Pianificazione e Gestione Servizi Sanitari

Pianificazione e Gestione Sale Operatorie

Prof. Domenico Conforti – Lezione 24 – 22.11.23 – Autori/Revisionatori: Salvati e Luciani

VARIABILI DECISIONALI

Decisione di assegnamento di tipo tattico del blocco temporale al team della specialità chirurgica. L'obiettivo specifico è quello di integrare il livello tattico e quello operativo, quindi colorare ogni blocco temporale disponibile con il colore associato alle varie specialità chirurgiche.

x^s_{tb} = 1 se il team t £ T_s verrà assegnato al blocco b, altrimenti 0

Decisione di schedulazione dell'intervento nel blocco temporale. Decidere, quindi, quali interventi effettuare inserendoli all'interno del blocco temporale corrispondente all'assegnamento con la relativa specialità chirurgica.

y^s_{tpb} = 1 se il team t £ T_s andrà ad eseguire l'intervento chirurgico sul paziente p nella lista d'attesa della specialità chirurgica nel blocco b, altrimenti 0

Decisione di utilizzo del blocco di overtime, cioè la possibilità di poter utilizzare questo blocco di lavoro straordinario nel caso in cui si dovesse sforare la capacità del blocco temporale.

o^s_{tb} = 1 se il team t £ T_s utilizza il blocco di overtime associato al blocco b, altrimenti 0

Le variabili decisionali in questione sono tutte binarie.

FUNZIONE OBIETTIVO

Sulla base di queste variabili andremo a definire tre macro-obiettivi che si possono realizzare attraverso la specifica definizione di cinque funzioni obiettivo.

$$\max \sum_{\substack{t\in T_s \\ s\in S}} \sum_{p\in L_s} \sum_{b\in B} y^s_{tpb}$$

$$\max \sum_{\substack{t\in T_s \\ s\in S}} \sum_{\substack{p\in L_s \\ s\in S}} \sum_{b\in B} pr^s_p y^s_{tpb}$$

Minimization of waiting time of patients

$$\max \sum_{\substack{t\in T_s \\ s\in S}} \sum_{\substack{p\in L_s \\ s\in S}} \sum_{b\in B} w^s_p y^s_{tpb}$$

Maximization throughput of scheduled patients

$$\min \sum_{\substack{t\in T_s \\ s\in S}} \sum_{\substack{p\in L_s \\ s\in S}} \sum_{b\in B} (d_b x^s_{tb} + d^b_{ov} o^s_{tb}) - \sum_{s\in S} \sum_{t\in T_s} \sum_{\substack{p\in L_s \\ s\in S}} \sum_{b\in B} sd^s_p y^s_{tpb}$$

Resource efficiency maximization

$$\min \sum_{s\in S} \sum_{t\in T_s} \sum_{b\in B} C_{ov} o^s_{tb}$$

Minimization of overtime cost

1. Il **primo macro-obiettivo** (in azzurro) è legato al rendere efficace il sistema, in termini di qualità del servizio erogato e nei confronti del paziente in lista di attesa. Ciò è possibile garantendo ai pazienti in lista d'attesa di avere l'intervento giusto al momento giusto entro i termini giusti, rispettando condizioni quali priorità clinica e tempo massimo di permanenza in lista d'attesa. Questo macro-obiettivo viene implementato attraverso tre specifiche funzioni obiettivo, che sono tutte e tre da massimizzare.

    - La prima è relativa alla variabile decisionale y (schedulazione degli interventi) che stabilisce, assumendo valore 1, se l'intervento verrà schedulato nell'orizzonte temporale di pianificazione: per massimizzare questa somma totale bisogna far assumere alla y il valore 1 per tutte le combinazioni degli indici presenti.

    - La seconda riguarda la priorità clinica che viene tipicamente definita per ogni tipologia di paziente.

    - La terza riguarda il parametro ω_p relativamente alla definizione di quello specifico sottoinsieme di pazienti in lista d'attesa a cui, all'interno dell'orizzonte temporale corrente di pianificazione, sta scadendo il tempo massimo d'attesa.

