Esercizi MOBD

Indice

Esercizi KKT per problemi vincolati

Problema. Dato il problema nella forma:

$$\begin{cases} \min f(x) \\ g(x) \leqslant 0 \quad (\lambda_i) \\ h(x) = 0 \quad (\mu_i) \end{cases}$$

1. Il problema ammette soluzione?
2. Il problema è convesso?
3. Scrivere KKT e risolvere KKT.
4. Verificare se un punto x è regolare, soddisfa KKT e le conclusioni che si possono trarre.
5. Verificare se un punto è ottimo.

Answer.

(1) Per verificare che il problema ammette soluzioni:

- **Teorema di Weierstrass**(Condizione Sufficiente):

  Se $f(x)$ continua e S compatta$\longrightarrow\exists$minimo o massimo globale. In particolare, l'insieme S è compatto se chiuso e limitato.

- **Coercività**(Condizione sufficiente):

  Nel caso in cui si voglia utilizzare la coercività, occorre o trovare una successione per cui la $f(x)$ diventa illimitata o verificare gli insiemi di livello della funzione:

  $$\mathcal{L}(\alpha) = \{x \in S: f(x) \leqslant \alpha\} \quad \alpha = f(x_0), x_0 \in S \text{ iniziale}$$

  Quindi, se la funzione $f(x)$ è continua e coerciva$\rightarrow\exists$ minimo o massimo globale.

  Nel caso in cui ne Wierstrass ne la coercività vengono soddisfatti, allora non si può concludere nulla.

(2) Affinché il problema sia convesso occorre che la $f(x)$ sia convessa e l'insieme ammissibile S sia anch'esso convesso.Nel caso in cui queste condizioni vengano soddisfatte le condizioni successive diventano necessarie e sufficienti, altrimenti solo necessarie. A tale scopo:

- Verifica, se possibile, grafica;
- Tutto ciò che è lineare è convesso;
- La funzione non lineare $f(x)$ è convessa su S se la sua relativa matrice Hessiana:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial^2 x_1} \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} \cdots\cdots \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} \frac{\partial^2 f(x)}{\partial^2 x_2} \cdots\cdots \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \; \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial} \cdots\cdots\cdots\cdots \frac{\partial^2 f(x)}{\partial^2 x_n} \end{pmatrix}$$

  è convessa. Questa verifica viene effettuata tramite il criterio dei minori di Nord-Ovest o più in generale:

- Applico criterio di nord-ovest:

  A) tutti i minori hanno determinante >0 $\longrightarrow$STOP $\quad Q \succ 0$

  B) Si ha un minore con determinante =0 $\longrightarrow$**Criterio minori principali**:

    i. tutti i minori principali hanno determinante $\geqslant$0 $\longrightarrow$ STOP $\longrightarrow Q \succcurlyeq 0$

    ii. trovo minore con determinante <0:

      a) se i minori principali hanno determinante >0 e un determinante <0

      $$\longrightarrow \text{STOP} \Longrightarrow Q \text{ indefinita}$$

      b) Ritorno al passo 1 considerando $-Q$ per dedurne la (semi)definita negatività.

(3) KKT fornisce condizioni necessarie e sufficienti o solo necessarie in base, rispettivamente, alla convessità o meno del problema. Quindi ne deriva che, se il problema è non convesso, allora i punti che soddisfano KKT sono candidati all'ottimo globale e quindi sono minimi locali. Per scrivere KKT occorre:

- **Lagrangiana generalizzata:**

$$L(x, \lambda_0, \lambda, \mu) = \lambda_0 f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- **Condizioni Necessarie di KKT:**

$$\begin{aligned} &\nabla f(x^*) + \nabla g(x^*) \lambda^* + \nabla h(x^*) \mu^* = 0 \\ &g(x^*) \leqslant 0, h(x^*) = 0 \\ &\lambda^{*T} g(x^*) = 0 \\ &\lambda^* \geqslant 0 \end{aligned}$$

A questo punto trovo dei candidati x^* ad essere ottimi globali.

N.B.: Si consiglia la risoluzione di KKT e FJ partendo dalla complementarietà discutengo i casi in cui i moltiplicatori $\boldsymbol{\lambda_i}$ siano $\geqslant$0

(4)+(5) L'iter da seguire per verificare se un un punto x^* candidato è tale occorre:

1. Se il punto soddisfa KKT, allora quest'ultimo è un candidato. Altrimenti:

   a. Applico le condizioni di regolarità/qualificazione dei vincoli dei vincoli:

   - **LICQ:**x^*, candidato di minimo locale, soddisfa **LICQ** se per i vettori $\nabla g_i(x^*)$ con $i \in I(x^*) = \{i = 1,..., n: g_i(x^*) = 0\}$ (insieme dei vincoli attivi) e $\nabla h_j(x^*) \quad j = 1, \ldots, p$ sono **linearmente indipendenti** allora $\lambda_0^* \neq 0$.

