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

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 KARVINÉ

Obsah přednášky

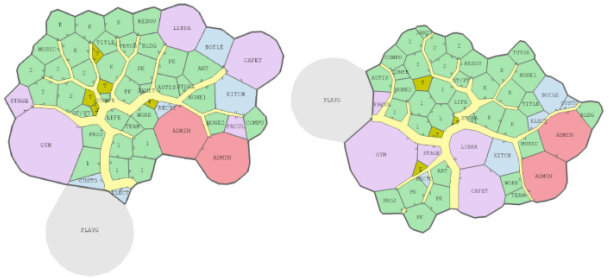
SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- Co jsou Genetické algoritmy (GA)
- Příklad
- Selekce
- 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 bewteen classes and material usage. Right: Also optimized for minimizing fire escape paths.

Genetické algoritmy

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

- 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.

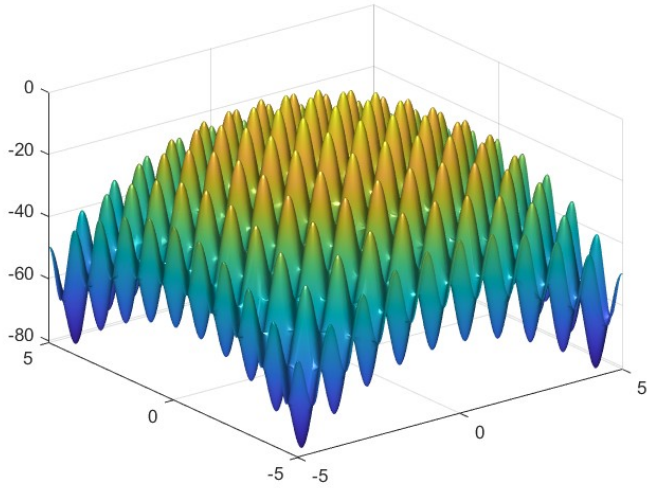
- Používají se především pro optimalizaci (nelineární regrese, neuronové sítě)
- Hledání **globálního** minima chybové funkce – podobně jako např. simulované žíhání, genetická optimalizace **umí překonat** uvíznutí v lokálním minimu (maximu)

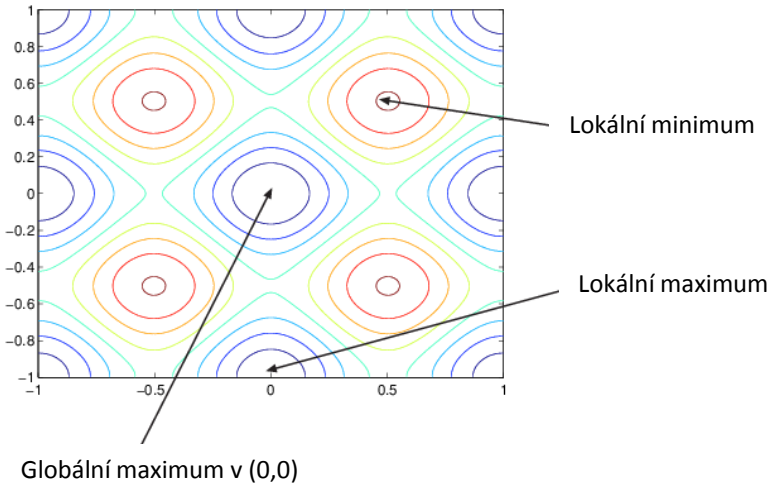
Příklad

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Hledání globálního maxima Rastriginovy funkce

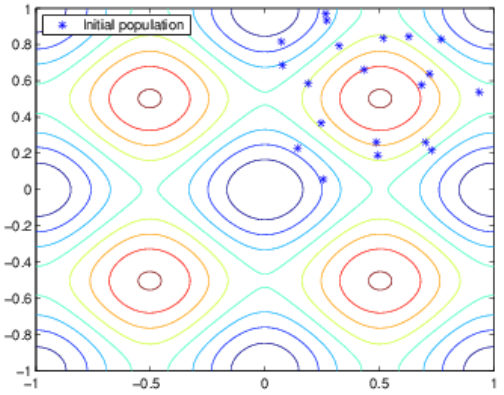
$$fit(x_1, x_2) = -20 - x_1^2 - x_2^2 + 10(\cos 2\pi x_1 + \cos 2\pi x_2)$$

