

Genetický algoritmus

Vladimír Kvasnička
Katedra matematiky
Chemickotechnologická fakulta STU
812 35 Bratislava
email: kvasnic@cvt.stuba.sk



Stretnutie so študentmi Gymnázia z Groesslingovej
ul., ktoré sa uskutočnilo dňa 13. Novembra 1998 na
CHTF STU

Obsah prednášky

1. Čo je DNA – biomakromolekula, ktorá je nosičom genetickej informácie
2. Čo je Darwinova evolúcia – princíp prirodzeného výberu
3. Formalizácia (algoritmizácia) Darwinovej evolúcie
4. Čo je genetický algoritmus – prenos základných predstáv Darwinovej evolúcie do informatiky (matematiky)
5. Prečo genetický algoritmus poskytuje rozumné výsledky?
6. Záver prednášky

1. DNA

Experimentálne skutočnosti

- každej bunke organizmu je obsiahnutá **úplná genetická informácia** daného organizmu.
- Genetická informácia je obsiahnutá vo forme **chromozómu** (chromozómov) v bunečnom jadre.
- Chromozóm je realizovaný biomakromolekulou nazývanou **DNA** (deoxyribonukleová kyselina).

3D štruktúra DNA

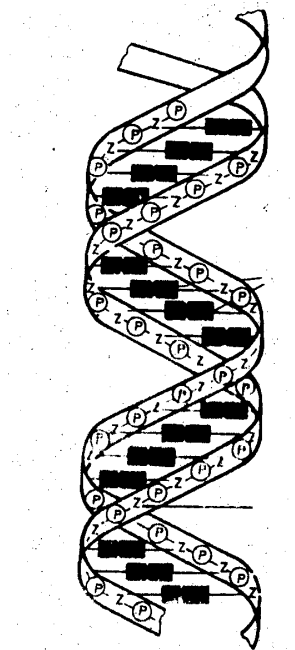
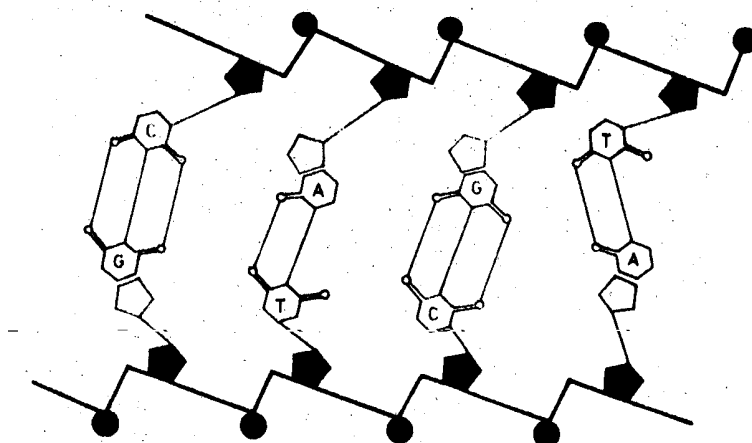


Schéma dvojvláknovej DNA



Čo si musíme zapamätať?

- V DNA je genetická informácia uložená lineárnym spôsobom pomocou reťazcov symbolov {A,T,G,C}

... **AATGGCATGGCCA** ...

- Molekulová biológia nám poskytuje poznatok o tom, že v priebehu reprodukcie (množenia) jedincov dochádza ku “kríženiu” a k “mutácii” genetickej informácie

Kríženie

... **TTGCAA** ...
... **CATTGC** ... → ... **TTGTGC** ...
... **CATTGC** **CATCAA** ...

Mutácia

... **CATTGC** ... → ... **CGTTGC** ...

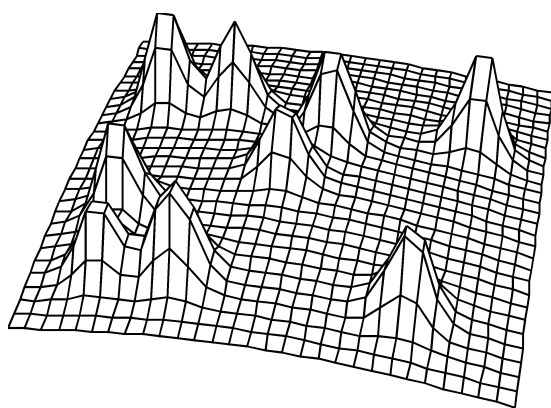
2. Darwinova evolúcia

Biologická evolúcia je progresívna zmena genetického obsahu jedincov populácie v priebehu mnohých generácií. Darwinov evolučný proces obsahuje tieto tri zložky:

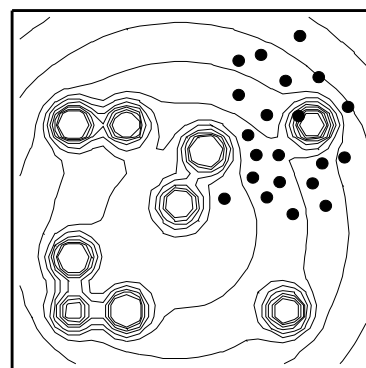
- (1) **Prirodzený výber** (Darwin), proces v ktorom "silnejší" jedinci populácie majú viac potomkov v nasledujúcej generácii, ako "slabší" jedinci.
- (2) **Náhodný genetický drift**, náhodné udalosti v živote jedincov populácie. Takéto udalosti sú napr. náhodná mutácia alebo náhodná smrť zjavne "silného" jedinca predtým, ako dostal príležitosť reprodukcie. Stochastické efekty náhodného genetického driftu sú významné hlavne pre malé populácie.
- (3) **Proces reprodukcie**, genetická informácia potomkov vzniká tak, že genetická informácia rodičov si vymení určité časti krížením, pričom môže dochádzať k jej poruche – mutácii.