    Il contributo di queste tre funzioni obiettivo, se le volessimo considerare in modo integrato, non è solo massimizzare l'efficacia complessiva del sistema ma anche minimizzare il tempo di attesa dei pazienti in modo da svuotare velocemente le liste.

2. Il **secondo macro-obiettivo** (in rosa) serve per garantire l'efficiente utilizzo delle risorse. Infatti, è necessario modulare il trade-off tra efficacia, efficienza ed equità in ambito sanitario: da una parte bisogna garantire qualità e servizi per i pazienti, dall'altra parte bisogna cercare di gestire complessivamente le risorse sanitarie nel modo migliore possibile. Questo si realizza sulla base di un utilizzo del blocco temporale della sala operatoria, che deve essere riempito il più possibile senza lasciare spazi vuoti.
Nella funzione obiettivo in questione il primo termine (primo membro della differenza) è la capacità massima prevista complessiva di tutti i blocchi temporali, considerando anche gli eventuali blocchi di overtime. In particolare, d_b è la durata di ogni singolo blocco temporale, x è la variabile decisionale di assegnamento e o è la variabile decisionale di utilizzo.
Nel secondo membro viene indicata la quantità della risorsa tempo che viene effettivamente utilizzata: qui troviamo la variabile decisionale y (di schedulazione) e ogni intervento è pesato dalla sua durata media stimata.
L'obiettivo, quindi, è minimizzare questa differenza.

3. Il **terzo macro-obiettivo** (in giallo) è relativo all'utilizzo del blocco di overtime, il cui costo viene definito dal coefficiente c_{ov} di costo o di penalità dell'overtime, il quale deve essere minimizzato. In parole povere, è possibile utilizzare il blocco di overtime ma senza eccedere.

La presenza di cinque funzioni obiettivo caratterizza il modello in questione come **multi-obiettivo**. In genere, un problema multi-obiettivo è valido quando gli obiettivi in gioco, considerato che fanno riferimento a differenti portatori di interessi con differenti priorità, tendono ad aiutarsi.

Tuttavia, questi obiettivi tra di loro sono abbastanza contrastanti e ciò fa sì che il relativo modello di ottimizzazione multi-obiettivo assuma un certo significato.

In queste situazioni, generalmente, si realizza una sorta di somma pesata tra le funzioni obiettivo per ottenerne una complessiva e ritornare alla classica formulazione del modello di ottimizzazione. Il problema, in questo caso, è decidere a quale delle cinque funzioni obiettivo dare maggiore priorità in termini di significato (che dipende dalle scelte specifiche del decision maker rispettando anche le caratteristiche specifiche del reparto chirurgico).

VINCOLI

inconsistent assignment

$$\sum_{b\in\mathcal{B}} y^s_{tpb} \le 1 \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s, p\in\mathcal{L}^s_t \qquad (2)$$

$$\sum_{b\in\mathcal{B}} x^s_{tb} \le 1 \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s, B\in\mathcal{B} \qquad (3)$$

$$\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}_s} x^s_{tb} \le 1 \qquad \forall b\in\mathcal{B} \qquad (4)$$

allotted OR time

$$\sum_{t\in\mathcal{T}_s}\sum_{b\in\mathcal{B}} (d_b x^s_{tb} + d^b_{ov} o^s_{tb}) \ge Tmin^s \qquad \forall s\in\mathcal{S} \qquad (5)$$

$$\sum_{t\in\mathcal{T}_s}\sum_{b\in\mathcal{B}} (d_b x^s_{tb} + d^b_{ov} o^s_{tb}) \le Tmax^s \qquad \forall s\in\mathcal{S} \qquad (6)$$

$$\sum_{b\in\mathcal{B}} (d_b x^s_{tb} + d^b_{ov} o^s_{tb}) \le Tmax^s_t \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s \qquad (7)$$

