

Рачунарска интелигенција

Еволутивна израчунавања

Александар Картељ

kartelj@matf.bg.ac.rs

Датум последње измене: 27.11.2019.

Еволутивна израчунавања

Уводни концепти

Еволуција

- Еволуција се може сматрати **оптимизационим процесом** са циљем побољшања **прилагођености** организма (или система) **динамичном и такмичарски** настројеном окружењу
- Домени:
 - Хемијски
 - Космички
 - Биолошки
 - Људске творевине
- Наш фокус је доминантно на биолошкој еволуцији
- Lamarck (1744-1829), подржавао теорију еволуције кроз наслеђивање стечених карактеристика познату већ 2000 година – касније оповргнута!
- Darwin (1809-1882), теорија еволуције кроз процес **природне селекције**

Природна селекција

- Свака јединка се **такмичи** са осталима у циљу преживљавања
- „**Најбоље**“ јединке имају већу шансу да преживе и оставе ПОТОМСТВО
- Због овога чешће преносе своје гене, тј. **карактеристике**
- Током времена, овакве „**погодне**“ карактеристике постају доминантне у популацији
- Током стварања потомака улогу играју и случајни догађаји:
 - Кроз укрштање (бира се ген оца или мајке)
 - Кроз процесе мутације (насумичне измене због спољних догађаја)
- Неки случајни догађаји могу додатно да унапреде организам

Еволутивна израчунавања

- Имитирају процес еволуције кроз:
 - Природну селекцију
 - Укрштање
 - Мутацију
 - итд.
- Уместо организама и њихове борбе за преживљавањем, јединке у популацијама, кодирају решења неког проблема
- Након неког времена, решења еволуирају у смеру побољшања

Уопштени еволутивни алгоритам (ЕА)

- Еволутивни алгоритми траже оптимална решења путем **стохастичке** претраге над **простором решења**
- Јединке (хромозоми) представљају појединачне тачке у простору решења
- Кључни аспекти ЕА су:
 1. Кодирање решења у виду хромозома – нпр. низ целих бројева
 2. Фитнес функција – оцена квалитета јединке
 3. Иницијализација почетног скупа јединки (почетних решења)
 4. Оператори селекције – како бирамо оне који се репродукују
 5. Оператори укрштања – како се врши стварање нових јединки

Псеудокод ЕА

Иницијализуј број генерација на $t = 0$;

Креирај и иницијализуј n_x -димензиону популацију $C(0)$ од n_s јединки;

while *нису задовољени услови за завршетак* **do**

 Израчунај фитнес $f(\mathbf{x}_i(t))$, сваке јединке $\mathbf{x}_i(t)$;

 Изврши укрштање и формирај потомке;

 Одабери нову популацију, $C(t + 1)$;

 Пређи у наредну генерацију, $t = t + 1$;

end

EA алгоритми

- Генетски алгоритми
 - еволуција над линеарним генотипом, низом.
- Генетско програмирање
 - еволуција над стаблоидним генотипом.
- Еволутивно програмирање
 - еволуција фенотипа тј. понашања; кодирање секвенцама понашања, а не низовима; нема укрштања; фитнес је релативан у односу на друге јединке.
- Еволутивне стратегије
 - еволуција еволуције: еволуција генотипа + еволуција параметера еволуције генотипа.
- Диференцијална еволуција
 - као стандардни EA само се мутација бира из унапред непознате случајне расподеле – прилагођене популацији; бирају се вектори помераја који су релативни у односу на остале јединке популације.
- Културна еволуција
 - еволуција културе у популацијама – јединке прихватају веровања из популације, али и утичу на популацију сразмерно својој прилагођености.
- Коеволуција
 - еволуција и преживљавање јединки кроз сарадњу и такмичење, нпр. биљке и инсекти (симбиоза).