V biológii je "sila" (fitness) definovaná ako relatívna schopnosť prežitia a reprodukovania sa v danom **prostredí** a v danej **populácii**. "Sila" môže byť chápaná ako atribút genotypu. Potomkovia rodičov s väčšou silou budú viac početnejší a "silnejší" ako potomkovia menej "silnejších" rodičov.

Povrch "sily" (A) - vhodný prístup ako vizualizovať akty selekcie v populácii, ktorá je v evolúcii. Populácia je reprezentovaná ako "oblak" bodov (B) na povrchu "sily".



A



B

3. Formalizácia evolúcie

(1) Jedinci populácie sú reprezentovaní **chromozómami** - lineárne reťazce symbolov

$$P = \{x_1, x_2, \dots, x_p\}$$

$$x \in \{\alpha, \beta, \dots\}^n$$

(2) Každý chromozóm populácie je ohodnotený pozitívnym reálnym číslom, ktoré sa nazýva "**sila**" (fitness)

$$f: P \rightarrow R_+$$

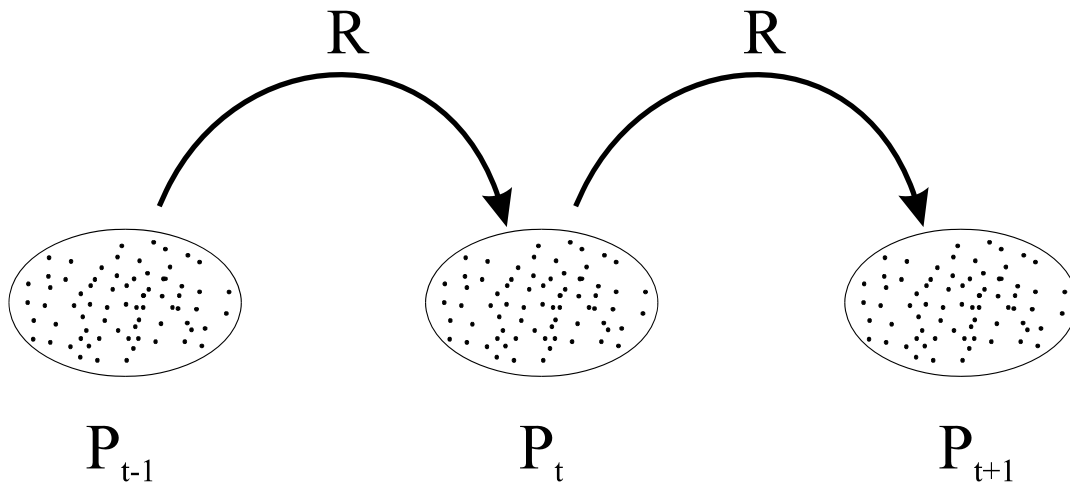
(3) **Reprodukcia** (proces množenia). Dvaja kvázináhodne vybraní jedinci z populácie (v závislosti od ich "sily", "silnejší" jedinci sú vybraní s väčšou pravdepodobnosťou) produkujú potomkov (deti). Hlavné časti reprodukcie sú:

- (1) **selekcia** rodičov,
- (2) **kríženie a mutácia**,
- (3) **návrat** potomkov do populácie.