Scheduling of patients (capacity of OR blocks)

$$\sum_{p\in\mathcal{L}^s_t} sd^s_{pt} y^s_{tpb} \le d_b x^s_{tb} + d^b_{ov} o^s_{tb} \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} \qquad (8)$$

$$o^s_{tb} \le x^s_{tb} \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} \qquad (9)$$

$$y^s_{tpb} \in \{0,1\} \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s, p\in\mathcal{L}^s_t, b\in\mathcal{B} \qquad (10)$$

$$x^s_{tb} \in \{0,1\} \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} \qquad (11)$$

$$o^s_{tb} \in \{0,1\} \qquad \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} \qquad (12)$$

2. Nell'orizzonte temporale di pianificazione corrente (settimana) il paziente p può avere al più un solo intervento.

3. L'assegnamento inconsistente: ogni gruppo è formato a sua volta da blocchi temporalmente sovrapposti (inconsistenti). Ogni team chirurgico può essere assegnato al più ad un solo gruppo.

4. Il blocco temporale non può essere condiviso: ad ogni blocco temporale viene assegnato al più un unico team chirurgico.

5. Il tempo totale di durata dei blocchi temporali assegnati alla specialità chirurgica s, considerando anche l'eventuale blocco di overtime, deve essere maggiore o uguale di una soglia minima e minore o uguale di una soglia massima.

6. Per ogni team chirurgico, il tempo complessivo di intervento deve essere minore o uguale di una soglia massima stabilita (fissata dall'ospedale o da protocolli).

7. La capacità del blocco temporale deve essere rispettata: la somma totale delle durate di tutti gli interventi schedulati non deve eccedere la durata (=capacità) del blocco.

8. Il blocco temporale di overtime può essere utilizzato solo se assegnato. Se rimane vuoto (x = 0), la o non può essere uguale ad 1 (condizione logica).

Le righe 9, 10, 11 e 12 sono le condizioni sulle singole variabili.

La struttura complessiva del modello di ottimizzazione è la seguente:

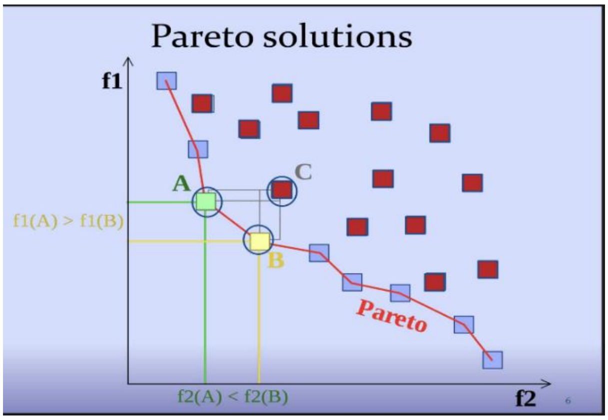
$$
\begin{aligned}
& opt\, f(x) = [f_1(x), f_2(x), f_3(x), f_4(x), f_5(x)] = \\
& = \max \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} \sum_{p\in\mathcal{L}_t^s} \sum_{b\in\mathcal{B}} y_{tpb}^s \qquad (1)\\
& \max \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} \sum_{p\in\mathcal{L}_t^s} \sum_{b\in\mathcal{B}} pr_p^s y_{tpb}^s \\
& \max \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} \sum_{p\in\mathcal{L}_t^s} \sum_{b\in\mathcal{B}} w_p^s y_{tpb}^s \\
& \min \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} \sum_{b\in\mathcal{B}} (d_b x_{tb}^s + d_{ov}^b o_{tb}^s) - \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} \sum_{p\in\mathcal{L}_t^s} \sum_{b\in\mathcal{B}} sd_p^s y_{tpb}^s \\
& \min \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} \sum_{b\in\mathcal{B}} C_{ov} o_{tb}^s
\end{aligned}
$$