     Quindi nei vincoli soddisfatti per il punto scrivo il gradiente $\nabla g_i(x^*)$ e questi, una volta sostituito il punto x^* devono avere rango pieno, quindi sono non singolari.

   - **Mangasarian-Fromovitz:**Sia $x^* \in S$ e supponiamo che g, h siano continuamente differenziabili in un intorno di x^*. Si dice che è soddisfatta in x^* la condizione di qualificazione dei vincoli di **Mangasarian-Fromovitz** se:

     i. I gradienti dei vincoli di eguaglianza $\{\nabla h_j(x^*), j = 1,\ldots, p\}$ sono linearmente indipendenti;

     ii. $\nabla g_i(x^*)^T d < 0 \quad \forall i \in I(x^*) \quad \nabla h_i(x^*)^T d = 0 \quad \forall i = 1, \ldots, p$;

   - **Condizioni di Slater:**Nelle **condizioni di Slater** consideriamo il caso in cui l'insieme ammissibile è definito attraverso vincoli convessi di diseguaglianza. Quindi:

     Supponiamo che le funzioni g_i siano convesso e continuamente differenziabili su un insieme aperto convesso contenente l'insieme ammissibile:

     $$S = \{x \in \mathbb{R}^n: g(x) \leqslant 0\}$$

     Si dice che è la condizione di Slater è soddisfatta se esiste $\hat{x} \in S$ tale che:

     $$g(\hat{x}) < 0$$

     Cioè che il punto $\hat{x}$ sia interno all'insieme.

   - **Linearità dei vincoli di eguaglianza e concavità:**Supponiamo che i vincoli di eguaglianza siano *lineari* e che i vincoli di diseguaglianza attivi siano concavi nel punto x^*. In tali ipotesi è possibile trovare un intorno $B(x^*, \rho)$ di x^* tale che, per ogni $x \in B(x^*, \rho)$ si abbia:

     $$\begin{aligned} h_i(x) &= h_i(x^*) + \nabla h_i(x^*)^T (x - x^*) \quad i = 1, \ldots p \\ g_i(x) &\leqslant g_i(x^*) + \nabla g_i(x^*)^T (x - x^*) \quad i \in I(x^*) \end{aligned}$$

     Un caso particolare è quando tutti i vincoli attivia in x^* siano lineari.

   b. Se il punto è regolare, allora si scarta. Altrimenti applico le condizioni di **Fritz-John:**

$$\begin{aligned} \nabla_x L(x^*, \lambda_0^*, \lambda^*, \mu^*) = 0 &= \lambda_0^* f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{j=1}^{p} \mu_j^* h_j(x^*) \\ \lambda_i^* g_i(x^*) &= 0 \quad i = 1, \ldots, m \\ \lambda_0^*, \lambda^* &\geqslant 0; (\lambda_0^*, \lambda^*, \mu^*) \neq 0 \\ g_i(x^*) &\leqslant 0, h_j(x^*) = 0 \end{aligned}$$

c. Se le condizioni di **Fritz-John** sono soddisfatte, allora il punto è candidato ad essere ottimo globale, altrimenti il punto viene scartato.

2. Le conclusioni che quindi si possono trarre è che il punto è solamente candidato all'ottimo globale se rispetta KKT e FJ e quindi sono ottimi locali. Nel caso convesso questi si traducono in candidati a ottimi globale che, sostituiti nella $f(x)$ hanno valore minimo: l'ottimo globale tra tutti questi punti è quello con valore di $f(x)$ minore. Altrimenti, non si può dire nulla.

Esercizi KKT per problemi con Vincoli di Box

Problema. Dato un problema nella forma:

$$\begin{cases} \min f(x) \\ l_i \leqslant x_i \leqslant u_i \end{cases} \quad i = 1, \ldots, n$$

$$l_i - x_i \leqslant 0 \quad i = 1, \ldots, n \quad (\lambda_i)$$

$$x_i - u_i \leqslant 0 \quad i = 1, \ldots, n \quad (\mu_i)$$

1. Il problema ammette soluzione?
2. Il problema è convesso?
3. Scrivere KKT e risolvere KKT.
4. Verificare se un punto x è regolare, soddisfa KKT e le conclusioni che si possono trarre.
5. Verificare se un punto è ottimo.