$$(x_1^{new}, x_2^{new}) = O_{repro}(x_1^{old}, x_2^{old})$$

Populácia jedincov $P(t)$ vzniká procesom

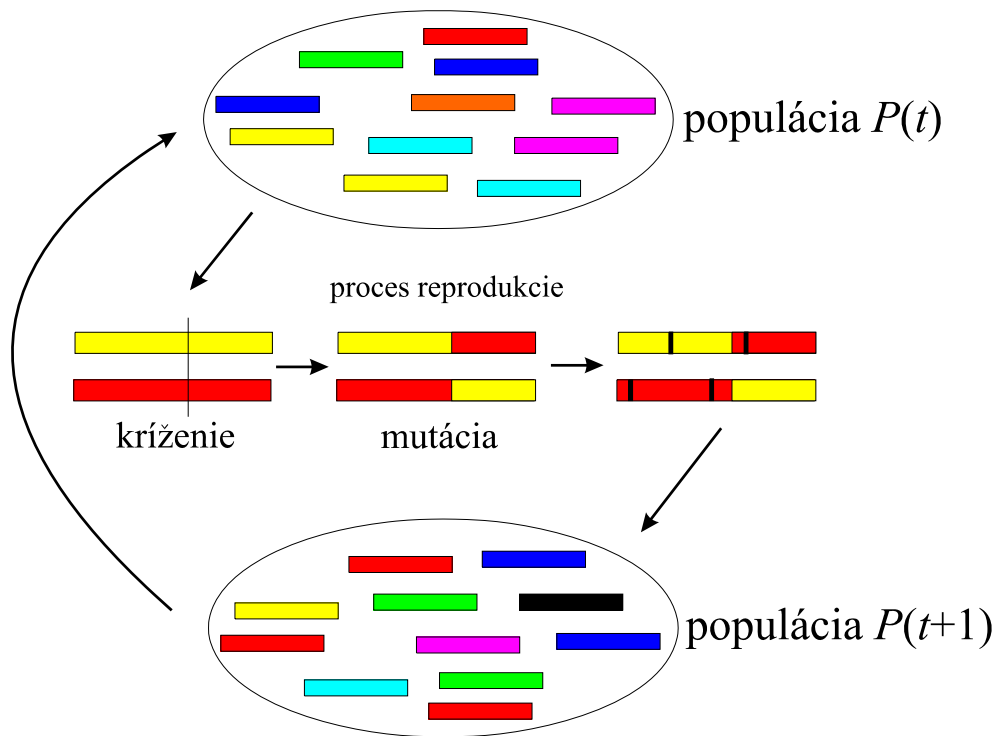
reprodukcie z populácie $P(t-1)$



Pascalovský pseudokód vyjadrujúci evolúciu populácie

```
P0 := náhodne vygenerovaná populácia  
Riešení (chromozómov);  
t := 0;  
while t < tmax do  
begin všetky riešenia z Pt sú  
ohodnotené silou;  
t := t + 1;  
Pt := R(Pt-1);  
end;
```

Znázornenie prirodzeného výberu



Pascalovský pseudokód prirodzeného výberu

```
P:=náhodne vygenerovaná populácia
  chromozómov;
t:=0;
while t<tmax do
begin t:=t+1;
      Q:=∅;
      while |Q|<|P| do
      begin vyber kvázináhodne dvoch
            rodičov  $x_1, x_2 \in P$ ;
             $(x_1', x_2') := O_{\text{repro}}(x_1, x_2)$ ;
            Q:=Q∪{ $x_1', x_2'$ };
      end;
      P:=Q;
end;
```

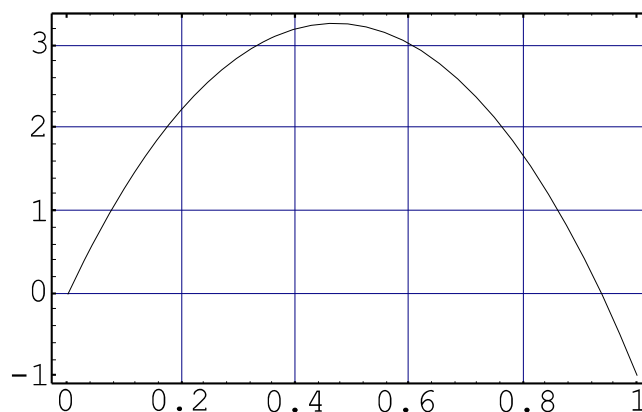
4. Genetický algoritmus

Genetický algoritmus (GA) je použitie **metafory Darwinovej evolúcie** v informatike a matematike na **zostrojenie optimálnych riešení**.

Jednoduchý ilustračný príklad

$$F(x) = 14x - 15x^2$$

pre $0 \leq x \leq 1$. Táto funkcia má globálne maximum pre $x_{\text{opt}} = 7/15$.



Binárna reprezentácia reálnych čísel.

Študujme bitový reťazec dĺžky n

$$\alpha = (\alpha_1 \alpha_2 \dots \alpha_n) \in \{0,1\}^n$$

Tento binárny reťazec α môže byť interpretovaný ako nezáporné celé číslo $\text{int}(\alpha)$

$$\begin{aligned}\text{int}(\alpha) &= \sum_{i=1}^n \alpha_i 2^{n-i} \\ &= \alpha_1 2^{n-1} + \alpha_2 2^{n-2} + \dots + \alpha_{n-1} 2 + \alpha_n\end{aligned}$$

Potom, k binárnemu reťazcu môžeme priradiť racionálne číslo $a \leq x \leq b$

$$\text{real}(\alpha) = a + \frac{b-a}{2^n - 1} \text{int}(\alpha)$$

Nech $0 \leq x \leq 1$ a $n=3$

α	$\text{int}(\alpha)$	$\text{real}(\alpha)$
000	0	0
001	1	1/7
010	2	2/7
011	3	3/7
100	4	4/7
101	5	5/7
110	6	6/7
111	7	1

