



EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



Název projektu	Rozvoj vzdělávání na Slezské univerzitě v Opavě
Registrační číslo projektu	CZ.02.2.69/0.0./0.0/16_015/0002400

**Dolování dat**

**Genetické algoritmy**

**Jan Górecki**



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

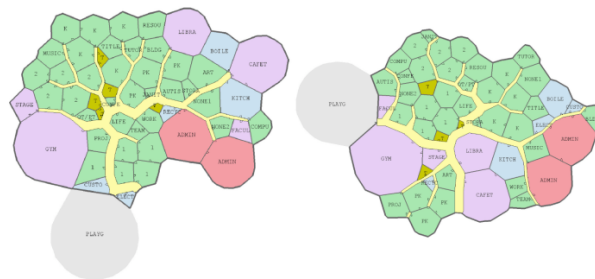
- Co jsou Genetické algoritmy (GA)
- Příklad
- Selekcce
- Křížení
- Mutace
- GA pro získávání pravidel z dat

Original layout



The original elementary school. Found somewhere in Maine.

'Optimized'



Left: Optimized for minimizing traffic flow between classes and material usage. Right: Also optimized for minimizing fire escape paths.

- Zdrojem inspirace se tentokrát stal mechanismus evoluce:

*„Nějaký živočišný druh se během svého vývoje zdokonaluje tak, že z generace na generaci se přenáší genetická informace jen těch nejsilnějších jedinců.“*

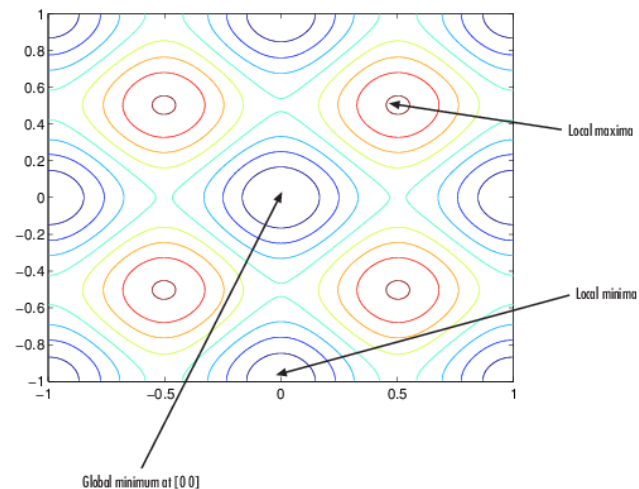
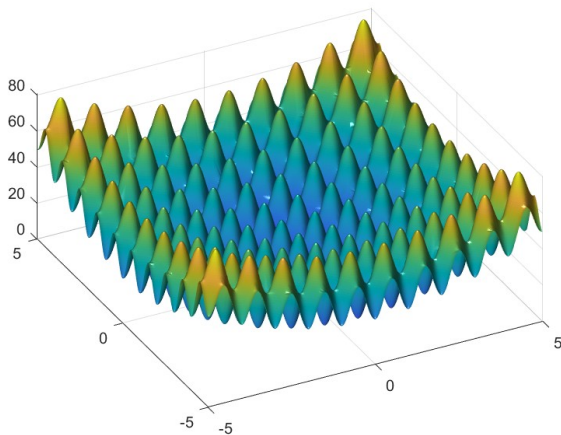
Genetický algoritmus (GA) je heuristický postup, vycházející z evolučního přístupu, který lze nasadit na řešení složitých problémů, pro které neexistuje použitelný exaktní algoritmus.

- Používají se především pro optimalizaci (nelineární regrese, neuronové sítě)
- Hledání **globálního** minima chybové funkce – podobně jako např. simulované žihání, genetická optimalizace **umí překonat** uvíznutí v lokálním minimu

