---------------------------------- **OPTIMIZATION AND GAME THEORY** ----------------------------------

Set Convesso
Il set $C \subseteq \mathbb{R}^n$ è **convesso** se $\forall x, y \in C$ and $\forall \alpha \in [0,1]$ ➔ $\alpha x + (1-\alpha)y \in C$

Set Affine
Il set $C \subseteq \mathbb{R}^n$ è **affine** se $\forall x, y \in C$ and $\forall \alpha \in \mathbb{R}$ ➔ $\alpha x + (1-\alpha)y \in C$

Combinazione Convessa

$$y = \sum_{i=1}^{k} \alpha_i x_i \quad con: \ 1)\ \alpha_1 \dots \alpha_k \in [0,1], \qquad 2) \sum_{i=1}^{k} \alpha_i = 1$$

Lemma 1
SE C è convesso ➔ $\forall x_1 \dots x_k \in C$ and $\alpha_1 \dots \alpha_k \in [0,1] : \sum_{i=1}^{k} \alpha_i = 1$,

$$\sum_{i=1}^{k} \alpha_i x_i \ \in C$$

Proposizione
SE $\{C_i\}_{i \in I}$ rappresenta ogni possibile famiglia di set convessi ➔ $\bigcap_{i \in I} C_i$ è convesso
Informalmente: l'**intersezione** di set convessi è anch'esso convesso

Convex Hull
$conv(C)$ di un set C è il più piccolo set convesso che contiene C

Proposizione
$conv(C) = \{tutte\ le\ combinazioni\ convesse\ dei\ punti\ in\ C\}$

Nota
$conv(C) = C$ per C convesso

Definizione di Poliedro
Un poliedro P è l'intersezione di un numero finito di semipiani in $\mathbb{R}^n$

Somma Di Set Convessi
Siano C_1 e C_2 set convessi ➔ $C_1 + C_2 \coloneqq \{x + y\!: \ x \in C_1, y \in C_2\}$ è convessa (Vale anche per la differenza)

Prodotto Di Set Convessi
Sia C un set convesso ➔ $\alpha \cdot C \coloneqq \{\alpha \cdot x\!: \ x \in \mathrm{C}\}$ è convesso

Closure
C convesso ➔ $cl(\mathrm{C})$ è convesso

Interior
C convesso ➔ $int(\mathrm{C}) \neq \emptyset$ è convesso

Affine Set
Sia $C \subseteq \mathbb{R}^n$ ➔ $aff(C)$ è il più piccolo set affine che contiene C

Relative Interior
Sia $C \subseteq \mathbb{R}^n$ convesso ➔ $ri(C) := \{x \in C: \exists\ \epsilon > 0:\ aff(C) \cap B(x,\epsilon) \subseteq C\}$

Teorema
Sia $C \subseteq \mathbb{R}^n$ set convesso non vuoto ➔ $ri(C)$ è convesso non vuoto

Teorema
Siano $A, B \subseteq \mathbb{R}^n$ due set convessi non vuoti ➔ $A\ e\ B$ sono linearmente separabili **SSE**

$$ri(A)\ \cap\ ri(B) = \emptyset$$

Funzioni Affine
Sia $f: \mathbb{R}^n \to\ \mathbb{R}^m$ affine, ovvero $f(x) = Ax + b$ con $A \in \mathbb{R}^{m\times n},\ b\ \in \mathbb{R}^m$.

SE $C \subseteq \mathbb{R}^n$ è convesso ➔ $f(C) = \{\ f(x):\ x \in C\ \}$ è convesso
SE $C \subseteq \mathbb{R}^m$ è convesso ➔ $f^{-1}(C) = \{x \in \mathbb{R}^n :\ \ f(x) \in C\ \}$ è convesso

Definizione Di Cono
Un set $C \subseteq \mathbb{R}^n$ è un cono **SE** $\forall x \in C,\ \forall \lambda > 0$ ➔ $\lambda x \in C$

Cono Di Recessione
Dato un poliedro $P = \{x: Ax \leq b\}$ il cono di recessione di P è

$$rec(P) := \{\ d:\ x + \alpha \cdot d \in P, \quad \forall x \in P, \quad \alpha > 0\ \}.$$

Cono Tangente •••
Dato $x' \in cl(C) \subseteq \mathbb{R}^n$ il cono tangente a x' è:

$$T_C(x') := \{\ d \in \mathbb{R}^n:\ \ \exists \{z_k\} \subset C,\ \ \exists \{t_k\} > 0,\ \ z_k \to x',\ \ t_k \to 0,\ \ \lim_{k\to\infty} \frac{z_k - x'}{t_k} = d\}$$

Funzione Convessa
Sia $C \subseteq \mathbb{R}^n$ set convesso. Una funzione $f: \mathbb{R}^n \to \mathbb{R}$ è convessa in C se

$$f(\alpha y + (1-\alpha)x) \leq\ \alpha f(y) + (1-\alpha)f(x) \qquad \forall x, y\ \in C, \quad \forall \alpha \in [0,1]$$

Teorema
$f: \mathbb{R}^n \to \mathbb{R}$ è convessa in $\mathbb{R}^n$ **SSE** $epi\ f_C := \{\ (x,y) \in C \times \mathbb{R}:\ y \geq f(x)\ \}$ è convesso

Funzione Concava
Sia $C \subseteq \mathbb{R}^n$ set convesso. Una funzione $f: \mathbb{R}^n \to \mathbb{R}$ è concava in C se $-f$ è convessa in C

$$f(\alpha y + (1-\alpha)x) \geq\ \alpha f(y) + (1-\alpha)f(x) \qquad \forall x, y\ \in C, \quad \forall \alpha \in [0,1]$$

Teorema
$f: \mathbb{R}^n \to \mathbb{R}$ è convessa in $C \subseteq \mathbb{R}^n$ set convesso ➔ f continua nel $ri(C)$

Strictly Convex
Dato $C \subseteq \mathbb{R}^n$ set convesso, $f: \mathbb{R}^n \to \mathbb{R}$ è **strictly** convex in C se

$$f(\alpha y + (1-\alpha)x) < \alpha f(y) + (1-\alpha)f(x), \qquad \forall x, y \in C, \qquad x \neq y, \qquad \forall \alpha \in (0,1)$$

Strongly Convex

Dato $C \subseteq \mathbb{R}^n$ set convesso, $f: \mathbb{R}^n \to \mathbb{R}$ è **strongly** convex in C se $\exists\, \tau > 0$:

$$f(\alpha y + (1-\alpha)x) < \alpha f(y) + (1-\alpha) f(x) - \frac{\tau}{2}\alpha(1-\alpha) \parallel y - x \parallel_2^2 \quad \forall x, y \in C, \quad \forall \alpha \in [0,1]$$

Teorema

f è strongly convex **SSE** $\exists\, \tau > 0$: $f(x) - \frac{\tau}{2} \parallel x \parallel_2^2$ è convessa

Da cui segue che:

f è strongly convex **SSE** esiste una funzione convessa ψ e $\tau > 0$: $f(x) = \psi(x) + \frac{\tau}{2} \parallel x \parallel_2^2$

Teorema - Convex

f è convessa in C **SSE** $f(y) \geq f(x) + (y-x)^T \nabla f(x) \quad \forall x, y \in C$

Teorema - Strictly Convex

f è strictly convex in C **SSE** $f(y) > f(x) + (y-x)^T \nabla f(x) \quad \forall x, y \in C, \;\; x \neq y$

Teorema - Strongly Convex

f è strongly convex in C **SSE** $\exists \tau > 0$: $f(y) \geq f(x) + (y-x)^T \nabla f(x) + \frac{\tau}{2} \parallel y - x \parallel_2^2 \quad \forall x, y \in C$

Teorema – Come capire i tipi di convessità •••

- f è convessa in C **SSE** $\forall x \in C$ la Hessiana $\nabla^2 f(x)$ è **semi-definita positiva** ($v^T\, \nabla^2 f(x) \geq 0$)
  Ovvero che gli autovalori λ di $\nabla^2 f(x)$ sono: $\lambda \geq 0 \;\; \forall x \in C$

- Se $\nabla^2 f(x) > 0 \; \forall x \in C$ ➔ f è strictly convex

- f è strongly convex in C **SSE** $\exists \tau > 0$: $\nabla^2 f(x) - \tau I$ è **semi-definita positiva** $\forall x \in C$. Ovvero:

$$v^T\, \nabla^2 f(x)\, v \geq \tau \parallel x \parallel_2^2 \quad \forall x \in C, \;\; \forall v \in \mathbb{R}^n$$

  Ovvero che gli autovalori λ di $\nabla^2 f(x)$ sono: $\lambda \geq \tau \;\; \forall x \in C$

Teoremi •••

- **SE** f è convessa e $\alpha > 0$ ➔ $\alpha \cdot f$ è convessa
- **SE** f_1, f_2 sono convesse ➔ $f_1 + f_2$ è convessa
- **SE** f è convessa ➔ $f(Ax + b)$ è convessa
- **SE** $f_1 \cdots f_n$ sono convesse ➔ $\max\{f_1(x), \cdots, f_n(x)\}$ è convessa
- **SE** $\{f_i\}_{i \in I}$ è una famiglia di funzioni convesse ➔ $\sup_{i \in I} \{f_i(x)\}$ è convessa

Teoremi

Siano $f: \mathbb{R}^n \to \mathbb{R}$ e $g: \mathbb{R} \to \mathbb{R}$

- Se f è convessa e g è convessa non-decrescente ➔ $f \circ g$ è convessa
- Se f è concava e g è convessa non-crescente ➔ $f \circ g$ è convessa
- Se f è concava e g è concava non-decrescente ➔ $f \circ g$ è concava
- Se f è convessa e g è concava non-crescente ➔ $f \circ g$ è concava

K-sublevel Set

Data una funzione $f: \mathbb{R}^n \to \mathbb{R}$ e dato $k \in \mathbb{R}$ definiamo il $k-sublevel\ set$:

$$S_k(f) = \{x \in \mathbb{R}^n : f(x) \leq k\}$$

Definizione – Quasi convex function
Dato un set $C \subseteq \mathbb{R}^n$, la funzione $f: \mathbb{R}^n \to \mathbb{R}$ si definisce **quasi convessa** in C **SE** il set

$$S_k(f) \cap C = \{x \in C: f(x) \leq k\}$$

È convesso $\forall\ k \in C$.