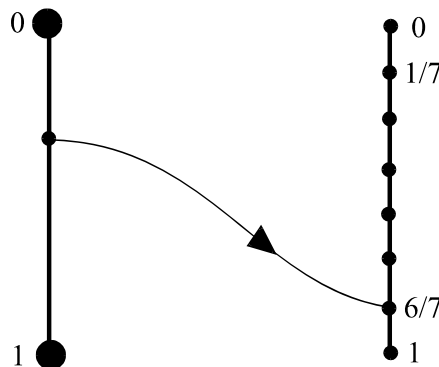
Znázornenie prechodu od binárneho reťazca k reálnemu (racionálnemu) číslu

$$\alpha \longrightarrow \text{int}(\alpha) \longrightarrow \text{real}(\alpha)$$

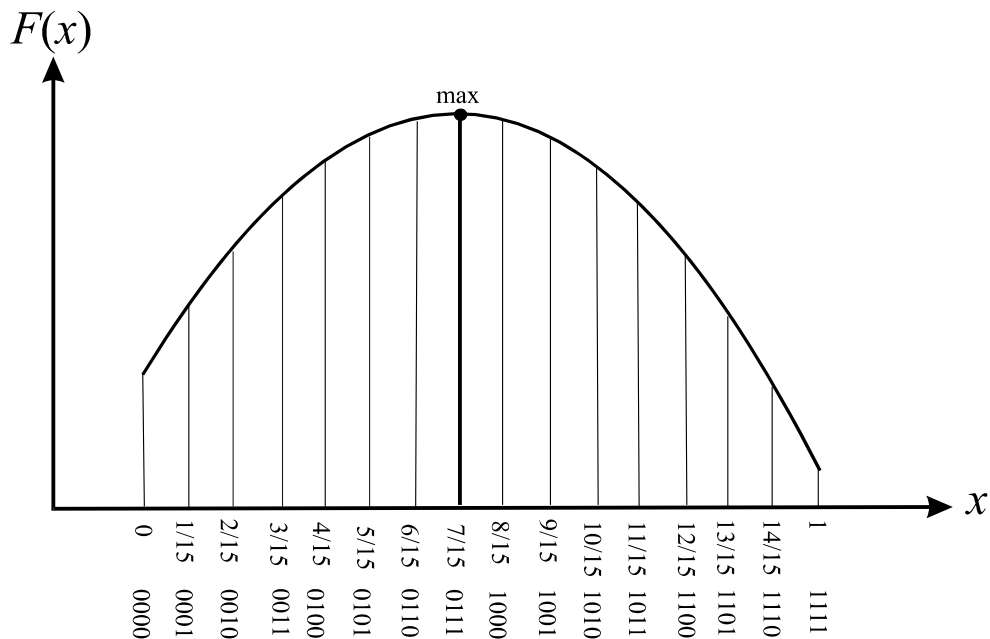
$$\downarrow \qquad \qquad \qquad \downarrow$$

$$0 \leq \text{int}(\alpha) \leq 2^n - 1 \quad a \leq \text{real}(\alpha) \leq b$$

V binárnej reprezentácii reálne čísla ležiace na úsečke $\langle a, b \rangle$ sú vyjadrené (aproximované) pomocou racionálnych čísel $i/(2^n - 1)$ (pre $i=0, 1, \dots, 2^n - 1$)



Priebeh funkcie $F(x)$ s vyznačením jednotlivých bodov pre binárnu reprezentáciu dĺžky $n=4$



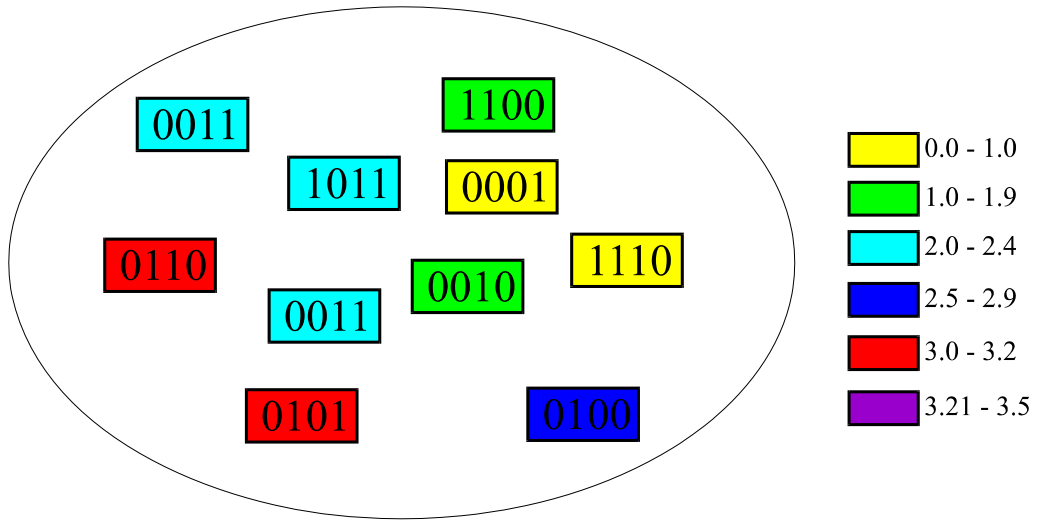
Pomocou metaforu Darwinovej evolúcie (genetického algoritmu) budeme hľadať v populácii binárnych reťazcov dĺžky $n=4$ také riešenie, v ktorom má funkcia $F(x)$ maximum

