Le condizioni di Karush-Kuhn-Tucker ed il metodo dei minimi quadrati

Giandomenico Mastroeni

Corso di Modelli Matematici Ambientali

A.A. 2020/2021

Contenuti della lezione

- Le condizioni di Karush-Kuhn-Tucker
- Il metodo dei minimi quadrati
- La retta di regressione lineare

Le condizioni di Karush-Kuhn-Tucker (KKT)

Consideriamo il problema:

$$\min \; f(x) \quad s.t. \quad x \in K := \{x \in \mathbb{R}^n : g(x) \leq 0\} \qquad (P)$$

ove

- $g(x) := (g_1(x), ..., g_m(x))$,
- $f, g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ sono funzioni di classe C^1, differenziabili con continuitá, per $i = 1, .., m$.

Essendo la regione ammissibile K definita mediante vincoli di disuguaglianza le condizioni di ottimalitá di tipo lagrangiano per il problema (P) risultano essere lievemente modificate: in questo caso particolare, prendono il nome di condizioni di Karush-Kuhn-Tucker.

Le condizioni KKT

Sia $\bar{x} \in K$ e poniamo $I(\bar{x}) := \{i \in [1, .., m] : g_i(\bar{x}) = 0\}$.

Teorema

Supponiamo che $\bar{x}$ sia un punto di minimo locale per il problema (P) e che valga una delle seguenti condizioni:
i) i vettori $\nabla g_i(\bar{x})$, $i \in I(\bar{x})$ siano linearmente indipendenti;
ii) g_i siano funzioni lineari (affini) per $i = 1, .., m$.
Allora esiste $\lambda = (\lambda_1, ..., \lambda_m)^T \in \mathbb{R}^m$ tale che la coppia $(\bar{x}, \lambda)$ sia soluzione del seguente sistema:

$$\begin{cases} \nabla f(x) \; + \; \sum_{i=1}^{m} \lambda_i \nabla g_i(x) \; = \; 0 \\ \lambda_i g_i(x) \; = \; 0, \quad i = 1, .., m \\ \lambda \geq 0, \; g(x) \leq 0 \end{cases} \tag{1}$$

Si noti che la condizione $\lambda_i g_i(x) = 0$, per $i = 1, .., m$ é equivalente a $\lambda^T g(\bar{x}) = 0$.

Definendo la funzione lagrangiana associata al problema (P)

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

il sistema (1) puo' essere equivalentemente formulato come

$$\begin{cases} \nabla_x L(x, \lambda) = 0 \\ \lambda_i g_i(x) = 0, \quad i = 1, .., m \\ \lambda \geq 0, \; g(x) \leq 0 \end{cases} \tag{2}$$

Osservazione: Se si considera (in luogo di (P)) il problema

$$\max\ f(x) \quad s.t. \quad x \in K := \{x \in \mathbb{R}^n : g(x) \leq 0\} \qquad (P')$$

ricordando che

$$max[f(x)] = -min[-f(x)]$$

basta considerare, in luogo di (1), il sistema

$$\begin{cases} \nabla f(x) \ - \ \sum_{i=1}^{m} \lambda_i \nabla g_i(x) \ = \ 0 \\ \lambda_i g_i(x) \ = \ 0, \quad i = 1, .., m \\ \lambda \geq 0, \ g(x) \leq 0 \end{cases} \qquad (3)$$

Il caso lineare

Consideriamo il caso particolare di un problema di programmazione lineare:

$$\max \; c^T x \quad s.t. \quad x \in K := \{x \in \mathbb{R}^n : Ax \leq b\} \qquad (LP).$$

In questo caso $f(x) = c^T x$ e $g(x) = Ax - b$. Considerando i gradienti come vettori riga avremo che $\nabla f(x) = c^T$ e $\nabla g_i(x) = A_i$, $i = 1, .., m$, ove A_i denota la riga i-esima della matrice A. Pertanto,

$$\nabla f(x) \; - \; \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = c^T - \sum_{i=1}^{m} \lambda_i A_i = c^T - \lambda^T A$$

da cui il sistema (3) diviene:

$$\begin{cases} c^T \; - \; \lambda^T A \; = \; 0 \\ \lambda^T(Ax - b) \; = \; 0, \\ \lambda \geq 0, \; Ax \leq b \end{cases} \qquad (4)$$