$$
\begin{array}{llrr}
\sum_{b\in\mathcal{B}} y_{tpb}^s & \le 1 & \forall s\in\mathcal{S}, t\in\mathcal{T}_s, p\in\mathcal{L}_t^s & (2)\\
\sum_{b\in\mathcal{B}} x_{tb}^s & \le 1 & \forall s\in\mathcal{S}, t\in\mathcal{T}_s, B\in\mathcal{B} & (3)\\
\sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} x_{tb}^s & \le 1 & \forall b\in\mathcal{B} & (4)\\
\sum_{t\in\mathcal{T}_s} \sum_{b\in\mathcal{B}} (d_b x_{tb}^s + d_{ov}^b o_{tb}^s) & \ge Tmin^s & \forall s\in\mathcal{S} & (5)\\
\sum_{t\in\mathcal{T}_s} \sum_{b\in\mathcal{B}} (d_b x_{tb}^s + d_{ov}^b o_{tb}^s) & \le Tmax^s & \forall s\in\mathcal{S} & (6)\\
\sum_{b\in\mathcal{B}} (d_b x_{tb}^s + d_{ov}^b o_{tb}^s) & \le Tmax_t^s & \forall s\in\mathcal{S}, t\in\mathcal{T}_s & (7)\\
\sum_{p\in\mathcal{L}_t^s} sd_{pt}^s y_{tpb}^s & \le d_b x_{tb}^s + d_{ov}^b o_{tb}^s & \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} & (8)\\
o_{tb}^s & \le x_{tb}^s & \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} & (9)\\
y_{tpb}^s \in \{0,1\} & & \forall s\in\mathcal{S}, t\in\mathcal{T}_s, p\in\mathcal{L}_t^s, b\in\mathcal{B} & (10)\\
x_{tb}^s \in \{0,1\} & & \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} & (11)\\
o_{tb}^s \in \{0,1\} & & \forall s\in\mathcal{S}, t\in\mathcal{T}_s, b\in\mathcal{B} & (12)
\end{array}
$$

SOLUZIONE DI PARETO

In un modello decisionale di questo tipo, in cui si hanno più funzioni obiettivo, non si avrà mai una soluzione univoca: avendo 5 funzioni obiettivo, trovare una configurazione delle alternative che consenta di massimizzare il valore di una di queste potrebbe renderne non significativa qualche altra. L'insieme delle decisioni da prendere può esser visto come una coperta corta: andando a coprire bene in qualche punto (andando a massimizzare il valore di una singola funzione obiettivo) si rischia di scoprire in qualche altro punto (cioè di trovare un valore in un’altra funzione obiettivo che non vada a massimizzare quella determinata decisione).

Quindi in questo tipo di modello decisionale si deve trovare la configurazione più equilibrata possibile per ottenere il miglior valore per ogni funzione obiettivo: questa configurazione viene chiamata **soluzione di Pareto** (economista e matematico). Le soluzioni di Pareto vengono rappresentate sull’asse cartesiano.

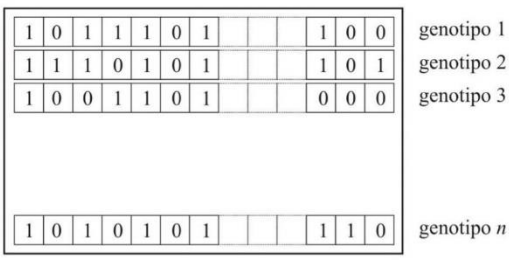


In questa immagine, ogni quadratino viene inteso come una istanza di variabile decisionale e per ognuno di essi si ha un valore di funzione obiettivo f1 (asse delle y) e un valore di funzione obiettivo f2 (asse delle x). Su questo grafico si deve andare a definire la cosiddetta **frontiera di Pareto**, cioè quella riga che indica che configurazioni migliori di quelle che si trovano al di fuori della frontiera non se ne trovano. Considerando l’esempio nell’immagine, si vede come la soluzione indicata con A sia migliore di quella indicata con C per le due funzioni obiettivo f1 ed f2 considerate. Questa frontiera mette a disposizione per l’operatore sanitario più scelte possibili e sta a lui decidere quale attuare in base al contesto in cui si trova.