$$F(\alpha_{opt}) = \max_{\alpha} F(\alpha)$$

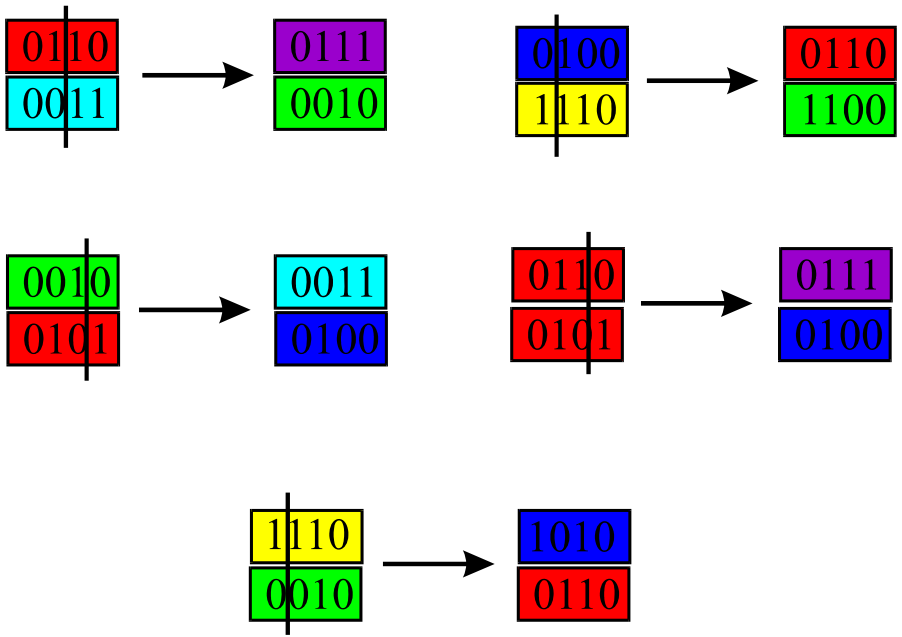
**Hodnoty kvadratickej funkcie $F(x)$ pre
racionálne premenné reprezentované
binárnymi reťazcami dĺžky $n=4$**

$\text{int}(\alpha)$	α	$x=\text{real}(\alpha)$	$F(x)$
0	0000	0.00000 (0/15)	0.00000
1	0001	0.06667 (1/15)	0.86667
2	0010	0.13333 (2/15)	1.60000
3	0011	0.20000 (3/15)	2.20000
4	0100	0.26667 (4/15)	2.66667
5	0101	0.33333 (5/15)	3.00000
6	0110	0.40000 (6/15)	3.20000
7	0111	0.46667 (7/15)	3.26667(*)
8	1000	0.53333 (8/15)	3.20000
9	1001	0.60000 (9/15)	3.00000
10	1010	0.66667 (10/15)	2.66667
11	1011	0.73333 (11/15)	2.20000
12	1100	0.80000 (12/15)	1.60000
13	1101	0.86667 (13/15)	0.86667
14	1110	0.93333 (14/15)	0.00000
15	1111	1.00000 (15/15)	-1.00000