PROBELI DI OTTIMIZZAZIONE - INTRO

Problema Di Ottimizzazione

$$f_* = \min\{\, f(x) : x \in X \,\} \qquad \rightarrow P$$

Con

$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$
$$X \coloneqq \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ \ i = 1 \ldots m,\ \ h_j(x) = 0,\ \ j = 1 \ldots p\} \subseteq \mathbb{R}^n$$

Valore Ottimo

Il valore ottimo di P è definito come $v(p) = \inf\{\, f(x):\ x \in X \,\}$. $\quad v(P) \in \mathbb{R}$

Ottimo Globale

$$x^* \in X : f(x^*) \leq f(x) \quad \forall x \in X$$

Ottimo Locale

$$x^* \in X : f(x^*) \leq f(x) \quad \forall x \in B(x^*, r) \quad \forall\, r > 0$$

Teorema di Weierstrass •••

Per f continua in X regione di ammissione chiusa e limitata ➔ Esiste (almeno) un ottimo globale

Corollario 2

Se la funzione obiettivo f è continua e la regione di ammissione X è chiusa e $\exists\ k \in \mathbb{R}$ tale che il $k - sublevel\ set$: $S_k(f) = \{\, x \in X : f(x) \leq k \,\}$ è non-vuoto e limitato ➔ Esiste (almeno) un ottimo globale

Funzione Coerciva

$$\lim_{\substack{\|x\| \to \infty \\ x \in X}} f(x) = +\infty$$

Corollario 3

Se la funzione f è **continua** e **coerciva**, e $X \neq \emptyset$ è chiuso ➔ Esiste (almeno) un ottimo globale

Teorema 1 •••

Sia f **convessa** in X **convesso** ➔ Ogni ottimo locale di P è anche globale

Proposizione 1

Sia f **strictly convex** in X **convesso**, e P ammette un ottimo globale x^* ➔ x^* è l'unica soluzione ottima di P

Teorema 2

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ è **strongly convex** in $\mathbb{R}^n$ e X è **chiuso** ➔ Esiste un ottimo globale

Corollario 1

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ è **strongly convex** in $\mathbb{R}^n$ e X è **chiuso** e **CONVESSO** ➔ Esiste un ottimo globale UNICO

Teorema 3

Sia X **APERTO** e sia f **differenziabile** in $x^* \in X$. Se x^* è un ottimo locale per P ➔ $\nabla f(x^*) = 0$

Teorema 4 - Second order necessary optimality condition
Sia X aperto e $x^* \in X$ ottimo locale per P ➔ $1)\ \nabla f(x^*) = 0. \quad 2)\ \nabla^2 f(x^*) \geq 0$

Teorema 5 – Second order sufficient optimality condition
Sia X aperto, preso $x^* \in X$ assumiamo le seguenti condizioni vere

1) $\nabla f(x^*) = 0$
2) $2)\ \nabla^2 f(x^*) \geq 0$

➔ x^* è un ottimo locale per P

Teorema 6 – Optimality condition for convex problems
Sia f differenziabile e convessa nel set X convesso e aperto ➔ $x^* \in X$ è un ottimo globale per P **SSE** $\nabla f(x^*) = 0$

Teorema 7
Sia f differenziabile e **STRICTLY CONVEX** nel set aperto e convesso X ➔ $x^* \in X$ è un ottimo globale **UNICO** per P **SSE** $\nabla f(x^*) = 0$

METODI DI OTTIMIZZAZIONE PER PROBLEMI NON VINCOLATI

Metodo del gradiente

1) Scegliere $x_0 \in \mathbb{R}^n$, $k = 0$
2) $SE\ \nabla f(x_k) = 0\ ALLORA\ stop$
3) $ALTRIMENTI$
    a. $d_k = -\nabla f(x_k)$
    b. $t_k \leftarrow \min_{t>0} f(x_k + t \cdot d_k)$
    c. $x_{k+1} = x_k + t_k \cdot d_k$
    d. $k = k + 1$
4) Step 2

Proposizioni

Sia f continua e differenziabile

1) $(d_k)^T d_{k+1} = 0,\ \forall k$
2) **SE** $\{x_k\} \to x^*\ ALLORA\ \nabla f(x^*) = 0$. Ovvero x^* è un punto stazionario per f

Teorema

Se f è coerciva ➔ $\forall x_0$ punto di partenza della sequenza $\{x_k\} \to x^*$, essa è limitata, e ogni punto di accumulo di $\{x_k\}$ è un punto stazionario per $f(x)$

Corollario

Se f è coerciva e convessa ➔ $\forall x_0$ punto di partenza della sequenza $\{x_k\} \to x^*$, essa è limitata, e ogni punto di accumulo di $\{x_k\}$ è un minimo globale per $f(x)$

Corollario

Se f è stongly convex ➔ $\forall x_0$ punto di partenza della sequenza $\{x_k\}$, essa converge all'unico minimo globale per $f(x)$

Metodo del gradiente – Caso Quadratico

Se $f(x) = \frac{1}{2} x^T Q x + c^T x$, con $Q > 0$.

Dove: $g_k = \nabla f(x_k) = Q x_k + c$

Per cui $t_k = -\frac{(d_k)^T \cdot g_k}{(d_k)^T \cdot Q \cdot d_k}$

Metodo del gradiente con la linea di ricerca inesatta di Armijo

1) Scegliere $\alpha, \gamma \in (0,1),\ t' > 0,\ x_0 \in \mathbb{R}^n,\ k = 0$
2) $SE\ \nabla f(x_k) = 0\ ALLORA\ stop$
3) $ALTRIMENTI$
    a. $d_k = -\nabla f(x_k)$
    b. $t_k = t'$
    c. $while\ f(x_k + t_k \cdot d_k) > f(x_k) + \alpha\, t_k\, (d_k)^T\, \nabla f(x_k)$
        i. $t_k = \gamma\, t_k$
    d. $x_{k+1} = x_k + t_k \cdot d_k$
    e. $k = k + 1$
4) Step 2

Metodo del gradiente coniugato

1) Settare $x_0 \in \mathbb{R}^n,\ k = 0$
2) $g_0 = \nabla f(x_0) = Qx_0 + c$
3) $SE\ g_k = \nabla f(x_k) = 0\ \ ALLORA\ \ stop$
4) $ALTRIMENTI$
    a. $SE\ k = 0\ ALLORA\ d_k =\ -g_k$
    b. $ALTRIMENTI$
        i. $\beta_k = \frac{(g_k)^T\ Q\ \ d_{k-1}}{(d_{k-1})^T\ Q\ \ d_{k-1}}$
        ii. $d_k = -g_k + \beta_k d_{k-1}$
    c. $t_k =\ -\frac{(g_k)^T\ d_k}{(d_k)^T\ Q\ \ d_k}$
    d. $x_{k+1} = x_k + t_k \cdot d_k$
    e. $g_{k+1} = Qx_{k+1} + c$
    f. $k = k + 1$
5) Step 3

Proposizioni

• Un modo alternativo di calcolare $t_k = \frac{\|g_k\|^2}{(d_k)^T\ Q\ \ d_k}$
• Un modo alternativo di calcolare $\beta_k = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$
• Se non troviamo il minimo globale dopo k iterazioni ➔ $\{g_1,\ g_2, \dots, g_k\}$ sono ortogonali
• Se non troviamo il minimo globale dopo k iterazioni ➔ $\{d_1,\ d_2, \dots, d_k\}$ sono coniugate rispetto a Q e x_k è il minimo di $f(x)$ su $x_0 + Span(d_1,\ d_2, \dots, d_k)$

Teorema

Il metodo del Gradiente Coniugato (CG) trova il minimo globale in, al massimo, n iterazioni

Teorema

Se Q ha $r\ autovalori\ distinti$ ➔ CG trova il minimo globale in, al massimo, r iterazioni

Metodo di Newton – Basic Version

1) Settare $x_0 \in \mathbb{R}^n$, $k = 0$
2) $SE\ \nabla f(x_k) = 0\ \ ALLORA\ \ stop$
3) $ALTRIMENTI$
    a. $d_k \leftarrow \nabla^2 f(x_k)\ d = -\nabla f(x_k)$
    b. $x_{k+1} = x_k + d_k$
    c. $k = k + 1$
4) Step 2

Teorema Di Convergenza

Se x^* è un minimo locale di $f(x)$ e $\nabla^2 f(x^*) > 0$ ➔ $\exists \delta > 0$: $\forall x_0 \in B(x^*,\ \delta)$ e la sequenza $\{x_k\} \to x^*$ e

$$\|\ x_{k+1} - x^*\ \| \leq C\ \ \|\ x_k - x^*\ \|^2 \qquad \forall k > k',\quad per\ C > 0\ e\ k' > 0$$

Metodo di Newton – con linea di ricerca (inesatta)

1) Settare $\alpha, \gamma \in (0,1),\ t' > 0,\ x_0 \in \mathbb{R}^n,\ k = 0$
2) $SE\ \nabla f(x_k) = 0\ \ ALLORA\ stop$
3) $ALTRIMENTI$
    d. $d_k \leftarrow \nabla^2 f(x_k)\, d = -\nabla f(x_k)$
    e. $t_k = t'$
    f. $while\ f(x_k + t_k \cdot d_k) > f(x_k) +\ \alpha\, t_k\, (d_k)^T\, \nabla f(x_k)$
        i. $t_k =\ \gamma\, t_k$
    g. $x_{k+1} = x_k + t_k d_k$
    h. $k = k + 1$
4) Step 2

Teorema Di Convergenza

Se f è strongly convex ➔ $\forall x_0 \in \mathbb{R}^n$ la sequenza $\{x_k\} \to x^*$ minimo globale di $f(x)$

Se $\alpha \in (0,\ 1/2)$ e $t' = 1$ ➔ la convergenza è quadratica

PROBLEMI VINCOLATI

Set Di Disequazioni Attive

Definiamo $\mathcal{A}(x^*) = \{j: g_j(x^*) = 0\}$ il set di disequazioni attive $x^* \in X$

First Order Feasible Direction Cone

$$D(x^*) = \left\{ d \in \mathbb{R}^n : \begin{matrix} d^T \nabla g_i(x^*) \leq 0 & \forall i \in \mathcal{A}(x^*) \\ d^T \nabla h_k(x^*) = 0 & \forall k = 1 \dots p \end{matrix} \right\}$$

Definizione – Abadie Constraint Qualification (ACQ)

ACQ è valida per un punto ammissibile $x^* \in X$ **SE:** $T_X(x^*) = D(x^*)$

Teoremi •••

1) **Vicoli Affini**
   SE g_j e h_k sono funzioni affini $\forall j, k$ ➔ ACQ è verificata
2) **Slatex Condition**
   SE g_j è convessa $\forall j = 1 \dots m$, e h_k è affine $\forall k = 1 \dots p$, e $\exists x' \in X: g(x') < 0$, $h(x') = 0$ ➔ ACQ è verificata $\forall x \in X$
3) **Dipendenza lineare dei gradienti dei vincoli attivi**
   SE $x^* \in X$ e i vettori
   i. $\nabla g_i(x^*) \leq 0, \forall i \in \mathcal{A}(x^*)$
   j. $\nabla h_k(x^*) = 0 \quad \forall k = 1 \dots p$

   sono linearmente dipendenti➔ ACQ è verificata in x^*

TEOREMA KKT •••

SE x^* è un minimo locale e **SE** l'ACQ è verificata ➔ $\exists \lambda^* \in \mathbb{R}^n$, $\exists \mu^* \in \mathbb{R}^n$: (x^*, λ^*, μ^*) soluzione del sistema:

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \cdot \nabla h_j(x^*) = 0 \\ \lambda_i^* \cdot g_i(x^*) = 0 \qquad \forall i = 1 \dots m \\ \lambda^* \geq 0 \\ g(x^*) \leq 0 \\ h(x^*) = 0 \end{cases}$$

LAGRANGIANA

$$\mathrm{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i \cdot g_i(x) + \sum_{j=1}^{p} \mu_j \cdot h_j(x) = 0$$

KKT Con la lagrangiana

$$\begin{cases} \nabla_x \mathrm{L}(x, \lambda, \mu) = 0. \\ \lambda_i \cdot g_i(x) = 0 \quad \forall i = 1 \dots m \\ \lambda \geq 0 \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

Con $\lambda_i \cdot \nabla g_i(x) = 0$, sostituibile con $\lambda^T \cdot \mathrm{g}(x) = 0$ o $< \lambda,\ g(x) >= 0$

Teorema – KKT per problemi convessi
SE il problema P è convesso e (λ^*, μ^*, x^*) è la soluzione del KKT ➔ x^* è un ottimo globale

Rilassamento Lagrangiano

$$\begin{cases} \inf(\ \mathrm{L}(x, \lambda, \mu)\ ) \\ x \in \mathbb{R}^n \end{cases}$$

Funzione duale Lagrangiana

$$\varphi(\lambda\ , \mu)\ := \inf(\ \mathrm{L}(x, \lambda, \mu)\ )$$

1) Concava
2) $\varphi(\lambda\ , \mu) \to -\infty$ per qualche punto
3) $\varphi(\lambda\ , \mu)$ non differenziabile per qualche punto

Teorema
$\forall\lambda \geq 0$ e $\forall\mu \in \mathbb{R}^p$ ➔ $\varphi(\lambda\ , \mu) \leq \mathrm{v(P)}$ minimo globale per P

Problema duale Lagrangiano

$$\begin{cases} \max\ (\ \varphi(\lambda\ , \mu)\ ) \\ \lambda \geq 0 \end{cases} = \boldsymbol{D}$$

D è per definizione **convesso** sempre.

Teorema – Weak Duality
Per qualsiasi problema di ottimizzazione P vale che: $v(D) \leq v(P)$

Teorema – Strong Duality
Supponiamo che f, g e h siano continue e differenziabili:

$$\begin{cases} \min f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases} = P$$

con P convesso, $\exists x^*$ ottimo globale e l' ACQ è soddisfatto ➔
1) La Strong Duality è verificata: $v(D) = v(P)$ e D ammette l'ottimo
2) (λ^*, μ^*) è l'ottimo per D **SSE** (λ^*, μ^*) è un vettore dei moltiplicatori KKT associato a x^*

Teorema
(x^*, λ^*, μ^*) è punto di **sella** per $L(x,\ \lambda\ ,\ \mu)$, ovvero

$$L(x^*, \lambda\ , \mu) \leq\ L(x^*, \lambda^*\ , \mu^*) \leq L(x, \lambda^*\ , \mu^*) \qquad \forall x \in \mathbb{R}^n,\ \ \forall\lambda \in \mathbb{R}^m_+,\ \ \forall\mu \in \mathbb{R}^p$$

SSE x^* è un ottimo per P, e (λ^*, μ^*) è un ottimo per D, e vale $v(D) = v(P)$

SUPPORT VECTOR MACHINE

Iperpiano Di Separazione

$$H = \{x \in \mathbb{R}^n : w^T x + b = 0\} : \begin{matrix} w^T x_i + b > 0 & \forall x_i \in A \\ w^T x_i + b < 0 & \forall x_i \in B \end{matrix}$$

Dove A e B sono i due labeled-set. Affinché H valga $conv(A) \cup conv(B) = \emptyset$

Funzione di Decisione

$$y_i = f(x) = sign(w^T x + b) = \begin{cases} 1 & se \quad w^T x + b > 0 \\ -1 & se \quad w^T x + b < 0 \end{cases}$$ ← classe

Definizione – Margine di Separazione

Se H è l'iperpiano di separazione dei due insiemi A e B, allora definiamo il margine di separazione di H come:

$$\rho(H) = \min_{x \in A \cup B} \frac{|w^T x + b|}{\| w \|}$$

Teorema

Trovare l'iperpiano H che ha il massimo margine di separazione $\rho(H)$ equivale a risolvere il seguente problema di programmazione quadratico convesso:

$$\begin{cases} \min\limits_{w,b} \frac{1}{2} \| w \|^2 \\ w^T x_i + b > 1 & \forall x_i \in A \\ w^T x_i + b < -1 & \forall x_i \in B \end{cases}$$

Linear SVM

$$\begin{cases} \min\limits_{w,b} \frac{1}{2} \| w \|^2 \\ 1 - y_i(w^T x_i + b) \leq 0 & \forall x_i = 1, \dots, l \end{cases}$$

Funzione Lagrangiana

$$L(x, b, \lambda) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{l} \lambda_i y_i w^T x_i - b \sum_{i=1}^{l} \lambda_i y_i + \sum_{i=1}^{l} \lambda_i$$

1) **SE** $\sum_{i=1}^{l} \lambda_i y_i \neq 0$ → $\min\limits_{w,b} L(x, b, \lambda) = -\infty$
2) **SE** $\sum_{i=1}^{l} \lambda_i y_i = 0$ → L non dipende da b,
   L è strongly convex,
   $\nabla_{\mathrm{w}} L(x, b, \lambda) = w - \sum_{i=1}^{l} \lambda_i \, y_i \, x_i = 0$

Funzione Duale

$$\varphi(\lambda) = \begin{cases} -\infty & se \quad \sum_{i=1}^{l} \lambda_i y_i \neq 0 \\ -\frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} y_i y_j x_i^T x_j \, \lambda_i \lambda_j + \sum_{i=1}^{l} \lambda_i & se \quad \sum_{i=1}^{l} \lambda_i y_i = 0 \end{cases}$$

Problema Duale

$$\begin{cases} \max_{\lambda} -\frac{1}{2}\sum_{i=1}^{l}\sum_{i=1}^{l} y_i y_j x_i^T x_j \,\lambda_i\lambda_j + \sum_{i=1}^{l}\lambda_i \qquad OPPURE \qquad \max_{\lambda} -\frac{1}{2}\lambda^T X^T X\lambda + e^T\lambda \\ \sum_{i=1}^{l}\lambda_i y_i = 0 \\ \lambda \geq 0 \end{cases}$$

Con $e^T = (1, \dots, 1)$ e $X = \{y_1 x_1, \dots, y_l x_l\}$

1) Il duale è convesso: $X^T X \geq 0$
2) λ^* ottimo del KKT associato all'ottimo del primale (w^*, b^*), è ottimo anche del duale
3) **SE** $\lambda_i^* > 0$ ➔ x_i è detto **support vector**
4) **SE** λ_i^* è ottimo del duale ➔ $w^* = \sum_{i=1}^{l}\lambda_i^*\, y_i\, x_i$
5) b^*è ottenuto da: $\lambda_i^*(1 - y_i(w^T x_i + b^*))$ ➔ $b^* = \frac{1}{y_i} - (w^*)^T x_i$

Linear SVM con SOFT MARGIN
Primale

$$\begin{cases} \min_{w,b} \frac{1}{2} \parallel w \parallel^2 + C\sum_{i=1}^{l}\xi_i \\ 1 - y_i(w^T x_i + b) \leq \xi_i \quad \forall i = 1, \dots, l \\ \xi_i \geq 0 \quad \forall i = 1, \dots, l \end{cases}$$

1) **SE** $\xi_i > 1$ ➔ il punto x_i è miss-classificato e $\sum_{i=1}^{l}\xi_i$ è un limite superiore al numero di punti miss-classificati
   a. **SE** $x_i \in A$ e $w^T x_i + b < 0$ ➔ $\xi_i > 1$ miss-classificato
   b. **SE** $x_i \in B$ e $w^T x_i + b > 0$ ➔ $\xi_i > 1$ miss-classificato

Duale

$$\begin{cases} \max_{\lambda} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j x_i^T x_j \lambda_i\lambda_j + \sum_{i=1}^{l}\lambda_i \\ \sum_{i=1}^{l}\lambda_i y_i = 0 \\ 0 \leq \lambda_i \leq C \end{cases}$$

1) **SE** λ^* è una soluzione ottima di questo duale ➔ w^*= $\sum_{i=1}^{l}\lambda_i^* y_i x_i$
2) Possiamo ottenere b^*, fissando $0 \leq \lambda_i \leq C$, e usando la condizione complementare (slackness): $\begin{cases} \lambda_i^*[1 - y_i((w^*)^T x_i + b^*) - \xi_i^*] = 0 \\ \mu_i^*\xi_i^* = (C - \lambda_i^*)\xi_i^* = 0 \end{cases}$ da cui: $b^* = \frac{1}{y_i} - (w^*)^T x_i$
3) ξ_i^* è l'errore nel classificare il punto $i-esimo$
4) **SE** $0 \leq \lambda_i^* < C$ ➔ $\xi_i^* = 0$
5) **SE** $0 < \lambda_i^* \leq C$ ➔ $\xi_i^* = 1 - y_i((w^*)^T x_i + b^*)$
6) **SE** $\xi_i^* > 0$ ➔ $\lambda_i^* = C$ → condizione necessaria per **mis-classificare** x_i

Non-Linear SVM
Primale

$$\begin{cases} \min\limits_{w,b} \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{l} \xi_i \\ 1 - y_i(w^T \phi(x_i) + b) \leq \xi_i \quad \forall i = 1, \dots, l \\ \xi_i \geq 0 \quad \forall i = 1, \dots, l \end{cases}$$

Duale

$$\begin{cases} \max\limits_{\lambda} -\frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} y_i y_j \phi(x_i^T) \phi(x_j)\, \lambda_i \lambda_j + \sum_{i=1}^{l} \lambda_i \\ \sum_{i=1}^{l} \lambda_i y_i = 0 \\ \lambda \geq 0 \end{cases}$$

$$w^* = \sum_{i=1}^{l} \lambda_i^* y_i\, \phi(x_i)$$

Troviamo b^*, fissando $0 < \lambda_i^* < C$:

$$y_i \left[\sum_{i=1}^{l} \lambda_i^* y_i\, \phi^T(x_i)\, \phi(x_i) + b^* \right] - 1 = 0$$

$$f(x) = sign\ ((w^*)^T \phi(x_i) + b^*)$$

Funzione Kernel
Definiamo $k: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, la funzione kernel: $k(x, y) = < \phi(x),\ \phi(y) >$

Teorema
Se k è una funzione Kernel, e se $x_1, \dots, x_l \in \mathbb{R}^n$ ➔ la matrice $K: k_{ij} = k(x_i, x_j) \geq 0$
Possiamo usare la funzione k per ridefinire il duale.

