

Рачунарска интелигенција

Генетско програмирање

Александар Картељ

kartelj@matf.bg.ac.rs

Ови слајдови представљају прилагођење слајдова:

A.E. Eiben, J.E. Smith, Introduction to Evolutionary computing: Genetic programming

Генетско програмирање

- Развијено у Америци 90-тих година, Ј. Koza
- Обично се примењује у:
 - машинском учењу (предикција, класификација...)
- Конкурентан неуронским мрежама и сличним методама
 - Али захтева огромне популације (хиљаде јединки)
 - Релативно спор
- Специјлане карактеристике:
 - Нелинеарни хромозоми: стабла, графови
 - Мутација могућа, али није неопходна

GP техничке карактеристике

Репрезентација	Стабло
Укрштање	Размена подстабала
Мутација	Случајна промена у дрвету
Селекција родитеља	Фитнес сразмерна
Селекција преживелих	Генерацијска замена

Уводни пример: одређивање кредитне способности

- Банка хоће да направи разлику између добрих и лоших кандидата за давање позајмица
- Потребно је узети у обзир историјске податке

ID	Број деце	Плата	Брачни статус	ОК?
ID-1	2	45000	Married	0
ID-2	0	30000	Single	1
ID-3	1	40000	Divorced	1
...				

Уводни пример: одређивање кредитне способности (2)

- Могући модел:

IF (број деце = 2) AND (плата > 80000) THEN добар ELSE лош

- Општи приступ:

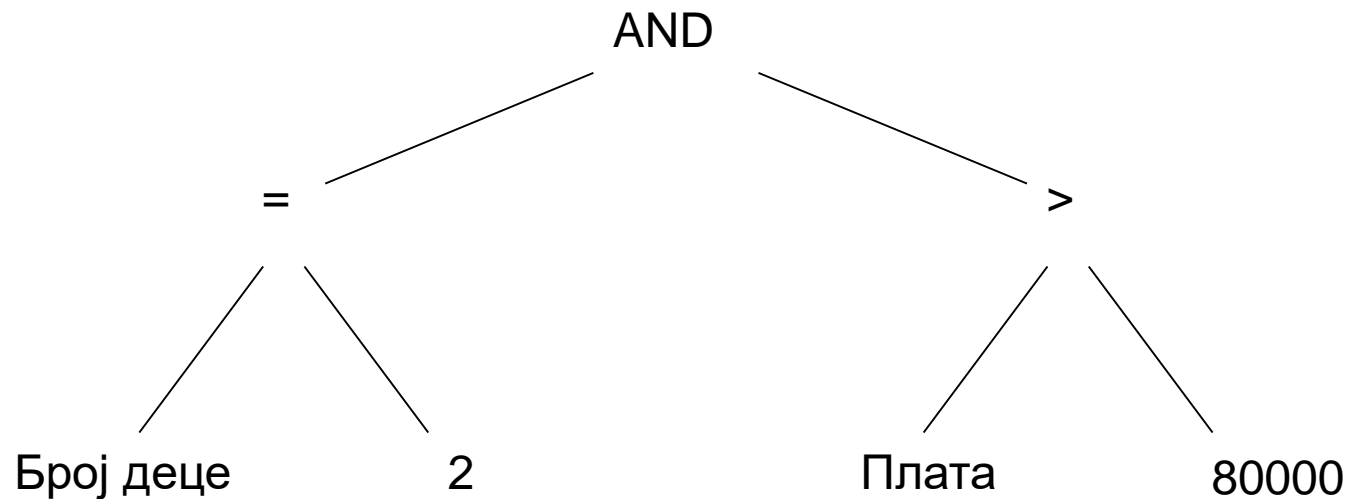
IF формула THEN добар ELSE лош

- Непозната је формула по којој се одређује?
- Простор претраге (фенотип) је скуп свих формула
- Фитнес формуле: проценат добро класификованих примера
- Природна репрезентација формуле (генотип) је стабло

Уводни пример: одређивање кредитне способности (3)

IF (број деце = 2) AND (плата > 80000) THEN добар ELSE лош

се може представити следећим стаблом:



Репрезентација стабла

- Стабла имају способност представљања великог броја формула

- Аритметичка формула

$$2 \cdot \pi + \left((x + 3) - \frac{y}{5 + 1} \right)$$

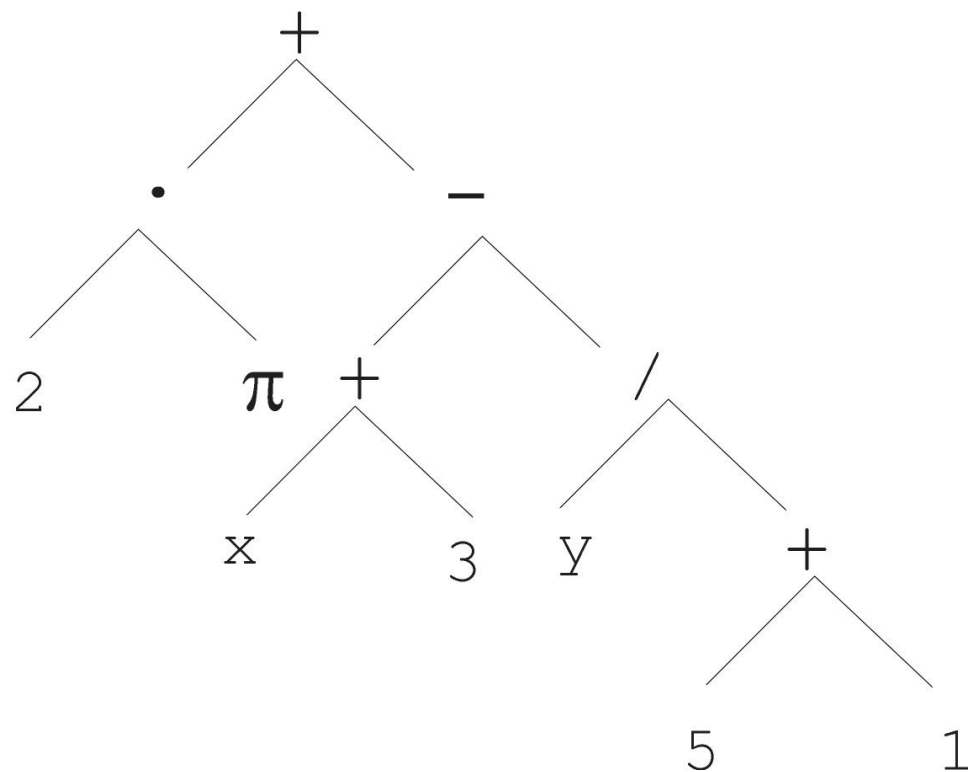
- Логичка формула

$$(x \wedge \text{true}) \rightarrow ((x \vee y) \vee (z \leftrightarrow (x \wedge y)))$$