Answer.

(1) Per verificare che il problema ammette soluzioni:

- **Teorema di Weierstrass**(Condizione Sufficiente):

  Se $f(x)$ continua e S compatta$\longrightarrow\exists$minimo o massimo globale. In particolare, l'insieme S è compatto se chiuso e limitato.

- **Coercività**(Condizione sufficiente):

  Nel caso in cui si voglia utilizzare la coercività, occorre o trovare una successione per cui la $f(x)$ diventa illimitata o verificare gli insiemi di livello della funzione:

  $$\mathcal{L}(\alpha) = \{x \in S : f(x) \leqslant \alpha\} \quad \alpha = f(x_0), x_0 \in S \text{ iniziale}$$

  Quindi, se la funzione $f(x)$ è continua e coerciva$\rightarrow\exists$ minimo o massimo globale.

  Nel caso in cui ne Wierstrass ne la coercività vengono soddisfatti, allora non si può concludere nulla.

(2) Affinché il problema sia convesso occorre che la $f(x)$ sia convessa e l'insieme ammissibile S sia anch'esso convesso.Nel caso in cui queste condizioni vengano soddisfatte le condizioni successive diventano necessarie e sufficienti, altrimenti solo sufficienti. A tale scopo:

- Verifica, se possibile, grafica;
- Tutto ciò che è lineare è convesso;

- La funzione non lineare $f(x)$ è convessa su S se la sua relativa matrice Hessiana:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial^2 x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \cdots\cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial^2 x_2} & \cdots\cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \vdots & \vdots & \vdots & \vdots \ \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial} & \cdots\cdots\cdots & & \frac{\partial^2 f(x)}{\partial^2 x_n} \end{pmatrix}$$

  è convessa. Questa verifica viene effettuata tramite il criterio dei minori di Nord-Ovest o più in generale:

- Applico criterio di nord-ovest:

  A) tutti i minori hanno determinante >0 $\longrightarrow$STOP $\quad Q \succ 0$

  B) Si ha un minore con determinante $=0$ $\longrightarrow$**Criterio minori principali**:

    i. tutti i minori principali hanno determinante $\geqslant 0 \longrightarrow$ STOP $\longrightarrow Q \succcurlyeq 0$

    ii. trovo minore con determinante <0:

      a) se i minori principali hanno determinante >0 e un determinante <0

      $$\longrightarrow \text{STOP} \Longrightarrow Q \text{ indefinita}$$

      b) Ritorno al passo 1 considerando $-Q$ per dedurne la (semi)definita negatività.

(3) Essendo vincoli lineari, possiamo applicare le condizioni di KKT per ottenere condizioni necessarie di minimo locale. A tale scopo definiamo la Lagrangiana:

$$L(x, \hat{\lambda}, \lambda) = f(x) + \sum_{i=1}^{m} \hat{\lambda}_i \, (l_i - x_i) + \sum_{j=1}^{n} \lambda_i (x_i - u_i)$$

$$L(x, \hat{\lambda}, \lambda) = f(x) + \sum_{i=1}^{m} \hat{\lambda}_i \, (l_i - x_i) + \sum_{j=1}^{n} \lambda_i (x_i - u_i)$$

Scriviamo le condizioni di KKT:

$$\exists (\hat{\lambda}^\star, \lambda^*) \, t.c. \nabla L(x^*, \hat{\lambda}^\star, \lambda^*) = 0 = \nabla f(x^*) - \hat{\lambda}^\star + \lambda^*$$

$$\frac{\partial f(x^*)}{\partial x_i} - \hat{\lambda}_i^\star + \lambda_i^\star = 0 \quad i = 1, \ldots, n$$

$$\hat{\lambda}_i^* \, (l_i - x_i^*) = 0 \quad i = 1, \ldots n$$

$$\lambda_i^* (x_i^* - u_i) = 0 \quad i = 1, \ldots, n$$

$$\hat{\lambda}^\star, \lambda_i^* \geqslant 0 \quad i = 1, \ldots, n$$

$$l_i \leqslant x_i^* \leqslant u_i$$

(4)+(5) Condizione necessarie (e sufficiente se il problema è convesso) sono:

$$\frac{\partial f}{\partial x_i} = \begin{cases} \geqslant 0 & \text{se } x_i^* = l_i \\ = 0 & \text{se } l_i < x_i^* < u_i \\ \leqslant 0 & \text{se } x_i^* = u_i \end{cases}$$

Le conclusioni che quindi si possono trarre è che il punto è solamente candidato all'ottimo globale se rispetta KKT e FJ e quindi sono ottimi locali. Nel caso convesso questi si traducono in candidati a ottimi globale che, sostituiti nella $f(x)$ hanno valore minimo: l'ottimo globale tra tutti questi punti è quello con valore di $f(x)$ minore. Altrimenti, non si può dire nulla.

