Tema 7: Algorismes de Cerca

Resum del tema 7

- Cerca línealCerca binària

- Cerca exhaustivaCerca aleatòriaCerca amb <u>algorisn</u>

- concepte de generació
 funció d'adaptació
 mutació i creuament
 selecció supervivents
 mètode estàndard
 mètode d'ordenació
 mètode diversitat
- - Distància de Hamming

Algorismes de cerca

El concepte de cerca inclou diferents idees:

- Cerca d'un determinat element en una llista (màx, x="a", el que compleix una certa condició, etc.).
 Cerca d'un determinat element en una llista ordenada.

- Cerca en un arbre. Cerca en un graf. Satisfacció de restriccions.

Ens centrarem en la cerca en llistes.

Algorismes de cerca: cerca lineal

força bruta es diu **cerca seqüencial o lineal**. La complexitat de l'algorisme és 0(n) en el L'algorisme que implementa la cerca d'un element en una llista amb una estratègia de pitjor cas!

```
i=0
while i < len(list) and list[i] != ele:
    i += 1
if i < len(list):
    return i</pre>
def linsearch(list,ele):
```

Algorismes de cerca: cerca lineal (millorat)

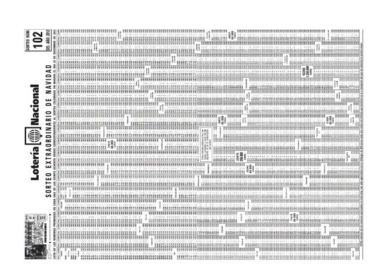
Podem fer una petita millora si afegim l'element que busquem al final de la llista:

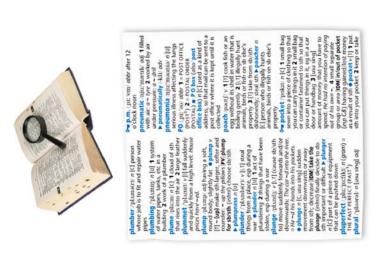
```
while list[i] != ele:
    i += 1
if i < len(list)-1:
    return i</pre>
def linsearch(list,ele):
    list.append(ele)
    i=0
                                                                                           else:
return -1
```

Recordeu de que list.append(ele) té una complexitat 0(1) i per això surt a compte fer-

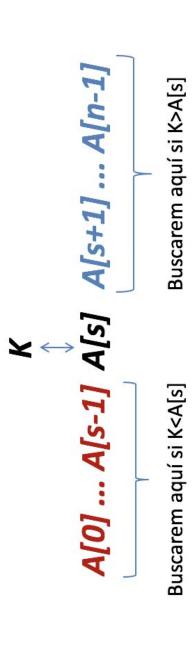
Si enlloc de ser 0(1) fos 0(n) (com per exemple list.insert(ele)) ja no valdria la pena.

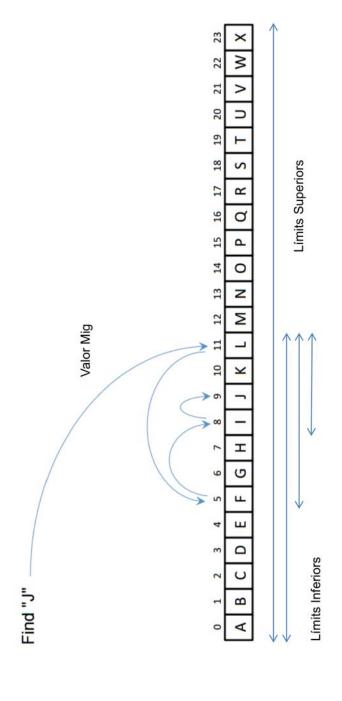
I si la llista està ordenada (un diccionari, els nombres de la loteria, etc.) ho podem fer millor?





La **cerca binària** ho fa comparant l'element cercat K a l'element central de la llista: si hi ha correspondència ja l'hem trobat, sinó, busquem a la subllista (de la dreta o de l'esquerra) que correspon.





Com acaba l'algorisme si J no hi és?

Anem a veure com funcionaria per K=70.

l (low) indica el límit inferior de la subllista a buscar, h (high) el límit superior, m (middle) el punt de comparació.

Índex	0	1	2	3	4	2	9	7	8	6	10	11	12
Valor	က	14	27	31	39	42	22	70	74	81	85	93	98
lt1	-						Ε						4
lt 2								-			Ε		_
lt3								-	Ε	٩			
It 4								l,m,h					

Per analitzar la seva complexitat calcularem el nombre de vegades que la clau de la cerca, K, es compara amb un element de la llista. En el pitjor dels casos (quan l'element no hi és), tenim aquesta relació de recurrència: T(n) = T(n/2) + 1.

