Dall’Odissea a Google : viaggio tra le tecniche di ottimizzazione

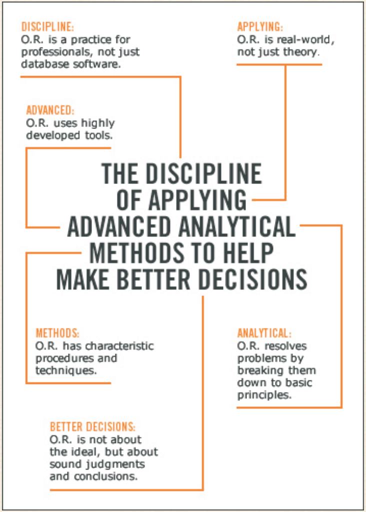
Laura Scrimali

Dipartimento di Matematica e Informatica

23 maggio 2019

La Ricerca Operativa

La Ricerca Operativa



Sommario

- L'Ottimizzazione matematica
- Il gioco dei Lego
- Ulisse, il più famoso commesso viaggiatore
- Google e le reti sociali: l'ottimizzazione dei grandi numeri

Nulla accade in natura che non possa essere ricondotto ad un problema di massimizzazione o minimizzazione.

Eulero (1707 -1783)

Tutti i giorni . . .

. . . accade di dover risolvere problemi di ottimizzazione. Si pensi al percorso migliore per recarsi al lavoro, fare la spesa senza spendere troppo o gestire attività senza spreco di tempo. Se è vero che si possono risolvere semplici problemi quotidiani senza ricorrere a strumenti matematici, è altrettanto vero che, nei casi più complessi, si rende necessario un approccio decisionale più sofisticato, quale quello offerto dall'Ottimizzazione Matematica.

Ottimizzazione Matematica

L'Ottimizzazione Matematica si occupa di problemi, detti **problemi di ottimizzazione**, della forma:

$$\begin{aligned} &\min \ f(x) \\ &x \in S, \end{aligned}$$

$f : \mathbb{R}^n \to \mathbb{R}$, $S \subset \mathbb{R}^n$

- f **funzione obiettivo**
- S **regione ammissibile**
- $x \in S$ **punto ammissibile**

Un problema di ottimizzazione consiste pertanto nel determinare, se esiste, un punto di minimo della funzione f sotto dei vincoli assegnati.

Può accadere che

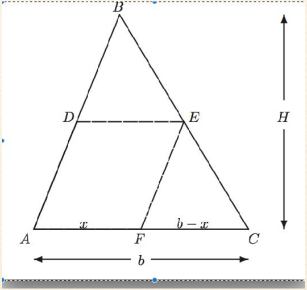
- l'insieme ammissibile S sia vuoto: non esistono punti ammissibili e di conseguenza non esistono soluzioni ottime;
- l'insieme ammissibile S sia non vuoto; allora
  - la funzione obiettivo non è limitata inferiormente e dunque non esiste un valore minimo;
  - la funzione obiettivo è limitata inferiormente, ma non esistono punti di minimo globale;
  - esistono punti di minimo globale di f su S e la funzione è, ovviamente, limitata inferiormente.

Un po' di storia

Il prima problema di massimo della storia si deve ad una leggenda narrata anche nel I Libro dell'Eneide. Didone, dopo la morte del marito, fugge con alcuni sudditi e approda sulle coste africane. Lì chiede al re della Libia un pezzo di terra su cui fondare una città. Il re non vuole dare asilo ai fuggiaschi, a meno che lei non acconsenta a sposarlo. La donna rifiuta le nozze e così ottiene dal re tanta terra quanta una pelle di bue ne potesse circondare. Didone riesce ad occupare la terra necessaria per fondare Cartagine: taglia in strisce sottilissime la pelle e recinta un bel pezzo di terreno a forma di semicerchio. Il problema di Didone è noto come problema isoperimetrico: fra tutte le curve piane di uguale perimetro, determinare quella di area massima.