Esercizio Iperpiano

Problema. Dato il problema nella forma:

$$\text{TS} = \left\{ \left(\begin{pmatrix} x_{1,1} \\ x_{1,2} \end{pmatrix}, y_1 \right), \ldots, \left(\begin{pmatrix} x_{n,1} \\ x_{n,2} \end{pmatrix}, y_n \right) \right\}$$

1. Dato un iperpiano di separazione $\begin{pmatrix} \hat{w} \\ \hat{b} \end{pmatrix}$ trovare l'iperpiano di separazione ottimo $\begin{pmatrix} w^* \\ b^* \end{pmatrix}$ e verificarne l'ottimalità.

Answer.

(1)

1. Verifica se l'iperpiano è di separazione occorre che:

$$\begin{cases} \hat{w}^T x^i + \hat{b} \geqslant 1 \quad \forall x^i \in A \quad A = \{x^i \in \text{TS}: y^i = 1\} \\ \hat{w}^T x^i + \hat{b} \leqslant -1 \quad \forall x^i \in A \quad B = \{x^i \in \text{TS}: y^i = -1\} \end{cases}$$

2. Il problema è nella forma:

$$\begin{array}{lll} \max\limits_{w,b} \rho(w,b) & = & \min\limits_{x^i \in A \cup B} \left\{ \dfrac{|w^T x^i + b|}{\|w\|} \right\} \\ w^T x^i + b & \geqslant & 1 \quad \forall x^i \in A \\ w^T x^j + b & \leqslant & -1 \quad \forall x^j \in B \end{array}$$

Diventa:

$$\begin{array}{lll} \min & \dfrac{1}{2}\|w\|^2 & \\ w^T x^i + b & \geqslant & 1 \quad \forall x^i \in A \\ w^T x^j + b & \leqslant & -1 \quad \forall x^j \in B \end{array}$$

3. Prendo $x^i \in A$ e $x^j \in B$ e calcolo:

$$\hat{d}_i = \min_{x^i \in A} \left\{ \frac{|\hat{w}^T x^i + \hat{b}|}{\|\hat{w}\|} \right\}$$

$$\hat{d}_j = \min_{x^j \in B} \left\{ \frac{|\hat{w}^T x^j + \hat{b}|}{\|\hat{w}\|} \right\}$$

Occorre ricordarsi i punti $\hat{x}^i \in A$ e $\hat{x}^j \in B$ per cui si è ottenuto il minimo.

4. Definisco $\rho(\hat{w}, \hat{b}) = \min\{\hat{d}_i, \hat{d}_j\} = \frac{1}{\|\hat{w}\|}$. Verifico che:

$$\rho(\hat{w}, \hat{b}) = \min\{\hat{d}_i, \hat{d}_j\} \leqslant \frac{1}{2}\hat{d}_i + \frac{1}{2}\hat{d}_j = \frac{\hat{w}^T(\hat{x}^i - \hat{x}^j)}{2\|\hat{w}\|}$$

5. Ora modifico $(\hat{w}, \hat{b})$ con due scalari α, β per ottenere il nuovo iperpiano candidato all'ottimo $(\overline{w}, \bar{b})$. In cui:

$$\overline{w} = \alpha\hat{w} \quad \alpha := \text{fattore di rotazione}$$

$$\bar{b} = \beta \quad \beta := \text{fattore di traslazione}$$

Per avere i punti sul margine del nuovo iperpiano, i coefficienti α, β devono soddisfare:

$$\begin{cases} \alpha\,\hat{w}^T\hat{x}^i + \beta = 1 \\ \alpha\hat{w}^T\hat{x}^j + \beta = -1 \end{cases}$$

La soluzione α deve essere $0 < \alpha < 1$. Ed in generale i coefficienti si presentano nella forma:

$$\alpha = \frac{2}{\hat{w}^T(\hat{x}^i - \hat{x}^j)}; \beta = \frac{\hat{w}^T(\hat{x}^i + \hat{x}^j)}{\hat{w}^T(\hat{x}^i - \hat{x}^j)}$$