- Програм

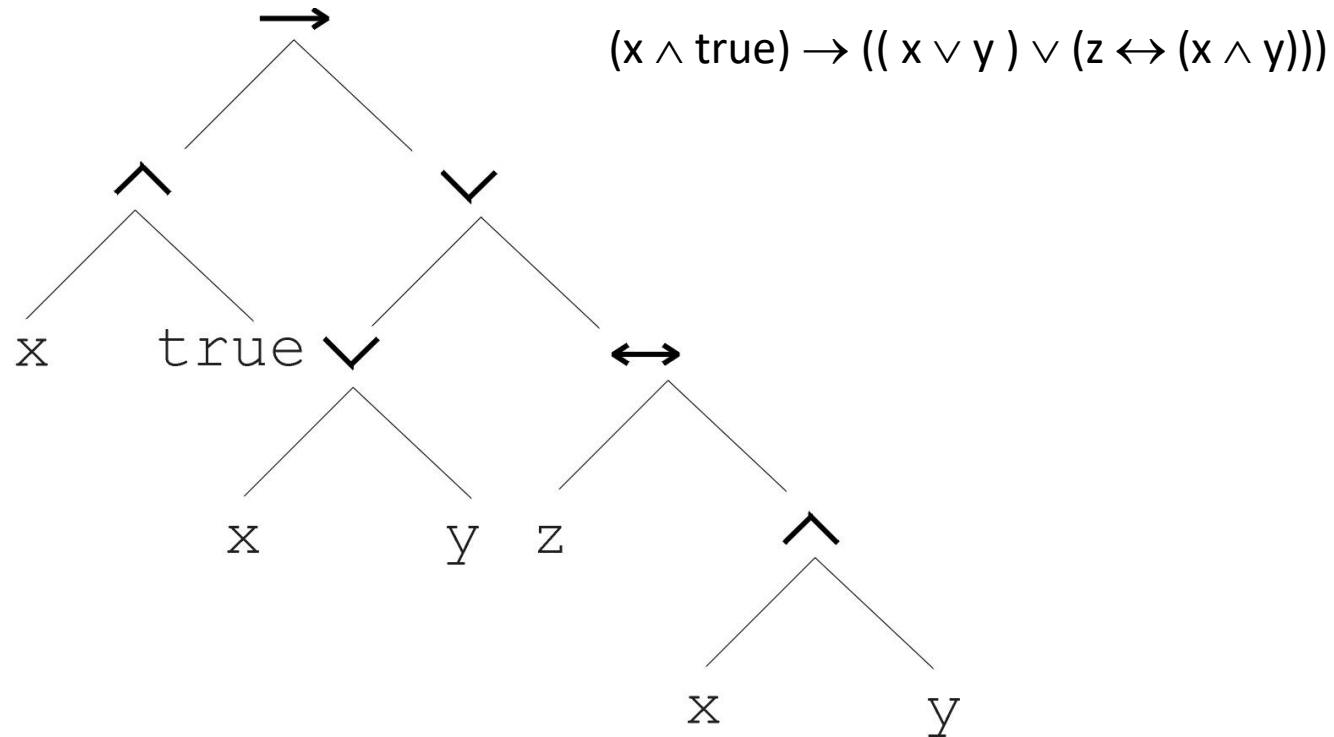
```
i = 1;
while (i < 20)
{
    i = i + 1
}
```

Репрезентација стабла (2)