Еволутивна израчунавања

Елементи еволутивног алгоритма

Кодирање (репрезентација) - хромозом

- Хромозоми сачињени од молекула DNA (налазе се у језгру ћелије)
- Сваки хромозом сачињен од великог броја гена
- Ген је јединица наслеђивања
 - Одређује анатомију и физиологију организма
 - Јер кодира и контролише процес изградње протеина
- Јединка је сачињена од секвенце гена
- Вредност (садржај) гена се зове генски алел
- У контексту ЕА – хромозом представља решење проблема
- Док су појединачни гени карактеристике решења

Кодирање

- Одабир погодног кодирања је кључно за решавање проблема
- Кодирање је најчешће засновано на низу вредности неког типа
 - Осим у случају генетског програмирања где је код нелинеарна (стабло)
- Класичан пример репрезентације: бинарни вектор фиксне дужине

- **Проблем 1:**

За задати граф $G=(V,E)$ пронаћи најмањи подскуп скупа чворова V' такав да свака од грана графа има бар један крај у подскупу V'

<https://www.nada.kth.se/~viggo/wwwcompendium/node10.html>

Кодирање (2)

- Код може бити засновано и на низу целих бројева фиксне дужине

- **Проблем 2:**

Трговачки путник (TSP) – Нека је задат скуп S од m градова и функција удаљености $d(c_i, c_j) \in \mathbb{N}$ за сваки пар градова.

Пронаћи пермутацију $p: [1..m] \rightarrow [1..m]$ такву да је укупна сума удаљености грана које се пролазе обиласком минимална.

<https://www.nada.kth.se/~viggo/wwwcompendium/node104.html>

Кодирање (3)

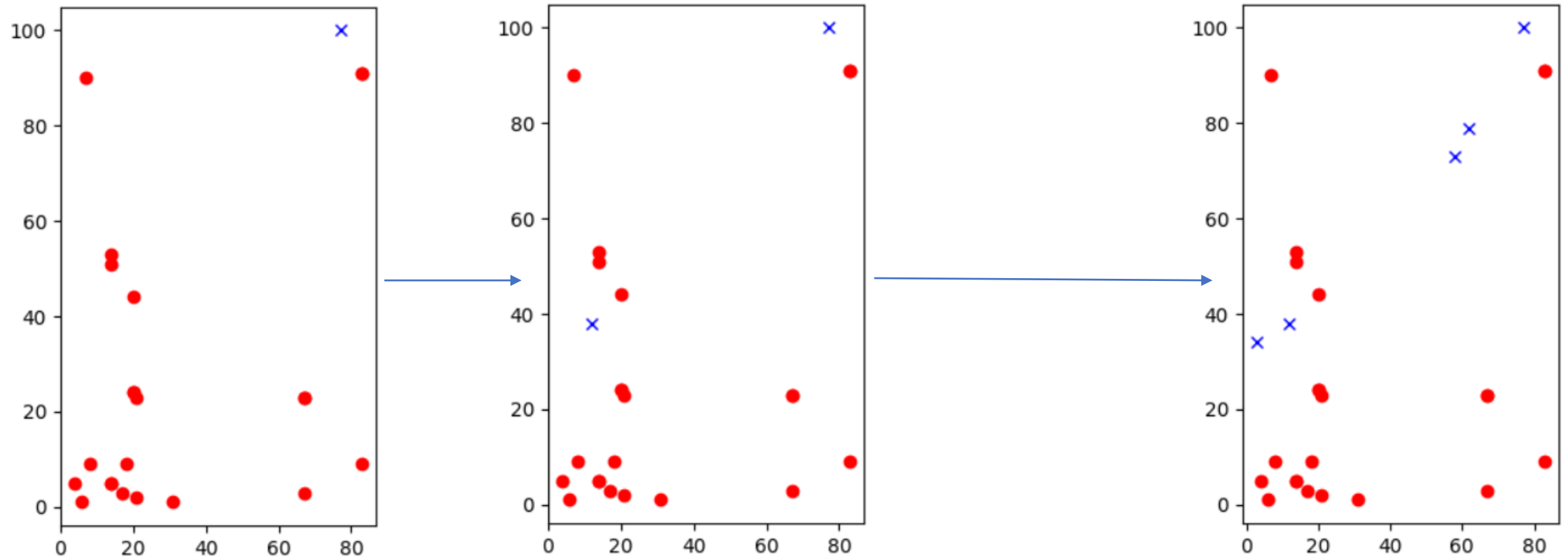
- Понекад се домен хромозома и домен решења не поклапају
- На пример, можемо користити низ реалних вредности за хромозом, а да решење буде бинарне природе
- Нпр. ако је код $\mathbf{x}=(x_1, \dots, x_n) \in [0, 1]^n$
- Како бисмо овакав код превели у решење проблема 1 и 2?
- Да ли је могуће обратно: да имамо бинарни/целобројни код хромозома, а да домен проблема буде реалан?