6. Il nuovo iperpriano per essere ammissibile deve soddisfare il passo 1.

7. Verifica che il nuovo margine sia maggiore di quello precedente;

8. Per verificare l'ottimalità di un iperpiano occorre studiare il problema primale o il suo duale:

   a. Scrivere il seguente problema con il vettore w incognito:

$$\begin{array}{rcl} \min_{w,b} & \frac{1}{2}\|w\|^2 & \\ y^i(w^Tx^i + b) & \geqslant & 1 \forall x^i \in T \end{array}$$

   b. Scrivere le condizioni di KKT (sono condizioni necessarie e sufficienti), introducendo gli opportuni moltiplicatori:

$$\begin{array}{lcl} \nabla_x L(x^*, \lambda_0^*, \lambda^*, \mu^*) = 0 & = & \lambda_0^* f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*) \\ \lambda_i^* g_i(x^*) & = & 0 \quad i = 1, \ldots, m \\ \lambda_0^*, \lambda^* & \geqslant & 0; (\lambda_0^*, \lambda^*, \mu^*) \neq 0 \\ g_i(x^*) & \leqslant & 0 \end{array}$$

   I vincoli corrispondenti ai punti che si trovano sul margine sono i vincoli di attivi. Di conseguenza, tutti gli altri avranno i corrispondenti moltiplicatori nulli.

c. Si risolve KKT imponendo $w = \overline{w}$ e $b = \bar{b}$. Se KKT è soddisfatto, allora è l'iperpiano ottimo altrimenti no.

d. Nel caso in cui l'iperpiano candidato non è dato, si risolve il punto b nelle incognite w e b.

Esercizio Iperpiano Ottimo Duale

Problema. Dato il problema in questa forma:

$$\begin{cases} \min \frac{1}{2} x^T Q\, x + c^T x \\ A\, x \geqslant b \\ A \in \mathbb{R}^{m\mathrm{x}n}, b \in \mathbb{R}^m \end{cases} \quad Q \succcurlyeq 0$$

1. Scrivere il duale di Wolfe;

2. Data soluzione del duale $\begin{pmatrix} \overline{x} \\ \overline{\lambda} \end{pmatrix}$ determinare e/o valutare soluzione del primale. Se fosse possibile, determiare valori dei moltiplicatori.

Answer.

(1) Per scrivere il duale di Wolfe, ricordiamo che il problema di minimo diventa, nel duale, di massimo e viceversa. Anche il problema duale può essere ricondotto ad un problema di minimo secondo la relazione:

$$\max f(x) = -\min - f(x)$$

Quindi si ottiene il seguente problema convesso:

$$\begin{cases} -\min_{x, \lambda} \frac{1}{2} x^T Q\, x + b^T \lambda = \theta(x, \lambda) \\ Q\, x + c + A^T \lambda = 0 \\ \lambda \geqslant 0 \end{cases}$$

(2) Sia $(\overline{x}, \overline{\lambda})$ soluzione del duale, allora $\exists x^*$, non necessariamente uguale a $\overline{x}$, tale che:

1. $Q(\overline{x} - x^*) = 0$;

2. x^* soluzione del problema primale (P);

3. $(\overline{x}, \overline{\lambda})$ è una coppia di minimo globale-moltiplicatore del primale (P);

Di conseguenza occorre:

1. Scrivere la Lagrangiana per il problema duale:

$$W(x, \lambda, v, z) = \frac{1}{2} x^T \mathrm{Qx} + b^T \lambda - v^T (Q\, x + c + A^T \lambda) - z^T \lambda$$

2. Scrivere le condizioni di KKT:

$$\begin{aligned}
\nabla_x W = Q\,x - Q\,v = 0 & \\
\nabla_\lambda W = b - A\,v - z = 0 & \quad (a) \\
z^T \lambda = 0 & \\
z \geqslant 0 & \\
Q\,\overline{x} + c + A^T \overline{\lambda} = 0 & \quad (\Delta) \\
\lambda \geqslant 0 &
\end{aligned}$$

3. Sono soddisfatte per ipotesi dal vettore $(\overline{x}, \overline{\lambda})$. Per trovare x^* del primale, scriviamo KKT per il primale:

$$\begin{aligned}
&\nabla_x L = 0 = -Q\,x + c + A^T \lambda \\
&\lambda^T (A\,x - b) = 0 \\
&\lambda \geqslant 0 \\
&A\,x \leqslant b
\end{aligned}$$

4. Da (a) $\nabla_\lambda W = b - A\,v - z = 0$ si ricava:$z = b - A\,v$ (con incognita v) che sostituito nella complementarietà del duale:

$$\overline{\lambda}^{\,T}(b - A\,v) = 0 \longrightarrow z \geqslant 0\,(\text{ammissibilità}) \longrightarrow A\,v \leqslant b(\text{ammissibilità primale di } v)$$