Il libro *Elementi* di Euclide (IV a.C.) è stato il primo libro di testo di matematica della storia, e contiene un problema di ottimizzazione.

In un dato triangolo ABC inscrivere un parallelogramma ADEF tale che EF//AB e DE//AC e tale che la sua area è massima.



Il problema può essere formulato come:

$$\max_{0<x<b} \frac{xH(b-x)}{b}$$

Si tratta di un problema non vincolato che si risolve ponendo uguale a zero la derivata della funzione obiettivo e risolvendo l'equazione corrispondente (Fermat, 1601-1665).
Lagrange ha poi esteso la tecnica al caso di problemi con vincoli di uguaglianza.
Isaac Newton combinando il metodo di Fermat e il metodo dei moltiplicatori di Lagrange, ha introdotto il primo algoritmo di ottimizzazione.
Nel XIX secolo Karl Weierstrass (1815 -1897) ha dimostrato il famoso risultato che fornisce, sotto opportune ipotesi, una condizione sufficiente per l'esistenza di una soluzione ottima.

Risolvere i problemi di ottimizzazione...

...è in generale difficile: spesso le computazioni sono molto lunghe e onerose e non sempre si trova una soluzione.

Eccezione: certi problemi possono essere risolti in modo efficiente (programmazione lineare, problemi di ottimizzazione convessa).

Approccio modellistico

1. **Formulazione del problema:** individuazione degli obiettivi, determinazione dei vincoli e raccolta dei dati.
2. **Costruzione del modello:** compromesso tra precisione e trattabilità matematica, individuazione delle variabili, funzione obiettivo, vincoli.
3. **Analisi del modello:** deduzione di proprietà (esistenza e unicità della soluzione, condizioni di ottimalità).
4. **Soluzione numerica:** costruzione dell'algoritmo e calcolo della soluzione.
5. **Validazione del modello:** collaudo e miglioramento del modello.

Problemi di ottimizzazione

- Problemi di natura industriale: *pianificazione della produzione*, *gestione ottima delle scorte*, *localizzazione e dimensionamento di impianti*.
- Problemi di progettazione ottima: *progettazione di reti e loro gestione*, *progettazione strutturale*, *progettazione di sistemi ottici*, *allocazione ottima di componenti elettronici*.
- Problemi di economia e finanza: *scelta di investimenti*, *composizione di un portafoglio*.
- Problemi di organizzazione: *determinazione dei turni del personale*, *manutenzione di beni*, *project planning*.

- Problemi scientifici: *studi sulla struttura del DNA*, *ricostruzione di immagini*.
- Problemi di diagnostica medica: *interpretazione e analisi dei dati ottenibili da strumenti di analisi clinica*.
- Problemi di controllo ottimo: *controllo di servomeccanismi e di sistemi di guida*, *controllo di traiettorie*.

Classificazione

Si possono distinguere due ampie classi di problemi:

Problemi di Programmazione Lineare (PL) La funzione obiettivo e le funzioni dei vincoli sono lineari, sono cioè del tipo: $a_1x_1 + \cdots + a_nx_n$.

Problema di Programmazione non Lineare Problema di Programmazione non Lineare: almeno una tra la funzione obiettivo e le funzioni vincolari non è lineare.

Problema generale di PL

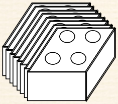
$$
\begin{aligned}
&\min \left(c_1 x_1 + c_2 x_2 + \cdots + c_n x_n \right) \\
&a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \gtreqless b_1 \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n \gtreqless b_m \\
&x_j \geq 0, \quad j = 1, \dots, n
\end{aligned}
$$

- x_j **variabili decisionali**
- c_j **coefficienti di costo (o di profitto)**
- b_i **termini noti**
- a_{ij} **coefficienti del sistema dei vincoli**

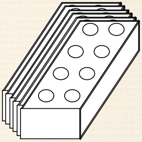
Il gioco dei Lego: un modello di produzione

Si vogliono determinare le quantità ottimali di tavoli e sedie in modo da massimizzare il profitto.