In Pratica

1) Scegliere una funzione Kernel k
2) Trovare la soluzione ottima del duale λ^*
3) Scegliere $i: 0 < \lambda_i^* < C$, e trovare b^*
4) Calcolare la decision function $f(x)$

REGRESSIONE

Sistema per trovare il polinomio di regressione

$$\begin{cases} \min \| Az - y \| \\ z \in \mathbb{R}^n \end{cases}$$

Dove: $A = \begin{pmatrix} 1 & x_1^1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_l^1 & \cdots & x_l^{n-1} \end{pmatrix} \in \mathbb{R}^{l \times n}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_l \end{pmatrix}$, $z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$

Residual Vector

$$r \in \mathbb{R}^l : r_i = p(x_i) - y_i$$

Dove r è la differenza fra valore reale e valore predetto.

Polinomio di regressione con $\| \bullet \|_2$

$$\begin{cases} \min \frac{1}{2} \| Az - y \|_2^2 \\ z \in \mathbb{R}^n \end{cases} = \frac{1}{2} z^T A^T A\, z - z^T A^T y + \frac{1}{2} y y^T$$

Con $rank(A) = n$, $A^T A > 0$

Soluzione: $z = (A^T A)^{-1} A^T y$

Polinomio di regressione con $\| \bullet \|_1$

$$\begin{cases} \min \frac{1}{2} \| Az - y \|_1 \\ z \in \mathbb{R}^n \end{cases} = \begin{cases} \min \sum_{i=1}^{l} u_i \\ u_i \geq A_i z - y_i \\ u_i \leq y_i - A_i z \end{cases}$$

Forma vettoriale:

$$\begin{cases} \min\limits_{z,u} \ (\emptyset_n^T, \ e_n^T) \begin{pmatrix} z \\ u \end{pmatrix} \\ D \begin{pmatrix} z \\ u \end{pmatrix} \leq d \end{cases}$$

Con:

$$D = \begin{pmatrix} A & -I_l \\ -A & -I_l \end{pmatrix}, \quad d = \begin{pmatrix} y \\ -y \end{pmatrix}$$

Polinomio di regressione con $\| \bullet \|_\infty$

$$\begin{cases} \min \| Az - y \|_\infty \\ z \in \mathbb{R}^n \end{cases} = \begin{cases} \min(u) \\ u_i \geq A_i z - y_i \\ u_i \leq y_i - A_i z \end{cases}$$

Forma vettoriale:

$$\begin{cases} \min\limits_{z,u} (0, 0, \dots, 0, 1) \begin{pmatrix} z \\ u \end{pmatrix} \\ D \begin{pmatrix} z \\ u \end{pmatrix} \leq d \end{cases}$$

Con:

$$D = \begin{pmatrix} A & -I_l \\ -A & -I_l \end{pmatrix}, \quad d = \begin{pmatrix} y \\ -y \end{pmatrix}$$

Regressione $\varepsilon - SV$ Lineare

$$\begin{cases} \min\left(\frac{1}{2} \parallel w \parallel^2\right) \\ y_i \geq w^T x_i + b + \varepsilon \\ y_i \leq w^T x_i + b - \varepsilon \end{cases}$$

Forma vettoriale:

$$\begin{cases} \min\limits_{z,u} \frac{1}{2}(w^T, \;\; b)\, Q \begin{pmatrix} w \\ b \end{pmatrix} \\ D \begin{pmatrix} w \\ b \end{pmatrix} \leq d \end{cases}$$

Con:

$$Q = \begin{pmatrix} I_l & \emptyset \\ \emptyset_l^T & \emptyset \end{pmatrix}, \;\; D = \begin{pmatrix} -X & -e_l \\ X & e_l \end{pmatrix}, \;\;\; d = \begin{pmatrix} \varepsilon e_l - y \\ \varepsilon e_l + y \end{pmatrix}$$

Regressione $\varepsilon - SV$ Lineare con variabili di rilassamento (Slack Variables)

$$\begin{cases} \min \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{l} (\xi_i^+, \xi_i^-) \\ y_i \leq w^T x_i + b + \varepsilon + \xi_i^+ \qquad \forall i = 1, \dots, l \\ y_i \geq w^T x_i + b - \varepsilon - \xi_i^- \qquad \forall i = 1, \dots, l \\ \xi^+, \, \xi^- \geq 0 \end{cases}$$

Forma vettoriale:

$$\begin{cases} \min\limits_{z,u} \frac{1}{2}(w^T, \;\; b, \;\; (\xi_i^+)^T, \;\; (\xi_i^-)^T)\, Q_1 \begin{pmatrix} w \\ b \\ \xi_i^+ \\ \xi_i^- \end{pmatrix} + c^T \begin{pmatrix} w \\ b \\ \xi_i^+ \\ \xi_i^- \end{pmatrix} \\ D_1 \begin{pmatrix} w \\ b \\ \xi_i^+ \\ \xi_i^- \end{pmatrix} \leq d_1 \\ \xi_i^+ \geq 0 \\ \xi_i^- \geq 0 \end{cases}$$

Con:

$$Q_1 = \begin{pmatrix} I_n & 0_n & \emptyset_{n \times 2l} \\ 0_n^T & 0 & 0_{2l}^T \\ \emptyset_{2l \times n} & 0_{2l} & 1_{2l \times 2l} \end{pmatrix}, \;\; D\_1 = \begin{pmatrix} -x & -e_l & -I_l & \emptyset_{l \times l} \\ x & e_l & \emptyset_{l \times l} & -I_l \end{pmatrix}, \;\;\; d = \begin{pmatrix} \varepsilon e_l - y \\ \varepsilon e_l + y \end{pmatrix}$$

Lagrangiana $\varepsilon - SV$ Lineare (Slack Variables)

$$L(w, b, \xi^+, \xi^-, \lambda^+, \lambda^-, \eta^+, \eta^-) =$$

$$\frac{1}{2} \parallel w \parallel^2 - w^T \left[\sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) x_i\right] - b \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) + \sum_{i=1}^{l} \xi^+ (C - \lambda_i^+ - \eta_i^+)$$
$$+ \sum_{i=1}^{l} \xi_i^- (C - \lambda_i^- - \eta_i^-) - \varepsilon \sum_{i=1}^{l} (\lambda_i^+ + \lambda_i^-) + \sum_{i=1}^{l} y_i \, (\lambda_i^+ - \lambda_i^-)$$

1) $\nabla_{\mathrm{w}} L = w - \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) x_i = 0$
2) $\nabla_{\mathrm{b}} L = -\sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) = 0$
3) $\nabla_{\xi_i^\pm} L = C - \lambda_i^\pm - \eta_i^\pm = 0$

Duale $\varepsilon - SV$ Lineare (Slack Variables)

$$\begin{cases} \max\limits_{\lambda^+,\lambda^-} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)(x_i)^T x_j - \varepsilon\sum_{i=1}^{l}(\lambda_i^+ + \lambda_i^-) + \sum_{i=1}^{l} y_i\,(\lambda_i^+ - \lambda_i^-) \\ \sum_{i=1}^{l}(\lambda_i^+ - \lambda_i^-) = 0 \\ \lambda_i^+ \in [0, C] \\ \lambda_i^- \in [0, C] \end{cases}$$

In formato vettoriale:

$$\begin{cases} \max\limits_{\lambda^+,\lambda^-} -\frac{1}{2}((\lambda^+)^T, (\lambda^-)^T)\ Q\ \begin{pmatrix}\lambda^+ \\ \lambda^-\end{pmatrix} + [-e(e_l^T,\ \ e_l^T) + (y^T, -y^T)]\begin{pmatrix}\lambda^+ \\ \lambda^-\end{pmatrix} \\ (e_l^T, -e_l^T)\begin{pmatrix}\lambda^+ \\ \lambda^-\end{pmatrix} = 0 \\ \lambda_i^+ \in [0, C] \qquad\qquad i = 1, \dots, l \\ \lambda_i^- \in [0, C] \qquad\qquad i = 1, \dots, l \end{cases}$$

Con:

$$Q = \begin{pmatrix} X & -X \\ -X & X \end{pmatrix},\ \ X = \left[(x_i)^T x_j\right]\ \forall i, j = 1, \dots, l$$

1) È un problema **convesso** di programmazione quadratica
2) **SE** $\lambda^+ > 0$ oppure $\lambda^- > 0$ ➔ x_i è un **support vector**
3) **SE** (λ^+, λ^-) ottimo del duale ➔ $w = \sum_{i=1}^{l}(\lambda_i^+ - \lambda_i^-)x_i$
4) b è ottenuto tramite la *complementary condition:* $\begin{cases} \lambda_i^+[\varepsilon + \xi_i^+ - y_i + w^T x_i + b] = 0 \\ \lambda_i^-[\varepsilon + \xi_i^- + y_i - w^T x_i - b] = 0 \\ \xi_i^+(C - \lambda_i^+) = 0 \\ \xi_i^-(C - \lambda_i^-) = 0 \end{cases}$
   i. $0 < \lambda^+ < C$ ➔ $b = y_i - w^T x_i - \varepsilon$
   ii. $0 < \lambda^- < C$ ➔ $b = y_i - w^T x_i + \varepsilon$

Regressione $\varepsilon - SV$ NON Lineare
Definiamo: $\phi: \mathbb{R}^n \to \mathcal{H}$, spazio delle caratteristiche, e cerchiamo una regressione lineare in $\{\phi(x_i), y_i\}$ nello spazio $\mathcal{H} \times \mathbb{R}$.

