

Evolucijsko računanje

Vaje 2025/2026

3 Simbolična regresija z genetskim programiranjem (GP)

Cilj naloge je spoznati proces reševanja simbolične regresije z uporabo genetskega programiranja (GP).

V okviru naloge lahko:

a) Rešujete obstoječe probleme

Poiščite en realen problem ali pa vsaj tri sintetične probleme simbolične regresije (npr. funkcije iz raziskovalnih člankov). Primere sintetičnih problemov lahko najdete v članku [Genetic Programming Needs Better Benchmarks](#) (Tabela 3). Pri obstoječih problemih poizkusite poiskati tudi identične funkcije.

b) Definirate lasten problem simbolične regresije.

Premislite katere realne probleme bi lahko modelirali z uporabo simbolične regresije. Primeri realnih problemov:

- **Učinkovitost porabe goriva avtomobila:** Poiščite enačbo, ki napoveduje porabo goriva glede na hitrost, število obratov motorja, težo vozila ipd.
- **Tekaške zmogljivosti:** Oblikujte enačbo, ki napove končni čas tekača na podlagi srčnega utripa, tempa in VO_2 max.
- **Vremenska napoved:** Ustvarite preprost model, ki na podlagi temperature, vlažnosti, tlaka in hitrosti vetra napove količino dežja.
- **Overclocking:** Razvijte enačbo za napoved frekvence procesorja (CPU) ali grafične enote (GPU) na podlagi napetosti in temperature.
- **Ocena cene nepremičnine:** Izpeljite model za napovedovanje cene nepremičnin na podlagi podatkov, kot so kvadratni metri, lokacija in število sob.
- **Obnašanje pospeška mobilnega telefona:** Oblikujte enačbo, ki na podlagi podatkov iz pospeškometra (x, y, z osi) napove skupno silo, ki deluje na telefon.
- **Optimizacija proizvodnje:** Poiščite model za napoved izkoristka stroja na podlagi spremenljivk, kot so hitrost stroja, temperatura in poraba energije.
- **Kmetijski pridelki:** Razvijte enačbo za napoved pridelka (npr. količine pšenice) glede na vnos gnojil, količino padavin in povprečno dnevno temperaturo.

Pri lastnem problemu lahko tudi uporabite resničen javno dostopen nabor podatkov (dataset) iz različnih virov (Kaggle, UCI Machine Learning Repository, OpenML, Google Dataset Search).

Praktični nasveti za izbiro in pripravo podatkov

- Izberite dataset z **numeričnimi** vhodnimi spremenljivkami in **numerično ciljno spremenljivko** (regresijski problem).
- Število primerov naj bo vsaj nekaj sto (idealno več tisoč), da je evolucija stabilnejša.
- Po potrebi odstranite manjkajoče vrednosti ali kategorijske spremenljivke (ali jih pretvorite v numerične z one-hot kodiranjem).
- Podatke razdelite na učno (70-80 %) in testno množico (20-30 %). GP izvajajte samo na učni množici, končno najboljšo enačbo pa ocenite na testni množici.

3.1 Implementacija in testiranje

Izbrane probleme boste implementirali in reševali v ogrodju [EARS](#) z uporabo algoritma [ElitismGPAAlgorithm](#). Primer reševanja simbolične regresije je na voljo v datoteki [SymbolicRegressionExample](#).

Za rešitev problema je potrebno:

- Konfigurirati GP z ustreznimi funkcijami (operacijami) v [baseFunctionNodeTypes](#) in terminali (spremenljivkami in konstantami) v [baseTerminalNodeTypes](#).
- Podati vhodne in izhodne podatke z uporabo razreda [Target](#).
- Po potrebi implementirati manjkajoče funkcije ali operatorje, tako da dedujete iz razredov [OperatorNode](#) in [Node](#).

Ko definirate problem, poskrbite za ustrezno pripravo vhodnih podatkov (npr. z uporabo naključnih ali enakomerno razporejenih točk funkcije). Ocenite rezultate in kakovost rešitve, nato pa po potrebi preizkusite različne nastavitve parametrov algoritma.

3.2 Oddaja poročila in kode

Oddati morate:

- **Kodo**, ki vsebuje implementacijo problema/ov v ogrodju EARS.
- **Poročilo** (do 2 strani), ki vključuje:
 - Kratek opis problema.
 - Pregled uporabljenih funkcij in terminalov.
 - Opis vhodnih podatkov.
 - Analizo rezultatov (npr. grafi, meritve napak, ugotovitve).

Točkovanje:

- Implementacija problema (14 točk): Popolnost in pravilnost implementacije v ogrodju EARS.
- Kreativnost problema (5 točk): Inovativnost pri izbiri ali oblikovanju problema/ov.
- Poročilo (6 točk): Jasnost in natančnost opisa ter analiza rezultatov.

* Naloga je **obvezna** in vredna **25 točk**.