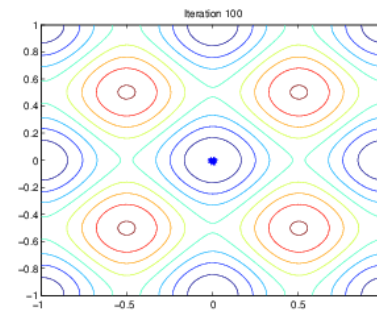
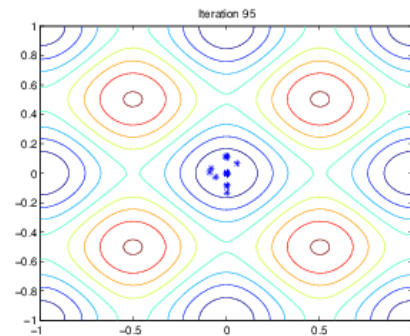
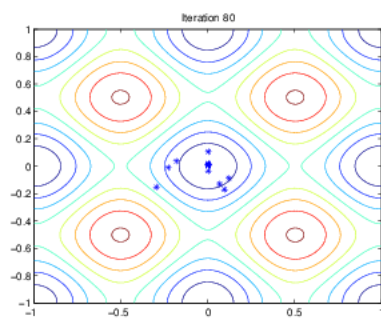
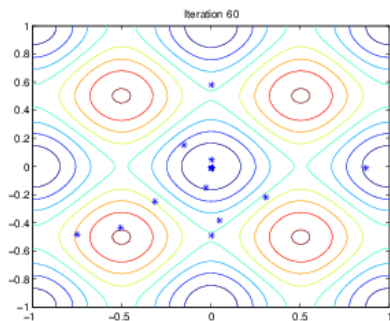
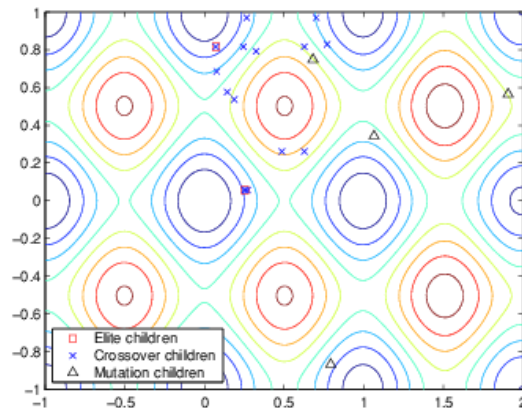
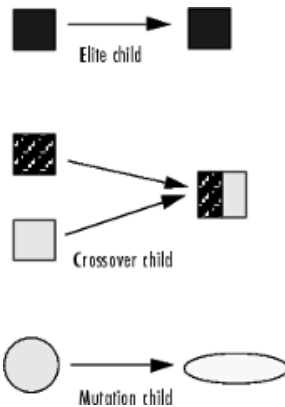
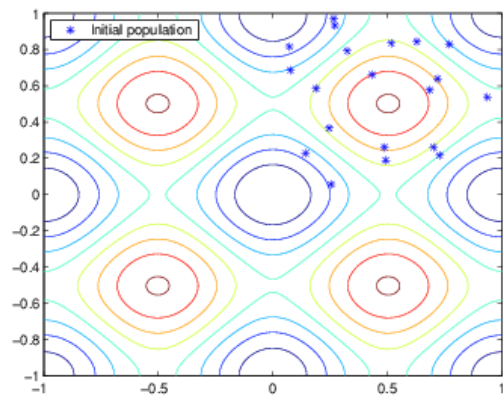


## Hledání globálního minima Rastriginovy funkce

$$fit_{Ras}(x_1, x_2) = 20 + x_1^2 + x_2^2 - 10(\cos 2\pi x_1 + \cos 2\pi x_2)$$



# Příklad



# Genetický algoritmus

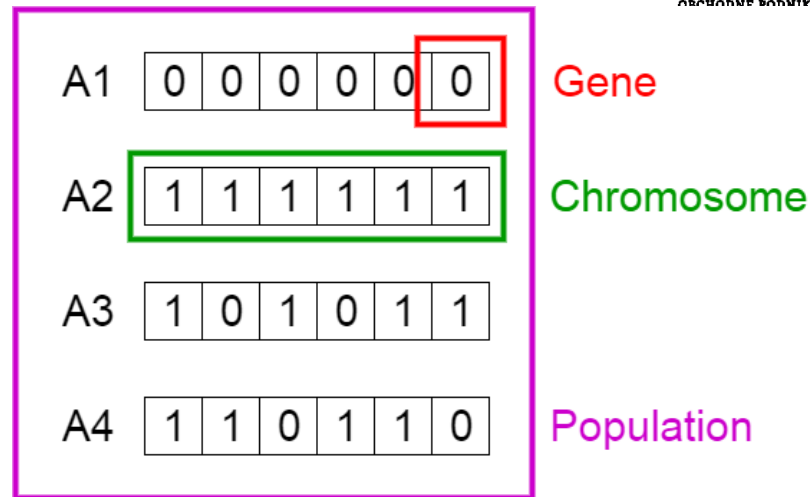
## Genetický algoritmus( $fit, N, K, M$ )