Segons el teorema Màster això és 0(log_2 n): per una llista de 1.000.000 elements són 20 comparacions!

Evidentment és un algorisme recursiu, però és pot implementar fàcilment de forma no

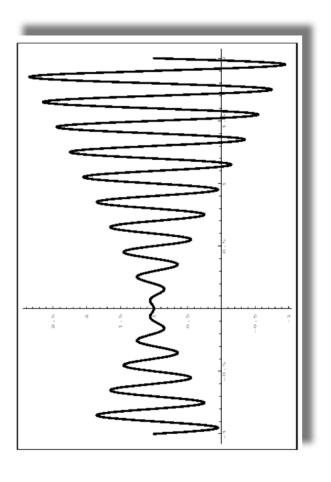
```
high = len(nums)-1
while low <= high:
    mid = (low + high) // 2
if nums[mid] > K: high = mid - 1
elif nums[mid] < K: low = mid + 1
else: return mid</pre>
def binsearch(nums, K):
```

El cas promig és més difícil d'analitzar, però es pot demostrar que és només una mica millor que el pitjor cas (tot i que del mateix ordre).

un element (o pocs), apliquem una cerca exhaustiva. Però si hem de fer moltes Observació: Si tenim una llista desordenada de mida n i només hem de buscar cerques (de l'ordre de n), val la pena ordenar-la primer i fer cerca binària dels elements després! En general considerem que **la complexitat de la cerca binària sobre una llista** ordenada és de 0(log_2 n).

Imaginem ara que tenim un vector no ordenat, com pot ser el corresponent als valors discrets d'una funció multimodal.

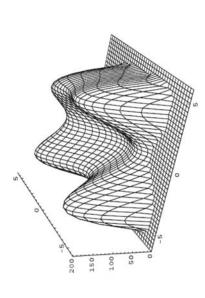
Com busquem el màxim? Quin tipus de cerca hi podem aplicar?



Podem aplicar-hi cerca exhaustiva (de complexitat 0(n), on n és el nombre de x de la

```
def linsearchfunc1d():
    x=0.0 # inicialització punt x, corresponent al valor màxim de func1d
    maxim=0.0 # inicialització valor màxim
    for i in frange(-1.0,2.0,0.01):
    if func1d(i)>maxim:
        maxim=func1d(i)
                                                                     def func1d(x): # funció multimodal de la que buscarem max
                                                                                                    import math
y = x * math.sin(10*math.pi*(x))+1.0
                                                                                                                                                                                                                def frange(start, stop, step):
                                                                                                                                                                                                                                                                           current = start
while current < stop:
    yield current</pre>
                                                                                                                                                                                                                                                                                                                                                                current += step
funció que conisderem):
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           print(maxim)
```

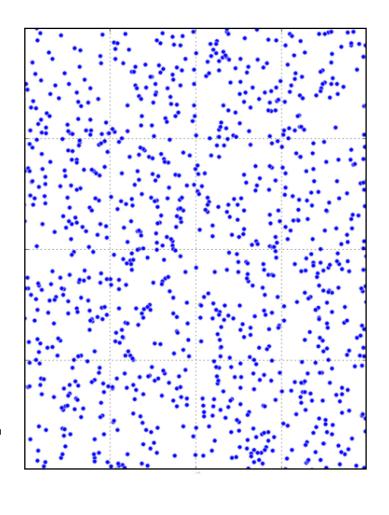
Si el nombre de punts a mostrejar és molt gran tenim un problema!



En aquest cas si analitzem els punts amb una precisió de 0.01 analitzaríem 300×300 punts. Si necessitéssim una precisió 0.000001 analitzaríem 3000000×3000000 punts.

No podem fer cerca exhaustiva!

Té sentit fer una cerca aleatòria? (= anar generant nombres de forma aleatòria dins del rang de les variables i quedar-se el màxim).



La cerca aleatòria és:

```
maxin=0.0
for i in range(1000):
    xtemp = (random.random()*3.0)-1.0
if func1d(xtemp)>maxim:
                                                                                                                             maxim=func1d(xtemp)
def rsearchfunc1d():
                                                                                                                                                x=xtemp
                    import random
```

En general, la cerca totalment aleatòria no és una bona solució: tenim el cost de la cerca afitat, però depèn molt de l'aleatorietat i té un resultat molt semblant, si no equivalent, a la cerca lineal, per força bruta.

