

Domača naloga - 1.del

Odkrivanje enačb in uporaba predznanja

Vito Rozman

7. junij 2023

1 Soočanje s problemom

Pri izbiri orodaj sem najprej preizkusil linearno regresijo, lasso regresijo, rige regresija in nazadnje še modele za iskane enačbe s pomočjo [PySR](#) orodij. Pri evalvaciji rezultata sem upošteval dve stavri, napako MSE in kompleksnost enačbe, ki sem jo ocenil z metodo "ostrega pogleda". Pri regresijah sem preizkusil različne parametre (meja in lambda parameter) in različno generirane spremenljivke. Prav tako sem sprobaj različne stopnje polinomov. Postopek iskanja enačbe sem izvedel v naslednji korakih.

- Prvo sem zagnam različne tipe regresij, ki smo jih obravnavali na vajah, brez da bi kakorkoli spremljal spremenljivke. Ta pristop se je bil obupen, saj je bila napaka zelo velika.
- Druga modifikacija je bila, da sem ustvaril nove spremenljivke glede na podano predznanje. Generiral sem: $T_w - T_a$, $\sqrt{T_w - T_a}$, $\sin(\theta)$, $\cos(\theta)$, $\frac{1}{\eta}$ in $-\eta$. Potem sem ponovil postopek kot pri prejšnji točki in dobil boljše rezultate. Najboljša napaka je znašala 0.029%, vendar je kompleksnost ostala enako salba.
- Tretji pristop je bil z drugimi metodami kot sta algoritem BACON in metoda iz knjižnice PySR. BACON se ni izkazal dobro, PySR pa presenetljivo dobro, saj sem dobil nizko napako in dokaj preprosto enačbo.

2 Model za iskanje enačbe

Kot že omenjeno sem uporabil orodje PySR in sicer funkcijo `PySRRegressor`.

Tabela 1: Nastavitev orodja

Binarne operacije	$+, -, *, /$
Druge operacije	$\cos(\cdot), \sin(\cdot), \text{inv}(\cdot), \cdot^2, \sqrt{\cdot}, \cdot^3, \cos^2(\cdot), \sin^2(\cdot)$
Parameter kompleksnosti	20
Največja globina gnezdenja	10
Napaka	MSE

3 Rezultat

Izbrana enačba je oblike:

$$\sin(\theta) \cdot \left(\frac{3}{2} - \eta\right) \cdot \sqrt{\frac{4(T_w - T_a)}{69}}.$$

Z dano enačbo sem dobil napako 0.0133%, kompleksnot modela pa 13. Dobljeni enačbi bi zaupal v približno 95% primerih, saj so dobljene konstante rezultat prileganje vhodnim podatkom. Ostale komponente se mi zdijo dokaj vredne, ker ustrezajo predpostavkam domenskega predznaja.

Opis datotek

- **preproces.py** skripta kjer generiram nove spremenljivke,
- **eq_regression.py** skripta s funkcijami regresije in algoritma BACON,
- **NAL1.ipynb** zvezek iskanja enačbe.