Università di Catania

L.M. in Informatica

Ottimizzazione

Laura Scrimali
DMI - Studio 349, piano II blocco I
Tel. 095 7383059 - E-mail: laura.scrimali@unict.it

Programmazione non lineare

La programmazione lineare si fonda sull'ipotesi che tutte le funzioni (obiettivo e vincoli) siano lineari. Esistono tuttavia alcuni casi nei quali questa assunzione non è valida. Si pensi a

- problemi di produzione con funzione prezzo non lineare;
- problema del trasporto con costo di trasporto variabile a seconda della quantità trasportata;
- problema del portafoglio titoli con funzione rischio non lineare.

Esistono poi alcuni problemi non lineari che possono essere linearizzati e quindi trattati come problemi di PL.

Problemi di ottimizzazione

Un problema di Programmazione non lineare si presenta nella forma:

$$\min f(x)$$
$$x \in X$$

$f : \mathbb{R}^n \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$.

Classificazione

- Se $X = \mathbb{R}^n$ il problema si dice non vincolato.
- Se $X \subset \mathbb{R}^n$ il problema si dice vincolato.

Un problema di Programmazione non lineare vincolata si presenta nella forma:

$$\begin{aligned} &\min f(x) \\ &g(x) \leq 0 \\ &h(x) = 0 \\ &x \in X \subseteq \mathbb{R}^n, \end{aligned}$$

$f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$. Il vincolo $x \in X$ è detto vincolo d'insieme, le funzioni g e h sono dette funzioni vincolari. Si pone sempre $p \leq n$ per evitare che l'insieme X possa risultare vuoto oppure ci siano vincoli ridondanti.

Definizione

Un problema viene detto di ottimizzazione con vincoli lineari se le funzioni $g_i(x)$, $i = 1, \ldots, m$, e $h_j(x)$, $j = 1, \ldots, p$, sono lineari.

Definizione

Un problema viene detto di ottimizzazione quadratica se la funzione $f(x)$ è quadratica e le funzioni $g_i(x)$, $i = 1, \ldots, m$, e $h_j(x)$, $j = 1, \ldots, p$, sono lineari.

Definizione

Data una soluzione ammissibile x, si dice che un vincolo g_i è attivo in x se viene soddisfatto come uguaglianza. Diversamente si dice inattivo in x.

Definizione

Un punto $x^* \in X$ si dice di **ottimo globale** se risulta $f(x^*) \leq f(x), \forall x \in X$.

Definizione

Un punto $x^* \in X$ si dice di **ottimo locale** se esiste $\varepsilon > 0$ tale che $f(x^*) \leq f(x), \forall x : |x - x^*| < \varepsilon$.

Definizione

x^* è un ottimo locale stretto se, in qualunque punto vicino a x^* differente da x^*, il valore della funzione obiettivo f è strettamente maggiore di $f(x^*)$. x^* è cioè localmente unico.

Gli algoritmi di programmazione non lineare non sono in grado di distinguere tra un minimo locale ed uno globale. Pertanto, è importante indagare sulle condizioni che assicurano che un minimo locale sia anche globale.

Richiami: gradiente

Sia data una funzione $f : \mathbb{R}^n \to \mathbb{R}$.

Definizione

L'insieme delle n derivate parziali calcolate in un generico punto $x^* = (x_1^*, \ldots, x_n^*)$ definisce le n componenti di un vettore che è detto vettore **gradiente** di f in x^* e si indica con

$$\nabla f(x^*)^T = \left(\frac{\partial f}{x_1}, \ldots, \frac{\partial f}{x_n}\right).$$

Se le derivate parziali sono continue, allora la funzione è differenziabile.

Richiami: matrice hessiana

Sia data una funzione $f : \mathbb{R}^n \to \mathbb{R}$ che possiede n derivate parziali prime $\dfrac{\partial f}{x_i}, i = 1, \ldots, n$ nel punto $(x_1^*, x_2^*, \ldots, x_n^*)$. Ogni derivata parziale $\frac{\partial f}{x_i}$ è una funzione che può essere derivata rispetto ad una variabile x_j, fornendo una derivata parziale seconda $\dfrac{\partial^2 f}{x_i x_j} = f_{x_i x_j}$.