Пример проблема

- Размотримо једноставан проблем који подсећа на NP тежак проблем звани p -Median (или k -Median) <https://www.nada.kth.se/~viggo/wwwcompendium/node132.html#8495>
- *p -Median проблем је дефинисан над графом, а његово решење је представљено подскупом од p одабраних чворова таквих да је укупна удаљеност до свих осталих чворова минимална*
- Проблем који ћемо размотрити је дефинисан слично с тим што није дат граф већ равн тј. правоугаоник одређених димензија
- Уместо удаљености изражене у броју грана дуж пута овде је удаљеност једноставно еуклидска
- Разлог је то што желимо да омогућимо лаку визуелизацију решења

Кодирање за 5-Median у равни

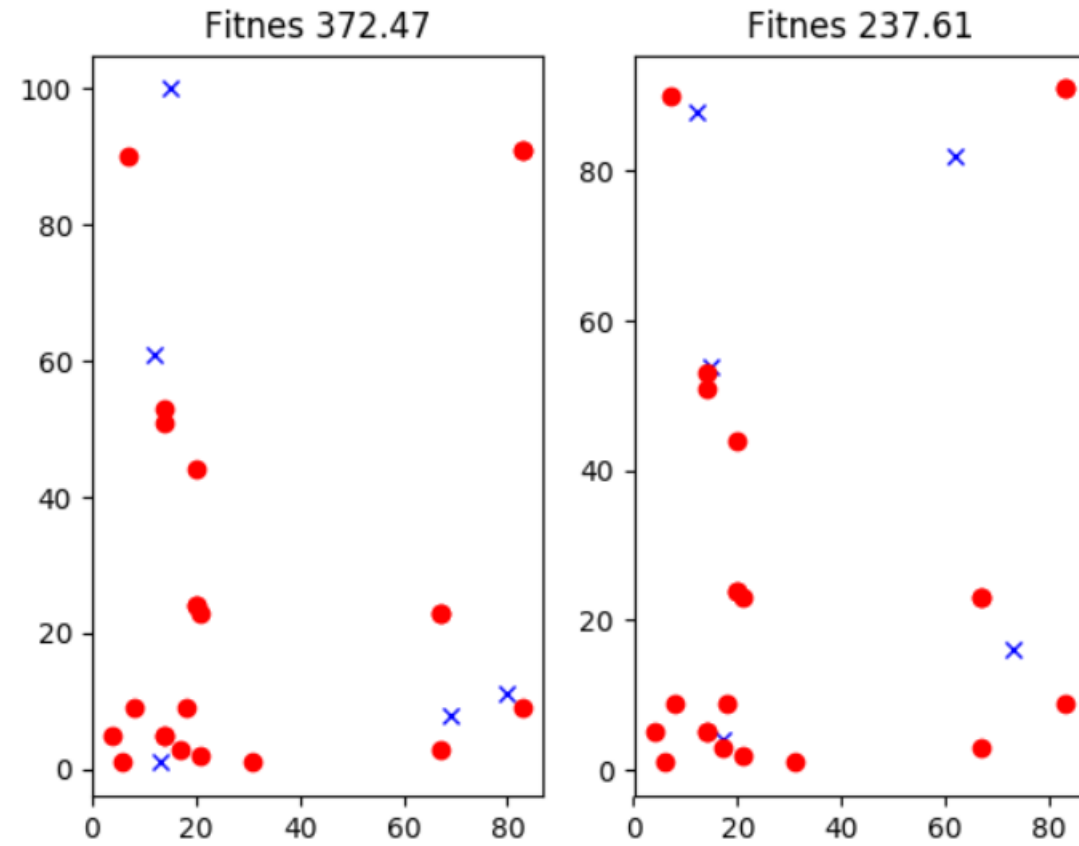
- Најједноставнија репрезентација је вектора реалних вредности
 - Неке методе не подржавају реалну репрезентацију



Фитнес функција

- Према Дарвиномо моделу еволуције, јединке са најбољим карактеристикама имају шансе да преживе и оставе потомство
- Квантификација ових карактеристика се изражава путем тзв. **фитнес функције**
- Фитнес функција се примењује над јединком
 - Ова вредност је обично апсолутна мера квалитета јединке
 - Међутим, некада може бити и релативна у односу на друге јединке
- Фитнес функција је обично, али не и нужно, једнака функцији циља
- У претходном примеру, за 5-Median у равни фитнес функцију дефинишемо на исти начин као и функцију циља, а то је укупну удаљеност свих тачака до најближе одабране тачке