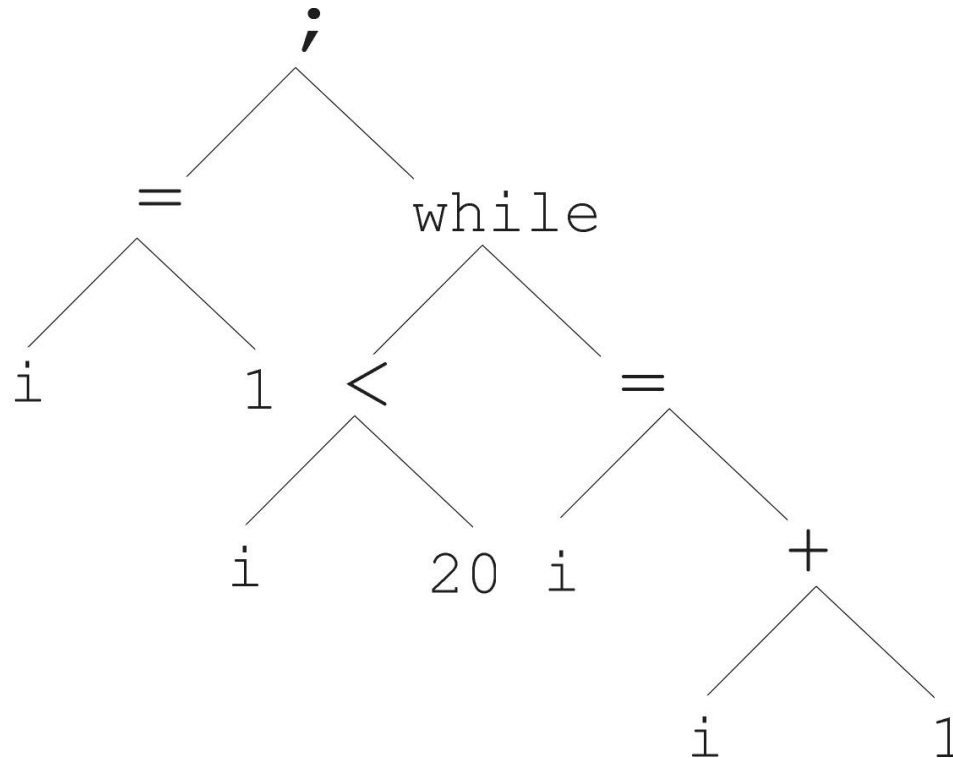


$$2 \cdot \pi + \left((x + 3) - \frac{y}{5 + 1} \right)$$

Репрезентација стабла (3)



Репрезентација стабла (4)



```
i = 1;  
while (i < 20)  
{  
    i = i + 1  
}
```

Репрезентација стабла (5)

- У генетским алгоритмима (GA), еволутивним стратегијама (ES), еволутивном програмирању (EP), хромозоми су линеарне структуре
 - Низови битова
 - Низови целих бројева
 - Низови реалних бројева
 - Пермутације
 - ...
- Стабло-хромозоми су нелинеарне структуре
- Код GA, ES, EP, величина хромозома је фиксна
- Стаблау GP могу да имају произвољну дубину и ширину

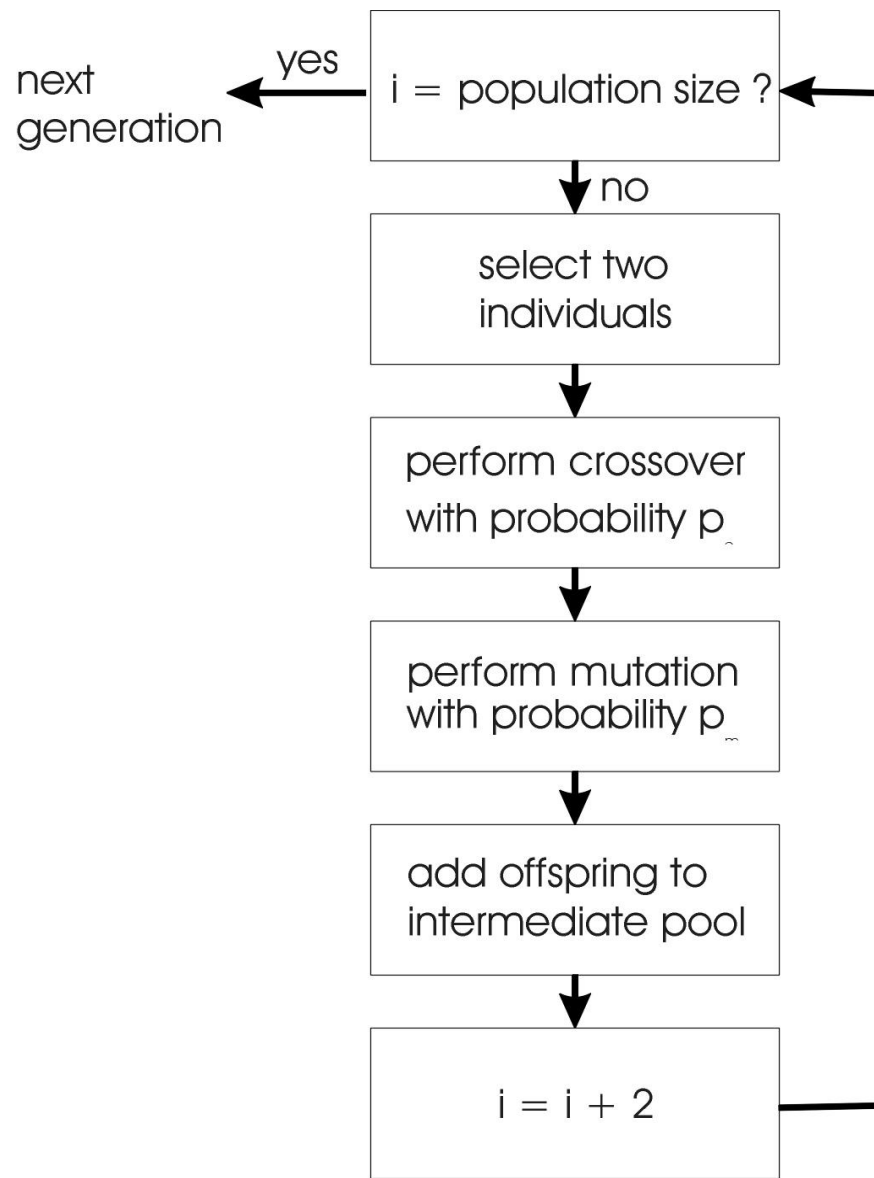
Репрезентација стабла (6)