Definizione

Le derivate parziali seconde possono essere raccolte in una matrice quadrata $n \times n$ detta **matrice hessiana**:

$$H(x_1^*, x_2^*, \ldots, x_n^*) = \begin{bmatrix} f_{x_1 x_1}(x^*) & f_{x_1 x_2}(x^*) & \cdots & f_{x_1 x_n}(x^*) \\ f_{x_2 x_1}(x^*) & f_{x_2 x_2}(x^*) & \cdots & f_{x_2 x_n}(x^*) \\ \cdots & & & \\ f_{x_n x_1}(x^*) & f_{x_n x_2}(x^*) & \cdots & f_{x_n x_n}(x^*) \end{bmatrix}$$

Richiami: matrice semidefinita

Definizione

Una matrice A di ordine $n \times n$ si dice semidefinita positiva (definita positiva) se $d^T A d \geq 0, \forall d \in \mathbb{R}^n$ ($d^T A d > 0$, $\forall d \in \mathbb{R}^n \setminus \{0\}$).

Teorema

Siano $\lambda_1 \leq \cdots \leq \lambda_n$ gli autovalori di una matrice A in ordine non decrescente.

- A è definita positiva se e solo se $\lambda_1 > 0$.
- A è definita negativa se e solo se $\lambda_n < 0$.
- A è semidefinita positiva se e solo se $\lambda_1 = 0$
- A è semidefinita negativa se e solo se $\lambda_n = 0$

Teorema di Sylvester

Sia A una matrice simmetrica $n \times n$. Siano d_k $k = 1, \dots, n$ i minori principali nord-ovest.

- A definita positiva se e solo se $d_k > 0$ per ogni k.
- A semidefinita positiva se e solo se $d_k \geq 0$ per ogni k.
- A è definita negativa se e solo se $-A$ è definita positiva.
- A è semidefinita negativa se e solo se $-A$ è semidefinita positiva.

Richiami: funzioni convesse

Definizione

Sia $K \subseteq \mathbb{R}^n$ un insieme convesso. Una funzione $f : K \to \mathbb{R}$ si dice convessa se

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall \lambda \in [0,1], \forall x, y \in K.$$

Se la diseguaglianza vale in senso stretto, f si dice strettamente convessa.

Il grafico di una funzione convessa si trova sempre al di sotto della corda sottesa a due suoi punti arbitrari.

Definizione

Sia $K \subseteq \mathbb{R}^n$ un insieme convesso. Una funzione $f : K \to \mathbb{R}$ si dice concava (strettamente concava) se $-f$ è convessa (strettamente convessa).

Caratterizzazione della convessità

Condizione di convessità del primo ordine

f è convessa se e solo se

$$f(x) \geq f(y) + \nabla f(y)^T (x - y), \quad \forall x, y \in K.$$

Una funzione differenziabile è convessa se e solo se il suo grafico si trova al di sopra della retta tangente in ogni suo punto.

Condizione di convessità del secondo ordine

Sia f differenziabile due volte. f è convessa (strettamente convessa) se e solo se la matrice hessiana è semidefinita positiva (definita positiva).

Ottimi locali e globali

Un problema di ottimizzazione convessa può

- minimizzare una funzione convessa su un insieme di vincoli convesso;
- massimizzare una funzione concava su un insieme di vincoli convesso.

Teorema

Per un problema di ottimizzazione convessa, ogni punto di minimo locale è anche un punto di minimo globale. Inoltre, sia f strettamente convessa, allora, se esiste la soluzione globale, essa è unica.

Per i problemi concavi, il teorema non vale.

Problemi non vincolati

Sia dato un problema di Programmazione non lineare non vincolata

$$\begin{aligned} &\min \; f(x) \\ &x \in \mathbb{R}^n. \end{aligned}$$

Una soluzione locale x^* deve soddisfare una condizione necessaria di ottimalità. Si osserva che i punti che soddisfano una condizione necessaria di ottimalità non sono forzatamente soluzioni del problema: potrebbero ad esempio essere di massimo e non di minimo.