Primale

$$\begin{cases} \min \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{l} (\xi_i^+, \xi_i^-) \\ y_i \geq w^T \boldsymbol{\phi}(\boldsymbol{x_i}) + b + \varepsilon + \xi_i^+ \\ y_i \leq w^T \boldsymbol{\phi}(\boldsymbol{x_i}) + b - \varepsilon - \xi_i^- \end{cases}$$

Duale

$$\begin{cases} \max\limits_{\lambda^+, \lambda^-} -\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} (\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)\, k(x_i, x_j) - \varepsilon \sum_{i=1}^{l} (\lambda_i^+ + \lambda_i^-) + \sum_{i=1}^{l} y_i\, (\lambda_i^+ - \lambda_i^-) \\ \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) = 0 \\ \lambda_i^+ \in [0, C] \\ \lambda_i^- \in [0, C] \end{cases}$$

Ricordando che $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$

Passi

1) Scegliere una funzione kernel: $k(x_i, x_j)$
2) Risolvendo il duale abbiamo: (λ^+, λ^-) ➔ $w = \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) \phi(x_i)$
   a. $w^T \phi(x_i) = \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) k(x_i, x_j)$
3) *complementary condition:* $\begin{cases} \lambda_i^+ [y_i - f(x_i) - \varepsilon - \xi_i^+] = 0 \\ \lambda_i^- [y_i - f(x_i) + \varepsilon + \xi_i^+] = 0 \\ \xi_i^+ (C - \lambda_i^+) = 0 \\ \xi_i^- (C - \lambda_i^-) = 0 \end{cases}$
4) $\boldsymbol{0 < \lambda^+ < C}$ ➔ $b = y_i - w^T \phi(x_i) - \varepsilon = y_i - \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) k(x_i, x_j) - \varepsilon$
5) $\boldsymbol{0 < \lambda^- < C}$ ➔ $b = y_i - w^T \phi(x_i) + \varepsilon = y_i - \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) k(x_i, x_j) + \varepsilon$

Funzione di regressione per $\varepsilon - SV$ NON Lineare

$$f(x) = w^T \phi(x_i) + b = \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) k(x_i, x_j) + b$$

CLUSTER

Modello

$$\begin{cases} \min \sum_{i=1}^{l} \min_{j=1,\dots,k} d(p_i, x_j) \\ x_j \in \mathbb{R}^n \quad \forall j = 1, \dots, k \end{cases} \quad : \; M$$

Dove $d(p_i, x_j)$ è la distanza fra il punto p_i e il centroide x_j

Problema di Clustering

1) **SE** $k = 1$ il problema è convesso
2) **SE** $k > 1$ il problema è non convesso e non differenziabile

Teorema

Il problema $\begin{cases} \min \sum_{i=1}^{l} \min_{j=1,\dots,k} \| p_i - x_j \|_2^2 \\ x_j \in \mathbb{R}^n \quad \forall j = 1, \dots, k. \end{cases}$ è equivalente al seguente problema non convesso e non differenziabile:

$$\begin{cases} \min_{x,\alpha} f(x, \alpha) := \sum_{i=1}^{l} \sum_{j=1}^{k} \alpha_{ij} \| p_i - x_j \|_2^2 \\ \sum_{j=1}^{k} \alpha_{ij} = 1 \qquad \forall i = 1, \dots, l \\ \alpha_{ij} \geq 0 \quad \forall i = 1, \dots, l \quad \forall j = 1, \dots, k. \\ x_j \in \mathbb{R}^n \quad \forall j = 1, \dots, k \end{cases} \quad : \; C_2$$

K-Means

1) Inizializzazione
   a. Set $t = 0$
   b. Setta i centroidi $x_1^0 \dots x_k^0$
   c. Setta $\alpha_{ij}^0 = \begin{cases} 1 & j: \| p_i - x_j \|_2 = \min_{h=1,\dots,k} \| p_i - x_h^0 \|_2 \\ 0 & altrimenti \end{cases}$
      i. $\forall i = 1 \dots l$
2) Aggiornamento centroidi
   a. $\forall j = 1, \dots, k. \qquad x_j^{t+1} = \frac{\sum_{i=1}^{l} \alpha_{ij}^t \cdot p_i}{\sum_{i=1}^{l} \alpha_{ij}^t}$
3) Aggiornamento clusters
   a. $\forall i = 1, \dots, l. \qquad \alpha_{ij}^{t+1} = \begin{cases} 1 & j: \| p_i - x_j^{t+1} \|_2 = \min_{h=1,\dots,k} \| p_i - x_h^{t+1} \|_2 \\ 0 & altrimenti \end{cases}$
4) Stopping condition
   a. Se $f(x^{t+1}, \alpha^{t+1}) = f(x^t, \alpha^t)$ ➔ STOP
   b. Altrimenti $t++$ e vai al passo 2

Teorema

L'algoritmo di K-Means termina dopo un numero finito di iterazioni in una soluzione (x^*, α^*) del sistema KKT del problema C_2 tale che

$$f(x^*, \alpha^*) \leq f(x^*, \alpha) \qquad \forall \alpha \geq 0 \colon \sum_{j=1}^{k} \alpha_{ij} = 1. \quad \forall i = 1, \dots, l$$
$$f(x^*, \alpha^*) \leq f(x, \alpha^*) \qquad \forall x \in \mathbb{R}^{kl}$$

Teorema $\| \bullet \|_1$

Il problema $\begin{cases} \min \sum_{i=1}^{l} \min\limits_{j=1,\dots,k} \| p_i - x_j \|_1 \\ x_j \in \mathbb{R}^n \quad \forall j = 1, \dots, k \end{cases}$ è equivalente al seguente problema non convesso e non differenziabile:

$$\begin{cases} \min\limits_{x,\alpha} f(x, \alpha) \coloneqq \sum\limits_{i=1}^{l} \sum\limits_{j=1}^{k} \alpha_{ij} \| p_i - x_j \|_1 \\ \sum\limits_{j=1}^{k} \alpha_{ij} = 1 \qquad \forall i = 1, \dots, l \\ \alpha_{ij} \geq 0 \qquad \forall i = 1, \dots, l \quad \forall j = 1, \dots, k \\ x_j \in \mathbb{R}^n \qquad \forall j = 1, \dots, k \end{cases} \quad : \; C_1$$

Teorema •••

Il problema C_1 è equivalente al seguente problema non convesso e non differenziabile:

$$\begin{cases} \min\limits_{x,\alpha,u} \sum_{i=1}^{l} \sum_{j=1}^{k} \sum_{h=1}^{n} \alpha_{ij} u_{ijk} \\ u_{ijk} \geq (p_i)_h - (x_j)_h \quad \forall i = 1, \dots, l, \;\; \forall j = 1, \dots, k, \;\; \forall h = 1, \dots, n \\ u_{ijk} \geq (x_j)_h - (p_i)_h \quad \forall i = 1, \dots, l, \;\; \forall j = 1, \dots, k, \;\; \forall h = 1, \dots, n \\ \sum_{j=1}^{k} \alpha_{ij} = 1 \qquad \forall i = 1, \dots, l \\ \alpha_{ij} \geq 0 \qquad \forall i = 1, \dots, l, \;\; \forall j = 1, \dots, k \\ x_j \in \mathbb{R}^n \qquad \forall j = 1, \dots, k. \end{cases}$$

K-Median

1) Inizializzazione
    a. Set $t = 0$
    b. Setta i centroidi $x_1^0 \dots x_k^0$
    c. Setta $\alpha_{ij}^0 = \begin{cases} 1 & j: \| p_i - x_j \|_1 = \min_{h=1,\dots,k} \| p_i - x_h^0 \|_1 \\ 0 & altrimenti \end{cases}$
        i. $\forall i = 1 \dots l$
2) Aggiornamento centroidi
    a. $\forall j = 1, \dots, k. \quad x_j^{t+1} = mediano(p_i{:}\ \alpha_{ij}^t = 1)$
3) Aggiornamento clusters
    a. $\forall i = 1, \dots, l. \quad \alpha_{ij}^{t+1} = \begin{cases} 1 & j{:}\ \| p_i - x_j^{t+1} \|_1 = \min_{h=1,\dots,k} \| p_i - x_h^{t+1} \|_1 \\ 0 & altrimenti \end{cases}$
4) Stopping condition
    a. Se $f(x^{t+1}, \alpha^{t+1}) = f(x^t, \alpha^t)$ ➔ STOP
    b. Altrimenti $t++$ e vai al passo 2

Teorema

L'algoritmo di K-Median termina dopo un numero finito di iterazioni in una soluzione (x^*, α^*) punto stazionario tale che:

$$f(x^*, \alpha^*) \le f(x^*, \alpha) \qquad \forall \alpha \ge 0{:} \sum_{j=1}^{k} \alpha_{ij} = 1. \quad \forall i = 1, \dots, l$$

$$f(x^*, \alpha^*) \le f(x, \alpha^*) \qquad \forall x \in \mathbb{R}^{kl}$$

METODI PER PROBLEMI DI OTTIMIZZAZIONE VINCOLATI

(Exact) Penality Method

Trasformiamo il problema P (convesso, e sono con le $g_i(x)$) in un problema senza vincoli aggiungendo una penalità a f

$$\begin{cases} \min f(x) + \frac{1}{\varepsilon}\rho(x) := \rho_\varepsilon(x) \\ x \in \mathbb{R}^n \end{cases} \qquad = P_\varepsilon$$

Dove

$$\rho(x) = \sum_{i=1}^{m} (\max\{0, g_i(x)\})^2$$

Proposizione

1) SE f e g_i sono continue e differenziabili ➔anche ρ_ε lo è, e il gradiente è:
$$\nabla\rho_\varepsilon(x) = \nabla f(x) + \frac{2}{\varepsilon}\sum_{i=1}^{m} \max\{0, g_i(x)\}\,\nabla g_i(x)$$
2) SE f e g_i sono convesse ➔ anche ρ_ε lo è
3) Ogni problema P_ε ha una relazione con P ➔ $v(P_\varepsilon) \leq v(P)\ \ \forall\varepsilon > 0$
4) SE x_ε^* risolve il problema P_ε ➔ x_ε^* è ottimo anche per P
5) SE $0 < \varepsilon_2 < \varepsilon_1$ ➔ $v(P_{\varepsilon_1}) \leq v\left(P_{\varepsilon_2}\right)$

Passi del metodo

1) Settare $\varepsilon_0 > 0$
2) Settare $\tau \in (0, 1)$
3) Settare $k = 0$
4) Trovare l'ottimo x_k del problema P_{ε_k}
5) SE $x_k \in X$
   a. STOP
6) ALTIMENTI
   a. $\varepsilon_{k+1} = \tau\varepsilon_k$
   b. $k++$
   c. Passo 1

Teoremi

- **SE** f è coerciva ➔ $\exists\ \{x_k\}$, sequenza limitata, e ogni punto stazionario è un'ottima soluzione di P
- **SE** $\{x_k\} \to x^*$ ➔ x^* è un'ottima soluzione di P
- **SE** $\{x_k\} \to x^*$ e i gradienti dei vincoli attivi sono linearmente indipendenti ➔ x^* è un'ottima soluzione di P e la sequenza $\{\lambda_k\}:\ \lambda_k^i = \frac{2}{\varepsilon}\max\{0, g_i(x_k)\} \to \lambda^*$ vettore dei moltiplicatori del KKT associati a x^*

(In quello esatto, dato che il problema P convesso, abbiamo una sola soluzione e quindi un unico ε.)