- Симболички изрази могу бити дефинисани помоћу
 - Скупа термова T
 - Скупа функција F (са придруженим арностима)
- Потом се може користити следећа рекурзивна дефиниција:
 1. Сваки $t \in T$ је коректан израз
 2. $f(e_1, \dots, e_n)$ је коректан израз ако $f \in F$, $\text{arity}(f)=n$ и e_1, \dots, e_n су коректни изрази
 3. Не постоје друге коректне форме израза
- У општем случају, изрази у GP нису типизирани (сваки $f \in F$ може узети било који $g \in F$ као аргумент)

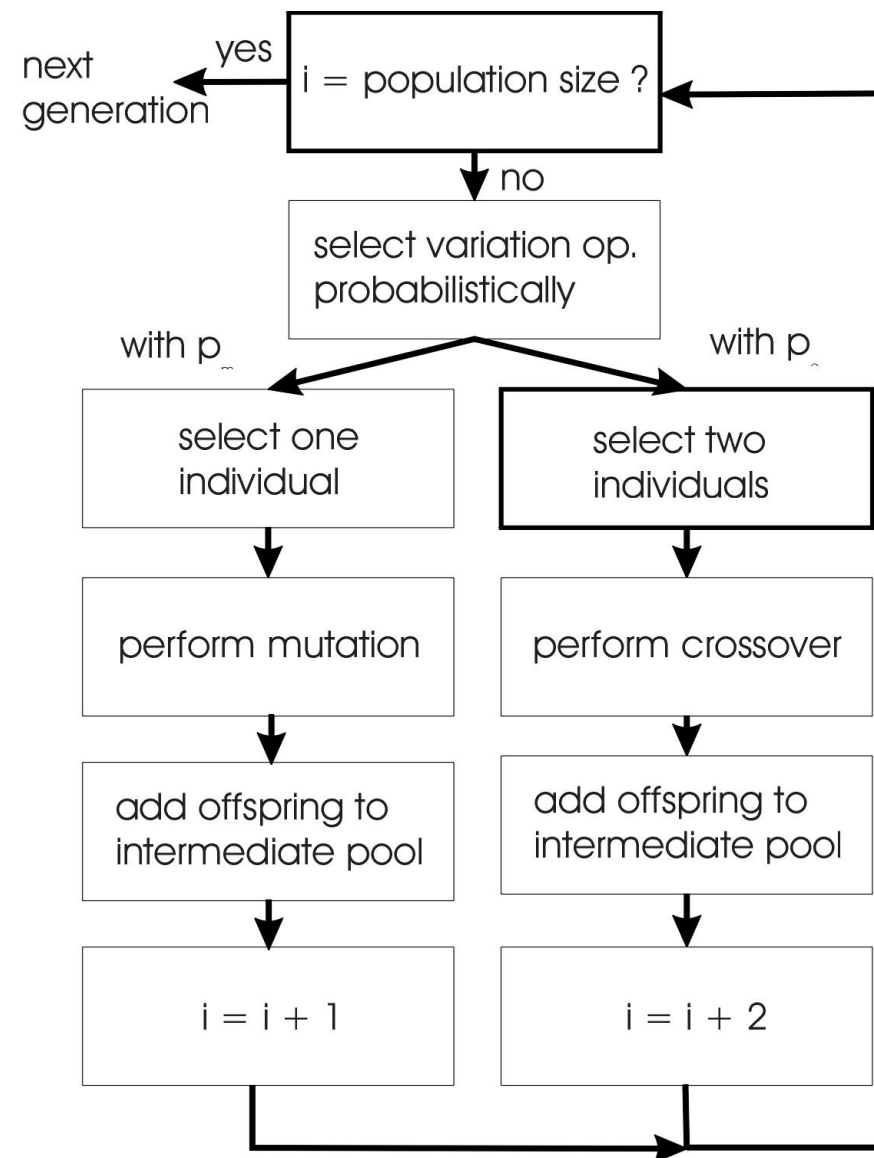
Генерисање потомака

Поређење:

- GA користи укрштање **И** мутацију секвенцијално (случајно)
- GP користи укрштање **ИЛИ** мутацију (случајно)



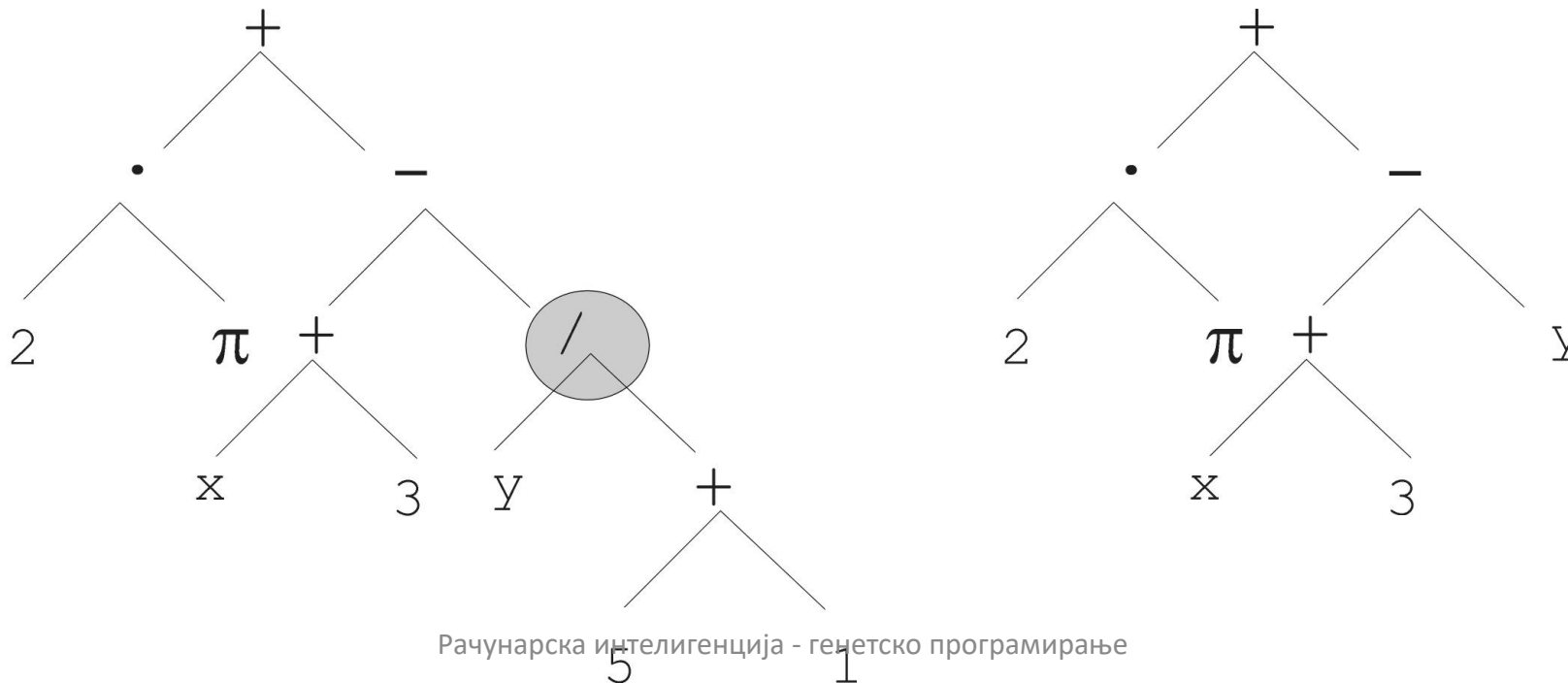
GA дијаграм тока



GP дијаграм тока

Мутација

- Најчешћи оператор мутације:
замени случајно одабрано подстабло
новим случајно генерисаним стаблом