Disponibilità settimanale di materie prime:

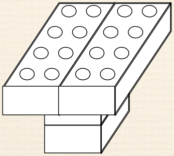


8 mattoni piccoli

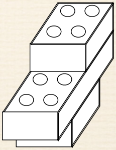


6 mattoni grandi

Prodotti:



Tavoli
Profitto = euro 20/Tavolo



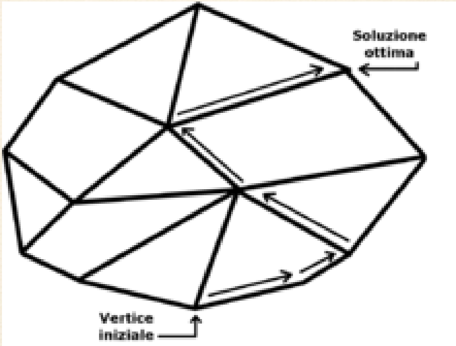
Sedie
Profitto = euro15/Sedia

Formulazione matematica

$$\begin{cases} \max(15x_1 + 20x_2) \\ x_1 + 2x_2 \leq 6 \\ 2x_1 + 2x_2 \leq 8 \\ x_1, x_2 \geq 0 \end{cases}$$

Il metodo del simplesso

Il metodo è stato elaborato da George B. Dantzig (1914-2005). Sfrutta le proprietà geometriche dell'insieme dei vincoli, mediante la loro rappresentazione algebrica.



La programmazione lineare intera

Un larghissimo numero di problemi reali può essere rappresentato da modelli di programmazione lineare intera. Si tratta di applicazioni caratterizzate dall'indivisibilità delle risorse e dalla necessità di scegliere tra un numero finito di alternative: la distribuzione di beni, il sequenziamento delle attività produttive, la gestione ottima del portafoglio titoli, la localizzazione degli impianti, la progettazione di sistemi automatici di produzione (robotica) e problemi di Ottimizzazione Combinatoria.

L'algoritmo del Branch and Bound: il bound

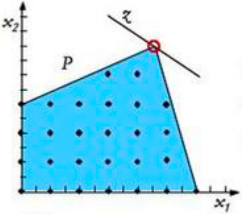
Max $z = 2x_1 + 3x_2$

s.a

$-1.3x_1 + 3x_2 \leq 9$

$3x_1 + 0.9x_2 \leq 18$

$x_1, x_2 \geq 0$ interi



$z^*_c = 23.89$

$x^*_1 = 4.51$

$x^*_2 = 4.95$

(S_1)

Max $z = 2x_1 + 3x_2$

s.a

$-1.3x_1 + 3x_2 \leq 9$

$3x_1 + 0.9x_2 \leq 18$

$x_1 \leq 4$

x_1, x_2 interi

(S_2)

Max $z = 2x_1 + 3x_2$

s.a

$-1.3x_1 + 3x_2 \leq 9$

$3x_1 + 0.9x_2 \leq 18$

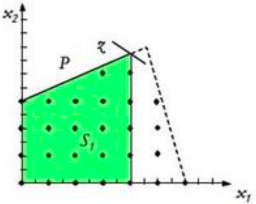
$x_1 \geq 5$

x_1, x_2 interi

L’algoritmo del Branch and Bound: il branch

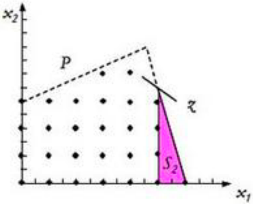
(S_1)

$$\begin{aligned} \text{Max } z &= 2x_1 + 3x_2 \\ \text{s.a} \\ -1.3x_1 + 3x_2 &\le 9 \\ 3x_1 + 0.9x_2 &\le 18 \\ x_1 &\le 4 \\ x_1, x_2 &\text{ interi} \end{aligned}$$