Anem a veure un tipus d'algorisme aproximat que ens fa una cerca, amb un cert component aleatori més intel·ligent, de l'espai de solucions: la cerca basada en algorismes genètics. El terme algorismes genètics s'utilitza per a referir-se a una família bastant àmplia de models computacionals de càlcul basats en els mecanismes d'evolució biològica.

La idea de selecció natural va ser introduïda per Charles Darwin el 1859 dins del seu llibre L'Origen de Les Espècies.

Aquesta idea pot servir d'analogia per a construir mètodes de cerca en problemes d'optimització combinatòria i mètodes d'aprenentatge.

Darwin va assentar les bases del principi d'evolució per selecció natural amb les següents idees:

- Cada individu tendeix a passar els seus trets característics a la seva descendència.
- Tot i així, la natura produeix individus amb trets diferents.
- Els individus més adaptats tendeixen a tenir més descendència, i a la llarga, la població tendeix a ser "millor".
- Al cap d'un llarg període, l'acumulació de canvis pot produir espècies totalment noves, adaptades al seu entorn.

Darwin va assentar les bases del principi d'evolució per selecció natural amb les següents idees:

- Cada individu tendeix a passar els seus trets característics a la seva descendència.
- Tot i així, la natura produeix individus amb trets diferents.
- Els individus més adaptats tendeixen a tenir més descendència, i a la llarga, la població tendeix a ser "millor".
- Al cap d'un llarg període, l'acumulació de canvis pot produir espècies totalment noves, adaptades al seu entorn.

A més a més la natura disposa d'una sèrie de mecanismes reguladors externs a aquest procés però igualment interessants: el mecanisme de diversitat, els paràsits, les organitzacions socials, etc.

Els mecanismes biològics que fan possible l'evolució són avui coneguts.

A la natura, podem veure com la transmissió de la informació genètica (genoma) es fa a través de la reproducció sexual. Aquest procediment permet als descendents ser diferents dels seus antecessors, tot i que conservant la majoria de trets.

consisteix en l'aparellament de cromosomes (lloc on trobem el genoma), l'intercanvi El mecanisme sobre el que està basada la reproducció es troba a nivell molecular, i d'informació, i la posterior partició. D'això n'hi direm creuament.

La probabilitat de que dos individus es creuin depèn de la seva adaptació al medi.

Per inspiració d'aquests mecanismes usarem terminologia de biologia per als nostre problemes:

- Gens
- Genoma
- Cromosomes
- Creuaments i mutacions. Funció d'adaptació.
- Mecanismes correctors/moduladors: diversitat, parasitisme, organització social,

El cicle normal d'un algorisme genètic és:

- avaluar l'adaptació de tots els individus de la població (amb la funció d'adaptació). Aquesta funció incorpora l'objectiu del problema.
- crear una nova població mitjançant reproducció fent servir:
- creuament
- d'assegurar que el resultat d'aplicar els operadors genera possibles solucions o mutació dels cromosomes dels individus + descartar la població antiga (hem al problema.Ĵ
- iterar sobre la nova població.

Cada una de les iteracions d'aquest cicle es coneix com a generació.

Quan dissenyem un algorisme genètic per resoldre un problema caldrà decidir algunes qüestions:

- Quina és la funció d'adaptació?
- Com representarem els individus/solucions?
- Com seleccionarem els individus per reproduir-se?
 - Com creuarem i mutarem els individus?
 - Quina és la probabilitat de mutació?
- Necessitem mecanismes moduladors (p.e. diversitat)?

La funció d'adaptació és pròpia de cada problema que volem resoldre.

En el problema que hem posat com a exemple (la funció multimodal), la funció d'adaptació és el valor de f(x): (math.sin(10*math.pi*(x))+1.0).

Per tant, la màxima adaptació correspon al màxim d'aquesta funció multimodal.

La **funció d'adaptació** és pròpia de cada problema que volem resoldre.

En el problema que hem posat com a exemple (la funció multimodal), la funció d'adaptació és el valor de f(x): (math.sin(10*math.pi*(x))+1.0).

Per tant, la màxima adaptació correspon al màxim d'aquesta funció multimodal. El problema del viatjant de comerç també és un problema candidat a ser resolt amb algorismes genètics.

La funció d'adaptació seria 1/d, on d és la distància recorreguda (i així un valor de la funció alt és una bona solució).