Barrier Method

Questo metodo cerca una sequenza di punti ammissibili che approssima la soluzione ottima di P, problema vincolato, approssimandolo con:

$$\begin{cases} \min f(x) - \varepsilon \sum_{i=1}^{m} \log\big(-g_i(x)\big) = \psi_\varepsilon(x) \\ x \in int(X) \end{cases} \qquad := P_B$$

Proprietà di $B(x) = \sum_{i=1}^{m} \log\big(-g_i(x)\big)$

1) $dom(B) = int(X)$
2) B convesso
3) B derivabile con:
    a. $\nabla B(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)} \nabla g_i(x)$
    b. $\nabla^2 B(x) = \sum_{i=1}^{m} \frac{1}{[g_i(x)]^2} \nabla g_i(x)\, \nabla g_i(x)^T + \sum_{i=1}^{m} \frac{1}{-g_i(x)} \nabla^2 g_i(x)$

Passi del metodo

1) Settare $\delta > 0$
2) Settare $\varepsilon_1 > 0$
3) Settare $\tau \in (0,1)$
4) Scegliere x_0
5) Settare $k = 1$
6) Trovare l'ottimo x_k del problema P_B
    a. Partendo da x_{k-1}
7) **SE** $m\, \varepsilon_k < \delta$
    a. **STOP**
8) **ALTIMENTI**
    a. $\varepsilon_{k+1} = \tau \varepsilon_k$
    b. $k++$
    c. Passo 1

Trovare x_0

Consideriamo il problema ausiliare:

$$\begin{cases} \min s \\ g_i(x) < s \quad \forall i = 1, \dots, m \end{cases}$$

1) Prendiamo $x^{\sim} \in \mathbb{R}^n$, e cerchiamo $s^{\sim} > \max\limits_{i=1,\dots,m} g_i(x^{\sim})$ ➔ $(x^{\sim}, s^{\sim})$
2) Cercare la soluzione ottima dell'ausiliare: (x^*, s^*) usando un Barrier Method
3) **SE** $s^* < 0$ ➔ $x^* \in int(X)$
    a. **ALTRIMETNI** $int(X) = \emptyset$

PROBLEMI MULTI OBIETTIVO

Definizione

$$\begin{cases} \min f(x) = (f_1(x), \dots, f_s(x)) \\ x \in X \end{cases}$$

Ordine di Pareto

Dati $x, y \in \mathbb{R}^s$

$$x \geq y \quad \Leftrightarrow \quad x_i \geq y_i \qquad \forall i = 1, \dots, s$$

1) Reflexive: $x \geq x$
2) Asimmetrico: se $x \geq y$ e $y \geq x$ ➔ $x = y$
3) Transitiva: se $x \geq y$ e $y \geq z$ ➔ $x \geq z$

Minimi di Pareto per un Sottoinsieme

1) **Ideal Minimun**
   a. $x^* \in A$ è un $IMin(A)$ se $y \geq x^*, \forall y \in A$
   b. $x^* \in A$ è un $IMin(A)$ se $A \subseteq (x^* + \mathbb{R}^s_+)$
2) **Minimun**
   a. $x^* \in A$ è un $Min(A)$ se $\nexists y \in A,\ y \neq x^*,\ x^* \geq y$
   b. $x^* \in A$ è un $Min(A)$ se $A \cap (x^* - \mathbb{R}^s_+) = \emptyset$
3) **Weak Minimun**
   a. $x^* \in A$ è un $WMin(A)$ se $\nexists y \in A,\ y \neq x^*,\ x^* > y$
   b. $x^* \in A$ è un $WMin(A)$ se $A \cap (x^* - int(\mathbb{R}^s_+)) = \emptyset$

Preposizione

Se $IMin(A) \neq \emptyset$ ➔ $IMin(A) = Min(A) = \{x^*\}$

Teorema – Esistenza del minimo

Se $\exists x^* \in A$: $A \cap (x^* - \mathbb{R}^s_+)$ è compatto ➔ $Min(A) \neq \emptyset$

Minimi di Pareto per un problema multi-obiettivo P •••

1) **Ideal Minimun**
   a. $x^* \in X$ è un $IMin(P)$ se $f(x^*) = IMin\big(f(X)\big)$ ovvero se $f(x) \geq f(x^*), \forall x \in X$
2) **Minimun**
   a. $x^* \in X$ è un $Min(P)$ se $f(x^*) = Min\big(f(X)\big)$ ovvero che $\nexists x \in X$:
      i. $f_i(x^*) \geq f_i(x)$. $\forall i = 1, \dots, s$
      ii. $f_j(x^*) > f_j(x)$. Per qualche $j \in \{1, \dots, s\}$
3) **Weak Minimun**
   a. $x^* \in X$ è un $WMin(P)$ se $f(x^*) = WMin\big(f(X)\big)$ ovvero che:
      i. $\nexists x \in X$: $f_i(x^*) > f_i(x)$. $\forall i = 1, \dots, s$

Teorema •••

Se f_i è continua $\forall i = 1, \dots, s$ e se X è compatto ➔ $\exists\ Min(P)$

Teorema •••

Se f_i è continua $\forall i = 1, \dots, s$ e se X è chiuso, e se $\exists v \in \mathbb{R}, \exists j = 1, \dots, s$:

$$\{x \in X: f_j(x) \leq v\}$$

è un insieme non vuoto e limitato ➔ $\exists\ Min(P)$

Corollario
Se f_i è continua $\forall i = 1, \dots, s$ e se X è chiuso, con f_j coerciva per qualche $j \in \{1, \dots, s\}$
➔ $\exists \; Min(P)$

Teorema
$x^* \in X$ è un $Min(P)$ **SSE** il problema ausiliario P_a ha soluzione 0

$$\begin{cases} \max \sum_{i=1}^{s} \varepsilon_i \\ f_i(x) + \varepsilon_i \leq f_i(x^*). \quad \forall i = 1, \dots, s \\ x \in X \\ \varepsilon \geq 0 \end{cases} \qquad := P_a$$

Teorema
$x^* \in X$ è un $WMin(P)$ **SSE** il problema ausiliario P_{wa} ha soluzione 0

$$\begin{cases} \max v \\ v \leq \varepsilon_i & \forall i = 1, \dots, s \\ f_i(x) + \varepsilon_i \leq f_i(x^*). & \forall i = 1, \dots, s \\ x \in X \\ \varepsilon \geq 0 \end{cases} \qquad := P_{wa}$$

PROBLEMI MULTI OBIETTIVO NON VINCOLATI

Definizione problema

$$\begin{cases} \min f(x) = \big(f_1(x), \dots, f_s(x)\big) \\ x \in \mathbb{R}^n \end{cases} \quad := P_u$$

con f_i continua e differenziabile $\forall i = 1, \dots, s$

Teorema

SE $x^* \in X$ è un $Min(P_u)$ ➔ il problema S_1 non ha soluzione.

$$\begin{cases} \nabla f_i(x^*)^T d < 0. \quad \forall i = 1, \dots, s \\ d \in \mathbb{R}^n \end{cases} \quad := S_1$$

Condizione Necessaria di Ottimalità

SE $x^* \in X$ è un $WMin(P_u)$ ➔ $\exists \theta^* \in \mathbb{R}^s$: (x^*, θ^*) è una soluzione di:

$$\begin{cases} \sum_{i=1}^{s} \theta_i \nabla f_i(x^*) = 0 \\ \theta \geq 0 \\ \sum_{i=1}^{s} \theta_i = 1 \\ x \in \mathbb{R}^n \end{cases} \quad := S$$

Condizione Sufficiente di Ottimalità

SE P_u è convesso (Ovvero ogni f_i è convessa) e (x^*, θ^*) è l'ottimo per S ➔

1) $x^* = WMin(P_u)$
2) **SE** vale anche $\theta^* > 0$ ➔ $x^* = Min(P_u)$

PROBLEMI MULTI OBIETTIVO VINCOLATI

Definizione problema

$$\begin{cases} \min f(x) = \big(f_1(x), \dots, f_s(x)\big) \\ x \in X = \{x \in \mathbb{R}^n\colon\ g_j(x) \le 0,\ \ h_k(x) = 0\} \end{cases} \qquad := P$$

$\forall j = 1, \dots, m$ e $\forall k = 1, \dots, p$

Teorema KKT Multi-Obiettivo

SE x^* è un $WMin(P)$ e **SE** l'ACQ è verificata in x^* ➔ $\exists \theta^* \in \mathbb{R}^s, \exists \lambda^* \in \mathbb{R}^m,\ \exists \mu^* \in \mathbb{R}^p$: $(x^*,\ \theta^*,\ \lambda^*,\ \mu^*)$ è soluzione del sistema:

$$\begin{cases} \sum_{i=1}^{s} \theta_i \cdot \nabla f_i(x) + \sum_{j=1}^{m} \lambda_j \cdot \nabla g_j(x) + \sum_{k=1}^{p} \mu_k \cdot \nabla h_k(x) = 0 \\ \theta \ge 0 \\ \sum_{i=1}^{s} \theta_i = 1 \\ \lambda \ge 0 \\ \lambda_j \cdot \nabla g_j(x) = 0. \qquad \forall j = 1, \dots, m \\ g(x^*) \le 0 \\ h(x^*) = 0 \end{cases}$$

Teorema – Condizione Necessaria di Ottimalità

SE $x^* \in X$ è un $WMin(P)$ ➔ il problema S non ha soluzione.

$$\begin{cases} \nabla f_i(x^*)^T d < 0. \quad \forall i = 1, \dots, s. \\ d \in T_X(x^*) \end{cases} \qquad := S$$

Corollario

SE x^* è un $WMin(P)$ e **SE** l'ACQ è verificata in x^* il problema S_1 non ha soluzione.

$$\begin{cases} \mathrm{v}^{\mathrm{T}} \nabla f_i(x^*)^T d < 0. \quad \forall i = 1, \dots, s \\ \mathrm{v}^{\mathrm{T}} \nabla g_j(x^*)^T d \le 0. \quad \forall j = 1, \dots, m \\ \mathrm{v}^{\mathrm{T}} \nabla h_k(x^*)^T d < 0. \quad \forall k = 1, \dots, p \\ v \in \mathbb{R}^n \end{cases} \qquad := S_1$$

Teorema – Condizione Sufficiente di Ottimalità

Assunti f_i e g_j sono convesse $\forall i\ e\ \forall j$, e h_k è affine $\forall k$:

1) **SE** $(x^*,\ \theta^*,\ \lambda^*,\ \mu^*)$ risolve il KKT ➔ $x^* = WMin(P)$
2) **SE** $(x^*,\ \theta^*,\ \lambda^*,\ \mu^*)$ risolve il KKT e $\theta^* > 0$ ➔ $x^* = Min(P)$

Proposizione

SE x^* è un unico minimo globale per f_k in X, per qualche $k = 1, \dots, s$ ➔ $x^* = Min(P)$

