

# Un problema di zaino (knapsack)

## - COME SCEGLIERE LE CANZONI DA MASTERIZZARE SUL CD -

### Problematica Riscontrata:

Mario vuole masterizzarsi un CD da mettersi in auto per le prossime vacanze. Ha fatto una lista delle canzoni che gli piacerebbe avere con se, ma la somma dei rispettivi ingombri (in Mb) eccede la capacità del CD. Si pone pertanto il problema di scegliere, tra questi preziosissimi brani, quali effettivamente masterizzare.

### Approcci quantitativi ed il percorso della modellazione matematica:

In realtà si possono spesso proporre e considerare diversi modi ed approcci per affrontare una stessa problematica, benchè semplice come quella di Mario. Badate bene: sto parlando di approcci tutti ragionevoli e di per sè coerenti, figurarsi poi gli altri (il mundu è bellu perchè avariu). Nello sforzo di formalizzarsi e rappresentarsi la situazione, per meglio decidere razionalmente, Mario ha già dei numeri che descrivono con grande precisione alcuni aspetti del suo problema: Mario ha un numero  $CD\_CAP$  che esprime la capacità del CD su cui masterizzerà le canzoni. Inoltre, per ogni brano  $b$  tra quelli individuati con la sua prima cernita, Mario conosce un ulteriore numero:  $INGOMBRO[b]$ . Possiamo pensare che tutti questi numeri siano degli interi, e siano tutti in nostro possesso, ossia forniti in input al processo di decisione. Tutti questi bei numeretti così precisi possono darci delirio di onnipotenza, illuderci che noi si abbia il controllo della situazione, ... e certo non possiamo esimerci dalla responsabilità di utilizzarli per una decisione responsabile!

Viene a questo punto molto naturale formulare e modellare la problematica di Mario secondo il seguente paio di lenti:

KNAPSACK 1

**INPUT:** Due numeri naturali  $n, B$  ed un insieme di  $n$  oggetti descritti ciascuno da un peso,  $p_i$  per ogni  $i = 1, \dots, n$ .

**OUTPUT:** Trovare un sottoinsieme  $S$  degli oggetti assegnati in input, a somma dei pesi non eccedente il budget assegnato  $B$ , e minimizzando lo scarto tra  $B$  e la somma dei pesi degli oggetti scelti (la parte di CD che non risulta utilizzata).

Questo è finalmente un problema ben definito, ossia un modello matematico. Si noti che nel modello matematico abbiamo preferito omettere di fare riferimento ai valori effettivi dei vari numeretti in gioco. È preferibile astrarre da questi in modo da consentire a Mario la possibilità di dirci che ha scoperto di nuove canzoni che gli piacciono, oppure se un domani, quando le capacità dei CD saranno raddoppiate, ci si dovesse ripresentare una simile situazione. Un problema o modello matematico è quindi una famiglia infinita di possibili istanze specifiche di problema concreto. Risolverlo significa fornire un metodo (algoritmo) grazie al quale noi si possa passare dalla specifica dei vari parametri ( $n, B$ , e tutti gli altri numeretti  $p_i$  per ogni  $i = 1, \dots, n$ ) che descrivono un'istanza specifica del problema, alla sua soluzione per quella specifica istanza (la scelta del sottoinsieme  $S$ ). Alla soluzione (scelta del sottoinsieme  $S$ ) chiediamo che sia ammissibile (ossia non si sfori la capacità  $B = CD\_DIM$  del CD) e che sia ottima (minimizzazione dello scarto). Si noti che il livello di generalità ed astrattezza del modello matematico cui siamo pervenuti lo rendono valido non solo per un numero infinito di istanze diverse, ma per un'infinità di applicazioni diverse, come ad esempio:

- la Banda Bassotti è entrata in un deposito di lingotti d'oro di zio Paperone, e trova lingotti di varie taglie (da chilo, da 3 etti, ...) e vuole portare via più oro che può,

sapendo però che il furgoncino di Nonno Bassotto non riuscirà a reggere un carico superiore ai 4,2 quintali;

- debbo ritagliare delle barre di prodotto finito da una barra di profilato;
- a Pierino la Peste piace fare dispetti alla maestra, ma oggi non può certo permettersi di farsi mandare un'altra volta dal direttore. Conosce perfettamente il grado di sopportazione della maestra, e ci sarebbero diversi scherzetti che avrebbe progettato tra cui scegliere oggi.

Questa è la ragione per la quale discipline come la Ricerca Operativa propongono e studiano modelli matematici, come il modello del knapsack non pesato (questo è il nome nella letteratura scientifica) cui siamo sopra pervenuti nel nostro sforzo di rappresentarci la problematica in essere per Mario.

A dire il vero, la ragione per cui sopra ci è venuto naturale considerare lo scarto del CD come la funzione obiettivo da minimizzarsi è sostanzialmente dovuta al fatto che questo era forse l'unico problema ben definito (modello matematico) che potessimo porci sulla base dei soli dati (numeretti) messi a nostra disposizione da Mario. Tuttavia è chiaro che non stiamo davvero rappresentando Mario e le sue vere esigenze se la nostra unica preoccupazione è quella di minimizzare lo scarto del CD. Entrare in questo, e certo dobbiamo farlo se ci preme la problematica di Mario, è tuttavia un bel casino, un vero e proprio vaso di Pandora, perchè potrebbe ad esempio emergere che Mario ambisca anche ad una certa diversificazione di generi, oppure che esibisca forme di gradimento per gruppi di canzoni più che non solo canzoni singole, o che vi siano altri vincoli inespressi come che almeno una canzone sia di genere "romantico". Diventa quindi nostra ulteriore responsabilità lo sforzo di mantenere un certo grado di semplicità nel modello che andremo a costruire, in modo che sia realistico e sensato, sia modellisticamente prima che computazionalmente dopo, affrontarne lo studio.

Vogliamo tener conto del fatto che, a Mario, certi brani sono più cari di altri, già questo introduce le sue belle difficoltà, non sarebbe sensato introdurne ulteriori in questa fase. Dobbiamo allora considerare che, per ogni brano, di numeri in input ne dobbiamo richiedere due (è una complicazione non banale perchè non so mica se il gradimento per una canzone si possa davvero esprimere con un numero, e, se optiamo per adottare questo secondo e più raffinato modello, dovremo poi chiedere a Mario di produrci, in qualche modo, questi numeri). Nonostante queste difficoltà di cui è bene restiamo consapevoli, il modello cui perveniamo è comunque, ancora una volta, un modello base molto importante della Ricerca Operativa:

#### KNAPSACK 2 (IL KNAPSACK PESATO)

**INPUT:** Due numeri naturali  $n, B$  ed un insieme di  $n$  oggetti descritti ciascuno da una coppia valore/peso,  $(v_i, p_i)$  per ogni  $i = 1, \dots, n$ .

**OUTPUT:** Trovare un sottoinsieme  $S$  degli oggetti assegnati in input, a somma dei pesi non eccedente il budget assegnato  $B$ , e massimizzando il valore totale raccolto.

Vi chiedo ora, come esercizio, di formulare come un modello di Programmazione Lineare Intera (PLI) questi due modelli classici della Ricerca Operativa. Siamo in un percorso guidato, nel documento `knapsack_mod.tex` forniremo una tale formulazione. Nelle directories `AMPL` e `GMPL` trovate la successiva traduzione di questi modelli di PLI in linguaggio AMPL e GMPL rispettivamente, ossia del codice che può essere utilizzato per trovare in automatico la soluzione ottima per la generica istanza di vostro interesse.