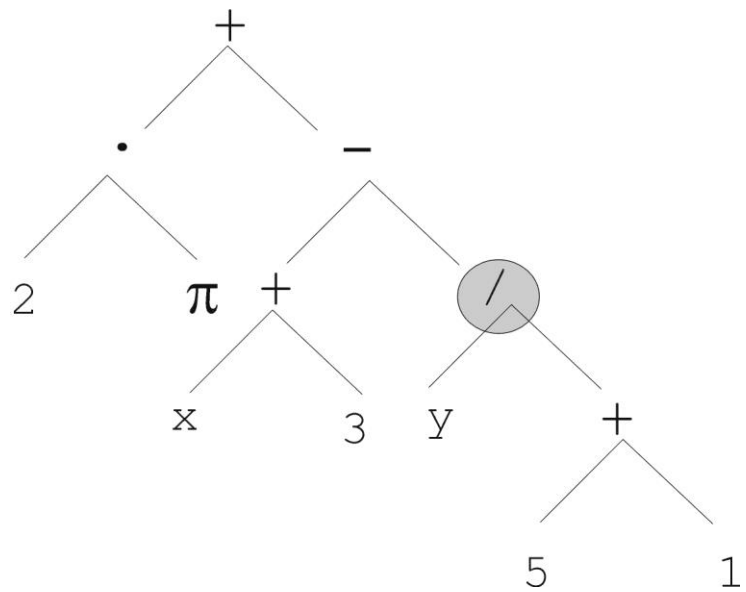


Мутација (2)

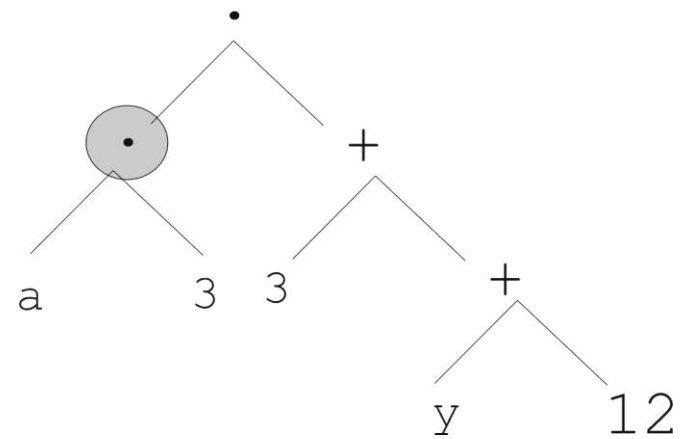
- Мутација има два параметра:
 - Вероватноћа p_m одабира мутације (у супротном укрштање)
 - Вероватноћа одабира унутрашње тачке (корена подстабла)
- Савет је да p_m буде 0 (Koza'92), или јако блиско 0, нпр. 0.05 (Banzhaf et al. '98)
- Величина детета може да буде већа од величине родитеља

Укрштање

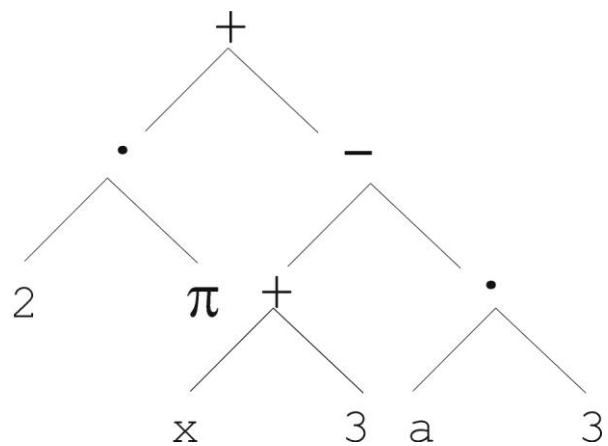
- Најчешћи оператор укрштања:
замена два случајно одабрана подстабла између родитеља
- Укрштање има два параметра:
 - Вероватноћа p_c за одабир укрштања (или мутације у супротном)
 - Вероватноћа одабира унутрашње тачке (корена подстабла)
као позиције за укрштање код сваког од родитеља
- Величина детета може да буде већа од величине родитеља



