0 & \hat{A}^{(2)} \end{bmatrix} = \begin{bmatrix} k_1 & v_1^T \\ 0 & \hat{H}_2 \hat{A}^{(2)} \end{bmatrix},$$

questi processi si iterano fino al passo $n-1$, dove ci si ritrova con la matrice triangolare superiore R :

$$R = A^{(n)} = H_{(n-1)} A^{(n-1)} = \begin{bmatrix} k_1 & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & k_n \end{bmatrix}$$

Dato che $A = QR$, avendo l'espressione di R , è facile ricavare Q , che risulta:

$Q = H_1 H_2 \dots H_{n-1}$, ovvero la produttoria di tutte le matrici elementari di Householder.

FATTORIZZAZIONE QR NEI PROBLEMI AI MINIMI QUADRATI

La fattorizzazione QR è un altro metodo efficace per la risoluzione dei problemi ai minimi quadrati. In questo metodo, come per il metodo delle equazioni normali, è più pratico usare il quadrato della norma del residuo, sostituendo A con la sua fattorizzazione, $A=QR$, ottenendo:

$$\|Ax - b\|^2 = \|QRx - b\|^2 = \|Q(Rx - Q^T b)\|^2 = \|Rx - c\|^2$$

in cui $c = Q^T b$, e Q sparisce dai calcoli in quanto una matrice ortogonale non influenza la norma euclidea di un vettore.

La matrice R ha forma:

$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$, dove R_1 è una matrice quadrata triangolare superiore, e se A è a rango pieno, non singolare.

Dividendo anche c in modo coerente (facendo combaciare le dimensioni dei blocchi con R) si ottiene

$c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$, con $c_1 \in R^n$ e $c_2 \in R^{m-n}$, ottenendo dunque:

$$\|Ax - b\|^2 = \|Rx - c\|^2 = \left\| \begin{bmatrix} R_1 x \\ 0 \end{bmatrix} - \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right\|^2 = \|R_1 x - c_1\|^2 + \|c_2\|^2.$$

Se R_1 non è singolare, il sistema $R_1 x = c_1$, ammette una ed una sola soluzione, per la quale si ha che $\|R_1 x - c_1\|^2 = 0$. Da questa soluzione si ricava che:

$$\min_{x \in R^n} \|Ax - b\| = \|c_2\|,$$

dove nell'eventualità che $c_2 = 0$, ci si ritrova nel caso di un sistema classico, con n equazioni ed n incognite. Se non è nullo invece è la soluzione del problema di minimizzazione, e la norma di c_2 è la misura del residuo. Il vantaggio del metodo della fattorizzazione QR sta nell'operare direttamente sulla matrice A , il cui condizionamento è la radice quadrata di quello del metodo delle equazioni normali, ovvero di $A^T A$. Questo aumenta la stabilità del metodo, inoltre la sua complessità computazionale è di $O(\frac{n^3}{3})$ moltiplicazioni, usando l'algoritmo di Householder.

Matlab

I problemi ai minimi quadrati su Matlab

Matlab è un software di matematica, il cui nome è Matrix Laboratory, quindi è nato per le operazioni sulle matrici. Matlab risolve automaticamente i problemi ai minimi quadrati con l'istruzione $x=A\backslash b$, utilizzando, in caso di problemi sottodeterminati, la soluzione con la x col maggior numero di 0. È possibile però usare specifiche istruzioni per avere i risultati desiderati, invece di avere lo scarto tipo con il maggior numero di 0, in caso non ci interessi.

Ponendo A rettangolare, o quadrata con rango non pieno, non esiste l'inversa, come già detto infatti si usa la sua pseudo-inversa:

$$A^+ = (A^T A)^{-1} A^T;$$

in MATLAB avremo $X=\text{pinv}(A)$, l'istruzione $\text{pinv}(A)$ calcola la pseudo-inversa di A . Una volta calcolata la pseudo-inversa è possibile quindi risolvere il problema:

$$Ax=b$$

anche se la matrice A non è invertibile.

Se il sistema di equazioni è sottodeterminato, allora esistono infinite soluzioni. Due possibili soluzioni si possono ottenere con gli operatori 'pinv' e '\'; il primo minimizza la norma della soluzione, il secondo il numero degli elementi non nulli, che risulta pari al rango della matrice. La soluzione con l'operatore '\' come già detto è la soluzione di default di MATLAB, in Figura A, si confrontano i risultati dei 2 metodi.

Il vettore x ha norma $\|x\|=0.0666$, mentre quella di y è $\|y\|=0.0700$, y però ha solo 3 elementi non nulli, x invece li ha tutti e 5.

```

New to MATLAB? Watch this Video, see Examples, or read Getting Started.
A =
    17    24     1     8    15
    23     5     7    14    16
     4     6    13    20    22

>> b=[1;0;1]

b =
     1
     0
     1

>> x=pinv(A)*b

x =
   -0.0366
    0.0480
    0.0061
    0.0119
    0.0246

>> y=A\b

y =
   -0.0376
    0.0429
         0
         0
    0.0406

```

Figura A

Se invece il sistema è irrisolvibile, ovvero il rango di A è minore di quello di [A b], il problema è cercare la migliore approssimazione della soluzione. L'operatore 'pinv' cercherà di minimizzare la norma dell'errore:

$$\min_{x \in R^n} \|Ax - b\|_2$$

L'operatore 'lsqov' cerca di minimizzare una norma opportunamente pesata:

$$\|Ax - b\|_{V^{-1}}^2 = (Ax - b)^T V^{-1} (Ax - b)$$

Infine l'operatore 'lsqnonneg' cerca una soluzione non negativa, ovvero risolve il problema ai minimi quadrati, ma non quello della minima norma, ma con l'ulteriore vincolo di una x sempre ≥ 0 . Nella Figura B, si producono i risultati di questi 3 metodi, nella Figura C invece si confrontano gli errori.

Figura B

```
Command Window
New to MATLAB? Watch this Video, see Examples, or read Getting Started.

>> A=rand(4,2);
>> b=rand(4,1);
>> x=pinv(A)*b

x =

    -0.3011
     0.7943

>> y=lsqnonneg(A,b)

y =

     0
    0.5953

>> V=[1 0 0 0;0 10 0 0; 0 0 100 0; 0 0 0 1000];
>> z=lsqcov(A,b,V)

z =

    -1.5836
     1.3337

fx >> |
```

Figura C

```
Command Window
New to MATLAB? Watch this Video, see Examples, or read Getting Started.

>> A*x-b

ans =

    -0.1479
     0.2893
    -0.0339
    -0.1531

>> A*y-b

ans =

    -0.2328
     0.2867
     0.0453
    -0.0937

>> A*z-b

ans =

    -0.0049
     0.0684
    -0.4565
    -0.6153

fx >> |
```