Un cromosoma representaria un circuit que és potencialment solució del

Normalment es considera que la millor **representació** dels cromosomes possible és la binària. El creuament, la mutació, i d'altres operacions que es poden utilitzar, són aleshores simples operacions a nivell de bit.

Suposem doncs que tenim una població inicial de quatre individus amb les següents característiques:

individu valor adaptació possibilitat de selecció

OUUTTOUTTI	×	27%
111010101100	9	24%
001110101001	9	24%
111011011100	נכ	20%

I suposem que definim la probabilitat de selecció en funció del valor d'adaptació (q)

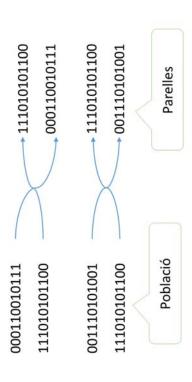
$$f_i = rac{q_i}{\sum_j q_j}$$

Com els seleccionem i els creuem?

Una primera alternativa per la selecció és triar parelles aleatòriament tenint en compte la seva probabilitat de selecció.

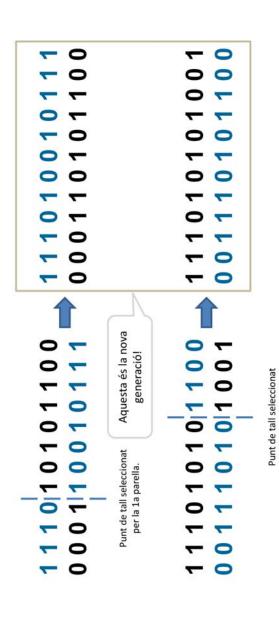
És com si fem rodar una ruleta i ens va donant individus; l'individu amb més probabilitat de selecció ocupa més posicions a la ruleta.

Imaginem que la selecció ens ha donat aquestes dues parelles:



I ara, com les creuem?

La forma més simple de creuament és generar un punt de tall aleatòriament i intercanviar:



Per a mutar-los canviarem el valor d'un quants bits de la població de forma aleatòria.

La probabilitat de que un bit canviï de valor és β i la que probabilitat de no canviar és (1- β), però sempre β < (1- β).

En resum, la **funció d'adaptació** depèn del problema que volem resoldre.

- La representació òptima, és en la majoria de casos i si no hi ha motius fonamentats per dubtar-ho, la binària.
 - La representació ha facilitar que el resultat d'aplicar els operadors genètics sigui

L'operació de **creuament** crea dos nous individus seleccionant punts de creuament en els cromosomes seleccionats i intercanviant les seves parts.

cromosoma i el canvi del seu valor. La probabilitat de mutació ha de ser petita (si no ho L'operació de **mutació** consisteix en la selecció aleatòria d'algun dels gens del convertim en cerca aleatòria!).

Probabilitat de supervivència:

- El valor d'adaptació de cada individu depèn del problema concret.
- valor d'adaptació, i es pot fer de diverses maneres: el mètode estàndard (que ja La probabilitat de supervivència a la següent generació és una ponderació del hem vist), el mètode d'ordenació, el mètode de la diversitat, etc.
- A partir d'ara suposarem que la generació següent es forma a partir d'una selecció cromosomes passen automàticament (així assegurem que una bona solució no es cromosomes descendents, però seguint una **estratègia elitista**: el(s) millor(s) entre els elements del conjunt format pels cromosomes progenitors i pels

Ponderació del valor d'adaptació.

solució al problema en qüestió. Com a resultat obté un valor d'adaptació q. Llavors Mètode Estàndard. Donat un cromosoma i, aquest és avaluat com a possible definim la seva probabilitat de selecció com

$$f_i = rac{q_i}{\sum_j q_j}$$

ividu valor adaptació possibilitat de selecció

8 32%	6 24%	6 24%	5 20%
000110010111	111010101100	0011101010001	111011011100

Un dels inconvenients associat al mètode estàndard és el poc pes que dóna al qualitativa: ordena de forma correcta però els seus valors no són precisos. generacions, i per tant, transmetre les poques coses que tinguin bones. Un altre possible inconvenient és que moltes vegades la funció d'avaluació és cromosomes "dolents", fet que els impedeix de passar a les futures

- d'adaptació del problema, podem ordenar els cromosomes segons aquest valor, i Mètode d'Ordenació: Per insensibilitzar el mètode de selecció respecte al valor els ponderem segons aquesta regla:
- Establim un valor aleatori p entre 0 i 1.
- Al primer cromosoma li assignem aquesta probabilitat
 - Al segon li assignem la probabilitat p*(1-p)
- La probabilitat de l'i-èssim cromosoma és p * (1 la probabilitat que s'hagi triat algun cromosoma anterior).