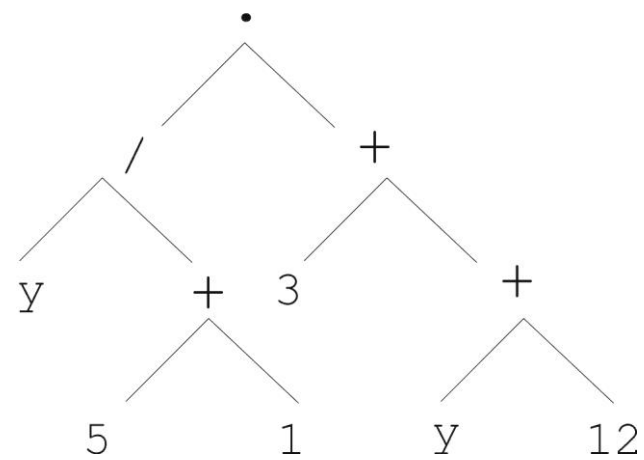
Родитель 1



Родитель 2



Дете 1



Дете 2

Селекција

- Селекција родитеља је обично фитнес-сразмерна
- Селекција у веома великим популацијама
 - Рангирај популацију према фитнесу и подели их у две групе:
 - група 1: најбољих $x\%$ популације
 - група 2 осталих $(100-x)\%$
 - 80% операција селекције изврши над групом 1, преосталих 20% над групом 2
 - За популације величине = 1000, 2000, 4000, 8000 $x = 32\%, 16\%, 8\%, 4\%$
 - Ови проценти су одређени емпиријски
- Селекција преживелих:
 - Стандардни приступ: генерацијски
 - Модел са стабилним стањем и елитизмом постаје популаран у последње време

Иницијализација популације

- Поставља се максимална дубина стабла D_{\max}
- Балансирани приступ (тежи се ка балансираном стаблу дубине D_{\max}):
 - Чвоорви на дубини $d < D_{\max}$ се случајно бирају из скупа функција F
 - Чворови на дубини $d = D_{\max}$ се случајно бирају из скупа термова T
- Ограничени приступ (тежи се ка стаблу ограничене дубине $\leq D_{\max}$):
 - Чворови на дубини $d < D_{\max}$ се случајно бирају из скупа $F \cup T$
 - Чворови на дубини $d = D_{\max}$ се случајно бирају из скупа T
- Стандардна GP иницијализација:
комбиновани приступ који користи и балансирани и ограничени приступ
(сваки по пола популације)

Приступ заснован на повећавању

- Bloat = „тенденција ка удебљавању”, стабла унутар популације током времена расту
- Дебата у научним истраживањима
 - Сетимо се „окамове бритве“
- Потребне су контрамере, e.g.
 - Спречавање употребе оператора који доводе до „превелике“ деце
 - Пенализација „превеликих“ јединки

Пример примене: симболичка регресија

- За дате тачке у \mathbf{R}^2 , $(x_1, y_1), \dots, (x_n, y_n)$
- Пронаћи функцију $f(x)$ такву да $\forall i = 1, \dots, n : f(x_i) = y_i$
- Могуће GP решење:
 - Дати функцијски симболи $F = \{+, -, /, \sin, \cos\}$, и термови $T = \mathbf{R} \cup \{x\}$
 - Фитнес представља грешку $err(f) = \sum_{i=1} (f(x_i) - y_i)^2$
 - Стандардни оператори
 - Величина популације= 1000, употреба приступа пола-пола

Дискусија

Да ли се GP:

Може користити за еволуцију рачунарских програма?

Шта је са другим репрезентацијама поред стабла?