(S_2)

$$\begin{aligned} \text{Max } z &= 2x_1 + 3x_2 \\ \text{s.a} \\ -1.3x_1 + 3x_2 &\le 9 \\ 3x_1 + 0.9x_2 &\le 18 \\ x_1 &\ge 5 \\ x_1, x_2 &\text{ interi} \end{aligned}$$



(S_1)

$\text{Max } z = 2x_1 + 3x_2$

s.a

$$-1.3x_1 + 3x_2 \leq 9$$
$$3x_1 + 0.9x_2 \leq 18$$
$$x_1 \leq 4$$
$$x_1, x_2 \quad \text{interi}$$

(S_2)

$\text{Max } z = 2x_1 + 3x_2$

s.a

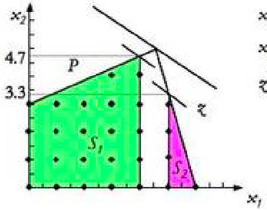
$$-1.3x_1 + 3x_2 \leq 9$$
$$3x_1 + 0.9x_2 \leq 18$$
$$x_1 \geq 5$$
$$x_1, x_2 \quad \text{interi}$$

(S_1)

$x_1^* = 4$

$x_2^* = 4.73$

$\boxed{z^* = 22.2}$



(S_2)

$x_1^* = 5$

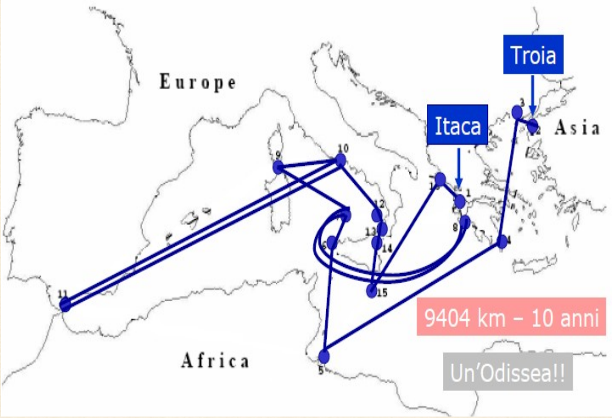
$x_2^* = 3.3$

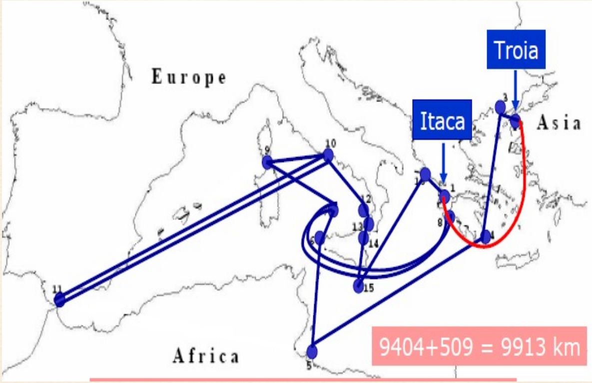
$z^* = 20.0$

Il viaggio di Ulisse

Sarebbe dovuto andare da Troia ad Itaca. . .

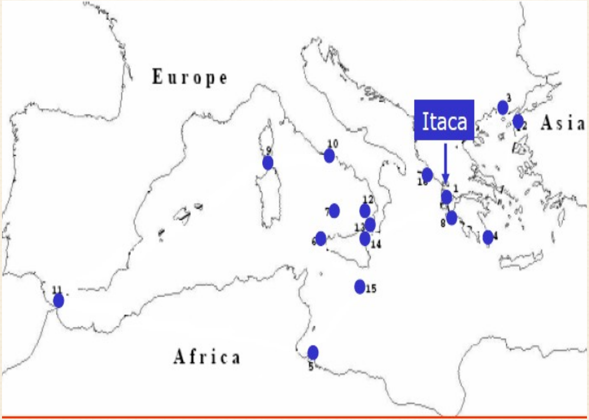
Avrebbe dovuto percorrere solo 509 Km. Ne ha percorsi 9404, visitando 16 diverse città.