### Inicializace

1. přiřaď  $t := 0$  (počítadlo generací)
2. náhodně vytvoř populaci  $P(t)$  velikosti  $N$
3. urči hodnoty funkce  $fit$  pro každého jedince v  $P(t)$

### Hlavní cyklus

1. dokud není splněna podmínka pro zastavení
    - 1.1. proved' selekci:
      - 1.1.1. vyber z  $P(t)$  jedince kteří se přímo přenesou do  $P(t+1)$
    - 1.2. proved' křížení:
      - 1.2.1. vyber z  $P(t)$  jedince určené k reprodukci
      - 1.2.2. aplikuj na každou dvojici  $[h_1, h_2]$  z výběru operaci křížení
      - 1.2.3. zařaď potomky do  $P(t+1)$
    - 1.3. proved' mutaci:
      - 1.3.1. vyber jedince z  $P(t+1)$  určené k mutaci
      - 1.3.2. aplikuj na každé vybrané  $h$  operaci mutace
  - 1.4. přiřaď  $t := t + 1$  (nová populace má opět velikost  $N$ )
  - 1.5. spočítej pro každé  $h \in P(t)$  hodnotu  $fit(h)$
2. vrať hypotézy  $h$  s nejvyšší hodnotu  $fit(h)$



## Podmínka pro zastavení:

- počet generací
- kvalita nejlepšího jedince v populaci
- změna kvality nejlepšího jedince mezi generacemi

$(1 - K) * N$  jedinců přímo

Nyní předpokládáme, že silnější jedinec má větší hodnotu  $fit$  než slabší jedinec ( $fit = 1/fit_{Ras}$ )

**ruletové kolo**

- pravděpodobnost, že bude vybrán jedinec  $h$  je úměrná poměru  $\frac{fit(h)}{\sum_i fit(h_i)}$

**pořadová selekce**

- nejprve jsou jedinci v populaci uspořádáni podle hodnoty  $fit$ , selekce se pak provádí na základě pravděpodobnosti, která je úměrná pořadí jedince v tomto uspořádání,

**turnajová selekce**

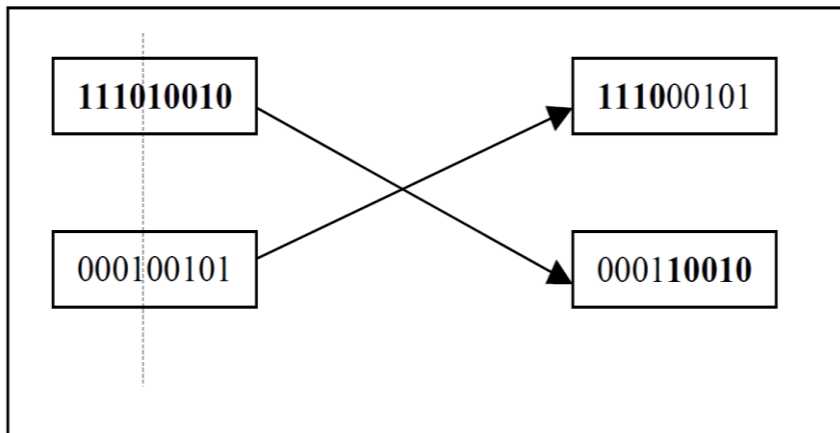
- nejprve se náhodně vyberou dva jedinci (nebo i jeden či více než dva), s předdefinovanou pravděpodobností  $p$  ( $> 0,5$ ) se pak vybere z těchto dvou jedinec s vyšší hodnotou  $fit$ , s pravděpodobností  $1-p$  se vybere z těchto dvou jedinec s nižší hodnotou  $fit$