Per exemple, suposem que p=0.667. Llavors:

0.333 = 1 - (0.667) és la probabilitat de que no hagi sortit el primer cromosoma.

Ordre Prob M. Estànd. Prob M. Ordenació	0.667	$0.222 = 0.667 \times 0.333$ $0.073 = 0.667 \times 0.111$	0.025=	0.012	0.111 = 1 – (0.667 + 0.222) és la probabilitat de que no hagi sortit ni el primer ni el segon cromosoma.
Prob M. Estànd.	0.22	0.15 0.125	0.075	0.0	0.111 = 1 – (0.667 + (
Ordre	H (7 %	4	2	m m
a _i	44	32 22,5	1,5	0	sts cromosom del pla i este màxim
Crom (x,y)	0001,0100	00011,0001	0001,0001	0111,0101	Imaginem que aquests cromosomes representen punts del pla i estem buscant el màxim

• Ponderació del valor d'adaptació: Mètode de Diversitat. Aquest mètode es basa en l'anomenat principi de diversitat: és quasi tan bo ser diferent com estar adaptat.

Definim la diversitat d'un grup de cromosomes com:



on d_i és una mesura de distància entre cromosomes.

Distància de Hamming

nombre de bits que s'han de canviar d'una cadena perquè passi a tenir el valor La distància de Hamming entre dues cadenes de la mateixa longitud és el nombre de posicions diferents. Si considerem cadenes de bits, correspon al d'una altra cadena.

Exemple: la distància de Hamming entre 011101 i 011011 és de 2 ja que per arribar de la primera cadena a la segona cal canviar 2 bits

Ponderació del valor d'adaptació: Mètode de Diversitat.

Com l'apliquem?

- El millor cromosoma passa automàticament a la següent generació (estratègia
- Calculem la diversitat de tots els cromosomes respecte als que han passat a la + següent generació.
- Ordenem els cromosomes segons la seva funció d'avaluació.
- Sumem els nombres que representen l'ordre obtingut per cada cromosoma als passos 2 i 3. I reordenem segons aquest valor.
 - Triem el cromosoma que passa a la següent generació segons el mètode d'ordenació i si queden cromosomes per triar, i tornem al punt 2.

Ponderació del valor d'adaptació: Mètode de Diversitat.

Exemple:

q	100	44	32	22,5	1	0
Cromosomes q _i	0100 0001	0001 0100	0011 0001	0001 0010	0001 0001	0111 0101

El cromosoma millor passa a la següent generació. En el nostre cas és el cromosoma $(0100\ 0001).$

Per tant resten per triar 2 cromosomes entre (0001 0100), (0011 0001), (0001 0010), (0001 0001), (0111 0101).

Construirem la taula segons el mètode de diversitat segons la diversitat de cada cromosoma amb respecte al que ja ha passat:

Cromosomes	diversitat ord div	ord div		ord est div+ord Probab.	Probab.
0001 0100	0.040	\vdash	1	2	0.667
0011 0001	0.250	2	2	7	0.073
0001 0010	0.059	3	က	9	0.222
0001 0001	0.062	4	4	∞	0.012
0111 0101	0.050	2	2	7	0.025

En aquest cas, la diversitat és només la inversa de la distància euclidiana 2D de cada cromosoma al que ja ha passat.

La distància que usem pot ser des del nombre de bits diferents entre cada cromosoma a una funció definida per l'usuari a partir del coneixement del problema.

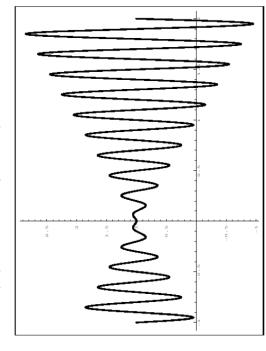
0100). A partir d'aquest moment repetim el procés anterior, però calculant la diversitat Llavors triem aleatòriament el següent que passa, i resulta que és el cromosoma (0001 respecte al dos cromosomes que ja han passat:

Prob	0.667	0.222	0.073	0.025
ord est div+ord	Ŋ	2	2	2
ord est	Н	2	က	4
ord div	4	က	2	Н
Cromosomes diversitat ord div	0.327	0.309	0.173	0.077
somes	00110001	00010010	00010001	01110101

En el cas d'empats per a l'ordenació resultat de la suma div+ord desempatem segons el valor d'ordenació pura.

