

Fattorizzazione QR e problemi ai minimi quadrati

Luca Gemignani

April 2025

1 Matrici elementari ortogonali

Sia $Q = I_n - \alpha \mathbf{x} \mathbf{x}^T$, con $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$. Se imponiamo $Q^T Q = I_n$ abbiamo

$$Q^T Q = I_n - \alpha \mathbf{x} \mathbf{x}^T - \alpha \mathbf{x} \mathbf{x}^T + \alpha^2 \|\mathbf{x}\|_2^2 \mathbf{x} \mathbf{x}^T = I_n$$

da cui

$$\alpha = \frac{2}{\|\mathbf{x}\|_2^2}.$$

Le matrici $Q = I_n - \frac{2}{\|\mathbf{x}\|_2^2} \mathbf{x} \mathbf{x}^T$, con $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$, sono dette matrici elementari di Householder. Tali matrici sono ortogonali e possono essere utilizzate in alternativa alle matrici elementari di Gauss per il calcolo della riduzione in forma tringolare.

Iniziamo con osservare che dato $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$, è possibile determinare una matrice elementare di Householder tale che $E\mathbf{v} = \gamma \mathbf{e}_1$, per un opportuno γ . Osserviamo che se E esiste allora

$$\|E\mathbf{v}\|_2^2 = \|\mathbf{v}\|_2^2 = \gamma^2$$

da cui $\gamma = \pm \|\mathbf{v}\|_2$. La condizione $E\mathbf{v} = \gamma \mathbf{e}_1$ si scrive

$$I_n \mathbf{v} - \frac{2}{\|\mathbf{x}\|_2^2} \mathbf{x} \mathbf{x}^T \mathbf{v} = \gamma \mathbf{e}_1$$

da cui

$$\mathbf{x} = \theta(\mathbf{v} - \gamma \mathbf{e}_1), \quad \theta \in \mathbb{R}.$$

Possiamo quindi prendere $\mathbf{x} = \mathbf{v} - \gamma \mathbf{e}_1$ con $\gamma = \pm \|\mathbf{v}\|_2$. Abbiamo allora

$$2\mathbf{x}^T \mathbf{v} = \|\mathbf{x}\|_2^2$$

e dunque

$$E\mathbf{v} = \gamma \mathbf{e}_1 + \mathbf{x} - \mathbf{x} = \gamma \mathbf{e}_1.$$

2 Fattorizzazione QR

Le matrici elementari di Householder possono essere utilizzate in analogia alle matrici elementari di Gauss per la riduzione di una matrice $A \in \mathbb{R}^{n \times n}$ in forma triangolare superiore.

Sia $A = A^{(0)}$. Al primo passo del processo di eliminazione determiniamo E_1 matrice elementare di Householder tale che $E_1 A^{(0)} \mathbf{e}_1 = \gamma_1 \mathbf{e}_1$. La matrice $A^{(1)} = E_1 A^{(0)}$ risulta pertanto partizionabile a blocchi come segue

$$A^{(1)} = \left[\begin{array}{c|c} \gamma_1 & \mathbf{a}_1^T \\ \mathbf{0} & A_2^{(1)} \end{array} \right], \quad A_2^{(1)} \in \mathbb{R}^{(n-1) \times (n-1)}.$$

Al secondo passo determiniamo una matrice ortogonale E_2 definita da

$$E_2 = \left[\begin{array}{c|c} 1 & \mathbf{0}^T \\ \mathbf{0} & \widehat{E}_2 \end{array} \right], \quad \widehat{E}_2 \in \mathbb{R}^{(n-1) \times (n-1)},$$

con $\widehat{E}_2$ matrice elementare di Householder tale che $\widehat{E}_2 A_2^{(1)} \mathbf{e}_1 = \gamma_2 \mathbf{e}_1$. Procedendo in questo modo per $n - 1$ passi si perviene ad una matrice triangolare superiore $A^{(n-1)} = R$ con

$$E_{n-1} E_{n-2} \cdots E_1 A = R$$

o, equivalentemente

$$A = (E_1^T E_2^T \cdots E_{n-1}^T) R.$$

Poiché il prodotto di matrici ortogonali è una matrice ortogonale, ponendo $Q = E_1^T E_2^T \cdots E_{n-1}^T$ si conclude che

$$A = QR$$

con Q matrice ortogonale ed R matrice triangolare superiore. Tale fattorizzazione è detta fattorizzazione QR di A . Ogni matrice ammette fattorizzazione QR. Il costo computazionale del calcolo della fattorizzazione QR è cubico nella dimensione della matrice.

La fattorizzazione QR riveste un'importanza particolare nel caso di matrici rettangolari. Se A è rettangolare, diciamo $A \in \mathbb{R}^{m \times n}$ con $m \geq n$, allora A si fattorizza nella forma QR come

$$A = QR, \quad Q \in \mathbb{R}^{m \times m}, \quad R \in \mathbb{R}^{m \times n},$$

dove

$$R = \left[\begin{array}{c} \widehat{R} \\ 0_{m-n,n} \end{array} \right]$$

con $0_{m-n,n}$ matrice nulla di ordine $(m-n) \times n$ e $\widehat{R}$ matrice triangolare superiore di ordine n . Se A ha rango massimo uguale ad n allora $\widehat{R}$ è invertibile. Il costo computazionale della fattorizzazione è $O(mn^2)$.