Il caso lineare

Dalla prima relazione nel sistema, segue che $\lambda^T A = c^T$ da cui la seconda equazione si puó scrivere come

$$\lambda^T (Ax - b) = \lambda^T Ax - \lambda^T b = 0 \iff c^T x - \lambda^T b = 0,$$

ossia $c^T x = \lambda^T b$. Quindi il sistema (4) coincide con il sistema che definisce le condizioni di ottimalitá di un problema di programmazione lineare, come visto nelle lezioni precedenti, ed il vettore λ soluzione del sistema (4) e' anche soluzione del problema duale associato al problema dato (LP):

$$\min \ \lambda^T b \quad s.t. \quad \lambda \in \Lambda := \{\lambda \in \mathbb{R}^m : \lambda^T A = c^T, \ \lambda \geq 0\} \qquad (D)$$

infatti si puó dimostrare che risulta:

$$\lambda^T b \geq c^T x, \quad \forall \lambda \in \Lambda, \ \forall x \in K.$$

Esempio. Consideriamo il problema

$$\begin{cases} \min(2x_1 + x_2) \\ -x_1 \quad \leq 0, \\ \quad -x_2 \leq 0, \\ x_1 + x_2 \leq 5 \end{cases}$$

La funzione lagrangiana é:

$$L(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) = 2x_1 + x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \lambda_3(x_1 + x_2 - 5)$$

Le condizioni KKT date dal sistema (2) sono:

$$\begin{cases} 2 - \lambda_1 + \lambda_3 = 0 \\ 1 - \lambda_2 + \lambda_3 = 0 \\ \lambda_1(-x_1) = \lambda_2(-x_2) = \lambda_3(x_1 + x_2 - 5) = 0 \\ \lambda_i \geq 0, \ i = 1, 2, 3 \\ x_1 \geq 0, \ x_2 \geq 0, \ x_1 + x_2 \leq 5 \end{cases}$$

Dalle prime due equazioni del sistema risulta evidente che λ_1 e λ_2 non possono essere contemporaneamente nulli, dalle condizioni $\lambda_1 x_1 = \lambda_2 x_2 = 0$ segue che $x_1 = 0$ oppure $x_2 = 0$. Verifichiamo le condizioni KKT nel punto $\bar{x} = (0,0)$ che fornisce il valore minore per la funzione obbiettivo.
Dalla condizione $\lambda_3(x_1 + x_2 - 5) = 0$ si ha $\lambda_3 = 0$ da cui il sistema precedente si riduce a

$$\begin{cases} 2 - \lambda_1 = 0 \\ 1 - \lambda_2 = 0 \\ \lambda_1 \geq 0, \;\; \lambda_2 \geq 0 \end{cases}$$

Segue che $\lambda_1 = 2$ e $\lambda_2 = 1$ e $\bar{x}$ e' soluzione ottima del problema proposto.

Il metodo dei minimi quadrati

Nell'ambito dei metodi dei minimi quadrati analizziamo il problema della determinazione della retta di regressione lineare mediante il seguente esempio.
Consideriamo una popolazione P di animali della quale possiamo calcolare la numerositá negli istanti di tempo $t_i = 1, 2, 3, 4$, pari a $P_i = 100, 140, 180, 200$, $i = 1, .., 4$, rispettivamente. Da indagini statistiche viene stimata l'esistenza di una dipendenza di tipo lineare tra la variabile P e la variabile indipendente t, della forma

$$P(t) = at + b.$$

Vogliamo determinare i parametri a e b in modo che risulti minimo l'errore che si commette rimpiazzando con il valore teorico $P(t_i)$ il valore osservato di P_i per $i = 1, .., 4$, ossia, il valore:

$$\sum_{i=1}^{4}[P(t_i) - P_i]^2 = \sum_{i=1}^{4}[at_i + b - P_i]^2 = f(a,b).$$

La retta che otterremo sara' detta *retta di regressione lineare di* P.
A tal fine risolviamo il problema

$$\min_{(a,b)\in\mathbb{R}^2} f(a,b) = (a+b-100)^2+(2a+b-140)^2+(3a+b-180)^2+$$

$$+(4a+b-200)^2,$$

le cui condizioni di ottimalitá sono date da $\nabla f(a,b) = 0$, cioe'

$$\frac{\partial f}{\partial a}(a,b) = 0 \quad \frac{\partial f}{\partial b}(a,b) = 0.$$