El problema és trobar la x dins del rang [-1 .. 2] que maximitza f:

$$f(x) = x \sin(10\pi x) + 1.0$$



- Utilitzarem un vector binari com a cromosoma per a representar el valor real de la variable x. La longitud del vector dependrà del domini i la precisió.
- En el cas estudiat, el domini de la variable \times és [-1,2], té longitud 3.
- necessitem mostrejar el rang en 3.000.000 posicions, o sigui, 22 bits: 2.097.152 Suposem que volem 6 decimals (1.000.000 de valors per cada unitat). Per tant, $2^{2}1 < 3.000.000 < 2^{2}2 = 4.194.304.$

La transformació d'una seqüència binària [b_21,...,b_0] a un nombre real x es fa en dos passos:

- Primer convertim la seqüència de base 2 a base 10.
- Després trobem el nombre real corresponent:

$$x = -1.0 + x' \frac{3}{2^{22} - 1}$$

on -1.0 és el límit esquerra de l'interval i 3 la longitud.

L'algorisme te els següents passos:

• Escollim com a població inicial 50 individus de forma aleatòria:

```
# Generem una poblacio de n cromosomes de longitud long.
pop = [[0] * long for x in range(n)]
for i in range(n):
   for j in range(long):
   if random.random()>0.5: pop[i][j] += 1
def initpop(n,long):
   import random
                                                                                                                                                                                                                        return pop
```

• La funció d'avaluació serà equivalent a la funció del gràfic f:

```
def cost(r):
   import math
    # Transformem els bits en un valor real a l'interval [-1,2]
                                                                     sum=0.0
for i in range(len(r)):
    sum = sum + r[i]*(2**i)
    x = -1.0 + sum * (3.0/(2.0**(len(r))-1.0))
    # Avaluem el cromosoma
    y = x * math.sin(10*math.pi*(x))+1.0
return y
```

Per fer la mutació imposem una probabilitat de mutació pm = 0.01 per a cada bit.

Per exemple,

- si tenim el cromosoma $\sqrt{3}$ = (1110000000111111000101), que té com a valor f(x3) 2.250650
- seleccionem el cinquè bit per mutar,
- obtindrem v3' = (1110100000111111000101).
- el nou cromosoma representa el valor $\times 3' = 1.721638$, i per tant $f(\times 3') = 2.343555$
- f(x3') s'ha incrementat respecte f(x3).

```
def mutacio(r,mutprob):
    import random
    for i in range(len(r)):
    if random.random() < mutprob:
        if r[i]==0: r[i]=1
        else: r[i]=0</pre>
```

Per al creuament d'una parella de cromosomes escollirem aleatòriament un punt de tall i intercanviarem informació per crear els dos descendents.

```
def creuament(r1,r2):
    import random
    i=random.randint(1,len(r1)-2)
    return r1[::]+r2[i:],r1[i:]+r2[::]
```

Si iterem l'algorisme 150 generacions trobem que el millor cromosoma és vªax = (1111001101000100000101), que correspon al valor xmax = 1.850773. L'evolució de l'algorisme es pot avaluar a partir del millor valor de la funció aconseguit en cada generació:

```
2.344251
2.733893
2.849246
      2.250283
2.250284
2.250363
2.150003
                              2.328077
                       10
12
```

Observació vàlida per a qualsevol problema de recerca amb algorismes genètics:

Quin és el factor que més pesa en el càlcul de la complexitat computacional de l'algorisme? El nombre d'avaluacions!

En el nostre problema hem fet 50×150=7.500 avaluacions!

complexitat de l'algorisme és el nombre d'avaluacions per la complexitat de Els operadors genètics tenen un cost computacional nul, per tant la

Possibles preguntes d'exàmen relacionades amb el tema 7

- 1. Quan té sentit aplicar algorismes genètics?
- 2. Tens aquests dos cromosomes: creua'ls i muta'ls.
- 0001110001
- 0 1100110011
- 3. Tens els següents cromosomes. Quins sobreviuran a la següent generació? perquè?
- 0001110000
- 0 1100110011
 - 100011000111011110001
- 11111110111
- 01000110010
- 4. Tens una llista ordenada amb 32 números i busques si conté el número 5. Quantes comparacions hauràs de fer com a màxim?
- 5. Tens una llista no necessàriament ordenada amb 32 números i busques si conté el número 5. Quin cost té en el cas pitjor? surt a compte ordenar-la?