5. Inoltre sappiamo che:

$$Q(\overline{x} - v) = 0 \longrightarrow Q\,\overline{x} = Q\,v$$

6. Sostituisco $Q\,v$ in $Q\,\overline{x} + c + A^T \overline{\lambda} = 0 \;\; (\Delta)$ deve soddisfare KKT.

7. Riepilogando, quindi:

   a. $x^* = v$ coppia $(x^*, \overline{\lambda})$ soddisfa KKT del primale;

   b. x^* è la soluzione ottima del primale$\longrightarrow x^* = v$ è la soluzione globale

   c. $Q(x - x^*) = 0$

Esercizio SVM Lineari

Problema. Dato il problema nella forma:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i$$

$$\begin{array}{rcllc}
y^i\,(w^T x^i + b) & \geqslant & 1 - \xi_i & i = 1, \ldots, l & \lambda_i \\
\xi_i & \geqslant & 0 & i = 1, \ldots, l & \mu_i
\end{array}$$

1. Applicare SVM Lineare e trovare Iperpriano Ottimo.

Answer.

In questo problema si può applicare la dualità di Wolfe poiché la funzione obiettivo è convessa: Q semidefinita positiva. Il termine $C\sum_{i=1}^{l} \xi_i$ rappresenta la penalità sugli errori del TS.

Quindi iniziamo con lo scrivere la Lagrangiana di questo problema:

$$L(w,b,\xi,\lambda,\mu)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\lambda_i\,(y^i\,(w^Tx^i+b)-1+\xi_i)-\sum_{i=1}^{l}\mu_i\,\xi_i$$

Il corrispondente duale di Wolfe, risulterà:

$$\max\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\lambda_i\,(y^i\,(w^Tx^i+b)-1+\xi_i)-\sum_{i=1}^{l}\mu_i\xi_i$$

$$\begin{aligned}
\nabla_w L(w,b,\xi,\lambda,\mu) &= w-\sum_{i=1}^{l}\lambda_i y^i x^i=0\\
\frac{\partial L}{\partial b}(w,b,\xi,\lambda,\mu) &= -\sum_{i=1}^{l}\lambda_i y^i=0\\
\frac{\partial L}{\partial \xi_i}(w,b,\xi,\lambda,\mu) &= C-\lambda_i-\mu_i \qquad i=1,\ldots,l\\
\lambda_i &\geqslant 0 \qquad i=1,\ldots,l\\
\mu_i &\geqslant 0 \qquad i=1,\ldots,l
\end{aligned}$$

Dall'annullamento della Lagrangiana rispetto a b otteniamo che $\sum_{i=1}^{l}\lambda_i y^i\,b=0$; mentre dall'annullamento della Lagrangiana rispetto a w otteniamo:

$$w=\sum_{i=1}^{l}\lambda_i\,y^i\,x^i$$

Sostituendo lìultima espressione nella funzione obiettivo, il problema può essere riscritto come:

$$\max\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y^i y^j\,(x^i)^Tx^j+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y^i y^j\,(x^i)^Tx^j-\sum_{i=1}^{l}\lambda_i y^i b+\sum_{i=1}^{l}\lambda_i-\sum_{i=1}^{l}\lambda^i\xi^i-\sum_{i=1}^{l}(C-\lambda_i)\,\xi_i$$

$$=\max\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y^i y^j\,(x^i)^Tx^j+\sum_{i=1}^{l}\lambda_i$$

Riscrivendo la funzione obiettivo in forma otteniamo il duale di Wolfe:

$$\begin{cases}
-\min\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y^i y^j\,(x^i)^Tx^j-\sum_{i=1}^{l}\lambda_i\\
\sum_{i=1}^{l}\lambda^i y^i \;=\; 0\\
\lambda_i \;\geqslant\; 0\\
C-\lambda_i \;\geqslant\; 0 \quad i=1,\ldots,l
\end{cases}$$

Gli ultimi due vincoli vengono riscritti come $0\leqslant\lambda_i\leqslant C \quad i=1,\ldots,l$

A questo punto defiscono la forma matriciale di questo problema introducendo la matrice:

$$X=[y^1x^1\ldots y^lx^l] \quad X\in\mathbb{R}^{n\times l}$$

In cui l'elemento $i\,j-$esimo risulterà nella forma:

$$[X^T X]_{ij} = y^i\, y^j\, (x^i)^T x^j$$

Quindi otteniamo il **duale in forma compatta**:

$$\min \frac{1}{2}\, \lambda^T X^T X\, \lambda - e^T \lambda$$

$$y^T \lambda = 0$$

$$0 \leqslant \lambda_i \leqslant C \quad i = 1, \ldots, l$$

in cui $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$.