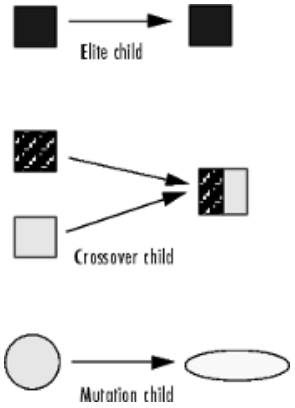


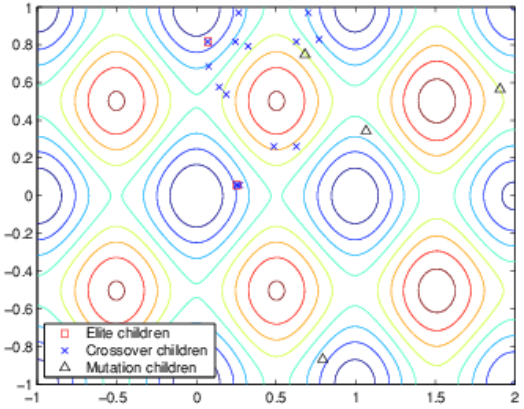


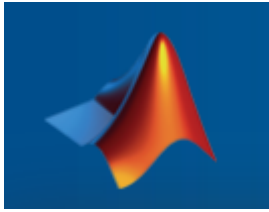
Příklad

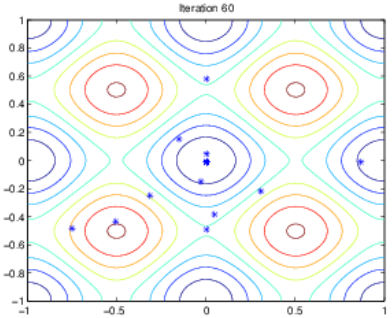
SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

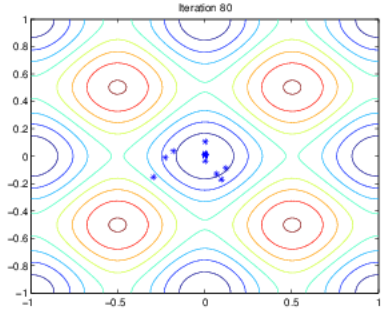


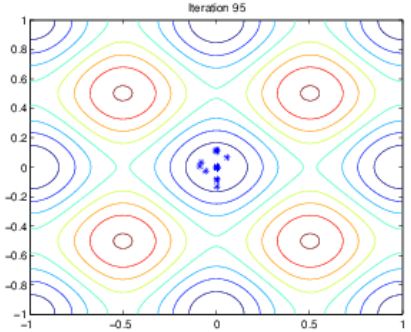


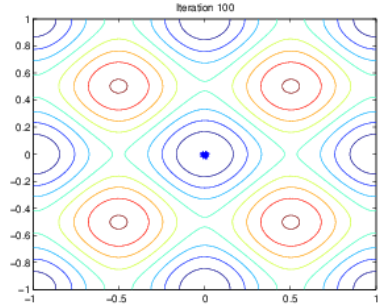












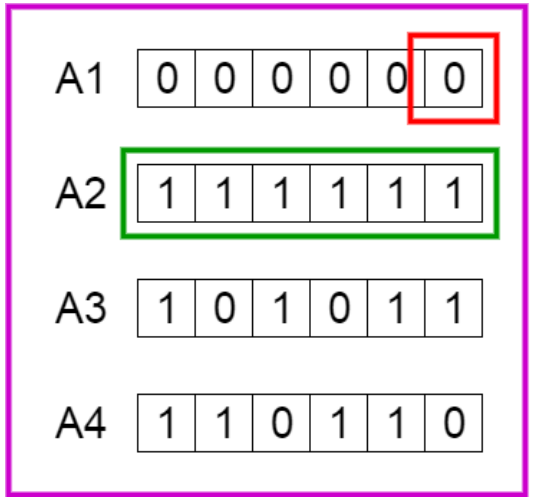
Genetický algoritmus

SLEZSKÁ UNIVERZITA OBCHODNĚ PODNIKATELSKÁ

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. proveď selekci:
      1.1.1. vyber z *P(t)* jedince kteří se přímo přenesou do *P(t+1)*
   1.2. proveď 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. proveď 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∈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