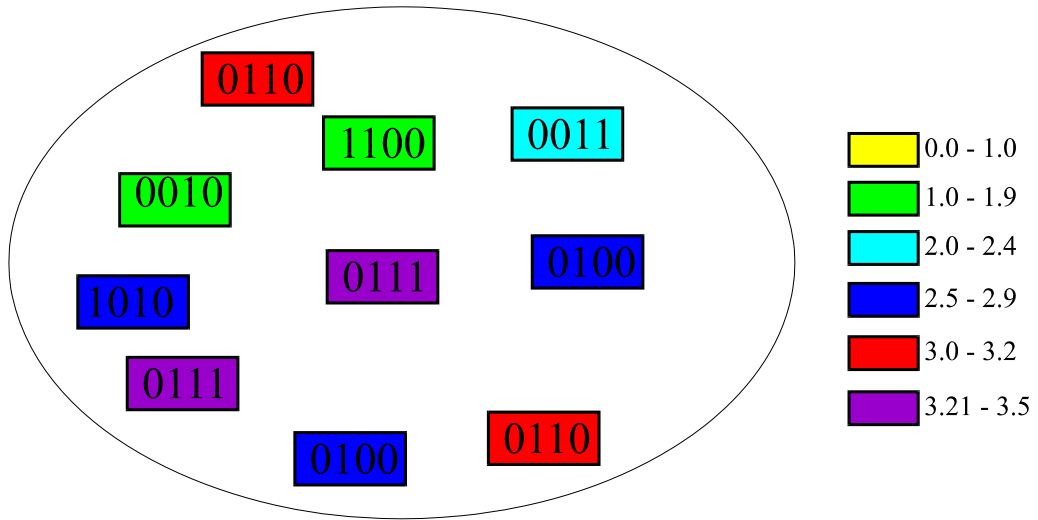
1. generácia



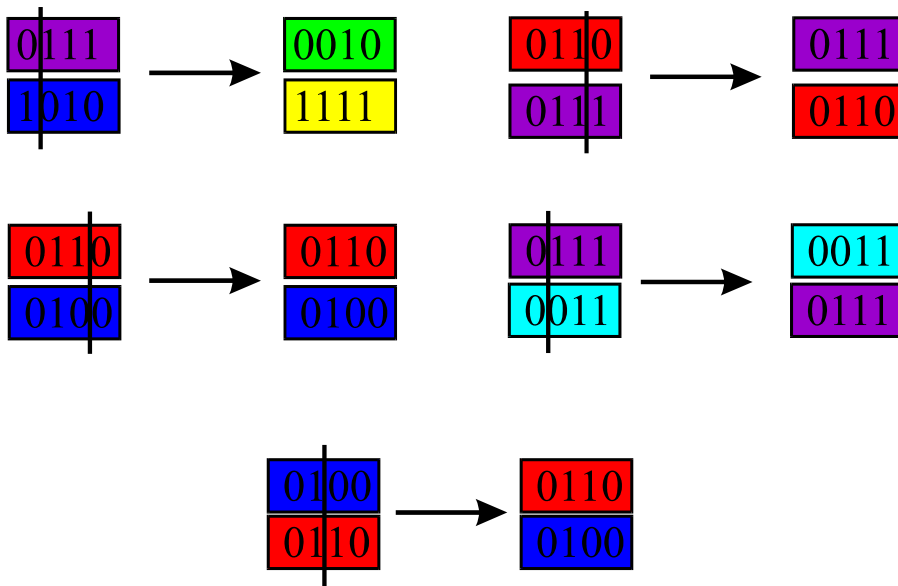
Tvorba novej populácie



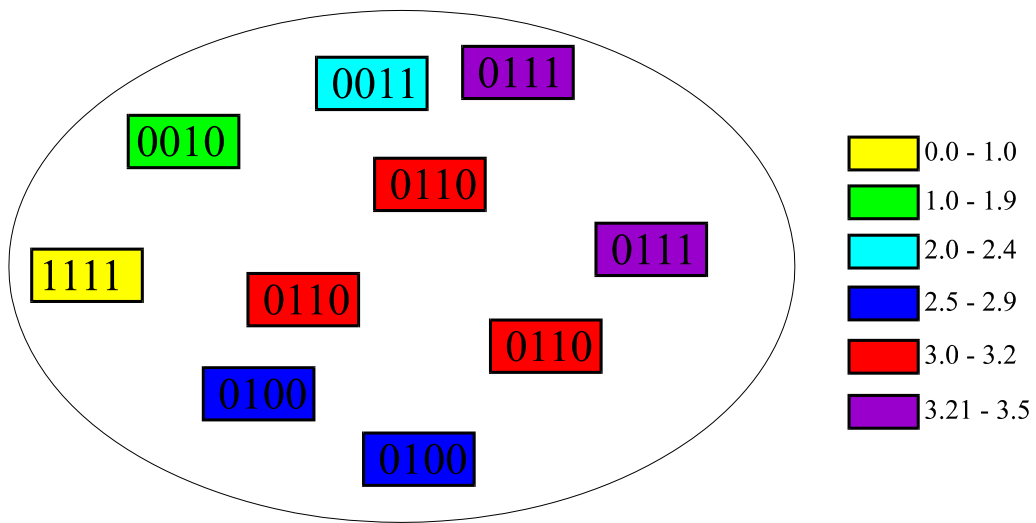
2. generácia



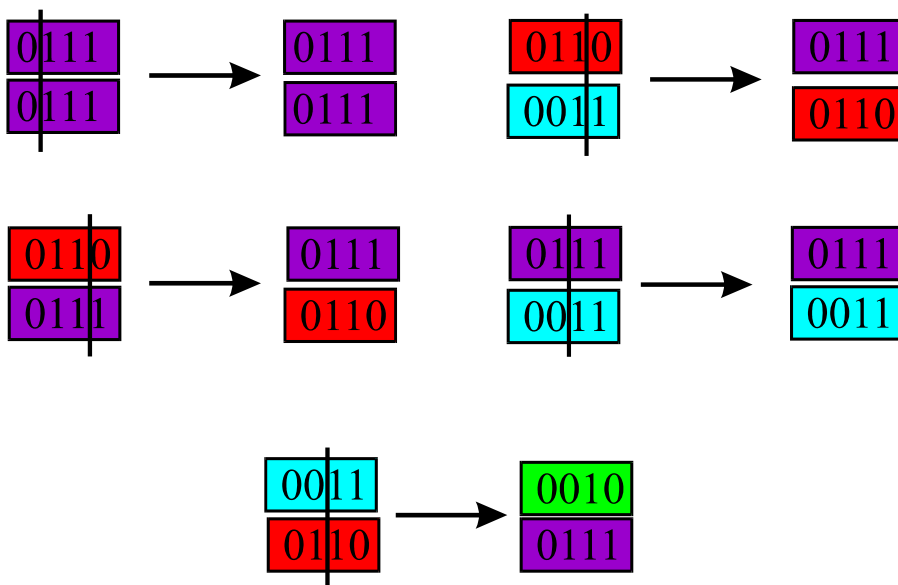
Tvorba novej populácie



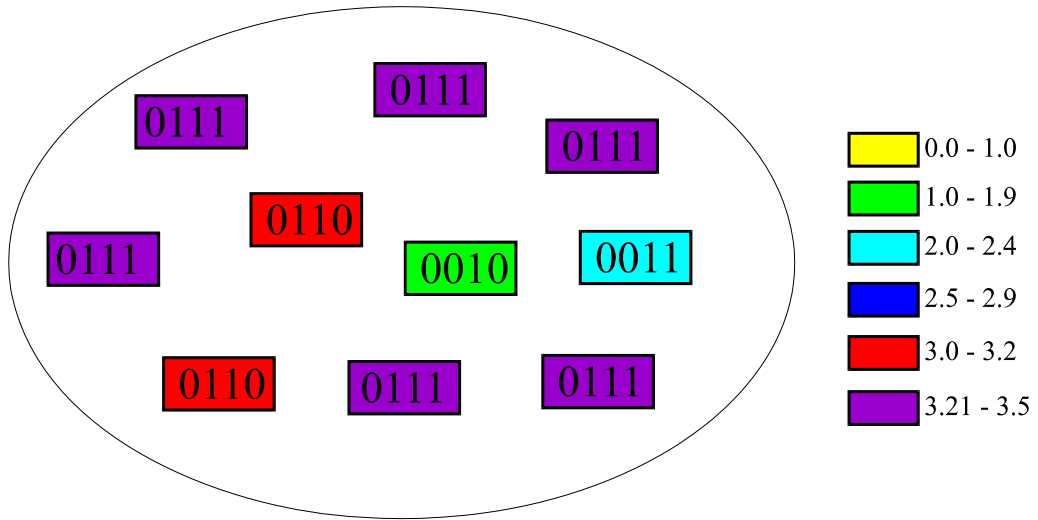
3. generácia



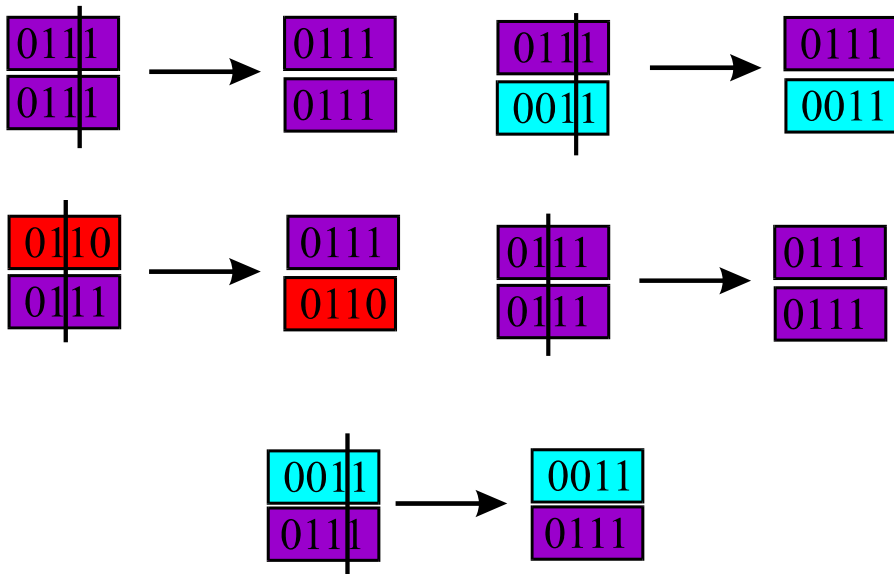
Tvorba novej populácie



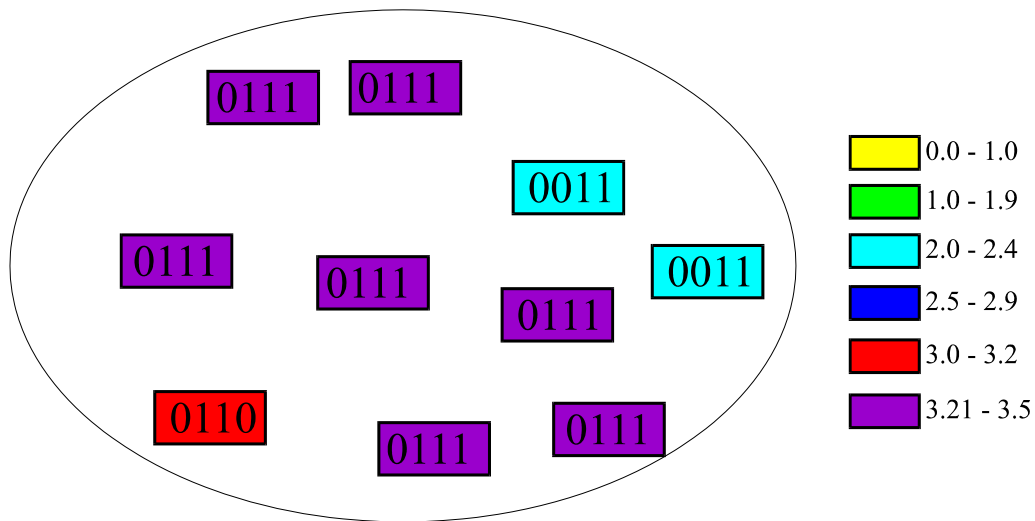
4. generácia



Tvorba novej populácie



5. generácia

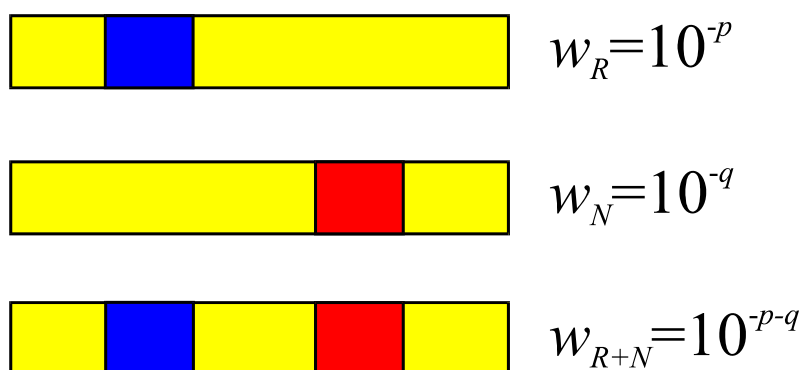


Populácia obsahuje 70% "fialových" chromozómov - riešení, ktoré zodpovedajú optimálnemu riešeniu, genetický algoritmus môžeme ukončiť.

5. Náznak teórie GA

Prečo GA pracuje? Ako to, že poskytuje rozumné výsledky? Pri hľadaní odpovedí na tieto a podobné otázky, obrátíme pozornosť na biologickú argumentáciu, ktorá vysvetľuje dôvody prečo je darwinovská evolúcia úspešná pri vzniku nových druhov lepšie prispôsobených k boju o prežitie v populácii.