ALGORITMI GENETICI

Per trovare la frontiera di Pareto si deve esplorare uno spazio di ricerca e determinare questa configurazione migliore: per fare ciò si possono usare degli algoritmi euristici, cioè algoritmi che non danno garanzia che arrivino esattamente alla soluzione ma che danno una soluzione che si avvicini il più possibile alla frontiera di Pareto. Uno degli algoritmi più usati è l’**Algoritmo Genetico (**AG). Gli AG sono algoritmi di ricerca che si ispirano ai meccanismi della selezione naturale e della riproduzione sessuata. Gli AG simulano l’evoluzione di una popolazione di individui, che rappresentano soluzioni candidate di uno specifico problema, favorendo la sopravvivenza e la riproduzione dei migliori (modello Darwiniano), per ottenere generazioni successive della popolazione che si spera siano migliori della generazione precedente, in modo tale da arrivare, generazione per generazione, ad una configurazione della popolazione che ci consenta di individuare la frontiera di Pareto.

Popolazione *P* dell’AG



Ogni individuo (che nel problema precedente è rappresentato da un quadratino, con l’insieme dei quadratini che costituisce la popolazione P) viene codificato come una stringa n di bit di lunghezza l fissata. Ogni stringa è composta da celle in cui vi sono inseriti i bit (0 e 1).

Per valutare la bontà di una generazione di individui si calcola la cosiddetta **funzione di fitness**: essa misura la bontà degli individui g_i della popolazione P nel risolvere il problema di ricerca dato.

Nel caso del modello decisionale trattato, la funzione di fitness misura il valore della funzione obiettivo.

L'insieme delle stringhe binarie di lunghezza l ha 2^l elementi; tale insieme rappresenta lo **spazio di ricerca** dell'AG, cioè lo spazio che l'AG deve esplorare per risolvere il problema di ricerca. I valori di fitness sui punti dello spazio di ricerca è detto **paesaggio d'idoneità.**

Se si considerano ad esempio delle stringhe binarie di lunghezza 2, il numero di stringhe è $2^2=4$. In questo caso lo spazio di ricerca S dell'AG è:

S= {(0,0) ; (0,1) ; (1,0) ; (1,1)}

I valori di fitness sui punti di S definiscono il paesaggio d'idoneità dell'AG, cioè il valore della funzione di fitness.

OPERATORI DELLA FUNZIONE DI FITNESS

Una volta che la funzione di fitness ha determinato il valore di bontà di ogni individuo della popolazione, una nuova popolazione di individui viene generata applicando alcuni operatori che si ispirano alla selezione naturale e alla genetica. Per passare alla generazione successiva abbiamo 3 operatori:

- **Selezione:** ispirato alla selezione naturale
- **Crossover**: ispirato alla genetica
- **Mutazione**: ispirato alla genetica

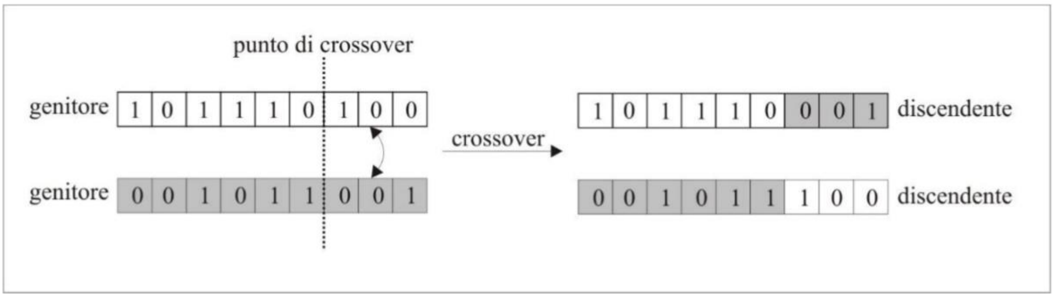
Selezione

L'operatore di **selezione** si ispira alla selezione Darwiniana, che sostiene che gli individui più forti abbiano maggiori probabilità di sopravvivere nell'ambiente in cui vivono, e quindi hanno maggiore probabilità di riprodursi. Nel contesto dell'AG, gli individui più forti sono quelli con fitness più alta, poiché risolvono meglio di altri il problema di ricerca dato; per questo essi devono essere privilegiati nella fase di selezione degli individui che potranno riprodursi dando luogo a nuovi individui.