Cio' equivale a risolvere il sistema

$$\begin{cases} 30a + 10b = 1720 \\ 10a + 4b = 620 \end{cases}$$

osservando che

$$\frac{\partial f}{\partial a}(a,b) = 2[a+b-100+2(2a+b-140)+3(3a+b-180)+4(4a+b-200)]$$

$$\frac{\partial f}{\partial b}(a,b) = 2[a+b-100+(2a+b-140)+(3a+b-180)+(4a+b-200)]$$

La soluzione del precedente sistema é $a = 34,\ b = 70$, da cui

$$P(t) = 34t + 70$$

é la retta di regressione lineare.

Ad esempio, supponendo che l'unitá di misura della variabile t sia il mese, dopo due mesi e mezzo, al tempo $t = 2.5$, deduciamo un valore stimato per la numerositá della popolazione pari a $P(2.5) = 155$.

Formule generali per la retta di regressione lineare

Siano X ed Y le variabili indipendente e dipendente rispettivamente e siano (X_i, Y_i) i valori osservati, $i = 1, .., n$. Per la retta di regressione lineare $Y = aX + b$ risulta

$$a = \frac{\sum_{i=1}^{n} X_i Y_i - n M_X M_Y}{\sum_{i=1}^{n} X_i^2 - n M_X^2} \quad b = M_Y - a M_X$$

ove

$$M_X = \frac{\sum_{i=1}^{n} X_i}{n}, \quad M_Y = \frac{\sum_{i=1}^{n} Y_i}{n}$$

Esempio. Considerando il problema precedente otteniamo

$$n = 4, \quad M_t = 2.5, \quad M_P = 155 \quad \sum_{i=1}^{4} t_i P_i = 1720, \quad \sum_{i=1}^{4} t_i^2 = 30$$

$$a = \frac{1720 - 4 \cdot 2.5 \cdot 155}{30 - 4 \cdot 6.25} = 34, \ \ b = 155 - 34 \cdot 2.5 = 70.$$

Stime statistiche

Siano X ed Y le variabili indipendente e dipendente rispettivamente e siano (X_i, Y_i) i valori osservati, $i = 1, .., n$. Definiamo:

- Media campionaria di X: $M_X = \dfrac{\sum_{i=1}^n X_i}{n}$;
- Media campionaria di Y: $M_Y = \dfrac{\sum_{i=1}^n Y_i}{n}$;
- Varianza campionaria di X: $S_X^2 = \dfrac{\sum_{i=1}^n (X_i - M_X)^2}{n-1}$
  $= \dfrac{\sum_{i=1}^n X_i^2 - nM_X^2}{n-1}$;
- Covarianza campionaria: $S_{XY} = \dfrac{\sum_{i=1}^n (X_i - M_X)(Y_i - M_Y)}{n-1}$
  $= \dfrac{\sum_{i=1}^n X_i Y_i - nM_X M_Y}{n-1}$;
- Coefficiente di correlazione: $\rho = \dfrac{S_{XY}}{S_X S_Y}$.

Esempio 2

Considerando il problema definito nell'esempio precedente. In questo caso $n = 4$ ed abbiamo giá calcolato le seguenti quantitá:

$$M_t = 2.5, \quad M_P = 155 \quad \sum_{i=1}^{4} t_i P_i = 1720, \quad \sum_{i=1}^{4} t_i^2 = 30.$$

Dalle formule viste otteniamo:

$$S_{tP} = \frac{\sum_{i=1}^{4} t_i P_i - n M_t M_P}{3} = \frac{170}{3},$$

$$S_t^2 = \frac{\sum_{i=1}^{n} t_i^2 - n M_t^2}{n-1} = \frac{30 - 4(2.5)^2}{3} = \frac{5}{3}$$

$$S_P^2 = \frac{\sum_{i=1}^{n} P_i^2 - n M_P^2}{n-1} = \frac{1020 \cdot 10^2 - 4(155)^2}{3} = \frac{59 \cdot 10^2}{3}$$

Calcoliamo il coefficiente di correlazione

$$\rho = \frac{S_{tP}}{S_t S_P} = \frac{\frac{170}{3}}{\sqrt{\frac{5}{3}} \cdot 10\sqrt{\frac{59}{3}}} = \frac{17}{\sqrt{295}} \approx 0.989.$$

Osserviamo che il coefficiente di correlazione ρ é molto prossimo ad 1, cosicché la retta di regressione lineare $P(t) = 34t + 70$ fornisce una stima abbastanza buona del comportamento della popolazione nell'arco temporale $[1, 4]$.