Predpoklad: Nech v populácii spontánne vznikli chromozómy - jedinci, ktorí majú buď "ruky" alebo "nohy", cieľom evolúcie nech je vznik jedincov, ktorí majú súčasne tak "ruky", ako aj "nohy".

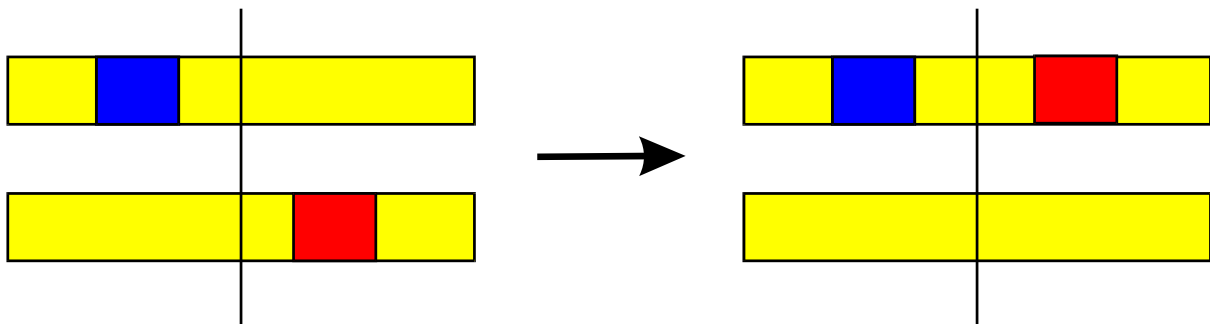


Pravdepodobnosť spontánneho vzniku chromozómu, ktorý súčasne obsahuje "ruky" aj "nohy" je v porovnaní s pravdepodobnosťami spontánneho

vzniku chromozómov obsahujúcich len "ruky" alebo len "nohy", veľmi malá

$$w_R = 10^{-p}, \quad w_N = 10^{-q}, \quad w_{R+N} = 10^{-p-q}$$

Táto nepriaznivá situácia sa dramaticky zmení, keď uvažujeme zapojenie "kríženia" v procese reprodukcie



Záver: Použitie "kríženia" pri reprodukcií jedincov podstatne urýchľuje evolúciu populácie.

Z tejto skutočnosti vyplýva, že v GA je kríženie chromozómov základnou "hnacou silou" evolúcie populácie smerom k populáciám obsahujúcich optimálny chromozóm s častou frekvenciou výskytu.

6. Záver

- Moderná informatika hľadá "nové"

nápady" v živej prírode. Tak napr., metafora ľudského mozgu viedla ku vzniku novej oblasti informatiky nazývanej "neurónové siete". Podobne, metafora Darwinovej evolúcie viedla ku vzniku tzv. evolučných algoritmov, kde GA patrí medzi ich najvýznamnejšieho reprezentanta.

- Metafora Darwinovej evolúcie poskytuje v informatike efektívny optimalizačný algoritmus, ktorý je schopný riešiť obtiažne úlohy.
- Evolučné algoritmy umožňujú počítačové simulácie evolučných procesov v biológii. Tieto výpočty umožňujú testovať hypotézy o význame a mechanizmoch evolúcie v biológii (a tiež aj v iných vedách) pri vzniku rôznych vlastností.