Фитнес функција за 5-Median у равни



Развој упоредног алгоритма

- За валидацију квалитета предложеног алгоритма добро је имати упоредни алгоритам
 - Идеално је ако је тај алгоритам егзактан, тј. ради тачно
 - У случају NP тешких проблема димензија коју решавамо упоредним алгоритмом је очекивано мала
 - Други начин је поређење са већ постојећим резултатима из литературе
- Често упоредни алгоритми користе исту функцију циља и кодирање
 - Из тог разлога, ово је добар моменат да се најпре размисли о истом
 - За решавање претходно поменутог проблема, имплементирати **алгоритам случајне претраге**
 - За домаћи размислити о неком виду **систематичне претраге?**

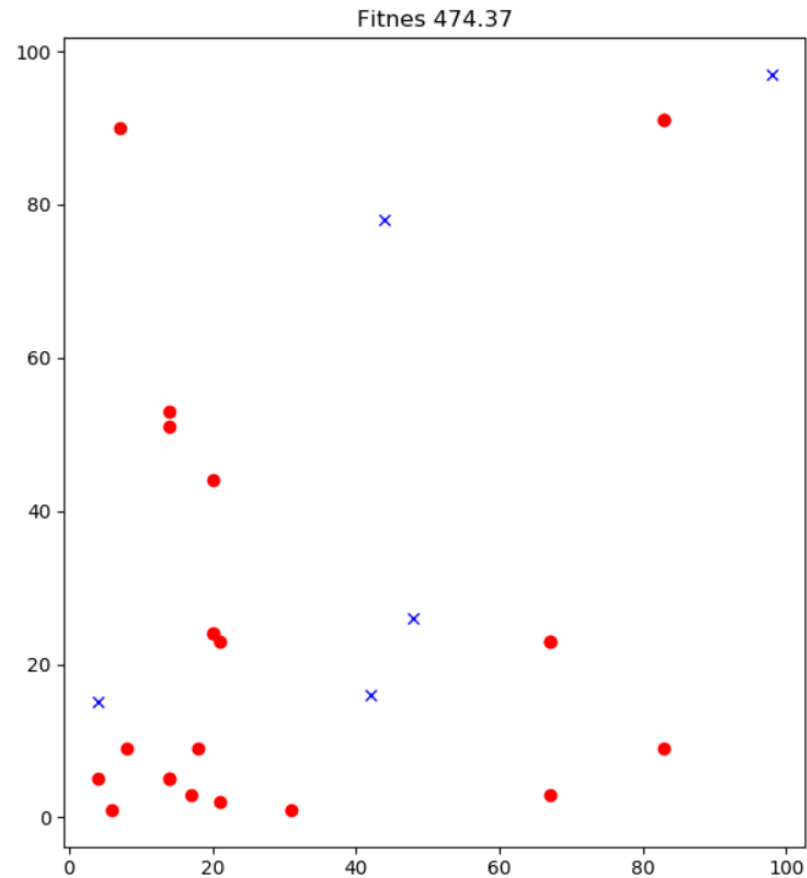
Иницијализација решења

- Стандардни приступ
 - Иницијалну популацију формирати од **насумично одабраних допустивих** решења
 - У случају недопустивих решења вероватно ће бити потребна поправка
- Зашто?
- Због бољег покривања целог скупа допустивих решења
- Довољно велик **случајан узорак** има добру репрезентативност
- У случају да неки део простора решења није покривен у почетку велика је шанса да се неће касније уопште обићи тај део

Иницијализација решења (2)

- Колика треба да буде величина популације и које су предности/мане велике/мале популације?
 - Величина се одређује обично емпиријски за конкретну методу
 - Велика популација омогућава већу „покривеност“ и повећава шансу за налажењем глобалног оптимума (**диверзификација**)
 - Мала популација је ефикаснија и омогућава бржу конвергенцију ка локалном оптимуму (**интензификација**)
- У случају разматраног проблема, поставити:
 - Величину популације на 10 јединки (пробати и варијанту са 20, 50, ...)
 - Иницијализацију спровести на случајан начин