Europe
Troia
Itaca
Asia
Africa
9404+509 = 9913 km

Il problema di Ulisse

Ulisse avrebbe dovuto determinare l'itinerario che gli avrebbe permesso di visitare le città una ed una sola volta, in modo da avere un percorso di lunghezza totale minima.

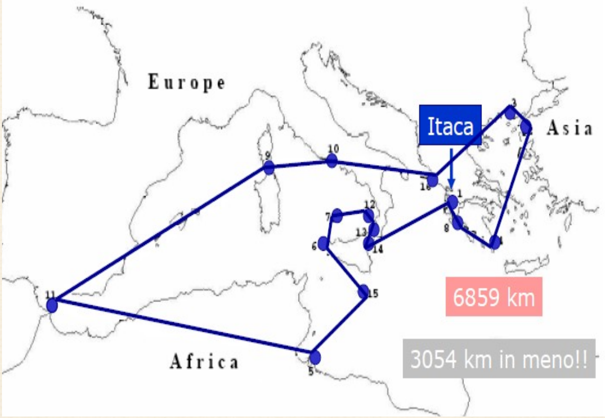


Problema del commesso viaggiatore (TSP)

Un commesso viaggiatore deve determinare l'itinerario che gli permette di visitare n città una ed una sola volta, in modo da avere un percorso di lunghezza totale minima (di costo totale minimo). Si consideri il grafo $G = (N, A)$, dove N è l'insieme degli n nodi (città) ed A è l'insieme degli archi (strade). Si vuole determinare un circuito hamiltoniano (che tocchi ogni nodo una ed una sola volta) di costo minimo. Sia c_{ij} il costo associato all'arco (i, j) e sia x_{ij} la variabile che vale 1 se l'arco appartiene al circuito hamiltoniano e zero altrimenti.

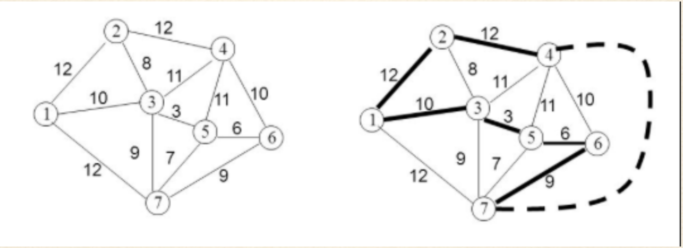
$$
\begin{cases}
\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \\
\text{(costo totale del circuito)} \\
\sum_{\substack{i=1 \\ i \neq j}}^{n} x_{ij} = 1, \quad j \in N \\
\text{(in ogni nodo entra un solo arco)} \\
\sum_{\substack{j=1 \\ i \neq j}}^{n} x_{ij} = 1, \quad i \in N \\
\text{(da ogni nodo esce uno ed un solo arco)} \\
\sum_{i \in S, j \in N \setminus S} x_{ij} \geq 1, \quad \forall S : S \subset N,\ 2 \leq |S| \leq n-2 \\
\text{(assenza di sottocircuiti)} \\
x_{ij} \in \{0,1\},\ \forall j \in \{1, \ldots, n\}, i \in \{1, \ldots, m\}.
\end{cases}
$$

La soluzione ottima

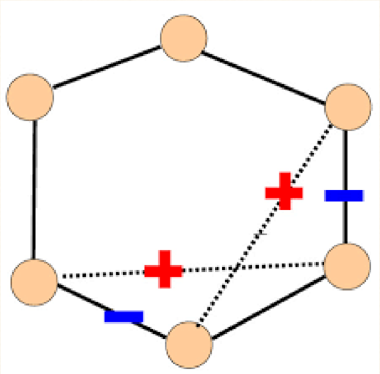


Il metodo del nodo più vicino

L'algoritmo del nodo più vicino costruisce un cammino contenente tutti i nodi, aggiungendo ad ogni iterazione il nodo che produce il minimo incremento di costo. Unendo poi gli estremi del cammino trovato, si ottiene un circuito hamiltoniano.