Una volta risolto il duale ed ottenedo i moltiplicatori λ^*:

$$w^* = \sum_{i=1}^{l} \lambda_i^*\, y^i\, x^i \qquad \lambda^* \geqslant 0$$

In cui gli unici punti del Training Set che contribuiscono ad individuare $w^{\cdot}$ sono gli x^i per cui $\lambda^{i^*} > 0$. I punti che soddisfano tale condizioni vengono detti **vettori di supporto.** Nel caso in cui $\lambda^{i^*} > 0$, allora il vincolo $i-$esimo del primale è attivo all'uguaglianza e quindi il punto si trova sul margine.

Per ricavare b^* si ricava per complementarietà. In particolare, (λ^*, μ^*) dove $\mu_i^* = C - \lambda_i^*$ sono i moltiplicatori associati alla soluzione primale. Allora questi moltiplicatori devono soddisfare la complementarietà:

$$\lambda_i^* (y^i\, (w^{*^T} x^i + b^*) - 1 + \xi_i^*) = 0 \quad i = 1, \ldots, l$$

$$\mu_i^*\, \xi_i^* = (C - \lambda_i^*)\, \xi_i^* = 0 \quad i = 1, \ldots, l$$

$$\xi_i^* (C - \lambda_i^*) = 0 \longrightarrow \text{Se}\, \lambda_i^* < C \longrightarrow \xi_i^* = 0 \longrightarrow x^i\, \text{è}\, \textbf{ben calssificato}$$

Per trovare b^* scelgo i: $0 < \lambda_i^* < C$. Allora $\xi_i^* = 0 \longrightarrow y^i (w^{*^T} x^i + b^*) = 1$.

I punti x^i per cui $\lambda_i^* = C$ non posso calcolare ξ_i^*. Allora sono candidati ad essere malclassificati. Infine, la nostra macchina SVN sarà del tipo:

$$y(x) = \text{sgn}(w^{*T} x + b^*)$$

Esercizio SVM Non Lineari

Problema. Il problema è nella forma:

$$\min \frac{1}{2} \sum_i \sum_j y^i y^j \lambda_i \lambda_j K(x^i, x^j) - \sum_{i=1}^{l} \lambda_i$$

$$\begin{aligned} \sum_i \lambda_i\, y^i &= 0 \\ 0 \leqslant \lambda_i &\leqslant C \end{aligned}$$

1. Utilizza un certo tipo di Kernel tra:

   a. Kernel Lineare: $K(x, y) = x^T y$;

b. Kernel Polinomiale: $k(x,y) = (x^T y + \gamma)^p \quad \gamma \geqslant 0, p \geqslant 1$

c. Kernel Gaussiano: $k(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \quad \sigma > 0$

d. Kernel a Tangente Iperbolica: $k(x,y) = \tanh(\beta x^T y + \gamma)$

Answer.

1. Funzione Kernel:

$$K = \begin{pmatrix} k(x^i, x^j) & \dots & k(x^1, x^l) \\ \vdots & \ddots & \vdots \\ k(x^l, x^1) & \dots & k(x^l, x^l) \end{pmatrix} \succcurlyeq 0$$

2. Scrittura del duale:

$$\min \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j K(x^i, x^j) - \sum_{i=1}^{l} \lambda_i$$

$$\begin{aligned} \sum_i \lambda_i y^i &= 0 \\ 0 \leqslant \lambda_i &\leqslant C \end{aligned}$$

3. Il problema è convesso, si risolve il duale trovando λ^* e si trova b^* per complementarietà:

$$b^* = \frac{1}{y^i} - \sum_{j=1}^{l} y^j \lambda_j^* k(x^i, x^j)$$

4. La SVM diventa:

$$y(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \lambda_i y^i k(x^i, x) + b^* \right)$$

Esercizi Decomposizione

Esercizio Decomposizione in generale ??

Esercizio SVM Light

Problema. Dato il problema nella forma:

$$\mathrm{TS} = \left\{ \left(\begin{pmatrix} x_{1,1} \\ x_{1,2} \end{pmatrix}, y_1 \right), \dots, \left(\begin{pmatrix} x_{n,1} \\ x_{n,2} \end{pmatrix}, y_n \right) \right\}$$

$$C = k$$

Applicare SVM Light.

Answer.