# Křížení

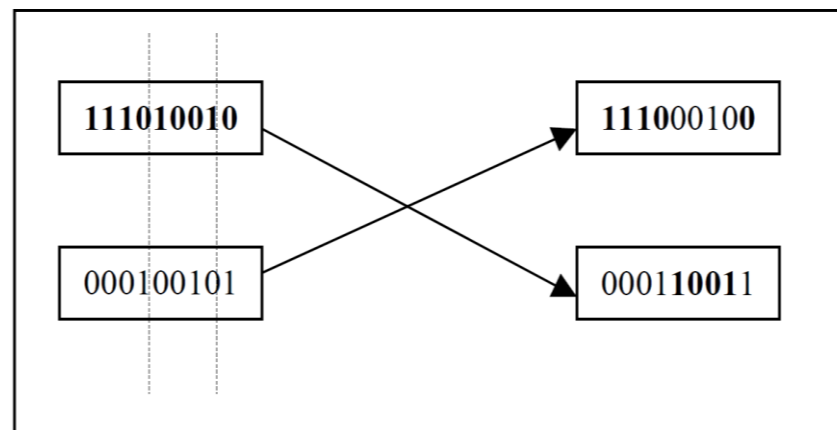


SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

$K * N/2$  dvojic



jednobodové



dvoubodové



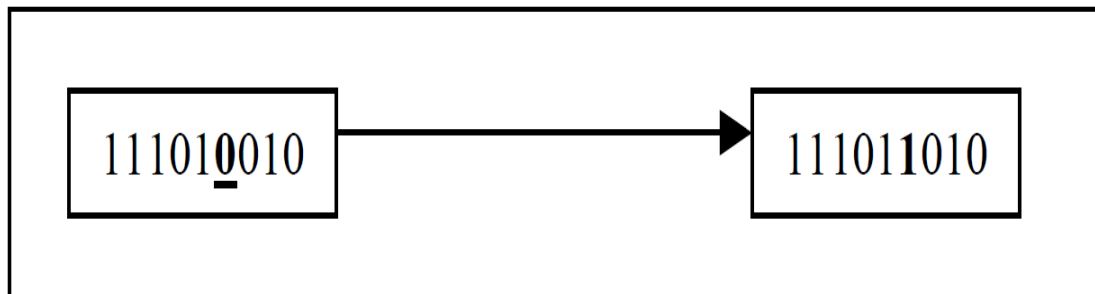
# Mutace

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVÍNĚ

$M \cdot N$  jedinců



# GA pro získávání pravidel z dat

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVÍNĚ

paralelní náhodné prohledávání

GABIL (deJong, 1993)

Jedinci jsou pravidla:

If konto(nízké)  $\wedge$  příjem(nízký) then úvěr(ne)  
If konto(vysoké) then úvěr(ano)

100 10 01

001 11 10

---

1. funkce  $\text{fit}(h)$  je druhou mocninou správnosti klasifikace

$$\text{fit}(h) = \left(\frac{a}{a+b}\right)^2,$$

2. počet jedinců v populaci je mezi 100 a 1000 v závislosti na konkrétní úloze,
  3. parametr  $K$  vyjadřující podíl křížení má hodnotu 0.6,
  4. parametr  $M$  vyjadřující podíl mutací má hodnotu 0.001,
  5. použitý operátor křížení je rozšířením výše uvedeného dvoubodového křížení; provedené rozšíření umožňuje křížit řetězce různých délek,
  6. mutace je použita tak, jak je uvedeno výše.
-

## Řešení ekonomických úloh

- Optimalizace výrobního plánu podniku (více výrobků a společné suroviny na jejich výrobu).
  - Výběr investic (investiční portfolio složeno z více investic s různou výnosností).
  - Problém obchodního cestujícího (častá úloha v podobě rozvážky zboží na jednotlivá distribuovaná místa, obdobou je i hledání spojnice optimální trasy mezi dvěma vzdálenými městy).
  - Úloha o bařohu (cílem je umístit výrobky o daných rozměrech do vymezeného prostoru, často využíváno u distribučních a zásilkových služeb při volbě vhodného obalového materiálu, potažmo krabice).
  - Umístění distribučního skladu (modifikace úlohy obchodního cestujícího, cílem je umístit distribuční sklad tak, aby bylo zvolené místo optimální vzhledem na propustnost distribučních cest a velikost zavážek).
-

# Děkuji za pozornost

Některé snímky převzaty od:  
prof. Ing. Petr Berka, CSc. [berka@vse.cz](mailto:berka@vse.cz)