Approssimazione esponenziale

Supponiamo di disporre dei dati (x_i, P_i), $i = 1, .., n$ e di voler determinare un'approssimazione della forma $P(x) = Ae^{Kx}$, ove $A, K \in \mathbb{R}$. Se utilizzassimo in modo diretto il metodo dei minimi quadrati saremmo portati a determinare A e K in modo da minimizzare la funzione

$$\sum_{i=1}^{n}(P(x_i) - P_i)^2 = \sum_{i=1}^{n}(Ae^{Kx_i} - P_i)^2 = f(A, K).$$

Applicando le condizioni KKT, che in questo caso coincidono con le classiche condizioni di stazionarietá della funzione f, ció comporta la risoluzione del sistema non lineare

$$\begin{cases} \dfrac{\partial f}{\partial A}(A, K) = 0 \\ \dfrac{\partial f}{\partial K}(A, K) = 0 \end{cases}$$

Il precedente sistema non ammette in generale una soluzione definita esplicitamente e puó essere risolto mediante metodi di approssimazione numerica.
Un' altra possibilitá per determinare la funzione $P(x) = Ae^{Kx}$ consiste nel considerare i valori $log(P_i)$ in luogo di P_i. In tal caso, prendendo i logaritmi di ambo i membri dell'espressione che definisce $P(x)$ otteniamo:

$$log(P(x)) = log(Ae^{Kx}) = log(A) + Kx.$$

Ponendo $y = log(P(x))$, la nostra approssimazione ora e' di tipo lineare nei dati $(x_i, log(P_i))$ e puó essere determinata utilizzando le formule definite per la retta di regressione lineare, ove poniamo:

$$a = K \quad b = log(A).$$

Ponendo $y_i = log(P_i)$, $i = 1, .., n$, le formule relative ai coefficienti a e b nella retta di regressione lineare dei dati (x_i, y_i) divengono:

$$a = \frac{S_{xy}}{S_x^2} = \frac{\displaystyle\sum_{i=1}^{n} x_i log(P_i) - nM_xM_y}{\displaystyle\sum_{i=1}^{n} x_i^2 - nM_x^2} \quad \text{ove } M_y = \frac{\displaystyle\sum_{i=1}^{n} log(P_i)}{n},$$

$$b = M_y - aM_x, \quad \text{da cui }, K = a, A = e^b = e^{M_y - KM_x}.$$

Osserviamo che, con questo procedimento, non stiamo minimizzando la funzione $\sum_{i=1}^{n}(P(x_i) - P_i)^2 = \sum_{i=1}^{n}(Ae^{Kx_i} - P_i)^2$, bensi' la funzione

$$\sum_{i=1}^{n}(log(P(x_i)) - log(P_i))^2 = \sum_{i=1}^{n}(log(A) + Kx_i - log(P_i))^2.$$

Esempio 3

Sia $n = 6$, per $i = 1, .., 6$ consideriamo i seguenti dati:

$(x_i, P_i) := (1, 374)\ (5, 1577)\ (9, 3858)\ (13, 7985)\ (17, 15000)\ (21, 25000)$

Ponendo $y_i = log(P_i)$, otteniamo:

$(x_i, y_i) := (1, 5.92)\ (5, 7.36)\ (9, 8.26)\ (13, 8.99)\ (17, 9.62)\ (21, 10.13)$

Applicando le formule relative alla regressione lineare per i dati (x_i, y_i), risulta: $a = K \approx 0.20$, $b = log(A) \approx 6.14$, $A = e^b \approx 464$.

L'approssimazione cercata é dunque:

$$P(x) = 464e^{0.2x}$$