SCALARIZATION METHOD

Associamo al problema P vincolato, il sistema P_α che associa pesi a ogni f_i:

$$\begin{cases} \min \sum_{i=1}^{s} \alpha_i f_i(x) \\ x \in X \end{cases} := P_\alpha$$

Con S_α insieme delle soluzioni ottime di P_α

$$\alpha \geq 0, \qquad \sum_{i=1}^{s} \alpha_i = 1$$

Teorema

- $\bigcup_{\alpha \geq 0} S_\alpha \subseteq \{WMin(P)\}$
- $\bigcup_{\alpha > 0} S_\alpha \subseteq \{Min(P)\}$

Teorema

SE che X è convesso e che f_i sono convesse in X, $\forall i = 1, \dots, s$ ➔

$$\bigcup_{\alpha \geq 0} S_\alpha = \{WMin(P)\}$$

Teorema

Sia P è lineare (ovvero che f_i è lineare $\forall i = 1, \dots, s$) e sia X un poliedro ➔

- $\bigcup_{\alpha \geq 0} S_\alpha = \{WMin(P)\}$
- $\bigcup_{\alpha > 0} S_\alpha = \{Min(P)\}$

Proposizione

SE x^* è un unico minimo globale per P_α, per qualche α, ➔ $x^* = Min(P)$

NOTA

Solo con $\alpha = 0$ si hanno $WMin(P)$, gli altri sono $Min(P)$ di Pareto; tranne che se con $\alpha = 0$ la funzione che rimane è strongly convex

NOTA

SE l'ottimo di P_α coincide sia con $\alpha = 0$ che con $\alpha = 1$ ➔ $x^* = IMin(P)$ perché minimizza entrambe le funzioni obiettivo

GOAL METHOD

Definiamo nello spazio obiettivo $\mathbb{R}^s$ un punto ideale z, così definito:

$$z_i = \min_{x \in X} f_i(x). \qquad \forall i = 1, \dots, s.$$

Teorema

Per avvicinarsi il più possibile a z, risolviamo:

$$\begin{cases} \min \| f(x) - z \|_q \\ x \in X \end{cases} \quad con \;\; q \in [1, +\infty] \qquad := G$$

- SE $q \in [1, +\infty)$ ➔ ogni soluzione ottima di G è un $Min(P)$
- SE $q = +\infty$) ➔ ogni soluzione ottima di G è un $WMin(P)$

Goal Method – Norma 2

$$\begin{cases} \min \frac{1}{2} \| Cx - z \|_2^2 = \frac{1}{2} x^T C^T C x - x^T C^T z + \frac{1}{2} z^T z \\ x \in X \end{cases} \qquad := G_2$$

GAME THEORY – Matrix Game

Definizione two-person non-cooperative game

$$P_1 \rightarrow \min_{x \in X} f_1(x, y), \qquad P_2 \rightarrow \min_{y \in Y} f_2(x, y)$$

Equilibrio di Nash

In un two-person non-cooperative game, una coppia di strategie $(x^*, \ y^*)$ è detta Equilibrio di Nash **SE**

$$f_1(x^*, y^*) = \min_{x \in X} f_1(x, y^*), \qquad f_2(x^*, y^*) = \min_{y \in Y} f_2(x^*, y)$$

Matrix Game

In un two-person non-cooperative game, dove:

- X e Y sono set finiti: $X = \{1, \dots, m\}$, $Y = \{1, \dots, n\}$
- $f_1 = -f_2$ (zero-sum game)

definiamo la Matrix Game C: $c_{ij} = f_1(i, j)$

Definizione – Strategia Strictly Dominated

Dato un two-person non-cooperative game, una strategia $x \in X$ è strettamente dominata da $x^* \in X$ se:

$$f_1(x, y) > f_1(x^*, y) \quad \forall y \in Y$$

Similmente $y \in Y$ è strettamente dominata fa $y^* \in Y$ se:

$$f_2(x, y) > f_2(x, y^*) \quad \forall x \in X$$

Le strategie dominate possono essere eliminate dal gioco

Mixed Strategies

SE C è un Matrix Game $m \times n$ ➔

• una Mixed Strategy per il giocatore 1 è un vettore grande m contenete probabilità.

$$X = \{x \in \mathbb{R}^m : x \geq 0, \ \sum_{i=1}^{m} x_i = 1\}$$

• una Mixed Strategy per il giocatore 2 è un vettore grande n contenete probabilità.

$$Y = \{y \in \mathbb{R}^n : y \geq 0, \ \sum_{j=1}^{n} y_j = 1\}$$

Pure Strategy

Sono i vertici di X e Y: $e_i = (0, \dots, 0, 1, 0, \dots, 0)$

Expected Cost

$$f_1(x, y) = x^T C y, \qquad f_2(x, y) = -x^T C y$$

Nota

$$x^T C y = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i c_{ij} y_j$$

Mixed Strategies Nash Equilibrium
SE C è un Matrix Game $m \times n$ ➔ $(x^*, y^*) \in X \times Y$ è un Mixed Strategies Nash Equilibrium **SE**
$$\max_{y \in Y} (x^*)^T C y = (x^*)^T C y^* = \max_{x \in X} x^T C (y^*)$$
O equivalentemente:
$$(x^*)^T C y \leq (x^*)^T C y^* \leq x^T C (y^*) \qquad \forall (x, y) \in X \times Y$$

(x^*, y^*) è un **punto di sella** per $f_1(x, y) = x^T C y$ su $X \times Y$

Punto Di Sella
Sia $X \subseteq \mathbb{R}^m$ e $Y \subseteq \mathbb{R}^n$. (x^*, y^*) è un punto di sella della funzione $F: X \times Y \to \mathbb{R}$ **SE**
$$F(x^*, y) \leq F(x^*, y^*) \leq F(x, y^*), \qquad \forall (x, y) \in X \times Y$$

Teorema
$(x^*, y^*) \in X \times Y$ soddisfa la condizione di punto di sella **SSE**
1) x^* è una soluzione ottima di: $\min_{x \in X} (\sup_{y \in Y} F(x, y))$
2) y^* è una soluzione ottima di: $\max_{y \in Y} (\inf_{x \in X} F(x, y))$
3) $\sup_{y \in Y} F(x, y) = \inf_{x \in X} F(x, y)$

Tutte e tre le condizioni sono equivalenti a: $\min_{x \in X} (\sup_{y \in Y} F(x, y)) = \max_{y \in Y} (\inf_{x \in X} F(x, y)) = F(x^*, y^*)$

Teorema – Esistenza del punto di sella
Sia $X \subseteq \mathbb{R}^m$ e $Y \subseteq \mathbb{R}^n$. Assumendo che
1) X e Y sono non vuoti, convessi e compatti
2) $F(\bullet, y)$ è continua e quasi-convessa in $X, \forall y \in Y$
3) $F(x, \bullet)$ è continua e quasi-convessa in $Y, \forall x \in X$

➔ F ammette un punto di sella in $X \times Y$

Corollario
1) Ogni Matrix Game ha almeno un Mixed Strategies Nash Equilibrium
2) (x^*, y^*) è un Mixed Strategies Nash Equilibrium **SSE**
$$\begin{cases} x^* \text{ è una soluzione ottima di: } \min_{x \in X} (\max_{y \in Y} x^T C y) \\ y^* \text{è una soluzione ottima di: } \max_{y \in Y} (\min_{x \in X} x^T C y) \end{cases}$$
   Con valore ottimo pari a $(x^*)^T C y^*$

Teorema
Il problema $\min_{x \in X} (\max_{y \in Y} x^T C y)$ è equivalente a risolvere il seguente problema di programmazione lineare:
$$\begin{cases} \min v \\ v \geq \sum_{i=1}^{m} c_{ij} x_i \qquad \forall j = 1, \dots, n \\ x \geq 0 \\ \sum_{i=1}^{m} x_i = 1 \end{cases} \qquad := P_1$$

Teorema
Il problema $\max_{y \in Y}(\min_{x \in X} x^T Cy)$ è equivalente a risolvere il seguente problema di programmazione lineare:

$$\begin{cases} \max w \\ w \geq \sum_{j=1}^{n} c_{ij} y_j \qquad \forall i = 1, \dots, m \\ y \geq 0 \\ \sum_{j=1}^{n} y_j = 1 \end{cases} := P_2$$

Proposizione
P_2 è il duale di P_1

Risoluzione Esercizi
I Pure Strategies Nash Equilibria possono essere cercati fra **i minimi di ogni colonna** della matrice per $Player_1$ e **i massimi di ogni riga** per $Player_2$.
I punti in comune sono Pure Strategies Nash Equilibria.

Nota
SE il P_1 è convesso ➔ Qualsiasi combinazione convessa fra due Pure Strategies Nash Equilibrium è una Mixed Strategies Nash Equilibria

Definizione Combinazione Convessa con due punti

$$\alpha x_1 + (1 - \alpha) x_2 \qquad \alpha \in [0, 1]$$

GAME THEORY – Bimatrix Games

Definizione

Un Bimatrix Game è un two-person non-cooperative game, dove:

- X e Y sono set finiti:
  - $X = \{x \in \mathbb{R}^m : x \geq 0,\ \sum_{i=1}^{m} x_i = 1\}$
  - $Y = \{y \in \mathbb{R}^n : y \geq 0,\ \sum_{j=1}^{n} y_j = 1\}$
- $f_1 \neq -f_2$ (non zero-sum game)
  - $f_1(x, y) = x^T C_1 y$
  - $f_2(x, y) = x^T C_2 y$

Teorema

Ogni Bimatrix Game ha almeno un Mixed Strategies Nash Equilibrium

Teorema

Definendo $B_1 : Y \to X$ e $B_2 : X \to Y$:

$$B_1(y) = \{Soluzioni\ ottime\ di\ \min_{x \in X} x^T C_1 y\}$$
$$B_2(x) = \{Soluzioni\ ottime\ di\ \min_{y \in Y} x^T C_2 y\}$$

➔ (x^*, y^*) è un Nash Equilibrium **SSE** $x^* \in B_1(y)$ e $y^* \in B_2(x)$

Teorema – KKT per Bimatrix Games

(x^*, y^*) è un Nash Equilibrium **SSE** $\exists \mu_1, \mu_2 \in \mathbb{R}$:

$$\begin{cases} C_1 y^* + \mu_1 e_m \geq 0 \\ x^* \geq 0 \\ \sum_{i=1}^{m} x_i^* = 1 \\ x_i^*(C_1 y^* + \mu_1 e_m) \geq 0. \quad \forall i = 1 \dots m \\ C_2 x^* + \mu_2 e_n \geq 0 \\ y^* \geq 0 \\ \sum_{j=1}^{n} y_j^* = 1 \\ y_j^*(C_2 x^* + \mu_2 e_n) \geq 0. \quad \forall j = 1 \dots n \end{cases} := KS$$