Il metodo del Branch and Bound per il TSP



Dati di grandi dimensioni

La maggior parte delle nostre attività quotidiane crea dei dati: ricerche su Google, acquisti con bancomat e carte di credito, foto, messaggi o tweet. Questi dati possono essere raccolti, analizzati e monetizzati.
Anche la produzione di dati all'interno delle aziende è aumentata considerevolmente, poiché è possibile acquisire un vantaggio competitivo sulla concorrenza lavorando sui dati. Conoscere la propria utenza amplia le possibilità di soddisfarla e di proporle nuovi prodotti.

Esempio

In un centro commerciale degli USA sono stati installati sistemi di riconoscimento facciale per la sicurezza. Dopo aver raccolto dati per qualche tempo, si è scoperto che la maggior parte dei clienti durante la pausa pranzo era di origine asiatica.

Esempio

In un centro commerciale degli USA sono stati installati sistemi di riconoscimento facciale per la sicurezza. Dopo aver raccolto dati per qualche tempo, si è scoperto che la maggior parte dei clienti durante la pausa pranzo era di origine asiatica.

Azioni

Sono stati cambiati i turni del personale in modo che la maggior parte degli impiegati asiatici fosse in servizio durante la pausa pranzo e sono stati assunti nuovi impiegati asiatici.

Esempio

In un centro commerciale degli USA sono stati installati sistemi di riconoscimento facciale per la sicurezza. Dopo aver raccolto dati per qualche tempo, si è scoperto che la maggior parte dei clienti durante la pausa pranzo era di origine asiatica.

Azioni

Sono stati cambiati i turni del personale in modo che la maggior parte degli impiegati asiatici fosse in servizio durante la pausa pranzo e sono stati assunti nuovi impiegati asiatici.

Risultato

Sono aumentate le vendite.

Reti sociali

Il numero di contenuti generati dagli utenti e riversati su piattaforme come Facebook, YouTube e Twitter cresce continuamente. Le informazioni in queste reti sociali possono essere generate da chiunque e in qualunque momento. Classificare una tale mole di dati usando approcci classici sarebbe troppo dispendioso in termini di tempo, computazioni e spazio di memoria.

La Scienza dei Dati

Si parla di *Big Data* quando le dimensioni dei database devono essere misurate in terabyte (2 alla quarantesima byte, $1.048.576$ megabyte) o in petabyte (2 alla cinquantesima byte).

Si tratta di un'area fortemente interdisciplinare, che richiede conoscenze di statistica, informatica, ottimizzazione e esperienze specifiche nell'ambito delle applicazioni in questione. L'ottimizzazione interviene anche per trasformare la conoscenza in decisioni.

Caratteristiche

- Volume: dimensione effettiva del dataset;
- Velocità: velocità di generazione dei dati; si tende all'effettuare analisi dei dati in tempo reale;
- Varietà: varie tipologie di dati, provenienti da fonti diverse (strutturate e non);
- Variabilità: possibile inconsistenza dei dati;
- Complessità: maggiore è la dimensione del dataset, maggiore è la complessità dei dati da gestire; il compito più difficile è collegare le informazioni, ed ottenerne di interessanti.
- Veridicità: qualità dei dati intesa come il valore informativo che si riesce ad estrarre.