Direzione di discesa e gradiente

Definizione

Sia $f : \mathbb{R}^n \to \mathbb{R}$, un vettore $d \in \mathbb{R}^n$ si dice che è una direzione di discesa per f rispetto ad un punto x se esiste $\delta > 0$ tale che

$$f(x + \lambda d) < f(x), \quad \forall \lambda \in (0, \delta).$$

Se d è una direzione di discesa nel punto x, spostandoci da questo punto di una quantità λ (spostamento) sufficientemente piccola, si è sicuri di ottenere un decremento della funzione f.

Il gradiente di una funzione $\nabla f(x)$ rappresenta la direzione di massimo accrescimento della funzione f nel punto x. Pertanto, $-\nabla f(x)$ rappresenta la direzione di massima discesa.

Teorema

Un vettore $d \in \mathbb{R}^n$ è una direzione di discesa per f se

$$\nabla f(x)^T d < 0.$$

Risulta così naturale esprimere le condizioni di ottimalità coinvolgendo in qualche maniera il gradiente della funzione.

Condizioni necessarie del primo e secondo ordine

Condizione necessaria del primo ordine

Sia f differenziabile. Se x^* è un punto di minimo locale, allora $\nabla f(x^*) = 0$.

Condizione necessaria del secondo ordine

Sia f differenziabile due volte. Se x^* è un punto di minimo locale, allora la matrice Hessiana è semidefinita positiva.

Le condizione necessaria fornisce i punti stazionari candidati ad essere punti minimali. La condizione del secondo ordine chiarisce se i punti sono di massimo, di minimo o altro.

Problemi convessi

Se il problema è convesso la condizione del primo ordine si può invertire, come mostra il teorema seguente.

Teorema

Se il problema è convesso, x^* è un punto di minimo locale. se e solo se $\nabla f(x^*) = 0$.

Condizioni sufficienti del secondo ordine

(C.S. del II ordine)

Sia x^0 un punto interno alla regione ammissibile e $\nabla f(x^0) = 0$.

- Se $H(x^0)$ è definita positiva, allora x^0 è un punto di minimo locale.
- Se $H(x^0)$ è definita negativa, allora x^0 è un punto di massimo locale.
- Se $H(x^0)$ è indefinita, allora x^0 è un punto sella.
- Se $H(x^0)$ è semidefinita non si può dire niente.

Occorre esaminare le derivate seconde e le condizioni del secondo ordine per capire se i punti stazionari trovati sono di massimo o di minimo o altro ancora.

Problemi con vincoli di uguaglianza

Consideriamo il problema:

$$\begin{aligned} &\min \; f(x) \\ &h(x) = 0 \\ &x \in X \subseteq \mathbb{R}^n, \end{aligned}$$

$f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^p$.

Regolarità dei vincoli di uguaglianza

Definizione

Un punto x^* che soddisfa il vincolo $h(x^*) = 0$ si dice punto regolare del vincolo se i vettori gradienti $\nabla h_1(x^*), \nabla h_2(x^*), \dots,$ $\nabla h_p(x^*)$ sono linearmente indipendenti.

La non regolarità dei vincoli presuppone una qualche forma di ridondanza dei vincoli. In tal caso la trattazione diventerebbe più complessa, quindi si limita l'attenzione al caso di vincoli regolari.

Condizione di ottimalità del primo ordine

Teorema (condizione necessaria)

Sia x^* un punto di minimo locale di f, soggetta ai vincoli $h(x) = 0$, e sia un punto regolare di questi vincoli. Allora esiste $\lambda \in \mathbb{R}^p$ tale che

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0.$$

Forma equivalente

Definizione

Si definisce funzione Lagrangiana la funzione

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x),$$

dove si chiamano moltiplicatori di Lagrange le variabili λ_j, $j = 1 \ldots, p$, associate ad ogni vincolo $h_j(x) = 0$.