È Verificato

Proposizione

(x^*, y^*, μ_1, μ_2) è una soluzione per KS **SSE** è una soluzione ottima del problema di programmazione quadratica seguente:

$$\begin{cases} \min \psi(x, y, \mu_1, \mu_2) = [x^T(C_1 y + \mu_1 e_m) + y^T(C_2 x + \mu_2 e_n)] \\ C_1 y + \mu_1 e_m \geq 0 \\ x \geq 0 \\ \sum_{i=1}^{m} x_i = 1 \\ C_2 x + \mu_2 e_n \geq 0 \\ y \geq 0 \\ \sum_{j=1}^{n} y_j = 1 \end{cases} := QP$$

E $\psi(x^*, y^*, \mu_1, \mu_2) = 0$

Formato Matriciale di QP

$$\begin{cases} \min \frac{1}{2}(w^T, \mu^T)\, H \binom{w}{\mu} \\ A_{in} \binom{w}{\mu} \leq b_{in} \\ A_{eq} \binom{w}{\mu} \leq b_{eq} \\ w \geq 0 \end{cases} \qquad := QP$$

Con: ($O == matrice\ di\ uni$)

$$H = \begin{pmatrix} O_{m\times m} & C_1 + C_2 & e_m & O_{m\times 1} \\ C_1^T + C_2^T & O_{n\times n} & O_{n\times 1} & e_n \\ e_m^T & O_{1\times n} & 0 & 0 \\ O_{1\times m} & e_n^T & 0 & 0 \end{pmatrix}$$

$$A_{in} = \begin{pmatrix} -C_2^T & O_{n\times n} & O_{n\times 1} & -e_n \\ O_{m\times m} & -C_1 & -e_m & O_{m\times 1} \end{pmatrix} \qquad b_{in} = \begin{pmatrix} O_{n\times 1} \\ O_{m\times 1} \end{pmatrix}$$

$$A_{eq} = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 \end{pmatrix} \qquad b_{in} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$w^T = (x^T, y^T)$$

GAME THEORY – Convex Games

$$P_1 \to \begin{cases} \min\limits_x f_1(x,y) \\ g_i^1(x) \leq 0. \quad \forall i = 1 \dots p \end{cases}, \qquad P_2 \to \begin{cases} \min\limits_y f_2(x,y) \\ g_j^2(y) \leq 0. \quad \forall j = 1 \dots q \end{cases}$$

Con $f_1, f_2, g^1\ e\ g^2$ continue e differenziabili

Teorema

SE le regioni di ammissione

- $X = \{x \in \mathbb{R}^m : g_i^1(x) \leq 0,\ \forall i = 1 \dots p\}$
- $Y = \{y \in \mathbb{R}^n :\ g_i^2(y) \leq 0,\ \forall j = 1 \dots q\}$

Sono chiuse, limitate e convesse e le funzioni di costo:

1) $f_1(\bullet, y)$ è quasi-convessa $\forall y \in Y$
2) $f_2(x, \bullet)$ è quasi-convessa $\forall x \in X$

➔ esiste almeno un Nash Equilibrium

Teorema – Condizione KKT

SE (x^*, y^*) è un Nash Equilibrium e ACQ sono verificati sia in x^* che in y^* ➔ $\exists \lambda^1 \in \mathbb{R}^p\ e$ $\exists \lambda^2 \in \mathbb{R}^q$:

$$\begin{cases} \nabla_x f_1(x^*, y^*) + \sum_{i=1}^{p} \lambda_i^1 \nabla g_i^1(x^*) = 0 \\ \lambda^1 \geq 0 \\ g^1(x^*) \leq 0 \\ \lambda_i^1\ g_i^1(x^*) = 0 \qquad \forall i = 1, \dots, p \\ \nabla_y f_2(x^*, y^*) + \sum_{j=1}^{q} \lambda_j^2 \nabla g_j^2(y^*) = 0 \\ \lambda^2 \geq 0 \\ g^2(y^*) \leq 0 \\ \lambda_j^2\ g_j^2(y^*) = 0 \qquad \forall j = 1, \dots, q \end{cases}$$

SE $(x^*, y^*, \lambda^1, \lambda^2)$ risolve il KKT di sopra, e il game è convesso ➔ (x^*, y^*) è un Nash Equilibrium

MATLAB

Funzione $linprog\ (f, A, b, Aeq, beq, lb, ub, opzioni)$

Dove. Il sistema che risolve, in generale è fatto:

$$\begin{cases} \min\limits_{x}(\ f^T x\ ) \\ Ax \leq b \\ Aeq\ x = b. \\ lb \leq x \leq ub \end{cases}$$

Funzione $quadprog\ (H, f, A, b, Aeq, beq, lb, ub, x_0, opzioni)$

Dove. Il sistema che risolve, in generale è fatto:

$$\begin{cases} \min\limits_{x}\left(\frac{1}{2}x^T H\ x + f^T x\right) \\ Ax \leq b \\ Aeq\ x = beq \\ lb \leq x \leq ub \end{cases}$$

Funzione

$[x, fval, exitflag, output, lambda] = fmincon\ (fun, x_0, A, b, Aeq, beq, lb, ub, nonlcon)$

Dove. Il sistema che risolve, in generale è fatto:

$$\begin{cases} \min\limits_{x} f(x) \\ Ax \leq b \\ Aeq\ x = beq \\ lb \leq x \leq ubq \\ c(x) \leq 0 \\ ceq(x) = 0 \end{cases}$$

$fun = @(x)$ ...

$nonlcon = @(x)\ const(x)$

```
function [c, ceq] = const(x)
        c = [...];
        ceq = [...];
end
```

Funzione $fminunc$ uguale ma senza vincoli

RISOLUZIONE ESERCIZI

ESERCIZIO 1: Mono-Objective

1) **Provare se il problema ammette un ottimo globale**
   a. Controllare se $f(x)$ è convessa tramite gli auto-vettori dell'Hessiana
   b. Controllare se $f(x)$ è continua nel set X di ammissione
2) **Applicare il metodo indicato tramite lo script MatLab**
3) **Il punto ottenuto è un minimo globale del problema?**
   a. Se $f(x)$ è strongly convex allora il minimo globale è unico
   b. Per farlo bisogna vedere se la soluzione trovata è anche soluzione del KKT associato al problema. Usando il teorema che dice: SE il problema P è convesso e (λ^*, μ^*, x^*) è la soluzione del KKT ➔ x^* è un ottimo globale

ESERCIZIO 2: SVM

1) **Scrivere il modello richiesto, copiandolo da sopra.**
2) **Risolvere usando MatLab, con i parametri descritti dalla traccia**
3) **Trovare i punti miss-classified**
   a. Bisogna considerare la "complementary slackness condition", in particolare se il punto è miss-classificato allora l'errore $\xi_i^* > 0$, per cui da $(C - \lambda_i^*)\xi_i^* = 0$ deriviamo $C = \lambda_i^*$. Cercare i punti con questo λ_i^*. Ma non basta, potrebbero esserci punti classificati correttamente anche con $C = \lambda_i^*$. Per notarlo basta calcolare il valore dell'Iperpiano con il punto indicato, e vedere se soddisfa:
      i. $w^T x_i + b > 0 \quad \forall x_i \in A$
      ii. $w^T x_i + b < 0 \quad \forall x_i \in B$
4) **Classifica il nuovo punto**
   a. Bisogna vedere se:
      i. $w^T x_i + b > 0 \quad \forall x_i \in A$
      ii. $w^T x_i + b < 0 \quad \forall x_i \in B$

ESERCIZIO 2: Regressione

1) **Scrivere il modello richiesto, copiandolo da sopra.**
2) **Risolvere usando MatLab, con i parametri descritti dalla traccia**
3) **Trovare i Support Vector:**
   a. Usare MatLab, i SV sono dati dalla seconda e dalla terza colonna
4) **Trovare i punti che cadono fuori dal $\varepsilon - tube$ usando la soluzione del duale**
   a. Guardando le condizioni del primale:
      i. $y_i - f(x_i) - \varepsilon - \xi_i^+ \leq 0$
      ii. $y_i - f(x_i) + \varepsilon + \xi_i^+ \leq 0$

   Se un punto è fuori dal $\varepsilon - tube$, allora l'errore è $\xi_i^+,\ \xi_i^- > 0$
   Bisogna cercare i punti a cui corrisponde un errore positivo, cerchiamo prima nella soluzione duale i punti (λ^+, λ^-), e cerchiamo i punti che hanno $C = \lambda_i^*$, per lo stesso principio del problema SVM.

ESERCIZIO 2: Clustering

1) **Scrivere il modello richiesto, copiandolo da sopra.**
2) **Risolvere usando MatLab, con i parametri descritti dalla traccia**

ESERCIZIO 3: Multi-Objective

1) **Prova l'esistenza del Minimo di Pareto, trovare i $\boldsymbol{WMin(P)}$ e i $\boldsymbol{Min(P)}$**
    a. Per farlo risolvere il P_α dello scalarizied problem, attraverso il KKT, provando che il problema P_α è convesso e che le ACQ sono soddisfatti:
        i. Per $\alpha_1 = 0$
        ii. Per $\alpha_1 = 1$
        iii. Per $0 < \alpha_1 < 1$
            1. Distinguere i casi per $\lambda = 0$ e $\lambda > 0$
    b. **SE** il problema P è lineare allora tutti i risultati trovati sono $WMin(P)$ o $Min(P)$ secondo il teorema di sopra descritto
    c. **SE** una delle f_i è Strongly Convex allora ha un unico minimo che è un $Min(P)$

ESERCIZIO 4: Mono-Matrix

1) **Trovare le strategie dominate per eliminarle:**
    a. Player 2: Eliminiamo le **colonne** che hanno tutti i valori ≤ agli elementi corrispondenti delle altre colonne (Mi tengo la colonna maggiore)
    b. Player 1: Eliminiamo le **righe** che hanno tutti i valori ≥ agli elementi corrispondenti delle altre righe (Mi tengo la riga minore)
2) **Trovare le Pure Strategies Nash Equilibiria**
    a. Cercare gli elementi che sono contemporaneamente **massimi per la propria riga** e **minimi per la propria colonna**.
    b. Ricordare di mantenere gli indici della matrice originale
3) **Trovare un Mixed Strategies Nash Equilibiria**
    a. Risolvere il sistema primale dato dalla definizione sopra, con MatLab
    b. **SE** il risultato del sistema è un Pure Strategies Nash Equilibiria, allora controllare se il problema è lineare, per cui ogni combinazione convessa fra due punti di Pure Strategies Nash Equilibiria è un Mixed Strategies Nash Equilibiria.

ESERCIZIO 4: Bi-Matrix

1) **Fa tutto MatLab** 😂