Selekce

SLEZSKÁ UNIVERZITA
ELSKÁ

$(1 - K)*N$ jedinců přímo

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,

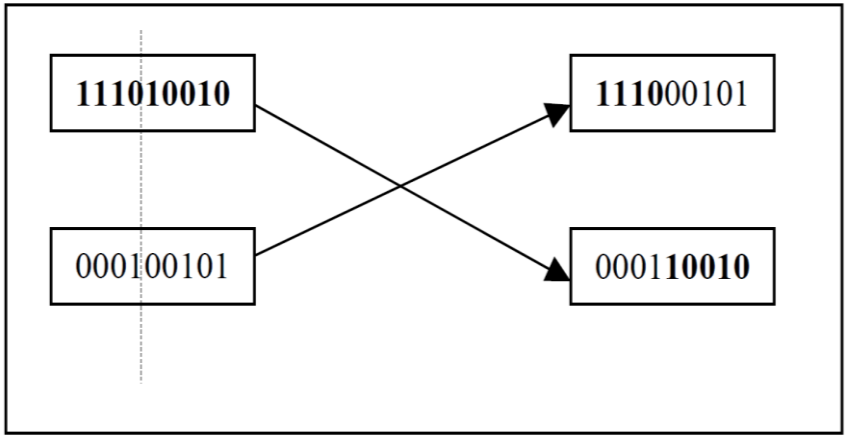
turnajová selekce

- nejprve se náhodně vyberou dva jedinci (nebo i jeden či více než dva), s předefinovanou 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*

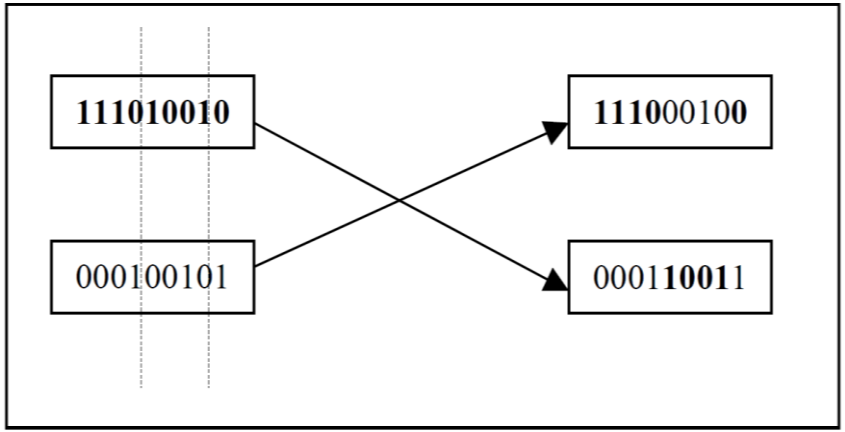
Křížení

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

K*N/2 dvojic



jednobodové

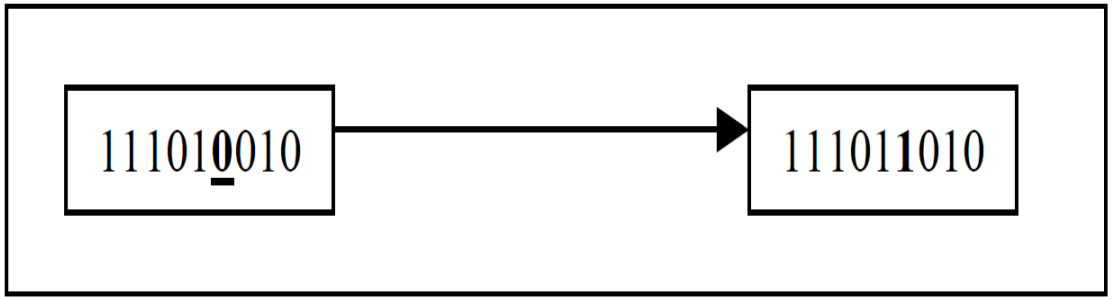


dvoubodové

Mutace

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

M*N jedinců



GA pro získávání pravidel z dat

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

paralelní náhodné prohledávání

GABIL (deJong, 1993)

Jedinci jsou pravidla:

If konto(nízké) ∧ příjem(nízký) then úvěr(ne)
If konto(vysoké) then úvěr(ano)

100 10 01
001 11 10

Zpřesnění základního algoritmu

1. funkce 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,

Aplikace

SLEZSKÁ UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ FAKULTA V KARVINÉ

Řešení ekonomických úloh

- Optimalizace výrobního plánu podniku (více výrobků a společné suroviny na jejich výrobu).
- 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 baťohu (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