Пример иницијалне популације за 5-Median у равни



Селекција

- У овој фази се врши одабир решења која треба да оставе потомство
- Начелна идеја: „дај већу шансу бољим решењима“
- **Селекциони притисак (енг. selection pressure):**
време потребно да се произведе униформна популација јединки односно да најбоље јединке оставе своје гене свуда
- Оператори селекције са високим селекционим притиском смањују разноврсност гена у популацији брже (преурањена конвергенција)

Селекција - приступи

- Случајна – свака јединка има исту шансу
 - Најнижи селекциони притисак, али спора конвергенција
- Пропорционална – даје већу шансу бољим јединкама

$$\varphi_s(\mathbf{x}_i(t)) = \frac{f_{\Upsilon}(\mathbf{x}_i(t))}{\sum_{l=1}^{n_s} f_{\Upsilon}(\mathbf{x}_l(t))}$$

- Рулетска селекција је стандарни начин имплементације овог механизма

Селекција – приступи (2)

- Турнирска селекција –турнир између случајног подскупа јединки
 - Ако је подскуп једнак популацији – стратегија је елитистичка
 - Ако је подскуп величине један – стратегија је случајна
 - Варирањем величине подскупа мења се селекциони притисак
- Ранговска селекција – уместо вредности фитнес функције користи се само редни број у уређењу популације
 - Смањује селекциони притисак,
јер се јако добрим решењима релативизује значај

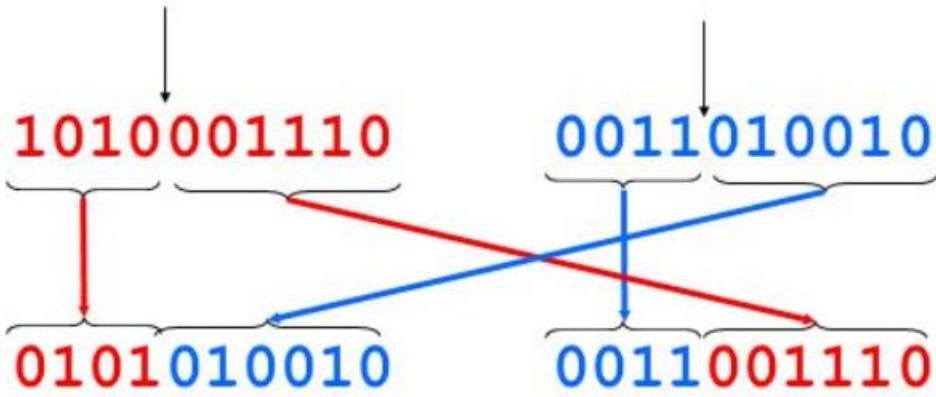
Елитизам

- Техника која спречава „губљење“ добрих јединки
 - Иако се најбоље јединке највероватније налазе у скупу одабраних родитеља, применом укрштања и мутације може доћи до тога да потомци имају лошији квалитет
- Приступ:
 - Изабери најбољу или неколико најбољих јединки и директно их прекопирај у наредну генерацију
 - Број елитистички одабраних јединки не сме бити превелики! (висок селекциони притисак)
- У случају разматраног проблема:
 - Тестирати случајну, елитистичку и рулетску селекцију

Укрштање

- Овим процесом се креирају нове јединке – потомци
- Подразумева употребу следећих оператора:
 - Оператори **укрштања** – рекомбинације гена
 - Оператори случајне **мутације** – промене насумичних гена – опционо
- О овим операторима ће бити више речи код генетских алгоритама
- У разматраном проблему:
 - Применити једнопозиционо насумично укрштање
 - Применити мутацију са променљивом вероватноћом (тестирати различите вредности 0.05, 0.1, 0.2...)