Per quanto riguarda gli operatori di crossover e di mutazione, si va a creare un pool di accoppiamento, detto **mating pool**. Ogni volta che un individuo della popolazione è selezionato ne viene creata una copia e tale copia viene inserita nel mating pool. Quando il mating pool è riempito con esattamente n (numero di individui della popolazione) copie di individui della popolazione, nuovi n discendenti vengono generati applicando gli operatori genetici.

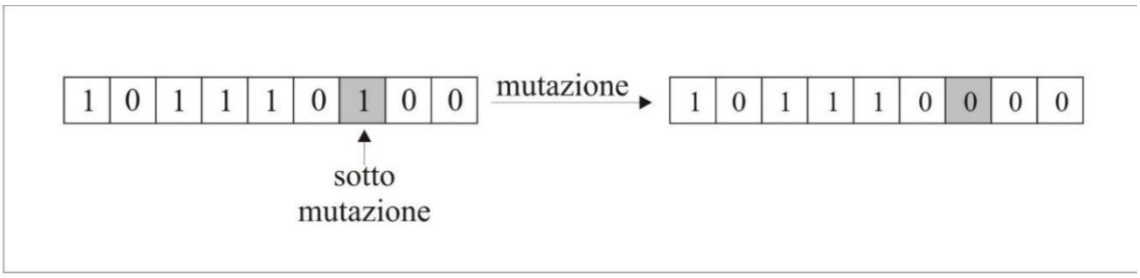
Crossover

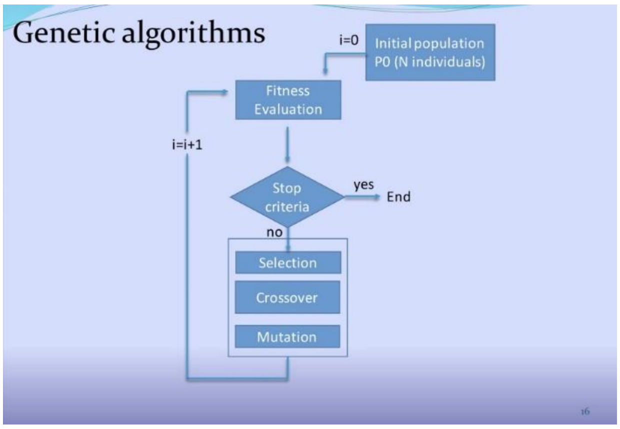
Per effettuare il crossover si scelgono a caso due individui (due stringhe) nel mating pool (che chiameremo genitori) e un punto di taglio, detto **punto di crossover**. Le porzioni di stringa che si trovano a destra rispetto al punto di crossover vengono scambiate tra i due genitori, generando due discendenti. Per la scelta del punto di crossover ci sono delle tecniche empiriche che consentono di indicare il punto di crossover più efficace per la popolazione che si sta studiando.



Mutazione

Per quanto riguarda la mutazione, una volta che due discendenti sono stati generati per mezzo del crossover, il valore dei bit dei nuovi individui è cambiato da 0 a 1 e viceversa.





In questa immagine è rappresentato lo schema complessivo dell'AG. Si ha una popolazione iniziale P0, composta da N individui. Per valutare la bontà di questa popolazione si va a valutare la funzione di fitness. L'esito della funzione di fitness ci permette di valutare l'eventuale criterio di stop: se sulla base del valore della funzione di fitness la popolazione P0 è buona, allora ci si ferma e si determina la frontiera di Pareto, altrimenti si utilizzano i tre operatori di selezione, crossover e mutazione per generare una nuova popolazione e valutare che sia buona. Una volta effettuate queste operazioni, si aumenta di 1 il valore dell'iterazione i e si ritorna a valutare il valore della funzione di fitness, ricominciando il ciclo.
Soltanto nel momento in cui il valore della funzione di fitness è buono si può uscire da questo ciclo.