1. Determinare $Q = k(x,y)^T k(x,y)$ in base al kernel scelto;

2. $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

3. Inizializzare $\alpha^\circ = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad \nabla f(\alpha^\circ) = -\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

4. Determinare gli insiemi:

$$R(\alpha) = L^+(\alpha^*) \cup U^-(\alpha^*) \cup \{i: 0 < \alpha^\circ < C\}$$

$$S(\alpha) = L^-(\alpha^*) \cup U^+(\alpha^*) \cup \{i: 0 < \alpha^\circ < C\}$$

5. Calcolare $-\frac{\nabla f(x)}{y}$

6. Verificare:

$$\max_{i \in R(\alpha)} \left\{ -\frac{\nabla_i f(x)}{y_i} \right\} \leqslant \min_{j \in S(\alpha)} \left\{ -\frac{\nabla_j f(x)}{y_j} \right\}$$

7. Se la condizione al punto 6 è violata, definire WS prendendo due indici i, j dai due insiemi e scrivere il problema ridotto:

$$\min_{\alpha_i, \alpha_j} \frac{1}{2} \left(\, 0 \ldots \alpha_i \ldots \alpha_j \ldots 0 \, \right) \begin{pmatrix} Q_{i,i} & Q_{i,j} \\ Q_{j,i} & Q_{j,j} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \\ 0 \end{pmatrix} - \alpha_i - \alpha_j$$

$$\alpha_i - \alpha_j = 0$$

$$0 \leqslant \alpha_i \leqslant 1 \quad 0 \leqslant \alpha_j \leqslant 1$$

8. Pongo $\alpha_i = \alpha_j$, sostituiso in funzione obiettivo e derivo rispetto alla variabile rimanente:

$$\frac{\partial f}{\partial \alpha_j} = 0$$

9. Pongo:

$$\alpha^1 = \begin{pmatrix} 0 \\ \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \\ 0 \end{pmatrix}$$

10. Il nuovo gradiente sarà:

$$\nabla f(x^1) = \nabla f(\alpha^\circ) + Q_{:,i}(\alpha_i^1 - \alpha_i{}^\circ) + Q_{:,j}\,(\alpha_j^1 - \alpha_j{}^\circ)$$

11. Ricomincio.

K-Means

Problema. Dato il problema nella forma:

$$\text{TS} = \left\{ \begin{pmatrix} x_{1,1} \\ x_{1,2} \end{pmatrix}, \ldots, \begin{pmatrix} x_{n,1} \\ x_{n,2} \end{pmatrix} \right\}$$

$$M = k$$

$$z = \left\{ \begin{pmatrix} z_{1,1} \\ z_{1,2} \end{pmatrix}, \ldots, \begin{pmatrix} z_{k,1} \\ z_{k,2} \end{pmatrix} \right\}$$

1. Effettuare un passo dell'algoritmo K-Means;
2. Calcolare la Silhoutte;

Answer.

(1) L'algoritmo di K-Means è un algoritmo di Clusterizzazione in cui:

- M:= numero di Cluster
- z:=posizione dei centroidi;
- TS:= insieme dei punti da assegnare ai cluster

Per risolvere questo tipo di esercizio occorre calcolare per ogni punto:

1. Assegnare il punto al cluster. A tale scopo si calola la distanza $d(x^i, z^j)$;
2. Si prend:

$$\arg\min_j \{d(x^i, z^j)\} \longrightarrow \delta_i = (\ e_j\ )(z_j)$$

Per esempio, la distanza minima per i esimo punto si ha per il j-esimo cluster. Allora:

$$\delta_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} (z_j)$$

3. Calcolare la nuova posizione dei centroidi dei cluster. Ricordandosi delle assegnazioni fatte:

$$z_j = \frac{\sum_i \delta_{i,j}\, x^i}{\sum_i \delta_{i,j}}$$

(2) Per calcolare la Silhouette, occorre:

1. Fissare il numero K;

2. Per ogni punto $i \in C(i)$ con $C(i)$ cluster i-esimo, calcolare:

$$a(i) = \frac{1}{|C(i)| - 1} \sum_{j \in C(i)} d(x^i, x^j)$$

$$b(i) = \min_{k \notin C(i)} \frac{1}{|C_k|} \sum_{j \in K} d(x^i, x^j)$$

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad \text{se } |C(i)| > 1$$

$$S(i) = 0 \quad \text{se } |C(i)| = 1$$

3. Al variare di K si calcola:

$$S_k = \frac{1}{N} \sum_{i=1}^{N} S(i)$$

4. Si prende il massimo di questi valori.