Једнопозиционо насумично укрштање



Мутација



Критеријуми заустављања

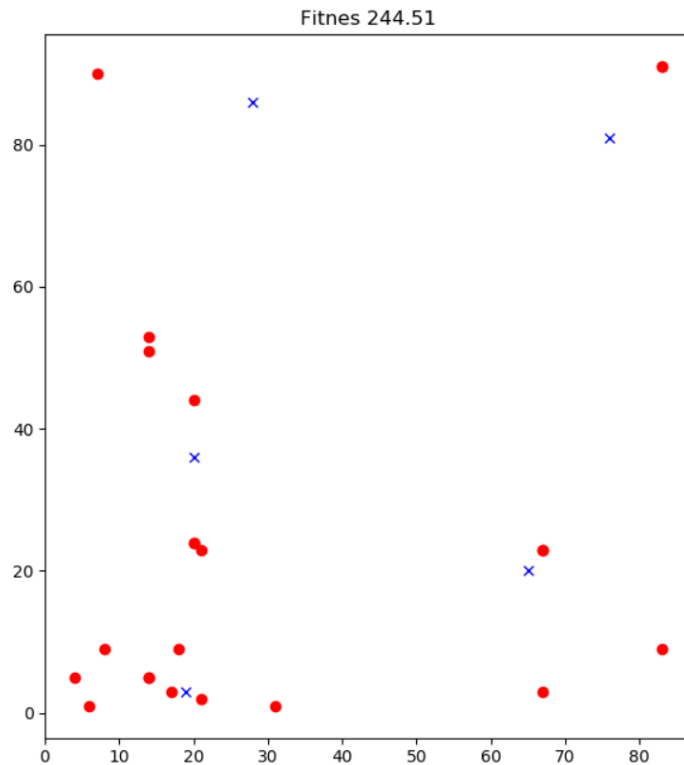
1. Након истека унапред фиксираног броја генерација
2. Након истека унапред фиксираног времена
3. Када нема унапређења у P последњих генерација
4. Када у P последњих нема промене у генотипу
5. Када је нађено прихватљиво решење:
само ако знамо шта нам је прихватљиво
6. Када се нагиб фитнес функције више не повећава:
потребно је пратити кретање фитнес функције кроз време

Упоређивање решења

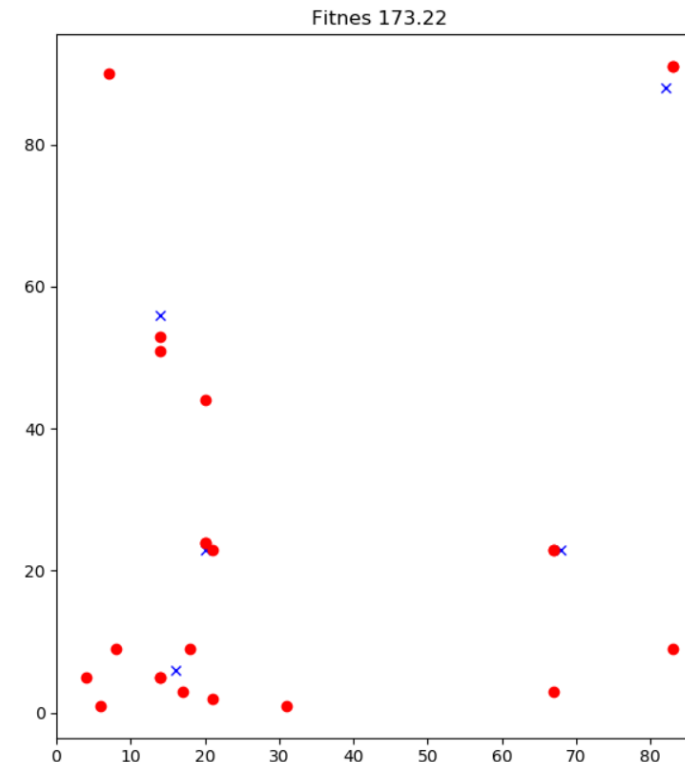
- Упоредити решења добијена:
 - Алгоритмом случајне претраге
 - Еволутивним алгоритмом
 - Егзактним алгоритмом:
могуће је егзактно извршити на мапи неке димензије
- Водити рачуна о томе да поређење буде фер:
 - Нпр. да се изврши исти број израчунавања фитнес функције
 - Фер је нпр. да се ЕА изврши 1000 пута са популацијом 10 јединки и да се случајни алгоритам изврши 10000 пута

Упоређивање решења (2)

Случајни алгоритам:
20000 генерација



Еволутивни алгоритам:
2000 генерација са 10 јединки
и елитистизмом



Напомена

- **Сви аспекти ЕА су демонстрирани кроз решавање проблема 5-Median проблема у равни**
- Можете преузети код са сајта и пробати нека унапређења или измене параметара:
 - Број итерација
 - Величину популације
 - Вероватноћу мутације
 - Број јединки које се бирају (колико пута се примењује селекција)
 - Другачију имплементацију неких оператора ЕА
 - ...