Internet of Things

Ogni oggetto intorno a noi si connette al mondo e raccoglie informazioni per aumentare la propria funzionalità e migliorare l'esperienza di utilizzo. Ogni dispositivo deve avere coscienza del suo stato e dell'ambiente in cui si trova e tutte queste rilevazioni generano a loro volta dati.


edureka!
BIG DATA
7 connected devices per person
COMMUNICATION
By 2020 each person will own an avg of 7 connected devices
71% of shoppers are multi channel
RETAIL
Retail companies are using IoT devices to manage their sales & customer acquisition
30% annual growth rate
MEDICAL
Medical data disclosure is the second most breached source of data
INDUSTRIAL
Project will increase in connected machine - to - machine devices over the next 5 yrs
VEHICLES
23.6 million cars is having internet access by 2016, raising from 8.7 million in 2010

Cosa fare con i dati?

Classificazione e predizione sono processi che consistono nel creare dei modelli che possono essere usati

- per descrivere degli insiemi di dati
- per fare previsioni future

Molti algoritmi sono stati sviluppati per risolvere problemi di classificazione e predizione: Apprendimento Automatico (machine learning), statistica e neurobiologia.

Regressione and Classificazione

Assegnati i dati a_i e gli output y_i associati ad alcuni elementi, possiamo costruire una funzione ϕ che mappi i dati ai suoi output $y_i \sim \phi(a_i)$. ϕ può essere applicata a futuri dati a per predire l'output $\phi(a)$. Può essere formulato come un problema di

- Regressione minimi quadrati
- Regressione logistica
- Support Vector Machine (SVM)

Ottimizzazione e Apprendimento Automatico

Le applicazioni suggeriscono l'integrazione tra Apprendimento Automatico e Ottimizzazione. In particolare, c'è una stretta connessione tra Apprendimento Automatico e Programmazione Mista-Intera (MIP). I dati grezzi sono trattati dall'Apprendimento Automatico per costruire modelli su cui applicare l'Ottimizzazione. L'integrazione non è però riservata unicamente a MIP. Ad esempio, in *deep learning* (apprendimento per stati mutlipli), si usano strutture più complicate in cui le osservazioni non sono fisse ma sono trasformate dal processo di apprendimento. Ciò implica la presenza di funzioni non convesse e decisioni discrete.

Tipi di apprendimento

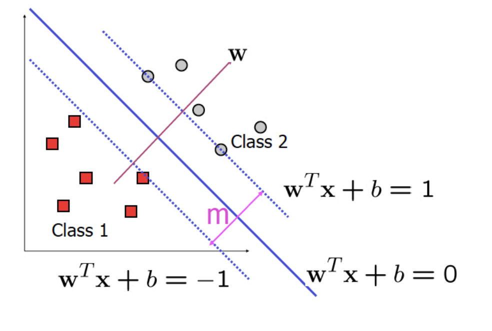
Apprendimento supervisionato: si sviluppano delle regole di classificazione e per l'addestramento si utilizzano dati già classificati. I più utilizzati sono:

- SVM;
- Reti neurali.

Apprendimento non supervisionato: cercano di individuare delle strutture (o informazioni) utili, all'interno di dati non classificati in precedenza.

- k-mean e clustering gerarchico
- Catene di Markov nascoste
- Reti neurali

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

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

È un problema convesso ed ammette ottimo che è unico in quanto la funzione obiettivo è strettamente convessa.

Condizioni KKT

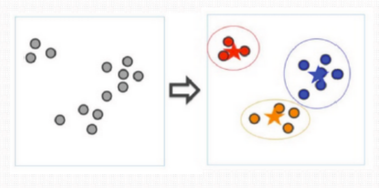
L'ottimo globale coincide con l'insieme delle soluzioni del sistema:

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

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

K-means

Algoritmo che permette di suddividere un insieme di dati di n osservazioni in K gruppi (cluster), individuando dei centroidi e minimizzando iterativamente la distanza delle osservazioni dai centroidi.



Algoritmo

per ogni elemento si ricalcola il centroide da associargli

ogni centroide si ricolloca in base alla media degli elementi ad esso associati.