Le condizioni necessarie del primo ordine con i vincoli $h(x^*) = 0$ costituiscono un sistema di $n + p$ equazioni in $n + p$ incognite, che può essere scritto anche nella forma

$$\begin{aligned} \nabla_x L(x, \lambda) &= \nabla f(x^*) + \lambda^T \nabla h(x^*) = 0 \\ \nabla_\lambda L(x, \lambda) &= h(x^*) = 0. \end{aligned}$$

Problemi con vincoli di uguaglianza e diseguaglianza

Consideriamo il problema:

$$\begin{aligned}
&\min \; f(x) \\
&g(x) \leq 0 \\
&h(x) = 0 \\
&x \in X \subseteq \mathbb{R}^n,
\end{aligned}$$

$f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$.

Regolarità dei vincoli

Supponiamo che le funzioni f, g e h siano continue e derivabili.

Definizione

Sia x^* un punto che soddisfa i vincoli $h(x) = 0$ e $g(x) \leq 0$. Sia I l'insieme degli indici per i quali $g_i(x^*) = 0$ (detti vincoli attivi). Allora x^* è detto punto regolare dei vincoli se i vettori gradienti $\nabla h_j(x^*)$, $j = 1, \dots, p$ e $\nabla g_i(x^*)$, $i \in I$, sono linearmente indipendenti.

Condizioni di ottimalità del primo ordine KKT

Teorema (Condizioni di Karush-Kuhn-Tucker)

Sia x^* un punto di minimo locale e sia un punto regolare dei vincoli. Allora esistono $\lambda \in \mathbb{R}^p$ e $\mu \in \mathbb{R}^m_+$ tali che:

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$
$$\mu^T g(x^*) = 0.$$

Definizione

Data la funzione Lagrangiana

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x),$$

con moltiplicatori di Lagrange λ_j, $j = 1 \ldots, p$, associati ad ogni vincolo $h_j(x) = 0$ e $\mu_i \geq 0$, $i = 1 \ldots, m$, associati ad ogni vincolo $g_i(x) \leq 0$, allora

$$\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \lambda^T \nabla h(x) + \mu^T \nabla g(x)$$

Problemi convessi

Se il problema è convesso le condizioni di KKT si possono invertire, come mostra il teorema seguente.

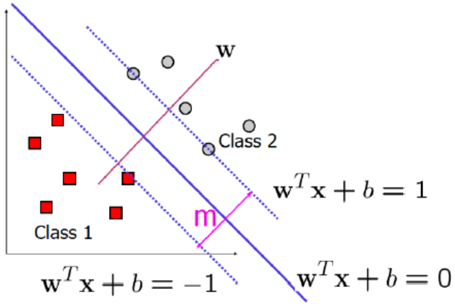
Teorema

Se il problema è convesso ed esistono $\lambda \in \mathbb{R}^p$ e $\mu \in \mathbb{R}^m_+$ tali che il vettore (x^*, λ, μ) soddisfa il sistema:

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$
$$\mu^T g(x^*) = 0,$$

allora x^* è un minimo globale.

Support Vector Machine



Si vuole massimizzare il margine m tra le classi. Risulta che $m = \frac{2}{\|w\|}$. Pertanto massimizzare m equivale a minimizzare $\|w\|$.

Siano $\{x_1, \ldots, x_n\}$ l'insieme dei dati e $y_i \in \{-1, 1\}$ l'indicatore della classe per ogni dato. Si deve risolvere il problema quadratico

$$\min \frac{1}{2}\|w\|^2$$
$$y_i(w^T x_i + b) \geq 1, \forall i.$$

Condizioni KKT per SVM

L'ottimo globale coincide con l'insieme delle soluzioni del sistema:

$$
\begin{aligned}
\nabla\left(\frac{1}{2}\|w\|^2\right) + \sum_{i=1}^{n} \lambda_i \nabla(-y_i(w^T x_i + b) + 1) = 0 \\
\lambda_i(-y_i(w^T x_i + b) + 1) = 0 \quad \forall i \\
-y_i(w^T x_i + b) + 1 \leq 0, \quad \forall i \\
\lambda_i \geq 0, \quad \forall i
\end{aligned}
$$

Soluzione: $w = \sum_i \lambda_i y_i x_i$ con $\sum_i \lambda_i y_i = 0$.