3 Problemi ai minimi quadrati

Iniziamo con il descrivere una applicazione.

”Intuitivamente, qual è il problema? Supponiamo di possedere dei dati, collezionati nel tempo, che raccolgono una serie di informazioni su delle abitazioni situate in una particolare zona degli Stati Uniti. Per ciascuna di queste abitazioni abbiamo, ad esempio, l’informazione sulla superficie che occupa, sul numero di stanze che possiede, sulla collocazione geografica rispetto alla quale è situata, sul prezzo di vendita e così via. A fronte di questi dati, vorremmo dunque rispondere alla seguente domanda: se possedessimo una casa di una determinata superficie, quale sarà il suo costo?

Le strade per rispondere sono essenzialmente due:

1. Nel dataset da cui partiamo abbiamo fortunatamente già mappato la superficie della nostra abitazione in relazione al suo costo, e allora la risposta sarà direttamente ottenibile dal valore risultante.
2. La superficie di nostro interesse non è presente nel nostro dataset. Dobbiamo pertanto lavorare, tramite la regressione lineare, per prevedere il costo dell’abitazione. ”

(tratto da <https://medium.com/matteotroia/machine-learning-la-regressione-lineare-7b2f096adc26>)

Sia $\mathcal{S} = \{(x_1, y_1), \dots, (x_m, y_m)\}$, $(x_i, y_i) \in \mathbb{R}^2$, il nostro dataset. Un modello di regressione lineare per $\mathcal{S}$ è una funzione lineare $f(x) = ax + b$ tale che $f(x_i) \simeq y_i$, $1 \leq i \leq m$. Per formalizzare questa condizione determiniamo i parametri a e b in modo da minimizzare la ”somma degli scarti quadratici”, ovvero risolviamo il problema

$$\min_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^m (ax_i + b - y_i)^2 = \min_{(a,b) \in \mathbb{R}^2} \|\mathbf{w}\|_2^2,$$

con $\mathbf{w}$ il vettore di componenti $w_i = ax_i + b - y_i$, $1 \leq i \leq m$. Denotando con $V \in \mathbb{R}^{m \times 2}$ la matrice ad elementi $v_{i,j} = x_i^{j-1}$ si ottiene che $\mathbf{w} = V \begin{bmatrix} a \\ b \end{bmatrix} - \mathbf{y}$. Pertanto il problema di minimizzazione può essere riformulato come

$$\min_{(a,b) \in \mathbb{R}^2} \left\| V \begin{bmatrix} a \\ b \end{bmatrix} - \mathbf{y} \right\|_2^2,$$

Se $x_i \neq x_j$ per $i \neq j$ allora V ha rango massimo $n = 2$. Sotto questa assunzione sia $V = QR$ la fattorizzazione QR di V . Si osserva che

$$\left\| V \begin{bmatrix} a \\ b \end{bmatrix} - \mathbf{y} \right\|_2^2 = \left\| Q^T V \begin{bmatrix} a \\ b \end{bmatrix} - Q^T \mathbf{y} \right\|_2^2 = \left\| R \begin{bmatrix} a \\ b \end{bmatrix} - Q^T \mathbf{y} \right\|_2^2.$$

Ricordando dalla sezione precedente che

$$R = \begin{bmatrix} \hat{R} \\ 0_{m-2,2} \end{bmatrix}$$

con $\hat{R} \in \mathbb{R}^{2 \times 2}$ invertibile, si conclude che il punto di ottimo è determinato risolvendo il sistema 2×2

$$\hat{R} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

con $z = Q^T \mathbf{y}$. Osserviamo inoltre che

$$\hat{R} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \iff \hat{R}^T \hat{R} \begin{bmatrix} a \\ b \end{bmatrix} = \hat{R}^T \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \iff V^T V \begin{bmatrix} a \\ b \end{bmatrix} = V^T \mathbf{y}.$$

Il sistema

$$V^T V \begin{bmatrix} a \\ b \end{bmatrix} = V^T \mathbf{y}$$

è noto come sistema "delle equazioni normali".

Il procedimento appena esposto si estende per la risoluzione dei sistemi rettangolari $A\mathbf{x} = \mathbf{b}$, $A \in \mathbb{R}^{m \times n}$, $m \geq n$, di rango pieno. Il problema viene riformulato come

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_2^2.$$

Questi sono detti problemi lineari ai minimi quadrati. L'algoritmo di scelta utilizza la fattorizzazione QR di A . L'approccio delle "equazioni normali" pone problemi di condizionamento al crescere di n e del condizionamento di A poiché si dimostra che

$$\mathcal{K}_2(A^T A) = (\mathcal{K}_2(A))^2.$$